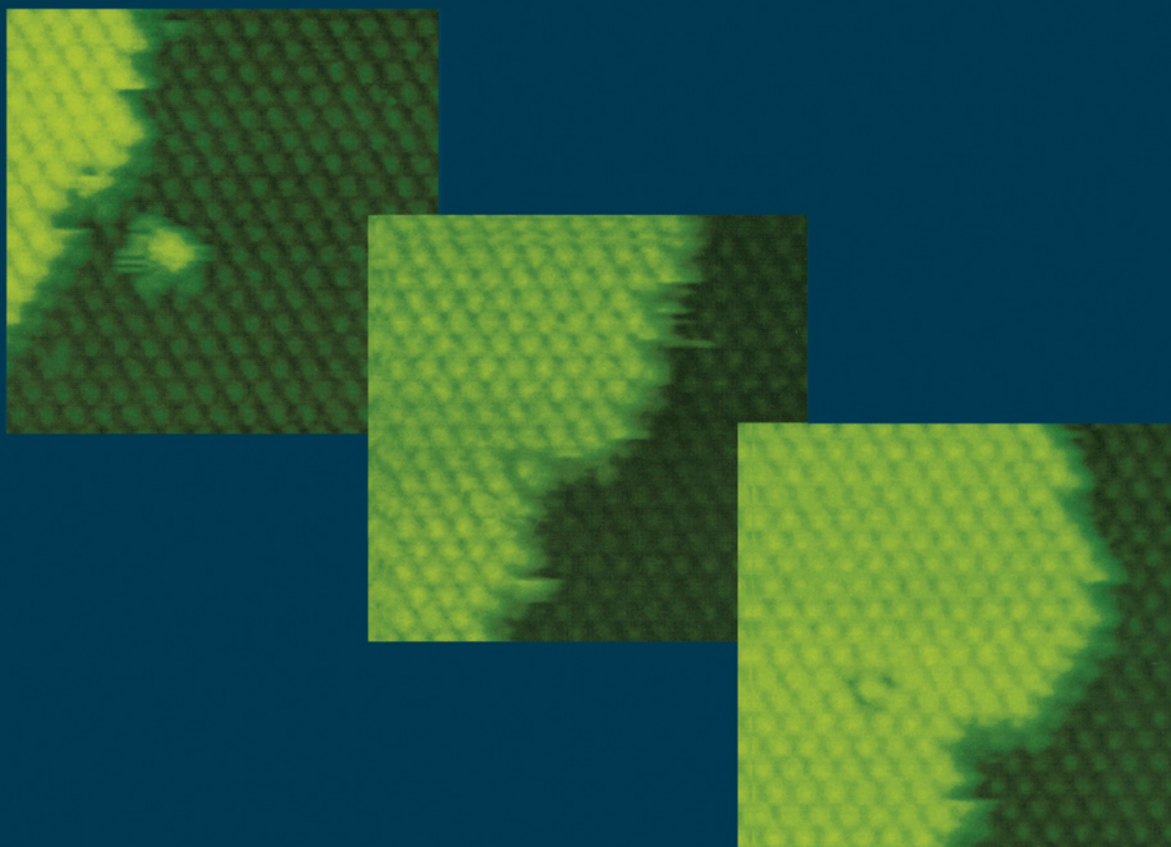# ADVANCES IN CRYSTAL GROWTH RESEARCH
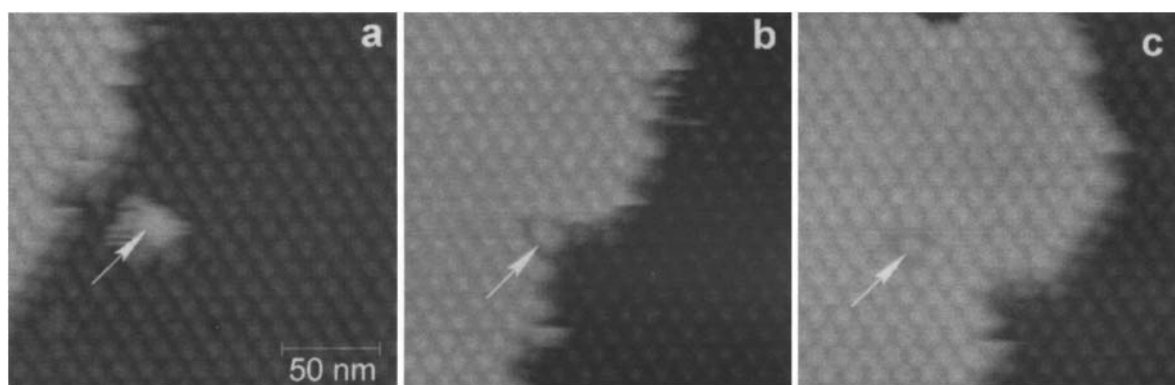
EDITED BY

K. SATO, Y. FURUKAWA AND K. NAKAJIMA

Advances in
Crystal Growth Research

About cover pictures

In-situ atomic force microscopy images showing incorporation of impurities by growing protein (ferritin) crystals. The dimers adsorb on the surface of growing crystals at concentrations largely exceeding those in the solution bulk (a), the adsorbed dimers appear as large clusters. Almost all adsorbed dimers are incorporated by the advancing steps (a, b and c). Monitoring the interaction of a step with a dimer molecule allows deconvolution of the dimer shape from imaging artifacts and we see in b and c that the dimers consist of two bound monomers. For further details, see S.-T. Yau, B.R. Thomas, O. Galkin, O. Gliko and P.G. Vekilov, Proteins: Structure, Function, Genetics 43 (2001) 343.

Prof. Kiyotaka Sato
Physical Chemistry Laboratory
Faculty of Applied Biological Science
Hiroshima University
Higashi-Hiroshima, 739-8528, Japan
tel:+81-824-247935
fax:+81-824-227062
e-mail:kyosato@hiroshima-u.ac.jp
lab-home page address;http://home.hiroshima-u.ac.jp/fdphys

Advances in
Crystal Growth Research

Edited by

K. SATO
Physical Chemistry Laboratory, Faculty of Applied Biological Science,
Hiroshima University,
Higashi-Hiroshima 739-8528, Japan

Y. FURUKAWA
Institute of Low Temperature Science, Hokkaido University,
Sapporo 060-0819, Japan

K. NAKAJIMA
Institute of Materials Research, Tohoku University,
Sendai 980-8577, Japan

Preface

The aim of this book is to provide a timely collection that highlights advances in current research of crystal growth ranging from fundamental aspects to current applications involving a wide range of materials.

This book is published on the basis of lecture texts of the 11th International Summer School on Crystal Growth (ISSCG-11) to be held at Doshisha Retreat Center, in Shiga Prefecture Japan, on July 24–29, 2001. This school is always associated with the International Conference of Crystal Growth (ICCG) series that have been held every three years since 1973; thus this school continues the tradition of the past 10 schools of crystal growth. In 2001, ICCG-13/ICVGE-11 (the 11th International Conference of Vapor Growth and Epitaxy) is held at Doshisha University in Kyoto right after ISSCG-11.

ISSCG-11 was organized with dual aims of providing an occasion to enjoy lectures and discussions in crystal growth research as well as to form closer ties with researchers throughout the world. Out of 24 lectures delivered in the school, 23 chapters are in this book.

The 22 chapters of this book are classified into three parts, together with Chapter 1 that reviews crystal growth history written by Prof. Ichiro Sunagawa. In Chapter 1, an overview of the science and technology of crystal growth dating back to the 17th century is presented with special attention to the significance of crystal growth for modern science and technology. Basic concepts of crystal growth overviewed in Chapter 1, such as equilibrium and growth forms, epitaxy, and interfaces, are newly highlighted in the following chapters that deal with cutting-edge research of crystal growth.

In Part 1, consisting of seven chapters, pertains to fundamental aspects of crystal growth. Following the theoretical arguments on phase field theory (Chapter 2) and heterogeneous nucleation (Chapter 3), recent microgravity experiments mostly performed in space are reviewed in relation to interface kinetics, mass transport, growth patterns, and crystal perfection (Chapter 4). The four chapters (Chapter 5 through Chapter 8) discuss basic phenomena of surface step and epitaxial growth involving computer simulations. All chapters in Part 1 will help readers to understand the related phenomena that occur in more complicated systems to be discussed in Parts 2 and 3.

Eleven chapters of Part 2 deal with crystal growth in the areas of bulk, thin film, and quantum dots for semiconductors, optoelectronics, magnetics, and optics. The growth of bulk crystals from the melt is discussed in Chapters 9 and 10 with emphasis on experiments and simulations. Growth features of thin film and quantum-dot sized crystals of various semiconducting, optoelectronic, and magnetic materials are elaborated in subsequent chapters; in particular, nitride thin film growth that has greatly advanced in recent years is described in Chapters 11-13. Characterization of epitaxial layers with X-ray scattering from crystal surfaces is discussed in Chapter 18. Chapter 19 describes growth features of optical crystals that have a transparent range from the near-ultraviolet to the infrared region.

In Part 3, basic mechanisms of crystal growth from solution are described. Dynamics at crystal-liquid interfaces as revealed by X-ray diffraction and atomic force microscopy are discussed in Chapters 20 and 21, respectively. Chapters 22 and 23 describe the control of crystal morphology and heterogeneous nucleation of organic crystals that is performed by putting specific additives into the growth media.

July, 2001

The editors

Kiyotaka Sato, Hiroshima University
Yoshinori Furukawa, Hokkaido University
Kazuo Nakajima, Tohoku University

This Page Intentionally Left Blank

Acknowledgements

K. Sato
Y. Furukawa
K. Nakajima
- -

This Page Intentionally Left Blank

Contents

Part 3   Dynamics of crystal-liquid interface

Crystal growth---Its significance for modern science and technology and its possible future applications

Ichiro Sunagawa
Yamanashi Institute of Gemmology and Jewellery Arts
Kofu, 400-0808 Japan
e-mail (home): i.sunagawa@nifty.com

An overview of the science and technology of crystal growth is presented, and the significance of crystal growth for modern science and technology is discussed. The views of the author on future targets of crystal growth studies in the 21st century are also presented in the hope of arousing interests by young people in the science and technology of crystal growth.

1. INTRODUCTION

In a treatise by Nicolous Steno published in 1669 [1] that claimed the corresponding interfacial angle is always constant regardless of the various forms exhibited by rock crystals, two important concepts were proposed. Firstly, he explained that rock crystals are formed by an inorganic process through the precipitation of tiny particles formed in a high-temperature aqueous solution, not through the action of bacteria in the soil. This claim by Steno was astonishing, since in those days, and until the 18th century, people, including academic scholars, had believed that mineral crystals were formed by the action of bacteria. Steno also explained in terms of growth rate anisotropy why rock crystals principally took hexagonal prismatic forms with six prism faces and terminated by alternating two types of triangular faces, and why the hexagonal prismatic forms vary considerably from crystal to crystal [1, 2].

The concept of growth rate anisotropy still holds true and forms the basis of modern science and technology of crystal growth. The present author respects Nicolous Steno as the pioneer of the science of crystal growth.

Figure 1. Nicolous Steno (1638-1686). Reproduced from "Historical Atlas of Crystallography", Ed. J.Lime-de-Faria, Kluwer Acad. Pub., Dordrecht, (1990).
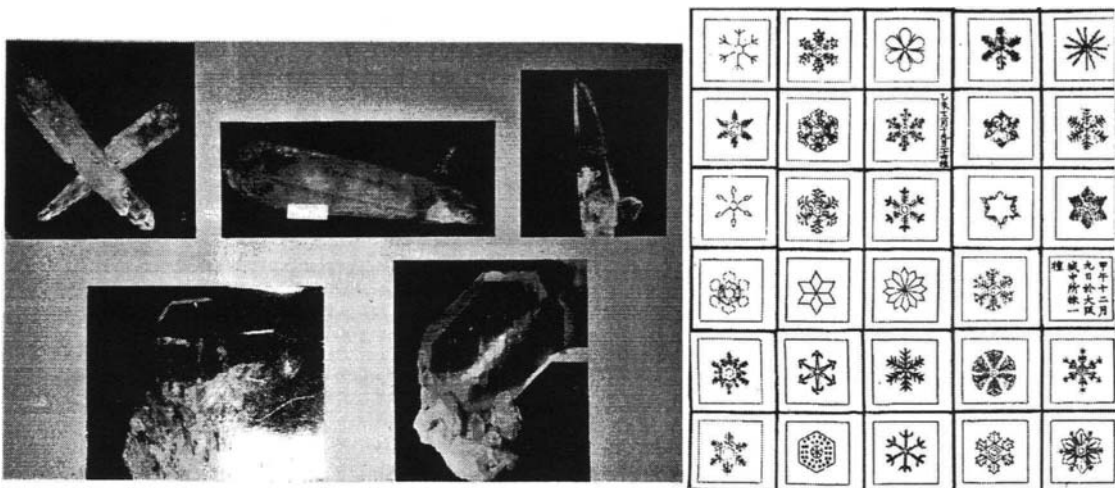
Figure 2. Both structural crystallography and science of crystal growth started from curiosity concerning a wide variety of forms exhibited by crystals. Polyhedral rock crystals (left) and dendritic snow crystals (right). Snow dendrites are sketches drawn by Lord T. Doi (published in 1833 and 1840). This pattern is used as a cover design of The Journal of Japanese Association of Crystal Growth (JACG).

Fifty years before Steno's publication, Kepler [3] who was fascinated by the elaborately varied dendritic forms of snow flakes, considered that snow crystals, although they exhibit thousands of different dendritic forms, are all composed of equal-sized spheres in a closely packed arrangement. This was the root of the concept of crystal structure. Interestingly, both structural crystallography and the science of crystal growth started from the curiosity as to why the same crystal species take elaborately varied forms, not only polyhedral but also dendritic forms.

This curiosity led to the development of studies along two different lines: one to understand and analyze how a crystal is constructed by elemental units, and the other to understand how a crystal nucleates or grows and its morphology is determined.

The concepts of unit cell, crystal lattice, 14 lattice types, 7 crystal systems, symmetry elements, 32 point groups, and 230 space groups were established by the end of the 19th century, and crystal structures were experimentally verified and analyzed by X-ray diffraction in 1912. The crystal structures of most solid materials were analyzed, and structural crystallography, crystal chemistry, and crystal physics developed rapidly. The most essential problem underlying in these studies was to know in detail the atomic arrangements, chemical bonding, and symmetry-property and structure-property relations. The process by which such a regular arrangement can be realized and how the process influences physical perfection and chemical homogeneity of the crystal were not the main concerns for structural crystallographers. Crystals were regarded as thermodynamically minimum

energy states that have ideally regular atomic arrangements. Although the concept of "mosaic structure" existed in the early days of structural crystallography, this was out of the main scope of study at that time. Haekel [4] considered structural crystallography to be an exact and mathematical science for analyzing a dead body, not a science to understand life or living activities.

2. MORPHOLOGY AND CRYSTAL GROWTH

The other interest on crystals developed to understand the origins of various forms of crystals, i.e., to understand why and how different morphologies of crystals appear.

The morphology of a crystal is determined by internal structural factors and external factors, i.e., growth parameters. Since external factors were too complicated to analyze in the early days of research, the first attempts to understand the origins of various forms of crystals focused on internal factors. If it is assumed that the morphology of a crystal can be completely determined by only internal factors, there should be a direct correlation with geometry of crystal lattices, i.e., types and parameters of unit cells, reticular density, reticular spacing of the crystal lattice, and therefore also with point groups and space groups or with anisotropy involved in chemical bonding.
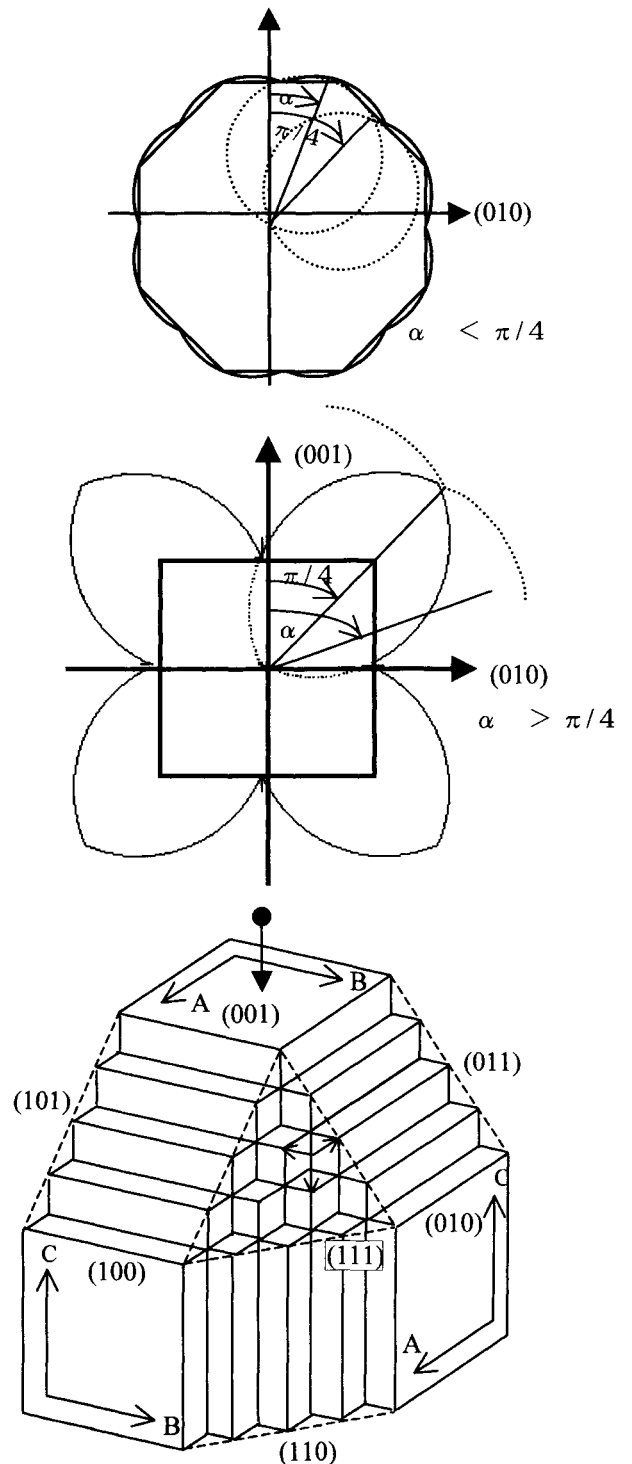
Figure 3. Equilibrium form (Wulff's plot, upper two) and Hartman-Perdok PBC theory (bottom figure; arrows indicate PBC's; {100} is an F face, {110} is an S face and {111} is a K face). See Ref. 5.

Bravais-Friedel's (B-F's) law, Donnay-Harker's extension of B-F's law, and Hartman-Perdok's periodic bond chain (PBC) theory were established on the basis of this assumption. Since these analyses entirely neglected the effects of external factors, the morphology deduced through these analyses was "structural" or "abstract" morphology, not "growth" morphology. However, the morphology deduced based on the basis of these theories could be used as criteria to analyze the deviation and the origin of morphology of real crystals. Gibbs-Curie-Wulf's thermodynamical "equilibrium" form is the morphology of a crystal that has reached the equilibrium state, i.e., minimum energy state. The "equilibrium form" is determined by the concept of minimum surface energy. It should be noted that all of these models dealt with bulk polyhedral crystals, not dendritic crystals. These treatments and their modern versions are summarized in Ref.. [5].

Until the 1920's, experimental investigations aimed at trying to understand the external factors that control "growth forms" as well as mechanisms of the nucleation and growth of crystals focused on the relations between bulk polyhedral crystals and bulk ambient phase. Through these studies, however, various important findings were obtained. As to the external factors that may have drastic effects on "growth forms" of polyhedral crystals, the effects of impurities, solvents and supersaturation were clarified, but the mechanisms by which these factors affect growth forms of polyhedral crystals were not determined.

Papapetrov [6] classified dendritic morphology

into two types, that with rounded tips and that with flat tips, but the origins that cause this difference were not clarified. Nakaya [7] constructed a morphodrom of snow crystals that relate the growth conditions (temperature and supersaturation) and various morphologies of snow crystals. In this so-called "Nakaya diagram", it was indicated for the first time that there are well-defined growth conditions under which specific crystal morphology having dendritic and polyhedral crystals with different crystal habits are formed.
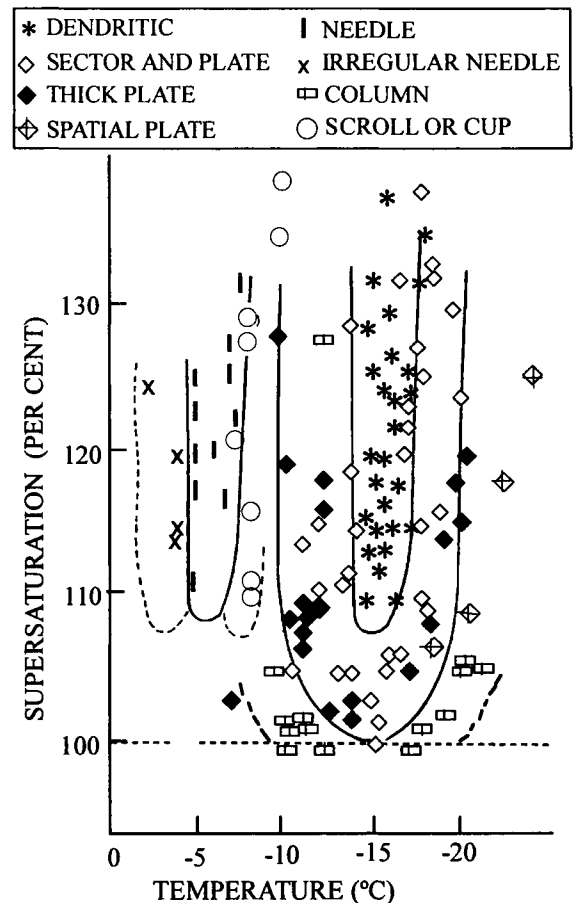
Figure 4. Nakaya' diagram. The first morphodrom of crystals.
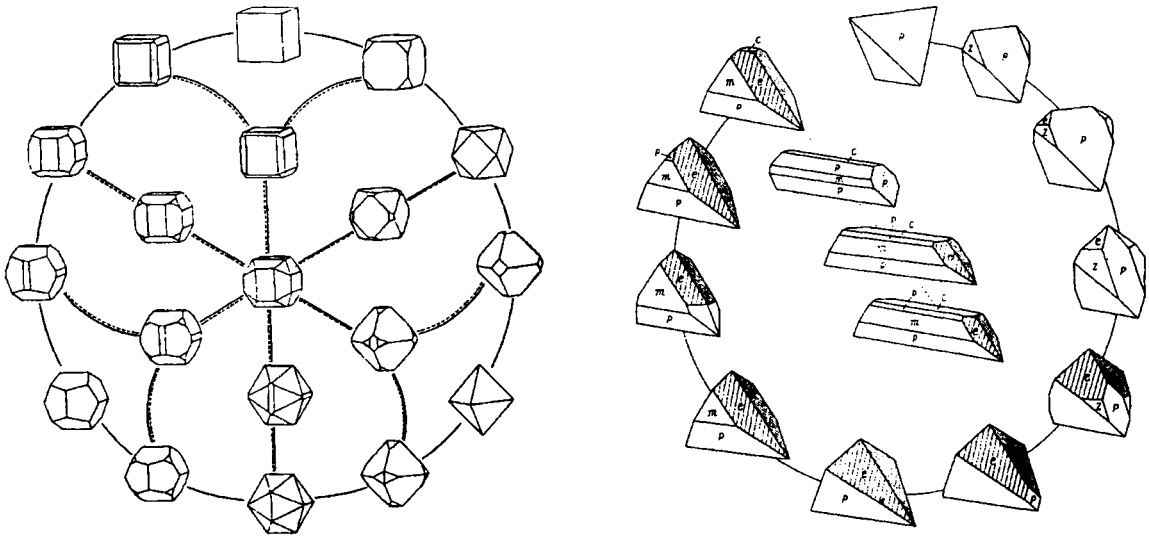
Figure 5. A crystal species may show a wide variety of polyhedral forms due to the relative growth rates of faces concerned (Left: This variation is called Tracht in German, and all of the forms belong to the same category in Habitus, pyrite FeS2) and anisotropic development of crystallographically equivalent faces (right: Habitus in German, chalcopyrite CuFeS2).

In relation to phenomena related to the environments (solution or melt phases) around growing crystals, the roles of diffusion, convection, concentration gradient and presence of a diffusion boundary layer, and a Berg effect were clarified. These effects, however, were found to be concerned with macroscopic relations between bulk crystal and growth media. The atomic processes of crystal growth were not understood in those days.

Therefore, it was still mysterious why a crystal could maintain a polyhedral form bounded by flat faces and why some crystals preferentially exhibited dendritic forms and others polyhedral forms. Key answers to these questions could not be obtained until the concepts of surface and interface and the atomic features of crystal growth had been presented.

3. SYNTHESES OF SINGLE CRYSTALS

Attempts to grow single crystals with high values, such as gemstones, in laboratories using less expensive and easily available raw materials may be regarded as an extension of alchemy. The first success in growing gemstones was achieved in the middle of the 19th century. Carat-sized emerald crystals were grown from high-temperature solutions. Growth of large single crystals of ruby from high-temperature solutions was attempted, but it was not possible to grow cuttable-sized ruby crystals until Verneuille made a breakthrough that enabled large ruby crystals to be grown from the melt phase.

This breakthrough caused a shock in the gem

market, and prices of natural rubies plummeted. Soon after, however, it was obvious that synthetic rubies were easily distinguished from natural rubies, despite the fact that both types of ruby crystal have the same chemical compositions and crystal structures. It was found that the diagnostic differences are due to differences in their growth processes. The prices of natural rubies shot up again, whereas those of synthetic rubies dropped sharply. This was simply because people prefer genuine natural stones to mass-produced synthetic ones and were willing to pay higher prices for natural rubies because of their rarity. In the case of gemstones, capability of mass production of large single crystals actually meant serious conflicts with people's images of gemstones.

On the other hand, there were many scientific and industrial fields that required the production of large single crystals. In turn, those industrial fields greatly developed partly due to the contribution of crystal growth science. Metallurgy is a good example. For metallurgists, it was necessary to grow single crystals of metals to investigate their physical properties, which must be examined in a single crystalline state.

Other examples are optic and piezoelectric crystals. Very large single crystals (at least several cm in diameter) must be grown so that these crystals are employed in optic and electric communication technology. For this purpose, the growth of large single crystals, for example, quartz, ADP and KDP for the piezoelectric application, was first achieved by using solution growth techniques.

However, the easiest and most efficient way to grow large single bulk crystals is to grow them from melt, as is used in Verneuille's method. In response to further demands in industries, various new growth methods, such as Bridgeman-Stockburger method, Czochralski method, and Floating Zone method were developed.

Table 1 Methods employed to grow synthetic gemstones (those having natural counterparts) and artificial gemstones (those having no natural counterparts).

Gem-stone	Method								
	Melt growth					Solution growth			
	V	Cz	BS	Sk	Zm	Ht	Hyd	Hp	Sin
Diamond Simulant								O	
TiO₂	O								
Si-titanat	O								
YAG, GGG		O							
c-ZrO₂				O					
Emerald						O	O		
Corundum									
Ruby	O	O					O		
Sapphire	O	O	O		O		O		
Star	O	O							
Chrysoberyl		O			O		O		
Spinel	O	O					O		
Peridot		O					O		
Quartz							O		
Opal							O		
Turquois									O
Lapis Lazulli									O
Coral									O
Ivory									O

V: Verneuille, Cz: Czochralski, BS: Bridgmann-Stockbarger, Sk: Skarl melt, Zm: Zone melting, Ht: High temperature solution (flux), Hyd: Hydrothermal, Hp: High pressure high temperature, Sin: Sintering (ceramics)

Crystallization performed in chemical industry is categorized as "industrial crystallization", the aim of which is mass production of tiny crystals with controlled size and morphology. Materials to be

synthesized in industrial crystallization range from seasonings to pharmaceuticals. Various types of industrial crystallization apparatuses were designed and constructed so that growth parameters influencing the growth of crystals with desired size and morphology could be controlled.

Developments in the technology of growing large single crystals and industrial crystallization are summarized in Ref. [8].

4. SURFACE AND INTERFACE

Volmer observed [9] growth layers spreading on a growing face of a crystal by detecting them with different interference colors in optical microscopic images. Based on these observations, Kossel and Stranski proposed a model (KS model) describing two-dimensional layers of atoms that spread over a growing surface [10]. This model was the first to

provide the concept of an atomic image of crystal growth and also to define solid and liquid interfaces. Kossel and Stranski categorized growing interface structures into complete and incomplete interfaces, and they defined kink, step and flat sites at which growing atoms are adsorbed and incorporated into a crystal. The concepts of "complete interface" and "incomplete interfaces" in the KS model correspond to what are now called smooth and rough interfaces, respectively. F-, S-, and K-faces proposed by Hartman-Perdok's theory are also of the same nature.

In the KS model, only perfect and ideal crystals (so-called Kossel crystals) were taken into account. The mechanism by which a complete interface is converted into an incomplete interface under the condition of varying thermodynamic and kinetic parameters was not considered by researchers at that time; this issue was treated much later. Another important concept related to crystal-liquid interfaces

| ADHESIVE TYPE | TWO-DIMENSIONAL NUCLEATION GROWTH | SPIRAL GROWTH |

$$R = A\,(\Delta\mu/kT)$$

$$R = A\,\exp\,(-B/\Delta\mu/kT)$$

$$R = A\,(\Delta\mu/kT)^2$$

ROUGH   SMOOTH

Thermodynamic roughening
Kinetic roughening
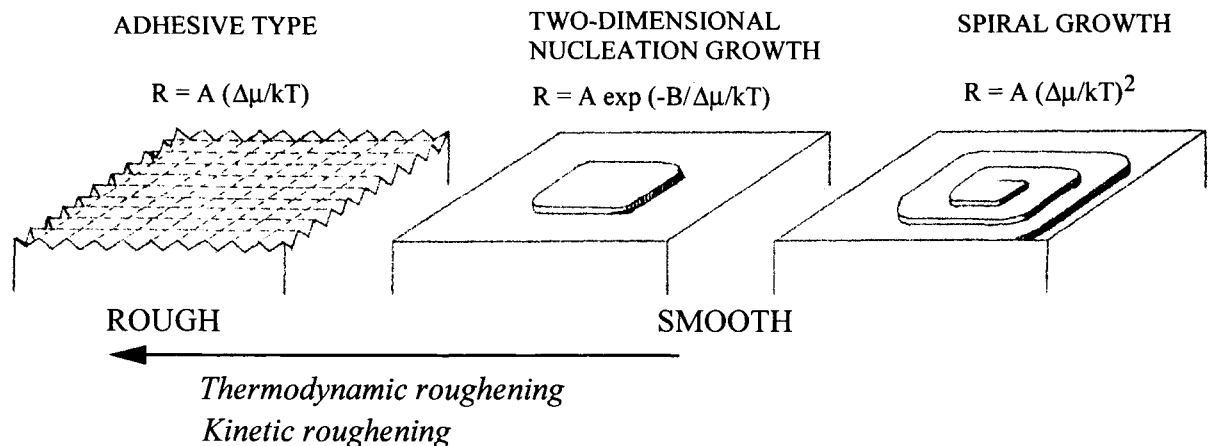
Figure 6. Rough (incomplete) and smooth (complete) interfaces, and growth mechanisms. Growth rate R vs. driving force $\Delta\mu/kT$ relations for respective mechanisms are also indicated. A smooth interface will transform into a rough interface with increase in temperature (thermodynamic roughening transition) or increase in driving force (kinetic roughening transition).
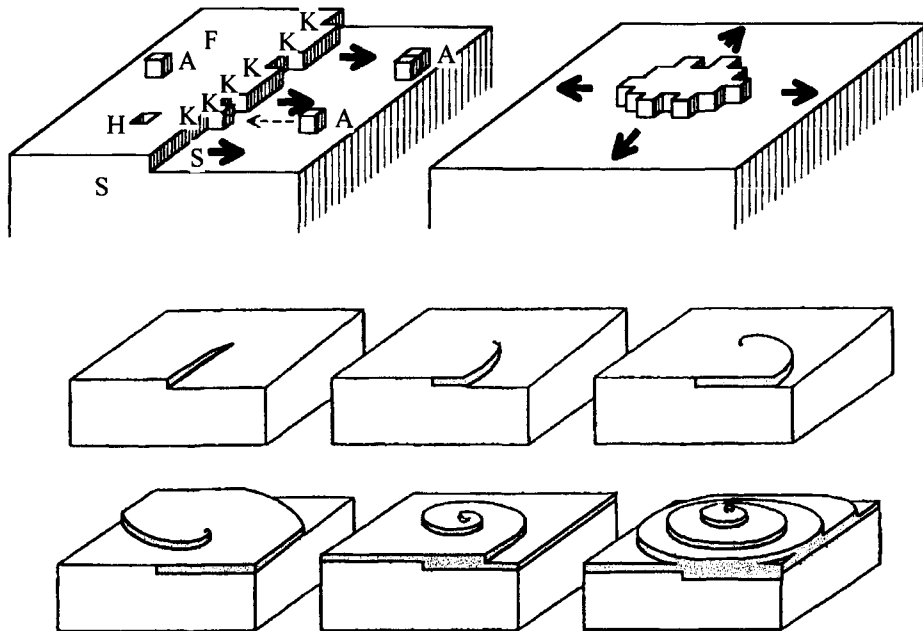
Figure 7. Two-dimensional nucleation growth (upper two) vs. spiral growth (lower six) mechanisms.

is morphological instability, which was dealt with much later in a quite elegant theory proposed by Mullins and Sekerka [11]. This concept enabled a deeper understanding of the mechanisms by which dendritic morphology and cellular textures are formed.

5. LATTICE DEFECTS

Around the same time as the K-S model was discussed, the concept of lattice defects, i.e., point, line and planar defects, was also introduced, and the physical properties of crystalline solids were studied in terms of intrinsic and structure-sensitive properties. It was realized that real crystals were not ideal, i.e., they were neither perfect nor homogeneous, and that they contained various types of lattice defects that greatly affect the physical properties of crystals.

The main concern of solid state physics was to understand the nature and behavior of lattice defects and their relations to structure-sensitive properties of crystals. This led to the development of semiconductor and optoelectronic industries in which single crystals were abondantly used. However, the importance of lattice defects in crystal growth was not realized for nearly 20 years until Frank proposed spiral growth theory in 1949 [12]

Frank's spiral growth theory was soon proved to be true by observations of growth spirals. The first observation of growth spiral steps with step heights of unit cell size was made on prism faces of natural beryl crystals, and numerous observations of growth spirals on a wide variety of crystals followed. These observations showed that the morphology of

growth spiral steps clearly reflects internal structural properties and crystal growth conditions and provide valuable information for assessing internal structural properties and external growth conditions of as-grown crystals. Therefore it became clear that observations of growth steps appearing on a crystal surface provide more fruitful information than do observations of the morphology of bulk crystals [13]. This was possible, because of the development at that time of new optical methods for observing surface steps with heights of unit cell or molecular scale, such as phase contrast microscopy and multiple-beam interferometry.

Nowadays, there are various techniques that enable *in situ* observation of the process of spiral growth with nanometer heights during crystal growth processes from vapor and solution phases. In the case of optical microscopy, phase shifting inter-ferometry enables *in situ* measurements of the spreading of elemental growth spiral steps, in combination with the measurement of supersaturation values of the solution very close to the growing interface. In the case of micro-probe techniques, AFM and STM are widely used. These two methods enable surface structures of crystals to be visualized with resolution accuracy as high as that obtained by optical methods along the vertical direction (normal to the crystal surface), and with very high resolution accuracy along the lateral direction. Essentially similar growth step patterns observed by optical microscopy on crystal faces in the earlier days have been observed by AFM and STEM. Use of the latter methods has revealed the occurrence of two-dimensional islands on the terraces of spiral steps, which were not seen in the earlier optical observations.
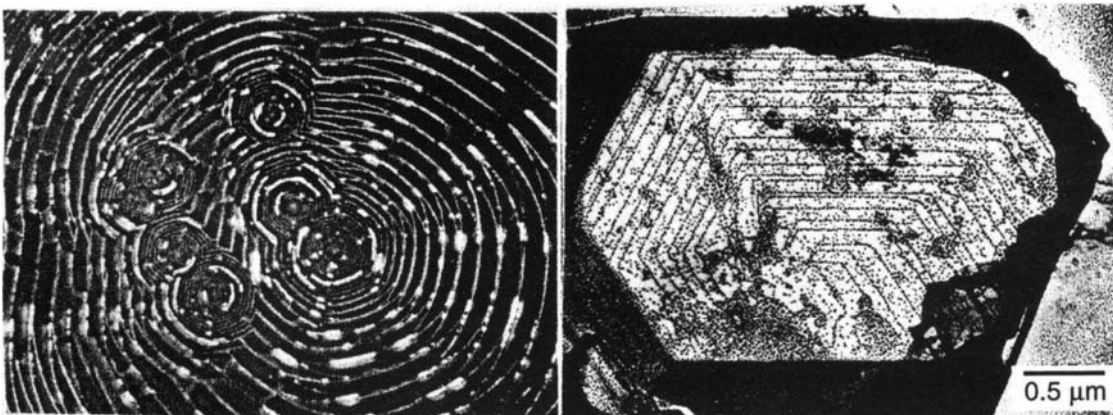
Figure 8. Growth spirals. Phase contrast (left; SiC, (0001) face) and TEM (right; kaolinite (001) face; decoration technique) photographs. Step heights of both spirals are unit cell heights. Also noticed is an interlacing pattern seen in SiC, which is due to zigzag stacking in the unit cell, and closer step separation at the spiral center, which is inferred to supersaturation shoot up at the final stage of growth. Morphology of growth spirals reflects the character of internal structure and changes in growth parameters more clearly than does the morphology of bulk crystals.
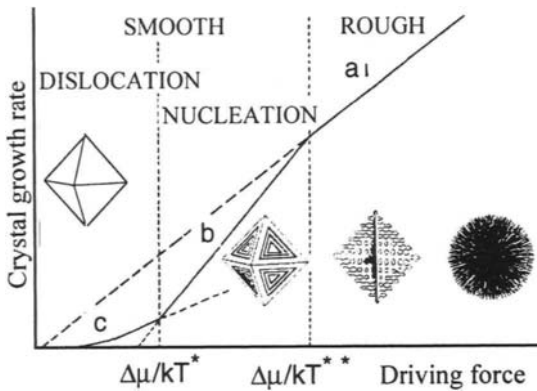
Figure 9. Changes in growth morphology of a crystal are now understood, as shown in this figure, on the basis of interface roughness and crystal growth mechanisms. "a" indicates the relationship of growth rate versus driving force for adhesive-type growth mechanism, "b" for two-dimensional nucleation growth mechanism, and "c" for spiral growth mechanism. Spherulite is the morphology of polycrystalline aggregate, others are for those of single crystal.

Whatever observation techniques are employed, it should be noted that surface microtopographic observation of crystal faces at an atomic or molecular level shows morphological features that are closely related to crystal growth processes

6. SYNETHESIS OF DISLOCATION-FREE CRYSTALS

To grow single crystals of high perfection and homogeneity that are useful for industrial purposes, a complete understanding of the atomic process of crystal growth is needed to assess how the perfection and homogeneity of crystals are controlled. As explained above, this task was first attempted by

metallurgists, since they had basic knowledge on texture formation in metal and alloy crystals: i.e., the issue of morphology involved in metallurgy. The necessity to grow large single crystals that had better controlled structural perfection and chemical homogeneity was further recognized by crystal growers dealing with optic, optoelectronic, semiconductor and piezoelectric crystals.

As an example, we may recall the case of ruby crystals. When the perfection of ruby crystals was not well controlled, ruby crystals did not have sufficient functional capability for them to be employed as stable and high-power laser-emitting devices. A laser beam emitted from such an imperfect ruby crystal was a "match flame". Only after the successful growth of single ruby crystals having high crystal perfection (low dislocation density) was a ruby laser put into practical use. This situation in the case of Si crystals was more dramatic. Numerous efforts were made to find ways for controlling the perfection and homogeneity of Si crystals. The growth of large dislocation-free single crystals had been one of the main targets in research on Si crystal growth, and it was finally achieved in the late 1950s.

In general, it is important to know how lattice defects and chemical inhomogeneity are introduced in crystalline materials during and after growth of crystals. For this purpose, information that relates mass and heat transports occurring during crystal growth to temperature distribution of a liquid phase near the growing crystal interface is essential to obtain a stable crystal/liquid interface and homogeneous impurity distribution. Stoichiometry

control is of highest importance in the case of growth of compound semiconductors such as GaAs. Mostly because of its high impact in the semiconductor industry, it should be stressed that crystal growth of dislocation-free semiconductor single crystals has been one of the most important driving forces in the development of the areas of crystal growth research.

In this respect, it is worth noting some historic remarks on international meetings/conferences held in the crystal growth community. Since the late 1940s, scientific and technological areas in crystal growth research started to have close co-operation, by holding international meetings to exchange knowledge on both fundamental aspects and applications of crystal growth. Faraday Society Discussion No. 5 (1949, United Kingdom), Cooperstown Conference on Growth and Perfection of Crystals (1958, U.S.A.), Nancy International Colloquium on Adsorption on Growing Crystals (1965, France) were some of the earlier events organized along this line,which eventually led to the 1st International Conference on Crystal Growth (ICCG-1) held in Boston, U.S.A. in 1966 and the publication of Journal of Crystal Growth. The ICCG has continued to be held every three years, and the 13th ICCG will be held in Kyoto, Japan in July 2001. Cooperation and mutual benefits among the researchers working on fundamental aspects and applications of crystal growth have formed the core philosophy of the ICCG meetings, and international meetings on specific topics, such as vapor growth and epitaxy, GaAs, II-VI compounds and biological macromolecules have also been held.

7. VAPOR GROWTH AND EPITAXY

As for crystal growth from the vapor phase, recent focus has been on epitaxial growth of thin films that enables fabrication of semiconductor devices as well as the growth of single bulk crystals.

The concept of epitaxy originated from observations of mineral parageneses. A wide variety of mineral parageneses with definite crystallographic relations between host and guest crystals were known in the mineral kingdom. It was Royer [14] who systematically investigated epitaxial relations in mineral crystals and proposed the concept of misfit ratio. Since then, extensive works have been carried out both on mechanisms of epitaxial growth and its applications, and the concepts of epitaxial temperature, surface cleanness, misfit dislocation and homo- and hetero-epitaxy, formation of a transition layer in hetero-epitaxy have emerged.

Three fundamental epitaxial growth mechanisms have been proposed: Frank-van der Merve, Stranski-Krastanov and Volmer-Weber mechanisms, which a re classified in terms of interface energy between a newly growing crystal (called a guest) and a substrate (called a host) and driving force [15]. Epitaxial relations were defined for crystal growth not only from the vapor phase but also from the liquid phase. It is well known that epitaxy, which has its liguistic roots in curiosity, has created a new industry created by innovation of semiconductor devices. Various types of semiconductor devices made of epitaxially grown thin film crystals are used in almost all computer-controlled equipments in our daily life.

As in the case of the growth of bulk single crystals, the results of morphological investigations of epitaxial layers provided useful information relevant to growth mechanisms and perfection and homogeneity control of epitaxially grown crystalline films.

Frank-van der Merve

Stranski-Krastanov

Volmer-Weber

(a)   (b)
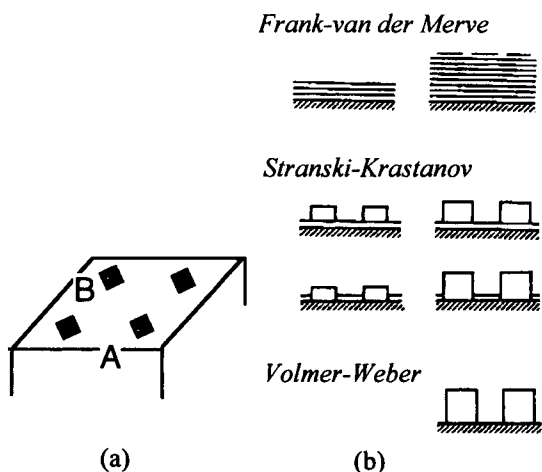
Figure 10.   (a) Epitaxial relation between host (A) and guest(B); (b) Three modes of epitaxial growth.

Epitaxial growth relationships between the host and guest combination in inorganic substances interacting through covalent and/or ionic bonds were well documented in the early stages of research. Further advances in epitaxial growth research revealed epitaxial growth occurring between crystalline materials interacting through more complicated atomic/molecular forces, such as metal (metallic bond) and inorganics (ionic and covalent bonds), inorganics and semiconductors (covalent bond), organics (van der Waals force) and inorganics. As will be discussed later, it will be possible to observe epitaxial growth of inorganic guest crystals on polymer or protein molecules.

8. COMPUTER SIMULATION

Since the 1970s, computer simulation has been extensively applied to crystal growth research. Typical subjects to which computer simulation using various techniques has been applied are as follows:

1)   Atomic processes of step propagation with and without impurities
2)   Thermodynamic and kinetic roughening transition of crystal-vapor and crystal-liquid interfaces
3)   Relations between growth kinetics and interface roughness
4)   Dendritic growth and pattern formation
5)   Convection and diffusion in melt and solution

Problems of "growth" forms have also been resolved by taking both internal and external factors into account, thanks to computer simulation.

9. BIOMINERALIZATION AND BIOCRYSTALLIZATION

As explained in a previous section, epitaxial relations have been observed in polymer and protein guest crystals on inorganic host crystals.   As a further extension along this line, an epitaxial relation of inorganic guest crystals growing on polymer or protein host crystals could be expected.   This would lead to further advances in the understanding of molecular-level mechanisms in biomineralization and even to the origin of life.   It has been indicated that, in biomineralization, growth of inorganic crystals takes place under precise control of proteins

involving direct epitaxial growth of inorganic crystals such as $CaCO_3$ at protein molecule surfaces, as will be elaborated below.

Inorganic and organic crystals are formed in various organs and cells through living activities. Crystal growth of this kind is driven either by the necessity for life, as revealed in the formation of apatite (bone and teeth), calcite and aragonite (pearls), magnetite (brain cell of pigeon, *algae*) crystals or in the excretion process of unnecessary substances for life, such as various calculi in many organs and wavellite and opal in plant tissues. There have been many observations of these biological inorganic and organic crystals since the beginning of the 19th century, and quite interesting morphologies, such as trowel-like calcite twins in cocolith have been observed.

An interesting point in the biomineral crystals that are formed due to the necessity for life is that the size and morphology of individual crystals are well controlled and quite different depending on the biological species. This may depend on the nature of relevant proteins that play key roles in nucleation and growth of crystals (e.g., apatite crystals in enamel and dentine in teeth). The texture of biological tissues that contain biomineral crystals is well-controlled, indicating that there is cooperation between protein molecules and inorganic crystals.

A series of international conferences on biomineralization have been held, but most participants are not specialists in the field of crystal growth. Among crystal growth researchers, the main interests in crystallization of biological molecules has been focused on protein crystals with

NORMAL

ABNORMAL

Aragonite, Pearl

Na-oxalate, Begonia

Inulin, Dahlia
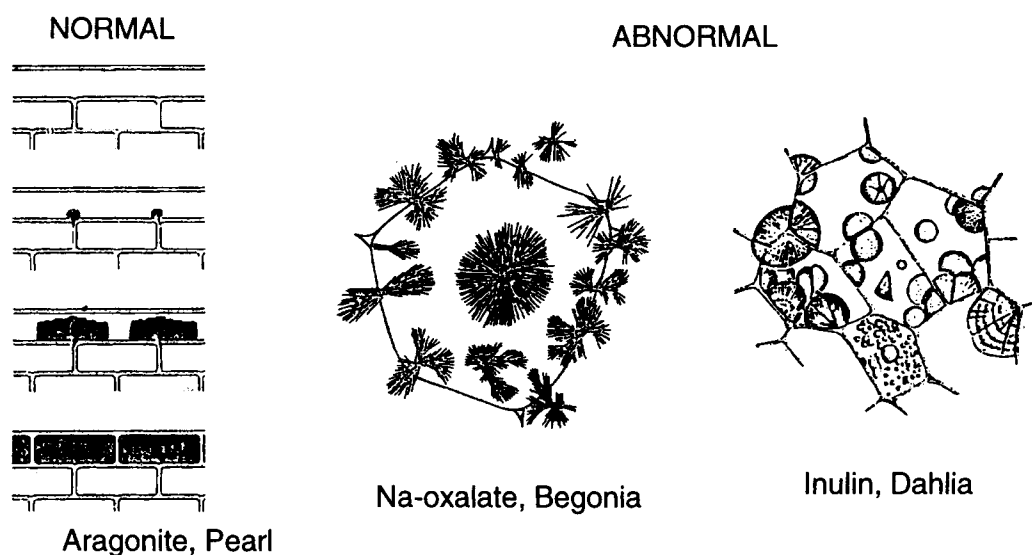
Figure 11. Inorganic and organic crystals formed by normal (left) biomineralization due to the necessity for living, and abnormal biomineralization (right) due to the process of excretion of unnecessary substance for living. (left) Pearl, a membrane film is formed in front, and aragonite crystals nucleate and grow within the film, resulting in controlled morphology, size and texture. (right) Na-oxalate in begonia and inulin in dahlia, both showing spherulitic and uncontrolled morphology.

14
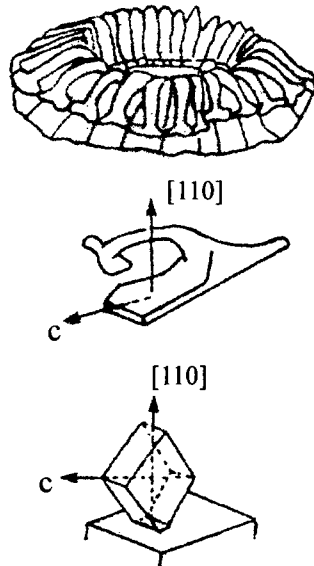
Figure 12. Exotic trowel-like calcite crystals in a cocolith, formed through its living activities and constituting its skeleton. One cocolith, ca. 4 μ m across, is composed of ca. 30 segments. A segment is a single crystalline calcite. Forms and sizes are all well controlled.

high perfection that are suitable for atomic-level structure determination by X-ray diffraction. Hence, the term biocrystallization has been used in this respect.

As a matter of fact, it has been proved that crystal growth mechanisms of inorganic and protein crystals are essentially the same. Recent *in-situ* AFM observations have revealed that protein crystal growth proceeds either through a spiral growth mechanism caused by screw dislocation or through a two-dimensional nucleation mechanism, both of which have been verified to occur in crystal growth of inorganic materials. However, further study is needed to elucidate many issues specifically inherent to protein crystal growth, such as growth unit size and its structure.

Cooperation between researchers in the fields of biomineralization and biocrystallization would be mutually beneficial, and the mechanism underlying the formation of well-controlled textures of biominerals would be better understood, through such cooperation, based on an understanding of the growth mechanism of single crystals.

10. CONCLUDING REMARKS (FUTURE TARGETS)

In the past 50 years, science and technology of crystal growth have greatly advanced, and the basic mechanisms of nucleation, growth and morphology control are now understood at atomic/molecular levels, enabling well-controlled synthesis of perfect and homogeneous bulk and thin film crystals, at least for mono-component single crystals grown in simple systems. This high level in the science and technology of crystal growth was reached through the cooperation of those researching both fundamental aspects of crystal growth and possible applications through theoretical and experimental approaches. We should keep this tradition for our future development.

The main contributions of crystal growth to modern science and technology can be summarized as follows:

(1) At least for growth of single crystals in simple systems, we now understand how and through what processes a cosmos (regular atomic arrangement) is formed from a chaotic phase at an atomic level, and how perfection, homogeneity and morphology of crystals are

controlled in the growth process. We can stress the importance of the kinetics process of crystal growth in modern science and technology, which require high-quality single crystals. Morphology control has been a core subject in these developments.

(2) We now have high-level technology to grow bulky single crystals and thin-film crystals with well-controlled perfection and homogeneity, which enables the production of various devices that support our affluent society.

With regard to future research, it should be noted that the above successes have been realized only for single crystals of simple and mono-component systems, such as Si. Further efforts are needed to study compound semiconductors and oxide crystals and more complex systems such as organic, polymer and protein crystals to control their perfection and
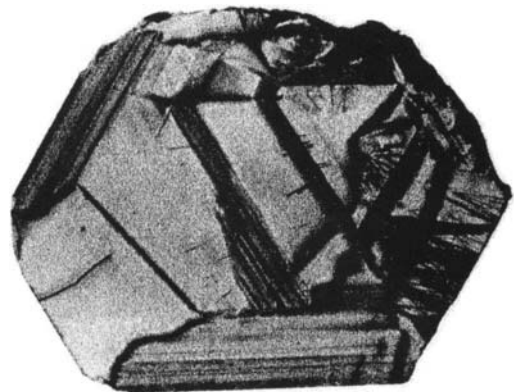
Figure 13. Growth process and history are recorded in forms such as growth sector, growth banding, spatial distribution of growth-induced dislocations, element partitioning. These are useful key codes to decipher "letters sent from the depth of the Earth". X-ray topograph of (0001) section of a rock crystal, taken by T. Yasuda

homogeneity at the same levels as those of, for example, Si. Controlling the homogeneity and perfection of these crystals would lead to the
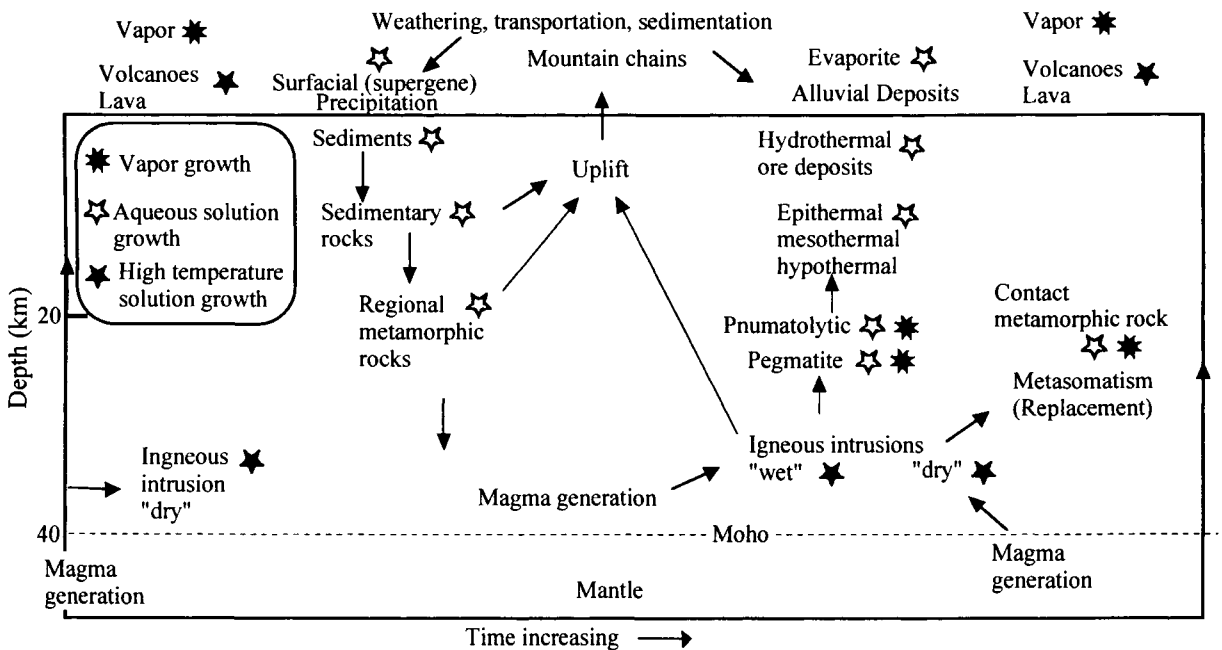
Figure 14. Schematic diagram showing crystal growth taking place in the Earth.

16

diversity of new materials and their applications.

In the past 50 years, the main concern was focused on single crystals, and texture formation has been largely neglected. Although texture formation in multi-component and complex systems is directly related to nucleation, growth and morphology of crystals, it has not been investigated as thoroughly as that of single crystal growth. This is therefore an important subject to be studied in near future.

Based on the understanding achieved in simple and mono-component systems, it would be possible to control the texture formation in complicated systems, such as ceramics, solid earth, planetary materials and biominerals. We will be able to understand the origins and the processes that occur in the formation of planets, earth and life. Those researches would lead us to discover the ways of utilizing wisdom of Nature for our life. Nature is still cleverer than us.

Another target for crystal growth research in the 21st century is to deepen our understanding of crystal growth mechanisms by taking quantum mechanics into account. Quantum mechanics has not been used in past crystal growth studies, since it has not be required for the elucidation of crystal growth processes so far. In the era of quantum well technology and atomic layer epitaxy, however, quantum mechanics will be essential for advanced research on crystal growth.

REFERENCES

1. N. Steno, De solido intra solidum naturaliter contento prodomus, Florence, 1669: English translation, J..W. Winter, The Prodromus of Nicolous Steno's dissertation concerning a solid body enclosed by process of nature within a solid, Hapner, New York, 1968

2. J.G. Burke, Origin of the Science of Crystals, Univ.California Press, Berkley and Los Angeles, 1966

3. J. Kepler, Strena seu de nive sexangula, Francofurti und Moem,Godfrey Tampach, Frankfurt,1611; English translation C.Hardie, "A new year's gift or on the six- cornered snowflakes, Oxford Univ. Press, Oxford

4. E. Haeckel, Kristallselen, Studies uber das Anorganische Leben, Alfred Kroner Verlag, Leipzig, 1917; English translation A. Mackay, Forma 14 (1999), 11

5. I. Sunagawa, Morphology of Crystal, Part A and B, Terra SciPub., Tokyo/ D.Reidel, Dordrecht, 1987

6. A. Papapetrov, Untersuchungen uber dendritisches Wachstum von Kristallen, Z. Krist., 92 (1935), 89-1

7. U. Nakaya, Snow Crystals, Natural and Artificial, Harvard Univ. Press, Cambridge, Mass., 1954

8. H.J. Scheel, Historical introduction, in Handbook of Crystal Growth, 1a, D.T.J. Hurle (Ed), North-Holland, 1-42., 1993

9. M. Volmer, Z. Phys. Chem., 102 (1923), 267,

8. W. Kossel, Nachur. Ges. Gottingen, 2 (1927), 135; I.N. Stranski, : Z. Phys. Chem., 136 (1928), 258

9. W.W. Mullins, and R.F. Sekerka, J. Appl. Phys., 34 (1963), 323

10. F.C. Frank, Disc. Faraday Soc., No.5 (1949), 48;

W.K. Burton, N. Cabrera,. and F.C.Frank, Phil. Trans., A243 (1951), 293

A.R. Verma, Crystal Growth and Dislocations, Butterworth, London.1953; W. Dekeyser, and S. Amelinckx, Les dislocations et la croissance des cristaux,Masson et Cie, Paris, 1955; I. Sunagawa and P. Bennema, P. : Morphology of growth spirals, Theoretical and experimental, in Preparation and Properties of Solid State Materials, 7, R.W. Wilcox,(Ed.), Marcel Dekker Inc., New York, 1982

14. L.. Royer, Bull. Soc. Fr. Min. Crist., 51 (1928), 7

15. A. A. Chernov, Modern Crystallography III, Springer-Verlag, Berlin, 1984

This Page Intentionally Left Blank

PART I

Understanding of Fundamental Aspects in Crystal Growth

This Page Intentionally Left Blank

Fundamentals of phase field theory

Robert F. Sekerka[a][*]

[a]University Professor, Physics and Mathematics, Carnegie Mellon University
Pittsburgh, Pennsylvania 15213, USA

Phase field theory provides an alternative method for solving dynamical problems involving crystallization from a melt. The sharp solid liquid interface of the classical model is replaced by a diffuse interface by introducing an auxiliary variable φ, the phase field, that indicates the phase. The variable φ varies from some value, say 0 in the solid, to another value, say 1 in the liquid. It changes continuously from 0 to 1 over a thin region of space, the diffuse interface. Equations for the time evolution of φ as well as those for temperature T and compositional fields ω (for alloys) are derived from postulated functionals for entropy, energy and chemical species. For a sufficiently thin diffuse interface, these equations account for the Gibbs-Thomson dependence of melting point on local interface curvature and also for linear interface attachment kinetics. They can be solved to generate complicated solidification patterns, such as occur during cellular and dendritic growth. They can also be generalized to include multicomponent diffusion, fluid convection, order-disorder transformations in crystals, and relative crystallographic orientation (grain boundaries) in polycrystalline materials.

1. INTRODUCTION

The phase field model [1–5] is based on a diffuse interface model [6] of crystallization from the melt. The model employs an auxiliary variable, the phase-field φ, that takes on constant values, e.g., 0 in the bulk solid and 1 in the bulk liquid phases, and varies continuously but rapidly over a thin spatial region, the diffuse interface. Its relationship to the classical sharp interface model has been studied by Caginalp [7–15] and others by means of asymptotics and distinguished limits. For sufficiently thin diffuse interfaces, it has been shown to incorporate the well-known Gibbs-Thomson boundary condition, according to which the melting point depends on the local curvature of the solid-liquid interface, even for anisotropic interfacial free energy. [16–18].

The phase field model can be derived on the basis of an entropy functional, first introduced to this problem by Penrose and Fife [19], or on the basis of continuum mechanics and a balance of microforces, as favored by Fried and Gurtin [20,21]. Wang et al. [22] introduced the concepts of *local* positive entropy production and nonclassical entropy fluxes. We shall follow this approach below in order to give a detailed derivation of the model for solidification from the melt of a rigid isotropic monocomponent material. The style in this section is tutorial. We will then use this information as background for discussion of more general models, but omitting many details of their derivation.

The phase field model has been shown to be useful for computations of rather complicated dendritic solidification patterns by the pioneering work of Kobayashi [23–25]. Since then, many others [26–35] have used the model successfully to study solidification morphologies. Karma and Rappel [36] have reexamined the relationship of the model to the classical sharp interface model and have shown that computational accuracy can be obtained for a thicker diffuse interface than previously used, provided that the kinetic coefficient is redefined in a new way. This even allows recovery of the local equilibrium limit. They have used this model to study dendritic growth and sidebranching [37,38].

The phase field model has been extended to study the solidification of alloys. Alloy phase field

models for rigid materials have been formulated by Wheeler-Boettinger et al. [39–42], Bi and Sekerka [43], and Charach and Fife [44,45]. They have been used to study solute trapping [41,46–51], dependence of surface tension on composition [50], isothermal dendritic growth and microsegregation [52–55], Ostwald ripening and coalescence [56], recalescence during dendritic solidification [57], and cell to plane front transitions during directional solidification [58]. Phase field models for solidification of eutectic alloys have also been developed [59–61].

Models that blend some hydrodynamics with the phase field model have been developed [62–67] and used for computations [68] to some extent. More recently, thermodynamically consistent diffuse interface models that include hydrodynamics have been formulated for a pure material and for a binary alloy by Anderson, Mc-Fadden and Wheeler [69,70]. They [71,72] also developed a phase field model, including convection and anisotropy, for solidification of a pure material and performed numerical computations for dendritic growth. Sekerka and Bi [73] have extended this model to multicomponent alloy solidification.

In the area of solid-solid transformations, Braun et al. [74–76] have used a phase field model with multiple order parameters to explore order-disorder phase boundaries and transformations in alloy crystals. Moreover, Kobayashi, Warren and Carter [77] have developed a phase field model involving two order parameters, expressed in terms of polar coordinates, that facilitate modeling of polycrystalline materials having grain boundaries.

2. MONOCOMPONENT SYSTEM

We proceed to derive in some detail a phase field model for crystallization (solidification) from the melt of a monocomponent material. To simplify the problem, we shall assume that the solid is isotropic, that both solid and liquid have the same constant density, and that convection in the melt is forbidden. Under these conditions, we postulate that the internal energy $\mathcal{U}$ and the entropy $\mathcal{S}$ in any subvolume $\mathcal{V}$ of our system are given by

$$\mathcal{U} := \int_{\mathcal{V}} [u + \frac{1}{2}\varepsilon_u^2 |\nabla\varphi|^2]\, d^3x \qquad (1)$$

and

$$\mathcal{S} := \int_{\mathcal{V}} [s(u, \varphi) - \frac{1}{2}\varepsilon_s^2 |\nabla\varphi|^2]\, d^3x \qquad (2)$$

where $u(\mathbf{r}, t)$ is the local density of internal energy, $\varphi(\mathbf{r}, t)$ is the phase field, $\mathbf{r}$ is the position vector, t is time, and ε_u^2 and ε_s^2 are constants. We regard these expressions to be *functionals* of $u(\mathbf{r}, t)$ and $\varphi(\mathbf{r}, t)$; in other words, $\mathcal{U}$ and $\mathcal{S}$ depend on functions, rather than just variables. The quantities u and s are internal energy densities that pertain to a homogeneous phase having a uniform value of φ. The terms involving $|\nabla\varphi|^2$ are corrections that are only important in the diffuse interface where φ changes from its value $\varphi = 0$ in bulk solid to its value $\varphi = 1$ in bulk liquid. The term $\frac{1}{2}\varepsilon_u^2 |\nabla\varphi|^2$ is called a gradient energy and $-\frac{1}{2}\varepsilon_s^2 |\nabla\varphi|^2$ is called a gradient entropy. They are analogous to the gradient free energy of Cahn-Hilliard theory [78] and go back to ideas of Rayleigh, van der Waals, Ginsburg and Landau. The signs of these terms are chosen for later convenience in accordance with thermodynamic principles that energy is minimized and entropy is maximized in an equilibrium state.

2.1. Equilibrium

We first investigate conditions for thermodynamic equilibrium of our system. We begin with Eqs. (1) and (2) but with $\mathcal{V}$ replaced by V, the volume of the entire system. We then proceed to maximize the entropy subject to the constraint of constant energy and no exchange with the exterior. To do this, we make small variations δu and $\delta\varphi$ throughout the system and require the variation $\delta\mathcal{S} = 0$, subject to the constraint $\delta\mathcal{U} = 0$. The constraint is handled by means of a Lagrange multiplier λ, so we require

$$\delta[\mathcal{S} - \lambda\mathcal{U}] = 0. \qquad (3)$$

We calculate the variations of $\mathcal{S}$ and $\mathcal{U}$ according to the calculus of variations. Thus

$$\delta\mathcal{S} = \int_V \left[\frac{\partial s}{\partial u}\delta u + \frac{\partial s}{\partial \varphi}\delta\varphi - \varepsilon_s^2(\nabla\varphi) \cdot \delta(\nabla\varphi)\right] d^3x$$

$$= \int_V \left\{ \frac{\partial s}{\partial u} \delta u + \left[\frac{\partial s}{\partial \varphi} + \varepsilon_s^2 \nabla^2 \varphi \right] \delta \varphi \right\} d^3 x$$
$$- \int_A \varepsilon_s^2 \delta \varphi \nabla \varphi \cdot \hat{n} \, d^2 x \qquad (4)$$

where we have used the divergence theorem to integrate by parts after using $\delta(\nabla \varphi) = \nabla(\delta \varphi)$. Here, A is the bounding area of V and $\hat{n}$ is its unit outward normal. The coefficients of δu and $\delta \varphi$ in the volume integral are called *functional derivatives*. Similarly,

$$\delta \mathcal{U} = \int_V [\delta u - \varepsilon_u^2 \nabla^2 \varphi] \, d^3 x + \int_A \varepsilon_u^2 \delta \varphi \nabla \varphi \cdot \hat{n} \, d^2 x. (5)$$

Hence

$$\delta[\mathcal{S} - \lambda \mathcal{U}] = \int_V \left[\frac{\partial s}{\partial u} - \lambda \right] \delta u \, d^3 x$$
$$+ \int_V \left[\frac{\partial s}{\partial \varphi} + (\varepsilon_s^2 + \lambda \varepsilon_u^2) \nabla^2 \varphi \right] \delta \varphi \, d^3 x$$
$$- \int_A (\varepsilon_s^2 + \lambda \varepsilon_u^2) \delta \varphi \nabla \varphi \cdot \hat{n} \, d^2 x. \qquad (6)$$

We assume that the integral over A can be made to vanish by application of suitable boundary conditions, such as vanishing of $\delta \varphi$ or $\nabla \varphi \cdot \hat{n}$ on A. Then for independent variations δu and $\delta \varphi$ within the volume, Eq. (3) requires the *integrands* in Eq. (6) to vanish. This yields[2]

$$\frac{1}{T} := \left(\frac{\partial s}{\partial u} \right)_\varphi = \lambda \qquad (7)$$

and

$$\left(\frac{\partial s}{\partial \varphi} \right)_u + (\varepsilon_s^2 + \lambda \varepsilon_u^2) \nabla^2 \varphi = 0. \qquad (8)$$

Eq. (7) is based on the fundamental differential

$$ds = \frac{1}{T} du + \left(\frac{\partial s}{\partial \varphi} \right)_u d\varphi \qquad (9)$$

where T is the absolute temperature. Thus the temperature is uniform throughout the system at equilibrium, in agreement with our intuition. Eq. (8) therefore takes the form

$$\left(\frac{\partial s}{\partial \varphi} \right)_u + \frac{\varepsilon_f^2(T)}{T} \nabla^2 \varphi = 0 \qquad (10)$$

---

where $\varepsilon_f^2(T) := \varepsilon_s^2 + T\varepsilon_s^2$. The remaining derivative can be rewritten by defining a Helmholtz free energy density $f = u - Ts$ so that

$$df = du - T \, ds - s \, dT = -s \, dT - T \left(\frac{\partial s}{\partial \varphi} \right)_u \qquad (11)$$

where Eq. (9) has been used. It follows that

$$\left(\frac{\partial s}{\partial \varphi} \right)_u = -\frac{1}{T} \left(\frac{\partial f}{\partial \varphi} \right)_T \qquad (12)$$

which allows Eq. (10) to be written

$$\left(\frac{\partial f}{\partial \varphi} \right)_T - \varepsilon_f^2 \nabla^2 \varphi = 0. \qquad (13)$$

For constant T, Eq. (13) can be obtained directly by defining a Helmholtz free energy functional

$$\mathcal{F} := \mathcal{U} - T\mathcal{S} = \int_\mathcal{V} \left[f(T, \varphi) + \frac{1}{2} \varepsilon_f^2 |\nabla \varphi|^2 \right] d^3 x. (14)$$

By means of the same calculus used to obtain Eq. (4) and Eq. (5), the expression on the left of Eq. (13) is recognized to be the functional derivative of $\mathcal{F}$ with respect to φ. Eq. (13) follows because the Helmholtz free energy is a minimum for equilibrium at constant T.

2.2. Choices of u, s and f

We can gain insight about these equilibrium conditions by making specific choices of the functions u, s and f. With T and φ as independent variables, we choose

$$u(T, \varphi) = [1 - p(\varphi)] u_S(T) + p(\varphi) u_L(T)$$
$$+ \frac{W_u}{2} g(\varphi) \qquad (15)$$

where $u_S(T)$ and $u_L(T)$ are the classical internal internal energy densities of solid S and liquid L and $p(\varphi)$ is a monotonically increasing function of φ such that $p(0) = 0$ and $p(1) = 1$, W_u is a positive constant and

$$g(\varphi) = \varphi^2 (1 - \varphi)^2. \qquad (16)$$

The quantity $(W_u/2)g(\varphi)$ is a 'double well' potential whose form is chosen to increase the energy of states for which $0 < \varphi < 1$ relative to the bulk solid ($\varphi = 0$) and the bulk liquid ($\varphi = 1$). One

could conceivably choose $p(\varphi)$ to be φ itself but we shall make a more judicious choice later (see Eq. (27)).

We substitute Eq. (15) into Eq. (9) and integrate at constant φ to obtain

$$s(T,\varphi) = [1-p(\varphi)]\int_0^T \frac{c_S(T')}{T'}dT' \qquad (17)$$
$$+ \; p(\varphi)\int_0^T \frac{c_L(T')}{T'}dT' - \frac{W_s}{2}g(\varphi)$$

where $c_S(T) = du_S(T)/dT$ and $c_L(T) = du_L(T)/dT$ are heat capacities of solid and liquid, respectively, and W_s is a positive constant. The term $-(W_s/2)g(\varphi)$ appears here as a 'constant' of integration which depends on φ, which is held constant during the integration. We assume that it has the form of an inverted double well potential, consistent with the notion that states other than bulk solid and liquid will have lower entropies. Since $g(0) = g(1) = 0$, Eq. (17) is consistent with the third law of thermodynamics which states that $s = 0$ at $T = 0$, independent of phase, provided there is internal equilibrium.

The corresponding Helmholtz free energy is therefore

$$f(T,\varphi) = [1-p(\varphi)]f_s(T) \qquad (18)$$
$$+ \; p(\varphi)f_L(T) + \frac{W_f(T)}{2}g(\varphi)$$

where

$$f_s(T) = u_S(T) - T\int_0^T \frac{c_S(T')}{T'}dT' \qquad (19)$$

and

$$f_L(T) = u_L(T) - T\int_0^T \frac{c_L(T')}{T'}dT' \qquad (20)$$

are Helmholtz free energies of solid and liquid, respectively, and $W_f(T) = W_u + TW_s$.

We are interested in temperatures in the vicinity of the melting point $T = T_M$ and would also like to avoid integration from $T = 0$. We therefore note that

$$\frac{\partial(f/T)}{\partial(1/T)} = f + \frac{\partial f}{\partial T}\frac{dT}{d(1/T)} = f + Ts = u \qquad (21)$$

which can be integrated to give

$$\frac{f_L(T) - f_s(T)}{T} = -\int_{T_M}^T \frac{L(T')}{(T')^2}dT' \qquad (22)$$

where $L(T) = u_L(T) - u_S(T)$. In deriving Eq. (22) we have used the fact[3] that $f_L(T_M) - f_S(T_M) = 0$ by definition of T_M. Hence Eq. (18) becomes

$$f(T,\varphi) = f_s(T) - p(\varphi)T\int_{T_M}^T \frac{L(T')}{(T')^2}dT'$$
$$+ \; \frac{W_f(T)}{2}g(\varphi). \qquad (23)$$

In particular,

$$f(T_M,\varphi) = f_S(T_M) + \frac{W_M}{2}g(\varphi) \qquad (24)$$

where $W_M := W_f(T_M) = W_u + T_M W_s$ is a constant. Partial differentiation of Eq. (23) with respect to T yields

$$s(T,\varphi) = s_S(T) + p(\varphi)\int_{T_M}^T \frac{L(T')}{(T')^2}dT'$$
$$+ \; p(\varphi)\frac{L(T)}{T} - \frac{W_u}{2}g(\varphi). \qquad (25)$$

At the melting point, Eq. (25) becomes simply

$$s(T_M,\varphi) = s_S(T_M) + p(\varphi)\frac{L_0}{T_M} - \frac{W_u}{2}g(\varphi) \qquad (26)$$

where $L_0 = L(T_M)$ is the latent heat of fusion.

We complete these functions by choosing

$$p(\varphi) = \frac{\int_0^\varphi g(y)\,dy}{\int_0^1 g(y)\,dy} = \varphi^3(10 - 15\varphi + 6\varphi^2) \qquad (27)$$

which varies smoothly and monotonically from 0 to 1 as φ varies from 0 to 1. It also has the property that its derivative $p'(\varphi) := dp/d\varphi = 30g(\varphi)$, so $p'(0) = p'(1) = 0$ and $p''(0) = p''(1) = 0$. This guarantees that the function $f(T,\varphi)$ has minima at exactly $\varphi = 0$ and $\varphi = 1$ for all positive values of T. Other choices of $p(\varphi)$ are possible, as discussed in [22].

To illustrate Eq. (23) and Eq. (25), one can make the often-used approximation that $L(T)$ is independent of T and therefore equal to L_0. Then

---

[3] Our phases have been assumed to have the same density, so the equality of f at the melting point is equivalent to the equality of chemical potential at uniform temperature and pressure.
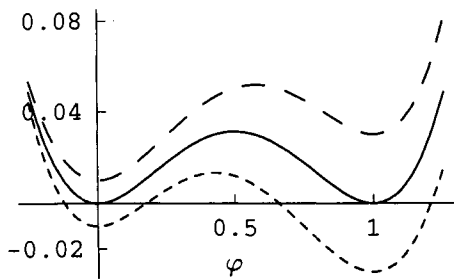
Figure 1. Plot of the biased double well potential $[f_S(T,\varphi) - f_S(T_M)]/L_0$ according to Eq. (28) for $T = T_M$ (full line), $T = 0.98T_M$ (large dashes) and $T = 1.02T_M$ (small dashes). The phase having the lower value of f is stable, and at T_M both phases are stable and can coexist.

we obtain

$$
\begin{aligned}
f(T,\varphi) &= f_S(T) - p(\varphi)L_0 \frac{T - T_M}{T_M} \\
&\quad + \frac{W_f(T)}{2} g(\varphi)
\end{aligned}
\tag{28}
$$

and

$$
s(T,\varphi) = s_S(T) + p(\varphi)\frac{L_0}{T_M} - \frac{W_s}{2} g(\varphi).
\tag{29}
$$

Eq. (28) has the form of a biased double well potential, as illustrated in Figure 1. For $T > T_M$, the liquid has the lower free energy and is the stable phase. Similarly, the solid is the stable phase for $T < T_M$. For $T = T_M$, bulk solid and liquid can coexist in equilibrium, which we now explore.

2.3. Solid-liquid coexistence

We can gain understanding of Eq. (13) by considering a one dimensional problem in a single spatial variable x in which there is a semi-infinite solid in the region $x < 0$, a semi-infinite liquid in the region $x > 0$ and a diffuse interface near $x = 0$. This will require $T = T_M$. Eq. (13) becomes

$$
\frac{df(T_M,\varphi)}{d\varphi} = \varepsilon_M^2 \frac{d^2\varphi}{dx^2}.
\tag{30}
$$

We multiply Eq. (30) by $d\varphi/dx$ to obtain

$$
\frac{df(T_M,\varphi)}{dx} = \frac{\varepsilon_M^2}{2} \frac{d}{dx}\left(\frac{d\varphi}{dx}\right)^2.
\tag{31}
$$

Integration on x then leads to

$$
f(T_M,\varphi) = f_S(T_M) + \frac{\varepsilon_M^2}{2}\left(\frac{d\varphi}{dx}\right)^2
\tag{32}
$$

where the constant of integration $f_S(T_M) = f_L(T_M)$ is chosen by noting that $d\varphi/dx = 0$ for x far from the diffuse interface. Substituting Eq. (24) into Eq. (32) gives

$$
\frac{W_M}{2}g(\varphi) = \frac{\varepsilon_M^2}{2}\left(\frac{d\varphi}{dx}\right)^2.
\tag{33}
$$

Taking the square root of Eq. (33) we obtain

$$
\varphi(1 - \varphi) = \ell \frac{d\varphi}{dx}
\tag{34}
$$

where

$$
\ell = \frac{\varepsilon_M}{\sqrt{W_M}}.
\tag{35}
$$

Separation of variables and integration of Eq. (34) subject to the condition $\varphi = 1/2$ when $x = 0$ yields $(x/\ell) = \ln[\varphi/(1 - \varphi)]$ which can be exponentiated and rearranged to give

$$
\varphi = \frac{1}{2} + \frac{1}{2}\tanh(x/2\ell).
\tag{36}
$$

Here, $\tanh(y) = (e^y - e^{-y})/(e^y + e^{-y})$ is the hyperbolic tangent function. Eq. (36) is illustrated in Figure 2. The thickness of the diffuse interface[4] is of the order of ℓ; more precisely, φ varies from about 0.05 to 0.95 over a distance of 6ℓ as x varies from -3ℓ to 3ℓ.

2.4. Solid-liquid interfacial free energy

For our one-dimensional problem, we can identify the solid-liquid interfacial free energy γ (sometimes called the surface tension) as an excess of the functional $\mathcal{F}$ per unit cross sectional

---

[4]In the somewhat simpler model of Wang et al. [22], $\varepsilon_u = 0$, $W_u = 0$, $\varepsilon_M = \epsilon\sqrt{T_M}$ and $W_s = 1/2a$ which gives $\ell = \epsilon\sqrt{2a} = \sqrt{2}\delta$ of that paper. Their interface thickness $\eta = 6\sqrt{2}\delta = 6\ell$.
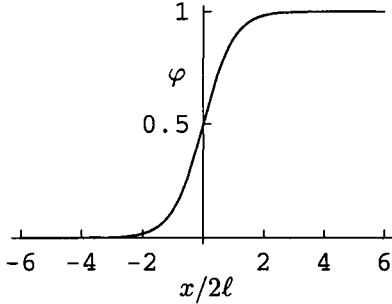
Figure 2. Plot of φ versus x/ℓ according to Eq. (24). φ varies from about 0.05 to 0.95 over a distance of 6ℓ, which can be considered to be the thickness of the diffuse interface.

area. Adapting Eq. (14) to this case gives

$$\gamma = \int_{-\infty}^{\infty} [f(T_M, \varphi) + \frac{\varepsilon_M^2}{2}\left(\frac{d\varphi}{dx}\right)^2 - f_S(T_M)] \, dx$$

$$= \int_{-\infty}^{\infty} \varepsilon_M^2 \left(\frac{d\varphi}{dx}\right)^2 \, dx \qquad (37)$$

where Eq. (32) has been substituted to obtain the second line. The integral can be done by converting to an integral over φ and substituting Eq. (34) to give

$$\int_{-\infty}^{\infty} \varepsilon_M^2 \left(\frac{d\varphi}{dx}\right)^2 \, dx = W_M \ell \int_0^1 \varphi(1-\varphi) \, d\varphi. \quad (38)$$

Thus we obtain[5]

$$\gamma = \frac{W_M \ell}{6} = \frac{\varepsilon_M \sqrt{W_M}}{6}. \qquad (39)$$

Since γ is an observable physical quantity, it is useful to solve for the parameters of the phase field model in terms of γ and ℓ. The result is

$$\varepsilon_M^2 = 6\gamma\ell; \qquad W_M = 6\frac{\gamma}{\ell}. \qquad (40)$$

The 'barrier height' (see Figure 1)

$$f_B := \frac{W_M}{32} = \frac{3\gamma}{16\ell} \qquad (41)$$

---

[5]For the simpler model of Wang et al. [22] this becomes $\gamma = \varepsilon_s T_M \sqrt{2}/(12\sqrt{a})$.

of the double well potential for f, which occurs at $\varphi = 1/2$ and $T = T_M$, is perhaps more physically meaningful.

2.5. Gibbs-Thomson equation

We can demonstrate in a simple way that the phase field model contains the Gibbs-Thomson equation, according to which the melting point depends on the local curvature of the solid-liquid interface. We reconsider Eq. (13) for a problem having spherical symmetry such that φ is a function of only the spherical radius r. We substitute Eq. (23) into Eq. (13) and use the spherical Laplacian to obtain

$$\frac{W_f(T)}{2} g'(\varphi) - p'(\varphi) T \int_{T_M}^{T} \frac{L(T')}{(T')^2} dT'$$

$$= \varepsilon_f^2(T) \left[\frac{d^2\varphi}{dr^2} + \frac{2}{r}\frac{d\varphi}{dr}\right]. \qquad (42)$$

We solve Eq. (42) approximately for a spherically symmetric diffuse interface located near $r = R$, with solid inside and liquid outside. The temperature will still be uniform but will shift from the melting point T_M for a planar interface by a small amount. We use a perturbation expansion of the form

$$\begin{pmatrix} \varphi \\ T \\ r \end{pmatrix} = \begin{cases} \varphi_0 + \epsilon\varphi_1 \\ T_M(1 + \epsilon T_1) \\ R(1 + \epsilon\rho) = R + \ell\rho \end{cases} \qquad (43)$$

where $\epsilon = \ell/R$ is a small quantity, T_1 is a dimensionless constant to be determined and ρ is a dimensionless coordinate. We multiply Eq. (42) by $2/W_M$ and expand to first order in ε. The left hand side becomes

$$g'(\varphi_0) + \epsilon \left[g''(\varphi_0)\varphi_1 - p'(\varphi_0)\frac{2L_0}{W_M}T_1 \right.$$
$$\left. + \frac{W_u T_M}{W_M} g'(\varphi_0) T_1\right] \qquad (44)$$

and the right hand side becomes

$$2\frac{d^2\varphi_0}{d\rho^2} + \epsilon\left[2\frac{d^2\varphi_1}{d\rho^2} + 4\frac{d\varphi_0}{d\rho} + \frac{d^2\varphi_0}{d\rho^2}\frac{2T_M\varepsilon_s^2}{\varepsilon_M^2}T_1\right]. (45)$$

We equate the terms that are independent of ϵ to obtain the leading order problem

$$g'(\varphi_0) - 2\frac{d^2\varphi_0}{d\rho^2} = 0. \qquad (46)$$

Eq. (46) is just a dimensionless form of Eq. (30) and has a solution

$$\varphi = \frac{1}{2} + \frac{1}{2}\tanh\left(\frac{\rho}{2}\right) \tag{47}$$

in agreement with Eq. (36). Then we equate the terms in Eq. (44) and Eq. (45) that contain ϵ to get the first order problem

$$g'(\varphi_0)\varphi_1 - 2\frac{d^2\varphi_1}{d\rho^2} = 4\frac{d\varphi_0}{d\rho} + \frac{2L_0}{W_M}p'(\varphi_0)T_1$$
$$+ \left[\frac{2T_M\varepsilon_s^2}{\varepsilon_M^2}\frac{d^2\varphi_0}{d\rho^2} - \frac{W_u T_M}{W_M}g'(\varphi_0)\right]T_1. \tag{48}$$

Differentiation of Eq. (46) with respect to ρ shows that $\varphi_1 = d\varphi_0/d\rho$ is a solution to Eq. (48) with the right hand side equal to zero, i.e., it satisfies the homogeneous equation. In order for Eq. (48) to have a particular solution, it is necessary for the right hand side to be 'orthogonal' to this homogeneous equation, a well-known solvability condition. Orthogonality means that if both sides of the equation are multiplied by $d\varphi_0/d\rho$ and integrated on ρ from $-\infty$ to ∞ the result is zero. This may be verified directly for the left hand side of Eq. (48). Multiplying the right hand side of Eq. (48) by $d\varphi_0/d\rho$ and integrating on ρ from $-\infty$ to ∞ leads to some integrals that may be evaluated:

$$\int_{-\infty}^{\infty}\left(\frac{d\varphi_0}{d\rho}\right)^2 d\rho = \int_0^1 \frac{d\varphi_0}{d\rho}d\varphi_0 = \frac{1}{6}$$

$$\int_{-\infty}^{\infty}\frac{d\varphi_0}{d\rho}p'(\varphi_0)d\rho = \int_0^1 p'(\varphi_0)d\varphi_0 = 1$$

$$\int_{-\infty}^{\infty}\frac{d\varphi_0}{d\rho}\frac{d^2\varphi_0}{d\rho^2}d\rho = \frac{1}{2}\int_{-\infty}^{\infty}\frac{d}{d\rho}\left(\frac{d\varphi_0}{d\rho}\right)^2 d\rho = 0$$

$$\int_{-\infty}^{\infty}\frac{d\varphi_0}{d\rho}g'(\varphi_0) = \int_0^1 g'(\varphi_0)d\varphi_0 = 0$$

Thus we obtain $T_1 = -W_M/3L_0 = -2\gamma/L_0\ell$ which gives

$$T = T_M - T_M\frac{\gamma}{L_0}\frac{2}{R}. \tag{49}$$

We recognize $2/R$ as the mean curvature of the diffuse interface, so Eq. (49) is just the Gibbs-Thomson equation. In the context of the phase field model, the Gibbs-Thomson effect arises because of the radial bending of the φ field, as formalized by the form of the Laplacian in spherical coordinates. For a more general interface having two principal radii of curvature R_1 and R_2, one would replace $2/R$ by $1/R_1 + 1/R_2$. The reader is referred to the literature [79] for further details.

2.6. Dynamical equations

We proceed to obtain dynamical equations for the time evolution of u and φ, or equivalently T and φ. These are based on the functionals given by Eq. (1) and Eq. (2) and the concepts of local energy conservation and local entropy production. There is no energy production so

$$\dot{U}_{\text{prod}} := \frac{d}{dt}\mathcal{U} + \int_{\mathcal{A}}[\mathbf{q} - \varepsilon_u^2\dot{\varphi}\nabla\varphi\cdot\hat{\mathbf{n}}]d^2x = 0. \tag{50}$$

The rate of entropy production is

$$\dot{S}_{\text{prod}} := \frac{d}{dt}\mathcal{S} + \int_{\mathcal{A}}[\frac{\mathbf{q}}{T} + \varepsilon_s^2\dot{\varphi}\nabla\varphi\cdot\hat{\mathbf{n}}]d^2x \geq 0. \tag{51}$$

Here, $\mathcal{A}$ is the area surrounding the arbitrary subvolume $\mathcal{V}$, $\hat{\mathbf{n}}$ is its unit outward normal, and a dot above a variable denotes partial differentiation with respect to time. The vector $\mathbf{q}$ is the classical heat flux, $\mathbf{q}/T$ is the classical entropy flux, and the additional fluxes in the area integrals are nonclassical fluxes associated with the gradient energy and gradient entropy corrections. These nonclassical fluxes arise whenever elements of the diffuse interface enter or leave a control volume, as explained by Fernandez-Diaz and Williams [80] and discussed in the context of the phase field model by Wang et al. [22]. Eq. (4) and Eq. (5) with $\delta u \to \dot{u}$ and $\delta\varphi \to \dot{\varphi}$ are used to compute the time derivatives in Eq. (50) and Eq. (51). The nonclassical energy and entropy fluxes are chosen to cancel the corresponding surface terms from the time derivatives. The remaining surface terms containing $\mathbf{q}$ are then converted to volume integrals by means of the divergence theorem.

Thus Eq. (50) becomes

$$\int_{\mathcal{V}}[\dot{u} + \nabla\cdot\mathbf{q} - \varepsilon_u^2\dot{\varphi}\nabla^2\varphi]d^3x = 0. \tag{52}$$

Since Eq. (52) holds in every arbitrary subvolume, the integrand itself must vanish and we obtain

$$\dot{u} + \nabla\cdot\mathbf{q} - \varepsilon_u^2\dot{\varphi}\nabla^2\varphi = 0. \tag{53}$$

For $\varepsilon_u^2 = 0$, Eq. (53) resembles the classical energy balance equation for a rigid material. In view of Eq. (15), however,

$$\dot{u} = \left[c_s(T) + p(\varphi) \frac{dL(T)}{dT} \right] \dot{T} + \dot{p}(\varphi)L(T) + \frac{W_u}{2} \dot{g}(\varphi). \tag{54}$$

The coefficient of $\dot{T}$ is a heat capacity that can depend on temperature and phase. The term $\dot{p}(\varphi)L(T) = p'(\phi)\dot{\varphi}L(T)$ represents the evolution of latent heat by the moving diffuse interface. The term $W_u \dot{g}(\varphi)/2 = W_u g'(\varphi)\dot{\varphi}/2$ arises from the double well potential for $u(T, \varphi)$. For the often used approximation that $L(T)$ and $c_S(T)$ are independent of T and for $W_u = 0$, Eq. (54) takes the simple form

$$\dot{u} = c_0 \dot{T} + \dot{p}(\varphi)L_0 \tag{55}$$

where c_0 is a constant. Eq. (55) with $p(\varphi) = \varphi$ was postulated for early phase field models. The term $\varepsilon_u^2 \dot{\varphi}\nabla^2\varphi$ in Eq. (53) is only non-vanishing in the interfacial region where it has order of magnitude $(\gamma/\ell)\dot{\varphi}$ and is often neglected relative to the term $\dot{p}(\varphi)L(T)$. In the simpler model of Wang et al. [22], which assumes $\varepsilon_u^2 = 0$, this term is completely absent, but it might be important in problems for which a diffuse interface makes up a substantial fraction of the whole system.

The remaining Eq. (51) for the entropy production is

$$\int_{\mathcal{V}} \left[\frac{\dot{u}}{T} + \nabla \cdot \left(\frac{\mathbf{q}}{T} \right) \right] d^3x$$
$$+ \int_{\mathcal{V}} \left[\left(\frac{\partial s}{\partial \varphi} \right)_u + \varepsilon_s^2 \nabla^2\varphi \right] \dot{\varphi} \, d^3x \geq 0. \tag{56}$$

Substitution of Eq. (53) into Eq. (56) yields

$$\int_{\mathcal{V}} \left[\mathbf{q} \cdot \nabla \left(\frac{1}{T} \right) \right] d^3x$$
$$+ \int_{\mathcal{V}} \left[\left(\frac{\partial s}{\partial \varphi} \right)_u + \frac{\varepsilon_f^2}{T} \nabla^2\varphi \right] \dot{\varphi} \, d^3x \geq 0. \tag{57}$$

Eq. (57) can be satisfied for every subvolume $\mathcal{V}$ by assuming linear constitutive laws of the form

$$\mathbf{q} = M_u \nabla \left(\frac{1}{T} \right) = -k\nabla T \tag{58}$$

where $M_u > 0$ and $k = M_u/T^2$ is the thermal conductivity, and

$$\tau \dot{\varphi} = \left(\frac{\partial s}{\partial \varphi} \right)_u + \frac{\varepsilon_f^2}{T} \nabla^2\varphi \tag{59}$$

where $\tau > 0$. With these constitutive laws, we have

$$\dot{S}_{\text{prod}} := \int_{\mathcal{V}} \left[\frac{1}{M_u} |\mathbf{q}|^2 + \tau |\dot{\varphi}|^2 \right] d^3x \geq 0. \tag{60}$$

Eq. (60) shows that there is a volumetric rate of entropy production

$$\frac{1}{M_u} |\mathbf{q}|^2 = M_u \left| \nabla \left(\frac{1}{T} \right) \right|^2 \tag{61}$$

due to the irreversible process of heat flow and a volumetric rate of entropy production

$$\tau |\dot{\varphi}|^2 = \frac{1}{\tau} \left[\left(\frac{\partial s}{\partial \varphi} \right)_u + \frac{\varepsilon_f^2}{T} \nabla^2\varphi \right]^2 \tag{62}$$

due to an irreversible process of interface motion. The 'driving forces' $\nabla(1/T)$ and $\partial s/\partial \varphi + (\varepsilon_f^2/T)\nabla^2\varphi$ for $\mathbf{q}$ and $\dot{\varphi}$ in Eq. (58) and Eq. (59) are precisely those quantities that vanish at equilibrium, as shown in section 2.1. Thus the phase field model with $\tau \neq 0$ requires a departure from the assumption of local equilibrium. Local equilibrium at the solid-liquid interface is frequently assumed for sharp interface models. By means of asymptotics [16,79], one can show that τ leads to a linear kinetic law of the form

$$v_I = \mu(T_E - T_I) \tag{63}$$

where v_I is the interface velocity (positive for freezing), T_E is the equilibrium temperature of the interface, including Gibbs-Thomson corrections, T_I is the actual temperature of the interface, and

$$\mu = \frac{6L_0\ell}{T_M^2\tau} \tag{64}$$

is a linear kinetic coefficient. For $\tau \to 0$, $\mu \to \infty$ and $T_I \to T_E$. This is the limit of local equilibrium at the solid-liquid interface.[6]

---

[6]Karma and Rappel [36] consider thicker interfaces and a perturbation analysis based on a different identification of small quantities. The variation of temperature across these interfaces requires this formula for the kinetic coefficient to be modified. See [79] for a comparison of results.

2.7. Scaling of the dynamical equations

For convenience, we summarize the dynamical phase field equations as follows:

$$\left[c_s(T) + p(\varphi)\frac{dL(T)}{dT}\right]\dot{T} + \dot{p}(\varphi)L(T) \qquad (65)$$

$$-\varepsilon_u^2\dot{\varphi}\nabla^2\varphi + \frac{W_u}{2}\dot{g}(\varphi) = \nabla \cdot (k\nabla T)$$

$$\tau\dot{\varphi} = p'(\varphi)\int_{T_M}^{T}\frac{L(T')}{(T')^2}\,dT' - \frac{W_f(T)}{2T}g'(\varphi) \qquad (66)$$

$$+\frac{\varepsilon_f^2(T)}{T}\nabla^2\varphi.$$

These equations are quite formidable and are usually solved only after some approximations are made. The quantities k and τ, which were introduced in the constitutive laws Eq. (58) and Eq. (59), could depend on temperature and phase. One often treats k as a constant and sets $\tau = 1/(MT)$ where M is a constant [79]. Similarly, it is frequently assumed that $L(T)$ and $c_S(T)$ are constants, which we denote by L_0 and c_0 respectively, as in Eq. (55). Since $|T - T_M|/T_M \ll 1$ for most problems of interest, we can set $\epsilon_f^2 \approx \epsilon_M^2$ and $W_f(T) \approx W_M$. With these simplifications, and with the notation $\Theta := (T-T_M)/(L_0/c_0)$, the phase field equations become

$$\dot{\Theta} + \dot{p}(\varphi) + \frac{W_u}{2L_0}\dot{g}(\varphi) - \frac{6\gamma\ell}{L_0}\dot{\varphi}\nabla^2\varphi = \kappa\nabla^2\Theta \quad (67)$$

and

$$t^*\dot{\varphi} = \frac{\ell}{6d_0}\Theta - \frac{1}{2}g'(\varphi) + \ell^2\nabla^2\varphi \qquad (68)$$

where $\kappa := k/c_S$ is the thermal diffusivity, $d_0 = (T_Mc_0/L_0)(\gamma/L_0)$ is a capillary length, and $t^* = T_M\ell\tau/6\gamma = \ell^2L_0/(T_M\gamma\mu)$ is a time associated with the kinetic coefficient. Alternatively one could define a diffusivity $\kappa_\varphi := \ell^2/t^* = \mu T_M\gamma/L_0$ associated with the phase field. The quantity γ/L_0 is also a capillary length, although it is smaller than d_0 by a factor of T_Mc_0/L_0 which is typically about 5.

Eq. (67) and Eq. (68) display a broad range of length and time scales. The smallest length is the interface thickness 6ℓ. Somewhat larger are the capillary lengths d_0 and γ/L_0, although just how much larger is a matter that can give rise to a hierarchy of length scales that lead to different asymptotic analyses of the thin interface limit [36,79]. Next comes the size R of typical morphological features (such as cell sizes or dendrite tips) that one would like to resolve. The largest length is the system size, which we denote by $\mathcal{L}$. We also need to consider a thermal length κ/v_I associated with interface motion of velocity v_I. This thermal length can be large or small compared to R depending on the rate of heat extraction or supercooling that drives the solidification. This disparity of length scales makes the problem difficult to solve numerically, because very fine grids and/or adaptive meshes are required. In order to approach the sharp interface model, one would like to take ℓ as small as computationally practical, and in order to approach local equilibrium at the interface, t^* should be taken as small as practical. Considerations of numerical stability can restrict the ratio $\Delta t/(\Delta x)^2$ of time steps ΔT to the square of spatial grid size Δx relative to t^*/ℓ^2 in finite difference computations.

3. ANISOTROPY

The model discussed in section 2 was based on the assumption that the solid is isotropic, which of course is not true for a crystal. Insofar as heat flow is concerned, this is not a serious issue for cubic crystals because their conductivity tensors are isotropic. For crystals with lower symmetry, we would have to make the appropriate modifications. Nevertheless, even for cubic crystals, the surface free energy and the kinetic coefficient are known to be anisotropic, and such anisotropies play a strong role in determining solidification morphologies, e.g., dendritic sidebranching. Consequently, phase field models have been modified to accommodate interfacial anisotropies.

We begin by allowing for anisotropic interfacial free energy because this can be treated even for equilibrium. We could conceivably go back to Eq. (1) and Eq. (2) and allow ε_u and ε_s to depend on interface orientation. But what is interface orientation for a diffuse interface? Presumably it can be characterized by the unit vec-
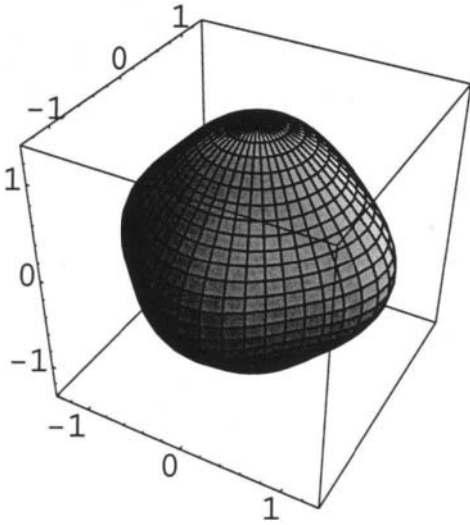
Figure 3. Polar plot of the cubic anisotropy function $[1+a_4Q_4(\hat{\mathbf{N}})]$ with Q_4 given by Eq. (72) and $a_4 = .3$, a large value selected for the sake of illustration.

Figure 4. Polar plot of the equilibrium shape for the cubic anisotropy of Figure 3 according to Eq. (88). Because of the large anisotropy, the shape is on the verge of having sharp corners.

tor $\hat{\mathbf{N}} = \nabla\varphi/|\nabla\varphi|$. This vector is undefined in bulk material where φ is essentially constant but we are concerned here only with interfacial anisotropy. Although this procedure will work [16] in two dimensions, the computations become quite cumbersome in three dimensions.

We therefore follow a somewhat abstract but elegant alternative, introduced to this problem by Wheeler and McFadden [17,18]. It is based on the ξ-vector formalism of Cahn and Hoffman [81,82] that was developed to handle anisotropy for sharp interface models. We retain ε_u and ε_s as constants but replace $\nabla\varphi$ by an anisotropy function $A(\nabla\varphi)$, which is a homogeneous function of degree 1 in the components of $\nabla\varphi$. To simplify the notation, we denote $\mathbf{P} := \nabla\varphi$ with cartesian components $P_\alpha = \partial\varphi/\partial x_\alpha$. The homogeneity condition is

$$A(\lambda P_1, \lambda P_2, \lambda P_3) = \lambda A(P_1, P_2, P_3) \qquad (69)$$

for any λ. An example of such a function is

$$A_4(\mathbf{P}) := \frac{P_1^4 + P_2^4 + P_3^4}{P^3} \qquad (70)$$

where $P := |\mathbf{P}| = (P_1^2 + P_2^2 + P_3^2)^{1/2}$ is the magnitude of $\mathbf{P}$. Such a function can be formed by multiplying a function $Q(\hat{\mathbf{N}})$ by P, i.e.,

$$A(\mathbf{P}) = P\,Q(\hat{\mathbf{N}}). \qquad (71)$$

Thus Eq. (71) extends a function of orientation $\hat{\mathbf{N}}$ to a vector space $\mathbf{P}$ that is no longer of unit length. The function $A_4(\mathbf{P})$ in Eq. (70) corresponds to

$$Q_4(\hat{\mathbf{N}}) = N_1^4 + N_2^4 + N_3^4. \qquad (72)$$

where $Q_4(\hat{\mathbf{N}})$ is the leading *anisotropic* term in the interfacial free energy anisotropy of a cubic crystal. In other words, the interfacial energy of such a crystal would be of the form $\gamma_0[1+a_4Q_4(\hat{\mathbf{N}})]$ where γ_0 and a_4 are constants. A polar plot of the function $[1 + a_4Q_4(\hat{\mathbf{N}})]$ is shown in Figure 3.

We shall proceed to demonstrate that this modification will lead to an anisotropic interface thickness parameter

$$\tilde{\ell}(\hat{\mathbf{N}}) := \ell Q(\hat{\mathbf{N}}) \tag{73}$$

and an anisotropic interfacial free energy given by

$$\tilde{\gamma}(\hat{\mathbf{N}}) := \gamma Q(\hat{\mathbf{N}}) \tag{74}$$

where the scale factors ℓ and γ are given by Eq. (35) and Eq. (39) for the isotropic phase field model.

We first explore briefly some relations satisfied by A. The first comes from Euler's theorem of homogeneous functions and can be derived by differentiation of Eq. (69) with respect to λ and then setting $\lambda = 1$. The result is

$$\sum_\alpha P_\alpha \frac{\partial A}{\partial P_\alpha} = A. \tag{75}$$

If we define a vector $\boldsymbol{\Xi}$ with components

$$\Xi_\alpha := \frac{\partial A}{\partial P_\alpha}, \tag{76}$$

Eq. (75) becomes

$$\sum_\alpha P_\alpha \Xi_\alpha = \mathbf{P} \cdot \boldsymbol{\Xi} = A. \tag{77}$$

Alternatively $\hat{\mathbf{N}} \cdot \boldsymbol{\Xi} = Q$ where A and Q are related by Eq. (71). The differential of Eq. (77) is

$$dA = \sum_\alpha \Xi_\alpha dP_\alpha + \sum_\alpha P_\alpha d\Xi_\alpha \tag{78}$$

but we also have

$$dA = \sum_\alpha \frac{\partial A}{\partial P_\alpha} dP_\alpha = \sum_\alpha \Xi_\alpha dP_\alpha = \boldsymbol{\Xi} \cdot d\mathbf{P}. \tag{79}$$

Comparison of Eq. (78) and Eq. (79) shows that

$$0 = \sum_\alpha P_\alpha d\Xi_\alpha = \mathbf{P} \cdot d\boldsymbol{\Xi}. \tag{80}$$

It therefore follows that $\hat{\mathbf{N}} \cdot d\boldsymbol{\Xi} = 0$ and $dQ = \boldsymbol{\Xi} \cdot d\hat{\mathbf{N}}$. From its definition in Eq. (76), we see that $\boldsymbol{\Xi}$ is a homogeneous function of degree zero, so it depends only on $\hat{\mathbf{N}}$.

We can therefore relate $\boldsymbol{\Xi}$ to the $\boldsymbol{\xi}$ vector of Hoffman and Cahn by

$$\boldsymbol{\xi} = \gamma \boldsymbol{\Xi} \tag{81}$$

where γ is just a scale factor given by Eq. (39). Then

$$\boldsymbol{\xi} \cdot \hat{\mathbf{N}} = \gamma \boldsymbol{\Xi} \cdot \hat{\mathbf{N}} = \gamma Q(\hat{\mathbf{N}}) = \tilde{\gamma}(\hat{\mathbf{N}}) \tag{82}$$

where Eq. (74) is used in the last step. The $\boldsymbol{\xi}$-vector has the well-known additional properties $\hat{\mathbf{N}} \cdot d\boldsymbol{\xi} = 0$ and $d\tilde{\gamma}(\hat{\mathbf{N}}) = \boldsymbol{\xi} \cdot d\hat{\mathbf{N}}$. We note that $\boldsymbol{\xi}$ also depends only on $\hat{\mathbf{N}}$.

We have yet to demonstrate that all of these anisotropic properties work out correctly for the phase field model. To do this, we need to examine the functional derivative of terms like $(1/2)\varepsilon_u^2 A^2$ that will now replace $(1/2)\varepsilon_u^2 |\nabla\varphi|^2$. Thus

$$\begin{aligned} \delta[\frac{1}{2}A^2] &= A\boldsymbol{\Xi} \cdot \delta\mathbf{P} = A\boldsymbol{\Xi} \cdot \nabla\delta\varphi \\ &= \nabla \cdot (A\boldsymbol{\Xi}\,\delta\varphi) - \delta\varphi\nabla \cdot (A\boldsymbol{\Xi}). \end{aligned} \tag{83}$$

When this quantity is integrated over volume, the complete divergence is integrated to the surrounding area by means of the divergence theorem. A similar calculus arises when δ is replaced by $\partial/\partial t$. In this manner, all terms containing ε^2 in equilibrium or dynamical equations should be replaced according to the prescription

$$\varepsilon^2 \nabla^2 \varphi \rightarrow \varepsilon^2 \nabla \cdot (A\boldsymbol{\Xi}), \tag{84}$$

where ε^2 stands for ε_u^2, ε_s^2, ε_f^2 or ε_M^2.

In particular, Eq. (13) for equilibrium becomes

$$\left(\frac{\partial f}{\partial \varphi}\right)_T = \varepsilon_f^2 \nabla \cdot (A\boldsymbol{\Xi}). \tag{85}$$

We proceed to solve Eq. (85) for solid-liquid equilibrium at $T = T_M$ for a one dimensional problem in the variable $x_0 := \mathbf{r} \cdot \hat{\mathbf{N}}_0$ where $\mathbf{r}$ is the position vector and $\hat{\mathbf{N}}_0$ is a constant unit vector. Thus φ depends only on x_0 and consequently $A\boldsymbol{\Xi}$ is only a function of x_0. Therefore,

$$\begin{aligned} \nabla \cdot (A\boldsymbol{\Xi}) &= \frac{d(A\boldsymbol{\Xi})}{dx_0} \cdot \nabla x_0 = \frac{d(A\boldsymbol{\Xi})}{dx_0} \cdot \hat{\mathbf{N}}_0 \\ &= \frac{d[AQ(\hat{\mathbf{N}}_0)]}{dx_0} = [Q(\hat{\mathbf{N}}_0)]^2 \frac{dP}{dx_0} \\ &= [Q(\hat{\mathbf{N}}_0)]^2 \frac{d^2\varphi}{dx_0^2}. \end{aligned} \tag{86}$$
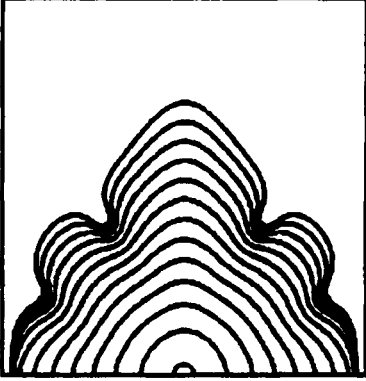
Figure 5. Time evolution of dendrite tip shapes computed from the phase field model for a four-fold sinusoidal interfacial free energy anisotropy of 3%.

Figure 6. Time evolution of dendrite tip shapes computed from the phase field model for a four-fold sinusoidal interfacial free energy anisotropy of 4%.

Hence Eq. (85) becomes

$$\frac{df(T_M, \varphi)}{d\varphi} = \varepsilon_M^2 [Q(\hat{N}_0)]^2 \frac{d^2\varphi}{dx_0^2}. \qquad (87)$$

Comparison of Eq. (87) with Eq. (30) shows that they have the same form, except for the change $\varepsilon_M \to \varepsilon_M Q(\hat{N}_0)$. Therefore, Eq. (35) and Eq. (36) apply with $\ell \to \ell Q(\hat{N}_0) = \tilde{\ell}(\hat{N}_0)$, in agreement with Eq. (73). Since the solid liquid interfacial free energy is also based on Eq. (13), the procedure in section 2.4 applies with $\gamma \to \gamma Q(\hat{N}_0) = \tilde{\gamma}(\hat{N}_0)$, in agreement with Eq. (74).

Treatment of the Gibbs-Thomson equation with anisotropy is, however, much more involved. As shown by Wheeler and McFadden [17], an asymptotic analysis for a very thin interface leads to

$$T = T_M - \frac{T_M \gamma}{L_0} \nabla_S \cdot \Xi = T_M - \frac{T_M}{L_0} \nabla_S \cdot \xi \quad (88)$$

where $\nabla_S \cdot$ is the surface divergence operator. Eq. (88) replaces Eq. (49) and depends on the directional derivatives of $\gamma(\hat{N})$ as well as its magnitude. Since the interface is diffuse, one must settle on some convention to define it more precisely, for example as the locus of $\varphi = 1/2$. Eq. (88) is equivalent to the Herring equation

$$T = T_M - \frac{T_M}{L_0} \left[\frac{1}{R_1} \left(\tilde{\gamma} + \frac{\partial^2 \tilde{\gamma}}{\partial \theta_1^2} \right) + \frac{1}{R_1} \left(\tilde{\gamma} + \frac{\partial^2 \tilde{\gamma}}{\partial \theta_1^2} \right) \right] \qquad (89)$$

where R_1 and R_2 are principal radii of curvature and θ_1 and θ_2 are angles of the interface normal vector, measured in the principal planes of curvature.

Eq. (88) can also be used to generate a parametric formula for the equilibrium shape taken on by a crystal at some uniform temperature $T < T_M$. If r is a position vector in three dimensional space, we have $\nabla_S \cdot r = 2$ for any surface divergence operator. Thus if we substitute $\Xi = Cr$ where C is a constant into Eq. (88), we obtain a solution for constant T. Solving for C and rearranging we obtain

$$r(\hat{N}) = \frac{T_M}{T_M - T} \frac{2\gamma}{L_0} \Xi(\hat{N}) = \frac{T_M}{T_M - T} \frac{2}{L_0} \xi(\hat{N}). (90)$$

Eq. (90) shows that r and Ξ are proportional at any $\hat{N}$. The orientation $\hat{N}$ can be specified by two independent angles made by $\hat{N}$ and some fixed axes. Eq. (90) therefore generates a surface r which is a function of these two angles that

specify the orientation of its *normal*. By using this formula, three dimensional plots of equilibrium shapes can be generated by using parametric plotting routines that are now readily available. An example is shown in Figure 4 for a large anisotropy $a_4 = 0.3$. Such a shape would have missing orientations for $a_4 > 1/3$. Note that the angles made by $\hat{\mathbf{N}}$ are not the same angles made by $\mathbf{r}$ with these fixed axes, except for the case of isotropy or for other special symmetry points where $\mathbf{r}$ happens to lie along $\hat{\mathbf{N}}$. Analytical conversion from $\hat{\mathbf{N}}$ to $\hat{\mathbf{r}}$ has been carried out to second order for small anisotropy by McFadden, Coriell and Sekerka [83]. The prefactor $[2\gamma T_M/(T_M - T)]L_0$ provides a length scale for the equilibrium shape, which becomes infinite as $T \to T_M$. For the case of isotropy, $\Xi = \hat{\mathbf{N}}$ and Eq. (90) agrees with Eq. (49) for a sphere.

As mentioned previously, anisotropy can play a strong role in determining solidification morphologies. This is illustrated in Figures 5 and 6 taken from the doctoral thesis of Shun-Lien Wang [84]. Time evolution of the tip shapes of a two-dimensional dendrite growing at high supercooling ($\Delta T = 0.8L_0/c_0$) are shown for 3% and 4% sinusoidal anisotropy of interfacial free energy and for a kinetic coefficient, μ, that is isotropic. In Figure 6 the tip is sharper and the side branching is more pronounced. See [29,30] for details of the computations.

For the dynamical phase field equations, τ can also be anisotropic. This can be handled by letting $\tau \to \tilde{\tau}(\hat{\mathbf{N}}) := \tau B(\hat{\mathbf{N}})$ where $B(\hat{\mathbf{N}})$ is an anisotropy function. Asymptotic analysis [16] then shows that the the kinetic coefficient μ in Eq. (63) gets its anisotropy from *both* $A(\hat{\mathbf{N}})$ and $B(\hat{\mathbf{N}})$. Eq. (64) is replaced by

$$\tilde{\mu}(\hat{\mathbf{N}}) = \frac{6L_0\ell A(\hat{\mathbf{N}})}{T_M^2 \tau B(\hat{\mathbf{N}})} = \frac{6L_0\tilde{\ell}(\hat{\mathbf{N}})}{T_M^2 \tilde{\tau}(\hat{\mathbf{N}})}. \tag{91}$$

4. ALLOYS

To develop a phase field model for a binary alloy having chemical components A and B, we need to add another variable b which we take to be the partial density (mass per unit volume) of component B. We retain the assumptions that

the solid is isotropic, that both solid and liquid have constant uniform density, and that convection in the melt is forbidden. Anisotropy can be added following the procedure of section 3. To simplify the treatment somewhat, we assume that the internal energy is

$$\mathcal{U} := \int_{\mathcal{V}} u \, d^3x \tag{92}$$

and the total mass of component B is

$$\mathcal{M}_B := \int_{\mathcal{V}} b \, d^3x. \tag{93}$$

But we take a more general entropy functional of the form

$$\mathcal{S} := \int_{\mathcal{V}} [s(u, b, \varphi) - \frac{1}{2}\varepsilon_0^2|\nabla u|^2 - \frac{1}{2}\varepsilon_b^2|\nabla b|^2 - \frac{1}{2}\varepsilon_\varphi^2|\nabla\varphi|^2] \, d^3x \tag{94}$$

where ε_0^2, ε_b^2 and ε_φ^2 are constants. We include all three gradient entropy corrections in order to exhibit the symmetry associated with u and b, which are both related to conserved quantities.

Indeed, since both energy and mass of B are conserved,[7] Eq. (92) and Eq. (93) lead to

$$\dot{u} + \nabla \cdot \mathbf{j}_u = 0; \qquad \dot{b} + \nabla \cdot \mathbf{j}_b = 0 \tag{95}$$

where $\mathbf{j}_u$ is an internal energy flux and $\mathbf{j}_b$ is the mass flux of component B. We reserve the symbol $\mathbf{q}$ for the heat flux which, for an alloy, can be different from the internal energy flux because energy can be transported by a diffusing species. We shall return to this subtle point later (see Eq. (105)).

For the entropy, we require positive local entropy production, subject now to two conservation laws, Eqs. (95). Thus

$$\dot{S}_{\text{prod}} := \frac{d}{dt}\mathcal{S} + \int_{\mathcal{A}} \mathbf{j}_s \cdot \hat{\mathbf{n}} \, d^2x \geq 0 \tag{96}$$

where $\mathbf{j}_s$ is the total entropy flux, including nonclassical terms due to gradient entropy corrections. In order to discuss the results in a more

---

[7]Of course the mass of component A is conserved as well, but since the density is is assumed to be constant this is not an independent condition.

succinct manner, we define the following functional derivatives:

$$\left(\frac{1}{T}\right)^{NC} := \frac{\delta s}{\delta u} := \left[\frac{\partial s}{\partial u} + \varepsilon_0^2\, \nabla^2 u\right] \quad (97)$$

$$-\left(\frac{\mu}{T}\right)^{NC} := \frac{\delta s}{\delta b} := \left[\frac{\partial s}{\partial b} + \varepsilon_b^2\, \nabla^2 b\right] \quad (98)$$

$$S_\varphi := \frac{\delta s}{\delta \varphi} := \left[\frac{\partial s}{\partial \varphi} + \varepsilon_\varphi^2\, \nabla^2 \varphi\right] \quad (99)$$

From the differential

$$ds = \frac{du}{T} - \frac{\mu}{T}db + \frac{\partial s}{\partial \varphi}d\varphi \quad (100)$$

we see that[8]

$$\left(\frac{\partial s}{\partial u}\right)_{b,\varphi} = \frac{1}{T}; \qquad \left(\frac{\partial s}{\partial b}\right)_{u,\varphi} = -\frac{\mu}{T} \quad (101)$$

where T is the classical temperature and $\mu := \mu_B - \mu_A$ is the so-called diffusion potential, where μ_A and μ_B are chemical potentials of components A and B, expressed per unit mass. The quantities on the left hand sides of Eq. (97) and Eq. (98) serve to define a non-classical temperature T^{NC} and a non-classical diffusion potential μ^{NC} whose gradients will play the roles of driving forces for diffusive transport of energy and chemical components. At equilibrium, T^{NC} and μ^{NC} will be uniform throughout the system [43,73]. If the total entropy flux is chosen to be

$$\mathbf{j}_s = \left(\frac{1}{T}\right)^{NC} \mathbf{j}_u - \left(\frac{\mu}{T}\right)^{NC} \mathbf{j}_b$$
$$+\varepsilon_0^2 \dot{u}\nabla u + \varepsilon_b^2 \dot{b}\nabla b + \varepsilon_\varphi^2 \dot{\varphi}\nabla\varphi, \quad (102)$$

the rate of entropy production takes the form

$$\dot{S}_{\text{prod}} = \int_V \mathbf{j}_u \cdot \nabla\left(\frac{1}{T}\right)^{NC} - \mathbf{j}_b \cdot \nabla\left(\frac{\mu}{T}\right)^{NC}$$
$$+ S_\varphi \dot{\varphi} \geq 0. \quad (103)$$

If $\varepsilon_0^2 = \varepsilon_b^2 = \varepsilon_\varphi^2 = 0$, this entropy flux becomes

$$\mathbf{j}_s = \frac{\mathbf{j}_u - \mu \mathbf{j}_b}{T} = \frac{\mathbf{q}}{T} + (\tilde{S}_B - \tilde{S}_A)\mathbf{j}_b \quad (104)$$

---

[8]Since we are assuming the density to be constant, it is also held constant in these derivatives but we do not indicate this explicitly.

with

$$\mathbf{q} := \mathbf{j}_u - (\tilde{H}_B - \tilde{H}_A)\mathbf{j}_b \quad (105)$$

where $\tilde{S}_A$ and $\tilde{S}_B$ are partial specific entropies and $\tilde{H}_A$ and $\tilde{H}_B$ are partial specific enthalpies of A and B, respectively. We recognize $\mathbf{q}$ as a flux of *conducted* heat [85–88] because the heat flux related to diffusive transport has been subtracted from $\mathbf{j}_u$. Eq. (104) is therefore in agreement with our intuition because $\mathbf{q}/T$ is the classical entropy flux due to heat conduction and the additional term represents an entropy flux due to diffusion.

We proceed to postulate linear constitutive laws that will lead to local positive entropy production. In doing this, we are guided by Curie's law [87], according to which tensors of rank differing by an odd integer cannot be coupled linearly. Adapted to the present case, Curie's law states simply that scalar causes cannot have vector effects, and *vice versa*. It follows that the vector fluxes $\mathbf{j}_u$ and $\mathbf{j}_b$ can be driven by *both* $\nabla(1/T)^{NC}$ and $-\nabla(\mu/T)^{NC}$ but not by S_φ, which alone will drive $\dot{\varphi}$. We therefore adopt constitutive laws of the form

$$\left[\begin{array}{c} \mathbf{j}_u \\ \mathbf{j}_b \end{array}\right] = \left[\begin{array}{cc} M_{uu} & M_{ub} \\ M_{bu} & M_{bb} \end{array}\right]\left[\begin{array}{c} \nabla(1/T)^{NC} \\ -\nabla(\mu/T)^{NC} \end{array}\right] \quad (106)$$

and

$$\dot{\varphi} = M_\varphi S_\varphi \quad (107)$$

where $M_\varphi > 0$ and the matrix in Eq. (106) leads to positive entropy production. This will be true if $M_{uu} > 0$, $M_{bb} > 0$ and $M_{uu}M_{bb} > \tilde{M}_{ub}^2$ where $\tilde{M}_{ub} = (M_{ub} + M_{bu})/2$. If the matrix M_{ij} is symmetric [89], the last condition becomes $M_{uu}M_{bb} > M_{ub}^2$ and M_{ij} can be diagonalized and will have positive eigenvalues. The local rate of entropy production per unit volume therefore becomes

$$\dot{s}_{\text{prod}} = M_{uu}|\nabla(1/T)^{NC}|^2 + M_{bb}|\nabla(\mu/T)^{NC}|^2$$
$$- (M_{ub} + M_{bu})\nabla(1/T)^{NC} \cdot \nabla(\mu/T)^{NC}$$
$$+ M_\varphi S_\varphi^2 \geq 0. \quad (108)$$

It is possible that the matrix M_{ij} could be diagonal, or alternatively that it could be diagonalized by using some linear combination of $\mathbf{j}_u$ and

$\mathbf{j}_b$. By appealing to Eq. (105), one might conjecture that choice of the non-classical analogs of the pair $\mathbf{q}$ and $\mathbf{j}_b$ would lead to a diagonal matrix. In other words, the conducted heat flux analogous to $\mathbf{q}$ would be driven only by $\nabla(1/T)^{NC}$. We explore this possibility later (see Eq. (126) *et seq.*). Further guidance in this matter must be left to microscopic models and ultimately to experiment.

4.1. Regular solution

To proceed, we must make a specific choice of the function $s(u, b, \varphi)$. To do this, we follow Warren and Boettinger [52] and adopt a regular solution model. We shall, however, generalize their model to allow for a molar volume, $1/n$, that is a linear function of the mole fraction X of component B. We do this because the density, ρ, must be a constant if we are to be consistent with the assumption of no convection. Otherwise, the continuity equation $\dot{\rho} + \nabla(\rho \mathbf{v}) = 0$ would require a non-vanishing fluid velocity $\mathbf{v}$. But the ratio

$$
\begin{aligned}
\frac{n}{\rho} &= \frac{1}{(1-X)m_A + X m_B} \\
&= (1-\omega)\frac{1}{m_A} + \omega\frac{1}{m_B}
\end{aligned}
\tag{109}
$$

where ω is the *mass* fraction of component B and m_A and m_B are the molecular weights of A and B. Thus if ρ is constant, n must vary with composition unless $m_A = m_B$. But we usually want to treat components for which $m_A \neq m_B$. We therefore dispense with the assumption of equal molecular weights and allow n to vary with composition.

For a regular solution [90], the excess[9] entropy of mixing is the ideal entropy of mixing, there is no excess volume change on mixing, and the excess heat (enthalpy) of mixing per mole is $\Omega_N(\varphi)X(1-X)$ where $\Omega_N(\varphi)$ can depend on the phase but is independent of the pressure p and the temperature T. The Helmholtz free energy per mole for such a regular solution can be written in the form

$$
f_N = (1-X)\bar{F}_A^0 + X\bar{F}_B^0
$$

$$
\begin{aligned}
&+ RT[(1-X)\ln(1-X) + X\ln X] \\
&+ \Omega_N X(1-X)
\end{aligned}
\tag{110}
$$

where $\bar{F}_A^0(p, T, \varphi)$ and $\bar{F}_B^0(p, T, \varphi)$ are partial molar Helmholtz free energies[10] corresponding to pure A and pure B, and R is the ideal gas constant. The quantity $\Omega_N(\varphi)$ is an interaction parameter for species A and B, positive for net repulsion and negative for net attraction. The chemical potentials per mole are

$$
\begin{aligned}
\mu_{AN} &= \mu_{AN}^0 + R\ln(1-X) + \Omega_N X^2 \tag{111} \\
\mu_{BN} &= \mu_{BN}^0 + R\ln X + \Omega_N(1-X)^2
\end{aligned}
$$

where μ_{AN}^0 and μ_{BN}^0 correspond to reference states, taken to be pure A and B, respectively. The chemical potentials per mass are

$$
\begin{aligned}
\mu_A &= \mu_A^0 + \frac{R}{m_A}\ln(1-X) + \frac{\Omega_N}{m_a}X^2 \tag{112} \\
\mu_B &= \mu_B^0 + \frac{R}{m_B}\ln X + \frac{\Omega_N}{m_B}(1-X)^2
\end{aligned}
$$

where $\mu_A^0 = \mu_{AN}^0/m_A$ and $\mu_B^0 = \mu_{BN}^0/m_B$ correspond to the reference states.

To treat an inhomogeneous system, we define a Helmholtz free energy per unit volume

$$
f(T, b, \varphi, \rho) = \rho\,(n/\rho)\,f_N(p, T, X, \varphi).
\tag{113}
$$

Since the factor n/ρ is not a constant, f takes the form[11]

$$
\begin{aligned}
f &= (1-\omega)f_A + \omega f_B \\
&+ \rho RT\left[\frac{(1-\omega)}{m_A}\ln(1-X) + \frac{\omega}{m_B}\ln X\right] \\
&+ \frac{\rho\Omega_N}{2}\left[\frac{(1-\omega)X}{m_A} + \frac{(1-X)\omega}{m_B}\right]
\end{aligned}
\tag{114}
$$

where $f_A := \rho\bar{F}_A^0/m_A$ and $f_B := \rho\bar{F}_B^0/m_B$ are free energy densities corresponding to the reference states. In obtaining Eq. (114) from Eq. (110), we have used the relations $(n/\rho)(1-X) = (1-\omega)/m_A$ and $(n/\rho)X = \omega/m_B$. Note that Eq. (114) is in a mixed notation, containing

---

[9]The word 'excess' in this context means: that in addition to the linear function of X obtained by mechanical mixture of the pure components.

[10]For clarity, we carry the pressure along for a while, although consistent with no convection we will assume it to be constant and drop it later.

[11]Strictly speaking, the function f depends on T, $b = \omega\rho$, ρ and φ, but we hereafter drop the dependence on ρ, consistent with constant ρ and constant p.

both ω and X; one of them could be eliminated but the resulting expression is unwieldy. In particular, $f_A(T,\varphi)$ and $f_B(T,\varphi)$ are given by forms similar to Eq. (18). Specifically,

$$
\begin{aligned}
f_A(T,\varphi) &= [1-p(\varphi)]f_{AS}(T) \\
&+ p(\varphi)f_{AL}(T) + \frac{W_{Af}(T)}{2}g(\varphi)
\end{aligned}
\tag{115}
$$

with a similar expression for $A \to B$. Here, $f_{AS}(T)$, $f_{AL}(T)$, $f_{BS}(T)$ and $f_{BL}(T)$ have forms similar to those given by Eq. (19) and Eq. (20).

From the function $f(T,b,\varphi)$ given by Eq. (114) we can calculate the driving forces. Since $(\partial s/\partial u)_{b,\varphi} = 1/T$, we have

$$
\nabla\left(\frac{1}{T}\right)^{NC} = \nabla\left[\frac{1}{T} + \varepsilon_0^2 \nabla^2 u\right]
\tag{116}
$$

where

$$
\begin{aligned}
u(T,c,\varphi) &= \frac{\partial[f(T,b,\varphi)/T]}{\partial(1/T)} \\
&= (1-\omega)u_A(T,\varphi) + \omega u_B(T,\varphi) \\
&+ \rho\Omega_N \frac{\omega(1-\omega)}{(1-\omega)m_B + \omega m_A}
\end{aligned}
\tag{117}
$$

with $u_A(T,\varphi)$ and $u_B(T,\varphi)$ given by equations of the form of Eq. (15). The last term in Eq. (117) is just another form of the last term in Eq. (114).

The driving force for the phase field requires computation of $(\partial s/\partial\varphi)_{u,b} = -(1/T)(\partial f/\partial\varphi)_{T,b}$. The result is

$$
\begin{aligned}
S_\varphi &= (1-\omega)\sigma_A(T,\varphi) + \omega\sigma_B(T,\varphi) \\
&- \frac{\Omega'(\varphi)}{T}\frac{\omega(1-\omega)}{(1-\omega)m_B + \omega m_A} + \varepsilon_\varphi^2\nabla^2\varphi
\end{aligned}
\tag{118}
$$

where

$$
\sigma_A(T,\varphi) = p'(\varphi)\int_{T_A}^T \frac{L_A(T')}{(T')^2}dT' - \frac{W_{Af}}{2T}g'(\varphi)
\tag{119}
$$

and T_A is the melting point of pure A. The quantity $\sigma_B(T,\varphi)$ is given by a similar expression with $A \to B$.

The remaining driving force can be calculated by using $(\partial s/\partial b)_{u,\varphi} = -(1/T)(\partial f/\partial b)_{T,\varphi} = -(1/\rho T)(\partial f/\partial\omega)_{T,\varphi}$. This computation is facilitated by using implicit differentiation and noting that $\omega/(1-\omega) = (m_B/m_A)X/(1-X)$ and

$dX/d\omega = X(1-\omega)/[\omega(1-X)]$. The result is

$$
\begin{aligned}
\left(\frac{\partial s}{\partial b}\right)_{u,\varphi} &= \frac{f_A(T,\varphi) - f_B(T,\varphi)}{\rho T} \\
&- R\left[\frac{\ln(1-X)}{m_A} - \frac{\ln X}{m_B}\right] \\
&- \frac{\Omega_N}{T}\left[\frac{(1-X)^2}{m_B} - \frac{X^2}{m_A}\right].
\end{aligned}
\tag{120}
$$

We observe that Eq. (120) is agreement with the second of Eqs. (101) after substituting Eqs. (112) and recognizing that $[f_A(T,\varphi) - f_B(T,\varphi)]/\rho = \mu_A^0(T,\varphi) - \mu_B^0(T,\varphi)$ because the pressure in the system is uniform. Finally, we compute the gradient of Eq. (120) in the form

$$
\begin{aligned}
\nabla\left(\frac{\partial s}{\partial b}\right)_{u,\varphi} &= -\nabla\left(\frac{\mu}{T}\right) \\
&= B_\omega\nabla\omega + B_\varphi\nabla\varphi + B_T\nabla\left(\frac{1}{T}\right)
\end{aligned}
\tag{121}
$$

where

$$
\begin{aligned}
B_\omega &= -\frac{R}{\omega(1-\omega)[(1-\omega)m_B + \omega m_A]} \\
&+ \frac{2\Omega_N(\varphi)}{T}\frac{m_A m_B}{[(1-\omega)m_B + \omega m_A]^3}
\end{aligned}
\tag{122}
$$

$$
\begin{aligned}
B_\varphi &= \frac{\sigma_B(T,\varphi) - \sigma_A(T,\varphi)}{\rho} \\
&- \frac{\Omega_N'(\varphi)}{T}\frac{\omega(1-\omega)}{(1-\omega)m_B + \omega m_A}
\end{aligned}
\tag{123}
$$

and

$$
\begin{aligned}
B_T &= \frac{u_A(T,\varphi) - u_B(T,\varphi)}{\rho} \\
&- \Omega_N(\varphi)\frac{\omega(1-\omega)}{(1-\omega)m_B + \omega m_A}.
\end{aligned}
\tag{124}
$$

For the special case $m := m_A = m_B$, one has $\omega = X$ and $(1-\omega)m_B + \omega m_A = m$ so these results simplify, in agreement with those of Bi and Sekerka [43].

4.2. Minimal model for alloy solidification

Obviously the phase field equations for alloy solidification are very complicated because they contain a great deal of detailed information. A

minimal model can be obtained by making numerous approximations. First we set $\varepsilon_0^2 = 0$ so that $T^{NC} = T$. Then we set $\varepsilon_b^2 = 0$ which makes $\mu^{NC} = \mu$. Note that this precludes treatment of spinodal decomposition [78,43]. Next, we assume that both solutions are ideal, so $\Omega_N(\varphi) = 0$. We also assume that L_A and L_B are constants.

Even with these simplifications, the alloy phase field equations are quite complicated. We have

$$\sigma_A(T, \varphi) = p'(\varphi) L_A \left(\frac{1}{T_A} - \frac{1}{T} \right) - \frac{W_{Af}}{2T} g'(\varphi) \quad (125)$$

and similarly for $\sigma_B(T, \varphi)$. The driving forces are now $\nabla(1/T)$ and $\nabla(\mu/T)$ instead of $\nabla(1/T)^{NC}$ and $\nabla(\mu/T)^{NC}$. Recognizing the breakup of $-\nabla(\mu/T) = \nabla(\partial s/\partial b)_{u,\varphi}$ given by Eq. (121), we return to Eq. (106) and do a linear transformation[12] to write it in the form

$$\begin{bmatrix} \mathbf{q} \\ \mathbf{j}_b \end{bmatrix} = \begin{bmatrix} \tilde{M}_{uu} & \tilde{M}_{ub} \\ \tilde{M}_{bu} & \tilde{M}_{bb} \end{bmatrix} \begin{bmatrix} \nabla(1/T) \\ B_\omega \nabla \omega + B_\varphi \nabla \varphi \end{bmatrix} \quad (126)$$

where $\mathbf{q}$ is given by Eq. (105) and $\tilde{M}_{uu} = M_{uu} + B_T(M_{ub} + M_{bu}) + B_T^2 M_{bb}$, $\tilde{M}_{ub} = M_{ub} + B_T M_{bb}$, $\tilde{M}_{bu} = M_{bu} + B_T M_{bb}$, and $\tilde{M}_{bb} = M_{bb}$. Consistent with Eq. (124), we identify $B_T = -(\tilde{H}_B - \tilde{H}_A)$. This transformation was inspired by a similar transformation in the classical case [89], and we note that symmetry of M_{ij} implies that $\tilde{M}_{ij}$ is also symmetric. We proceed to assume that the matrix $\tilde{M}_{ij}$ is diagonal, thus precluding heat fluxes driven by solute gradients and solute fluxes driven by temperature gradients (Dufour and Soret effects [86]). Then $\mathbf{q} = \tilde{M}_{uu} \nabla(1/T) = -k\nabla T$ as in the case of a pure material. With the notation $\tilde{M}_{bb} B_\omega = -\rho D$ where D is a diffusivity, the solute flux takes the form

$$\mathbf{j}_b = -\rho D \nabla \omega + \frac{D}{RT} \Phi(T, \omega, \varphi) \nabla \varphi \quad (127)$$

where

$$\Phi(T, \omega, \varphi) := T(\sigma_B - \sigma_A)\omega(1 - \omega) \quad (128)$$
$$\times [(1 - \omega)m_B - \omega m_A].$$

We recognize $D/(RT)$ as a classical mobility and the term containing it in Eq. (127) contributes to the diffusive flux in the interfacial region.

[12] It is possible to do this transformation even if $\Omega(\varphi) \neq 0$ but then Φ in Eq. (128) will be more complicated.

The alloy phase field equations therefore become

$$\dot{u} = \nabla \cdot [k\nabla T - (\tilde{H}_B - \tilde{H}_A)\mathbf{j}_b] \quad (129)$$

$$\dot{b} = -\nabla \cdot \mathbf{j}_b \quad (130)$$

$$\tau \dot{\varphi} = (1 - \omega)\sigma_A(T, \varphi) + \omega \sigma_B(T, \varphi) + \varepsilon_\varphi^2 \nabla^2 \varphi \quad (131)$$

with $u = (1 - \omega)u_A(T, \varphi) + \omega u_B(T, \varphi)$. Taking the time derivative of u gives

$$\begin{aligned} \dot{u} &= \dot{\omega}[u_A(T, \varphi) - u_A(T, \varphi)] \\ &+ (1 - \omega)\dot{u}_A(T, \varphi) + \omega \dot{u}_B(T, \varphi). \end{aligned} \quad (132)$$

The first term on the right of Eq. (132) is just

$$\dot{b}(\tilde{H}_B - \tilde{H}_A) = -(\tilde{H}_B - \tilde{H}_A)\nabla \cdot \mathbf{j}_b. \quad (133)$$

Thus Eq. (129) can be simplified and becomes

$$\begin{aligned} (1 - \omega)\dot{u}_A(T, \varphi) + \omega \dot{u}_B(T, \varphi) &= \nabla \cdot [k\nabla T] \\ &- \mathbf{j}_b \cdot \nabla(\tilde{H}_B - \tilde{H}_A). \end{aligned} \quad (134)$$

The quantity on the left hand side of Eq. (134) is just a compositionally weighted average of the time derivatives of the internal energies of A and B. If the heat capacities c_{AS}, c_{AL}, c_{BS}, and c_{BL} are all assumed to be equal to the same constant c_0 and $W_{uA} = W_{uB} = 0$, it may be simplified to the form $c_0 \dot{T} + \dot{p}(\varphi)[(1 - \omega)L_A + \omega L_B]$. With these same approximations, the quantity $\nabla(\tilde{H}_B - \tilde{H}_A) = (1/\rho)(L_B - L_A)\nabla p(\varphi)$. The energy equation Eq. (134) then becomes

$$\begin{aligned} c_0 \dot{T} &+ \dot{p}(\varphi)[(1 - \omega)L_A + \omega L_B] \quad (135) \\ &= \nabla \cdot [k\nabla T] \\ &- (1/\rho)(L_B - L_A)\mathbf{j}_B \cdot \nabla p(\varphi) \end{aligned}$$

Eq. (135) would even be simpler if $L_A = L_B$, but this might not be a good approximation because for materials of the same class, the latent heats L_A and L_B are approximately proportional (Richard's rule [90]) to the respective absolute melting temperatures, T_A and T_B, which generally differ considerably.

As yet, no one has solved simultaneously all three coupled phase field equations for alloys. Computations so far have been limited to isothermal solidification (driven by chemical diffusion)
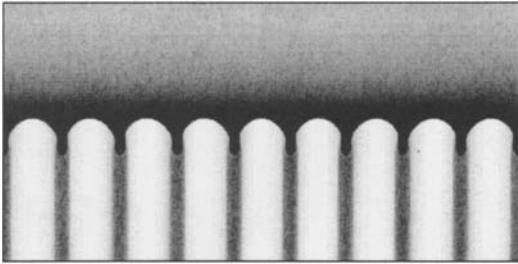
Figure 7. Solute field for a cellular interface growing at constant speed near the onset of morphological stability. The dark areas are rich in solute that segregates to the cell grooves.

Figure 8. Cellular interface growing at constant speed near the onset of morphological stability. The dotted curve is for $\varphi = 0.1$, the full curve for $\varphi = 0.5$ and the dashed curve for $\varphi = 0.9$. The cell height and width are shown in true ratio.

and cellular solidification in the 'frozen temperature' approximation in which Eq. (135) is ignored and a linear temperature gradient is imposed in a moving frame of reference. Two examples of the latter computations are shown in Figures 7 and 8 taken from the doctoral research of Bi [91]. Figure 7 shows the solute field in the vicinity of a cellular interface growing at constant speed near the onset of morphological stability. The dark areas are solute rich, and solute segregates strongly to the cell grooves. The solid behind the cell tips is relatively pure. Figure 8 shows the shape of a cellular interface in more detail. Three curves, corresponding to $\varphi = 0.1$, 0.5 and 0.9 are shown to illustrate the width of the diffuse interface in relation to the curvature of the cellular interface. The radii of the cell grooves are only a few times larger than the interface width but the radii of the cell tips are much larger than the interface width.

5. HYDRODYNAMICS

As mentioned previously, self-consistent phase field models that include hydrodynamics have been formulated [68–73] and used to some extent in computations. All of these theories are based on a model that treats the solid as a very viscous isotropic liquid. The results are quite complicated and lead to such things as Korteweg stresses [92] that can drive convection. It is important to note,

however, that the equations for the phase field and the accompanying thermal and solutal fields have essentially the same form as without convection, except that[13]

$$\frac{\partial}{\partial t} \to \frac{D}{Dt} := \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla \qquad (136)$$

where $\mathbf{v}$ is the barycentric velocity of the fluid. The quantity on the right hand side of Eq. (136) is recognized as the material derivative with respect to time. In other words, the phase field equations become invariant under a Galilean transformation, as would be expected. Anderson, McFadden and Wheeler [93] have recently used asymptotics to analyze the sharp interface limit of a model that includes convection. The reader is referred to the literature for details.

6. CONCLUSIONS

As we have shown, it is possible to formulate thermodynamically consistent phase field models for solidification from the melt, for pure materials and for alloys, even including hydrodynamics. The complexity of these models, however, increases drastically with the complexity of the system being modeled. Using these models for

---

[13]One must also be consistent with a formulation per unit mass, rather than per unit volume, because mass is conserved but local density is not [73].

computations is certainly possible, especially for pure materials, and has led to a number of interesting results. As computational power and algorithms improve, the more complex models should become tractable [94–96]. Exploring these models is a worthwhile goal because they contain a great deal of the known physics of solidification phenomena at length scales above the molecular level. This potentially includes the study of some aspects of nanophase solidification, although it is not clear how much of the correct physics is contained at the nanoscale level.

REFERENCES

1. B.I. Halperin, P.C. Hohenberg and S.-K. Ma, Phys. Rev. B 10 (1974) 139.
2. J.S. Langer, unpublished notes (August 1978).
3. G.J. Fix, in Free Boundary Problems: Theory and Applications, Vol. II, eds. A. Fasano and M. Primicerio (Pitman, Boston, 1983) p. 580
4. J.B. Collins and H. Levine, Phys. Rev. B 31 (1985) 6119.
5. J.S. Langer, Directions in Condensed Matter Physics, eds. G. Grinstein and G. Mazenko (World Scientific, Singapore, 1986) p. 165
6. J.W. Cahn, Acta. Met. 8 (1960) 554.
7. G. Caginalp, in Applications of Field Theory to Statistical Mechanics, ed. L. Garrido, Lecture Notes in Physics, Vol. 216 (Springer-Verlag, Berlin, 1985) p. 216
8. G. Caginalp, in Material Instabilities in Continuum Problems and Related Mathematical Problems, ed. J.M. Ball (Oxford Univ. Press, Oxford, 1988) p. 35
9. G. Caginalp and P.C. Fife, Phys. Rev. B 33 (1986) 7792.
10. G. Caginalp, Arch. Rat. Mech. Anal. 92 (1986) 205.
11. G. Caginalp, Ann. Physics (NY) 172 (1986) 136.
12. G. Caginalp and P.C. Fife, SIAM J. Appl. Math. 48 (1988) 506.
13. G. Caginalp, Phys. Rev. A 39 (1989) 5887.
14. G. Caginalp and E.A. Socolovsky, J. Comp. Phys. 95 (1991) 85.
15. G. Caginalp and X. Chen, in On the Evolution of Phase Boundaries, eds. M.E. Gurtin and G.B. McFadden, The IMA Volumes in Mathematics and Its Applications, Vol. 43 (Springer-Verlag, Berlin, 1992) p. 1
16. G.B. McFadden, A.A. Wheeler, R.J. Braun, S.R. Coriell and R.F. Sekerka, Phys. Rev. E 48 (1993) 2016
17. A.A. Wheeler and G.B. McFadden, Eur. J. Appl. Math 7 (1996) 367
18. A.A. Wheeler and G.B. McFadden, Proc. Roy. Soc. London A 453 (1997) 1611
19. O. Penrose and P.C. Fife, Physica D 43 (1990) 44.
20. E. Fried and M.E. Gurtin, Physica D 68 (1994) 326.
21. E. Fried and M.E. Gurtin, Physica D 72 (1994) 287.
22. S-L. Wang, R.F. Sekerka, A.A. Wheeler, B.T. Murray, S.R. Coriell, R.J. Braun and G.B. McFadden, Physica D 69 (1993) 189.
23. R. Kobayashi, Bull. Jpn. Soc. Ind. Appl. Math. 1 (1991) 22.
24. R. Kobayashi, Physica D 63 (1993) 410.
25. R. Kobayashi, in Pattern Formation in Complex Dissipative Systems, ed. S. Kai (World Scientific, Singapore, 1992) p. 121
26. A.A. Wheeler, B.T. Murray and R.J. Schaefer, Physica D 66 (1993) 243.
27. B.T. Murray, W.J. Boettinger, G.B. McFadden and A.A. Wheeler, in Heat Transfer in Melting, Solidification and Crystal Growth, eds. I.S. Habib and S. Thynell (ASME HTD-234, New York, 1993) p. 67
28. B.T. Murray, A.A. Wheeler and M.E. Glicksman, J. Crystal Growth 154 (1995) 386.
29. S.-L Wang and R.F. Sekerka, Phys. Rev. E 53 (1996) 3760.
30. S.-L Wang and R.F. Sekerka, J. Comp. Phys. 127 (1996) 110.
31. N. Provatas, N. Goldenfeld and J. Dantzig, Phys. Rev. Lett. 80 (1998) 3306.
32. N. Provatas, N. Goldenfeld and J. A. Dantzig, J. Comp. Phys. 148 (1999) 265.
33. Y.T. Kim, N. Provatas, N. Goldenfeld and J. A. Dantzig, Phys. Rev. E 59 (1999) 2549.
34. N. Provatas, N. Goldenfeld, J. A. Dantzig, J.C. LaCombe, A. Lupulescu, M.B. Koss, M.E. Glicksman and R. Almgren, Phys. Rev.

Lett. 82 (1999) 4496.

35. N. Provatas, N. Goldenfeld and J. A. Dantzig, in Modeling of Casting, Welding, and Advanced Solidification Processes, eds. B.G. Thomas and C. Beckermann (TMS, San Diego, 1998) p. 533

36. A. Karma and W.-J. Rappel, Phys. Rev. E 53 (1996) 3017.

37. A. Karma and W.-J. Rappel, Phys. Rev. Lett. 78 (1996) 4050.

38. A. Karma and W.-J. Rappel, Phys. Rev. E 57 (1998) 4323.

39. A.A. Wheeler and W.J. Boettinger, in On the Evolution of Phase Boundaries, eds. M.E. Gurtin and G.B. McFadden, The IMA Volumes in Mathematics and Its Applications, Vol. 43 (Springer-Verlag, Berlin, 1992) p. 127

40. A.A. Wheeler, W.J. Boettinger, and G.B. McFadden, Phys. Rev. A 45 (1992) 7424.

41. A.A. Wheeler, W.J. Boettinger, and G.B. McFadden, Phys. Rev. E 47 (1993) 1893.

42. W.J. Boettinger, A.A. Wheeler, G.B. McFadden and R. Kobayashi, in Modeling of Coarsening and Grain Growth, eds. S.P. Marsh and C.S. Pande (TMS, Warrendale, 1993) p. 45

43. Z. Bi and R. F. Sekerka, Physica A 261 (1998) 95.

44. Ch. Charach and P.C. Fife, SIAM J. Appl. Math. 58 (1998) 1826.

45. Ch. Charach and P.C. Fife, Open Systems, Information Dynamics 5 (1998) 99.

46. W.J. Boettinger, A.A. Wheeler, B.T. Murray and G. B. McFadden, Material Science and Engineering, A178 (1994) 217.

47. M. Conti, Phys. Rev. E 55 (1997) 701.

48. M. Conti, Phys. Rev. E 55 (1997) 765.

49. N.A. Ahmad, A.A. Wheeler, B.J. Boettinger and G.B. McFadden, Phys. Rev. E 58 (1998) 3436.

50. Ch. Charach, C.K. Chen and P.C. Fife, J. Stat. Phys. 95 (1999) 1141.

51. Ch. Charach and P.C. Fife, J. Crystal Growth 198/199 (1999) 1267.

52. J.A. Warren and W.J. Boettinger, Acta Metall. Mater. 43 (1995) 689.

53. M. Conti, Phys. Rev. E 56 (1997) 3197.

54. J.A. Warren and W.J. Boettinger, in Modeling of Casting, Welding and Advanced Solidification Processes VII, eds. M. Cross and J. Campbell (TMS, Warrendale, PA, 1995) p. 601

55. J.A. Warren and W.J. Boettinger, in Solidification Processing 1997, eds. J. Beach and H. Jones (Department of Engineering Materials, University of Sheffeld, UK, 1997) p. 422

56. J.A. Warren and B.T. Murray, Mod. Simul. Mater. Sci. Eng. 4 (1996) 215.

57. W.J. Boettinger and J.A. Warren, Met. Trans. A 27 (1996) 657.

58. W.J. Boettinger and J.A. Warren, J. Crystal Growth 200 (1999) 583.

59. A. Karma, Phys. Rev. E 49 (1993) 2245.

60. A.A. Wheeler, G.B. McFadden and W.J. Boettinger, Proc. R. Soc. Lond. A 452 (1996) 495.

61. K.R. Elder, J.D. Gunton, and M. Grant, Phys. Rev. E 54 (1996) 6476.

62. G. Caginalp and J. Jones, App. Math Lett. 4 (1991) 97.

63. G. Caginalp and J. Jones, in On the Evolution of Phase Boundaries, eds. M.E. Gurtin and G.B. McFadden, The IMA Volumes in Mathematics and Its Applications, Vol. 43 (Springer-Verlag, Berlin, 1992) p. 27

64. R. Tönhardt, Doctoral Thesis, Department of Mechanics, Royal Institute of Technology, Stockholm, Sweden (1998)

65. R Tönhardt and G. Amberg, J. Crystal Growth 194 (1998) 406.

66. C. Beckermann, H.-J. Diepers, I. Steinbach, A. Karma and X. Tong, J. Comp. Phys. 154 (1999) 468.

67. H.J. Diepers, C. Beckermann and I. Steinbach, in Solidification Processing 1997, eds. J. Beach and H. Jones (Department of Engineering Materials, University of Sheffeld, UK, 1997) p. 426

68. M.E. Gurtin, D. Polignone and J. Viñals, Math. Models. Methods Appl. Sci. 6 (1996) 815.

69. D.M. Anderson, G.B. McFadden and A.A. Wheeler, Annu. Rev. Fluid Mech. 30 (1998) 139.

70. D.M. Anderson, G.B. McFadden, Phys. Fluids 9 (1997) 1870.

71. D.M. Anderson, G.B. McFadden and A.A.

Wheeler, Physica D 135 (2000) 175

72. D.M. Anderson, G.B. McFadden and A.A. Wheeler, "A Phase-Field Model of Solidification with Convection: Numerical Simulations," to appear in Interfaces for the Twenty-First Century, (Imperial College Press, London 2001)

73. R. F. Sekerka and Z. Bi, "Phase Field Model of Multicomponent Alloy with Hydrodynamics," to appear in Interfaces for the Twenty-First Century, (Imperial College Press, London 2001)

74. R.J. Braun, J.W. Cahn, G.B. McFadden and A.A. Wheeler, Phil. Trans. R. Soc. Lond. A 355 (1997) 1787

75. R.J. Braun, J.W. Cahn, G.B. McFadden, H.E. Rushmeier and A.A. Wheeler, Acta mater. 46 (1998) 1

76. R.J. Braun, J. Zhang, J.W. Cahn, G.B. McFadden and A.A. Wheeler, "Model Phase Diagrams for an FCC Alloy," to appear in Interfaces for the Twenty First Century, (Imperial College Press, London 2001)

77. R. Kobayashi, J.A. Warren and W.C. Carter, Physica D 140 (2000) 141

78. J.W. Cahn and J.E. Hilliard, J. Chem. Phys. 28 (1958) 258.

79. G.B. McFadden, A.A. Wheeler and D.M. Anderson, Physica D 144 (2000) 154

80. J. Fernandez-Diaz and W.O. Williams, J. Appl. Math. and Physics 30 (1979) 749

81. D.W. Hoffman and J.W. Cahn, Surface Science 31 (1972) 368

82. J.W. Cahn and D.W. Hoffman, Acta. Met. 22 (1974) 1205

83. G.B. McFadden, S.R. Coriell and R.F. Sekerka, Acta. mater. 48 (2000) 3177

84. Shun-Lien Wang, "Computation of Dendritic Growth at Large Supercoolings by Using the Phase Field Model," Doctoral Thesis, Carnegie Mellon University 1995

85. L.D. Landau and E.M. Lifshitz, Fluid Mechanics (Pergamon Press, London, 1959)

86. S.R. de Groot and P. Mazur, Non-Equilibrium Thermodynamics (Dover Edition, New York, 1984)

87. D.D. Fitts, Non-equilibrium Thermodynamics (McGraw-Hill, New York, 1962)

88. R.F. Sekerka, "Notes on Entropy Production in Multicomponent Systems," unpublished but available at http://sekerka.phys.cmu.edu

89. R.F. Sekerka and W.W. Mullins, J. Chem. Phys. 73 (1980) 1413.

90. L.S. Darken and R.W. Gurry, Physical Chemistry of Metals, (McGraw-Hill, New York, 1953)

91. Zhiqiang Bi, Doctoral Research, Carnegie Mellon University, to be completed in 2001.

92. D.J. Korteweg, Arch. Néerl. Sci. Exactes Nat. Ser. II 6 (1901)

93. D.M. Anderson, G.B. McFadden and A.A. Wheeler, Physica D 2711 (2001) 1

94. M. Plapp and A. Karma, PRL 84 (2000) 1740

95. M. Plapp and A. Karma, J. Comp. Phys. 165 (2000) 592

96. A. Karma, "Phase-Field Formulation for Quantitative Modeling of Alloy Solidification, (arXiv:cond-mat/0103289) to be published

Generic mechanism of heterogeneous nucleation and molecular interfacial effects

Xiang Y. Liu

Department of Physics, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260
Email: phyliuxy@nus.edu.sg

A generic mechanism of heterogeneous nucleation is considered based on thermodynamic and kinetic theories along with applications of the theory to crystal growth and crystallization. In terms of a so-called interface correlation factor $f(m,R')$ and the entropic barriers associated with kink integration, both classic and non-classic interfacial effects and particle size effect are treated quantitatively. Within the framework of this mechanism, homogenous nucleation of the crystalline phase can be regarded as a special case of heterogeneous nucleation.

1. INTRODUCTION

The formation of a new phase in the body of the mother phase, such as gas, liquid, ions in general or solid, is one of the most fundamental aspects of phase transition, in particular, crystallization. The potential barrier which a system must overcome in order to create a (crystalline) nucleus in the mother phase and which determines the rate of nucleation is defined by the interface energy. Under a given condition, if the probability of creating a nucleus is homogeneous thorough out the system, the nucleation is defined as homogeneous nucleation. Otherwise, it is defined as heterogeneous nucleation. In heterogeneous nucleation on solid or liquid surfaces, microclusters, dusts and macromolecules, the properties of these foreign bodies are an additional factor upon which this barrier and rate depend. Since these foreign bodies occur extensively in various systems, in most cases, nucleation has the heterogeneous nature rather than the homogeneous nature. In this Chapter, heterogeneous nucleation will be discussed from both the macroscopic and microscopic points of view. Within such an approach, homogeneous nucleation can be treated as an upper limit of heterogeneous nucleation.

A general and simple picture of 3D nucleation can be described as follows. The constituent atoms or molecules in the solution may, on collision, join into groups of two, three, four, or more particles, forming dimers, trimers, tetramers, etc. The kinetics of nucleation is described by the nucleation rate J, which is defined as the number of nuclei created per unit volume-time, and determined by the nucleation barrier, kink integration rate, transport and other factors. J is an important characteristic of the process of new phase formation.

The central problem in the nucleation theory and experiment is to find J as a function of the parameters controlling the process. After the thermodynamic results of Gibbs [1], the paper by Volmer and Weber [2] in 1926 was the very first one devoted to J and it was followed by the pioneering studies of Farkas [3], Kaischew, Stranski and others [4-6]. To describe the kinetics of nucleation, many classic and non-classic theories have been published [7-25]. Many important results have been obtained since then and the aim of this Chapter is to help in getting a basic knowledge of the modern theory of heterogeneous nucleation. The Chapter will be arranged as follows: the thermodynamics (Section 2) and the kinetics of nucleation (Section 3) and then the effects of foreign bodies and additives on nucleation (Section 4), and two-dimensional nucleation controlled crystal growth (Section 5). The general remarks and conclusions are given in Section 6.

2. NUCLEATION BARRIER

2.1. Thermodynamic driving force

Nucleation is the process that the first-order phase transitions begin with. The driving force for nucleation of new phases (e.g., crystals) is $\Delta\mu$, which is defined as the difference between the chemical potentials μ_{mother} and $\mu_{crystal}$ of growth unit

in the mother and the crystalline phases:

$$\Delta\mu = \mu_{\text{mother}} - \mu_{\text{crystal}} \qquad (1)$$

When $\Delta\mu > 0$, it is said that the system is supersaturated. This is the thermodynamic pre-condition for nucleation and growth of the crystalline phase. Conversely, when $\Delta\mu < 0$, the system is undersaturated. Under such conditions, crystals will dissolve. In the case where $\Delta\mu = 0$, the mother phase is in equilibrium with the crystalline phase [27]. This implies that under the given temperature T, pressure P, concentration C_i, etc., one always has

$$\mu_{\text{mother}}^{\text{eq}} = \mu_{\text{crystal}} \qquad (2)$$

where $\mu_{\text{mother}}^{\text{eq}}$ is the chemical potential of solute molecules in the phase equilibrium (or coexistence) between the mother and the crystalline phases. It follows that for a given condition, μ_{crystal} can be expressed by $\mu_{\text{mother}}^{\text{eq}}$. Therefore, in many cases of practical importance $\Delta\mu$ can be expressed as

$$\Delta\mu = \mu_{\text{mother}} - \mu_{\text{mother}}^{\text{eq}} \qquad (3)$$

For condensation of vapors, the chemical potential of species i is given by [7,12,27]

$$\mu_i = \mu_i^\circ + kT \ln p_i \qquad (4)$$

where p_i is the actual partial pressure of vapor i, μ_i° denotes the standard state ($p_i = 1$) of the chemical potential, k is the Boltzmann constant, T is the absolute temperature. Based on Eqs.(2) and (4), it can be shown that the thermodynamic driving force is given by

$$\Delta\mu = kT \ln\left(p_i / p_i^{\text{eq}}\right) \qquad (5)$$

(p_i^{eq} is the equilibrium partial pressure of vapor i.) Similarly, for deposition of thin films from vapors, the thermodynamic driving force is given by [10,12, 27]

$$\Delta\mu = kT \ln(R/R_e) \qquad (6)$$

(R and $R_e(T)$ are the actual and equilibrium rates of impingement of molecules to the substrate).

For crystallization from solutions, the chemical potential of species i is given by [8,12,27]

$$\mu_i = \mu_i^\circ + kT \ln a_i \approx \mu_i^\circ + kT \ln C_i \qquad (7)$$

where a_i and C_i denote the activities and concentrations of solute, respectively, μ_i° denotes the standard state ($a_i = 1$) of the solute chemical potential. This then gives rise to the thermodynamic driving force

$$\Delta\mu = kT \ln\frac{a_i}{a_i^{\text{eq}}} \approx kT \ln\frac{C_i}{C_i^{\text{eq}}} \qquad (8)$$

(a^{eq}, C^{eq} are, respectively, the equilibrium activities and concentrations of solute.)

Notice that the thermodynamic driving force for crystallization is often expressed in terms of supersaturation. If we define supersaturation as

$$\sigma = \left(a_i - a_i^{\text{eq}}\right)/a_i^{\text{eq}} \approx \left(C_i - C_i^{\text{eq}}\right)/C_i^{\text{eq}} \qquad (9)$$

Eq.(8) can then be rewritten as

$$\Delta\mu = kT \ln(1+\sigma) \qquad (10)$$

If $\sigma < 1$, Eq.(10) can be approximated, after the Taylor series expansion, as

$$\Delta\mu/kT = \ln(1+\sigma) \cong \sigma. \qquad (11)$$

The crystallization of some ionic crystals from solutions often results from mixing the cationic and anionic components, such as

$$m\text{A}^{n+} + n\text{B}^{m-} \rightarrow \text{A}_m\text{B}_n \qquad (12)$$

(n and m are, respectively, the cationic and anionic valence.) In this case, Eq.(8) should be given as

$$\Delta\mu = kT \ln\frac{\left(a_{A^{n+}}\right)^m\left(a_{B^{m-}}\right)^n}{\left(a_{A^{n+}}^{\text{eq}}\right)^m\left(a_{B^{m-}}^{\text{eq}}\right)^n} = kT \ln\frac{\left(a_{A^{n+}}\right)^m\left(a_{B^{m-}}\right)^n}{K_{sp}} \qquad (13)$$

where $K_{sp}\left(=\left(a_{A^{n+}}^{\text{eq}}\right)^m\left(a_{B^{m-}}^{\text{eq}}\right)^n\right)$ is the participation constant. Similarly, if we define in this case supersaturation as

$$\sigma = \frac{\left(a_{A^{n+}}\right)^m\left(a_{B^{m-}}\right)^n - K_{sp}}{K_{sp}}, \qquad (14)$$

the relation between the thermodynamic driving force and supersaturation is the same as given by Eq.(11).

For crystallization from melts at temperatures not far below the melting temperature, we have the thermodynamic driving force by applying the similar thermodynamic principles, as [12, 19]

$$\Delta\mu = s_m \Delta T, \tag{15}$$

$$\Delta T = (T_m - T) \tag{16}$$

(s_m is the entropy of melting per molecule, ΔT is supercooling).

For electro-deposition, the thermodynamic driving force is given by [20]

$$\Delta\mu = z_i e (\phi^e - \phi^e_{eq}), \tag{17}$$

(z_i is the ion valence, e is the electron charge, ϕ^e is the actual electrode potential, ϕ^e_{eq} is the equilibrium potential of a bulk electrode of the deposit).

2.2. Nucleation barrier

A characteristic feature of nucleation process is that the substance with the properties of the new phase appears fluctuationally and is localized in nano-scale small spatial regions. These are occupied by atoms or molecules of various numbers which constitute the so-called clusters. The clusters staying in equilibrium with the surrounding mother phase are the critical nuclei, and the smaller or the larger clusters are the subnuclei or the supernuclei, respectively. Only the supernuclei are the clusters that can grow spontaneously to reach macroscopic sizes. (See Fig.1.) For simplicity, we call hereafter the subnuclei "clusters", and the supernuclei "nuclei".

Evidently, the nucleation rate J describing the number of nuclei successfully generated from the population of clusters per unit time, unit volume is determined by the height of a free energy barrier, so-called nucleation barrier (c.f. Fig.1). The occurrence of nucleation barrier is attributed to the following two-conservancy effects:
(1) since the crystalline phase is a stable phase, the occurrence of the new phase from the mother phase will lead to the lowering of the (Gibbs) free energy of the system;

(2) due to the interfacial (or surface) free energy, the increase in the size of the crystalline (new) phase leads to the increase of interface (or surface) area, consequently the interface (or surface) free energy. This will cause the increase of the Gibbs free energy of the system.

The combination of these two effects results in the formation of the nucleation barrier, as shown in Fig.1.
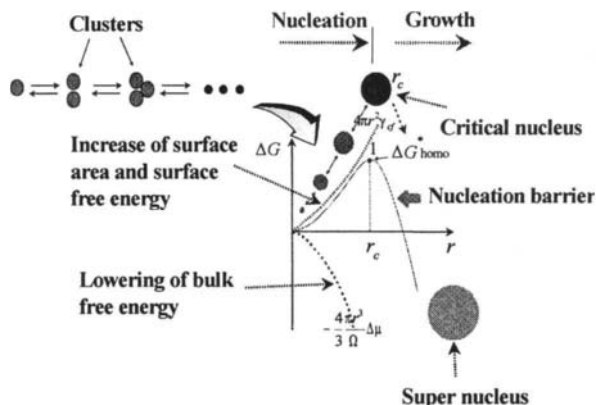
Figure 1. Schematic illustration of the formation of nucleation barrier.

The free energy change to form a cluster of $n = 1,2,3....$ molecules can be found by thermodynamic considerations, since it is defined as

$$\Delta G = G_{fin} - G_{ini} \tag{18}$$

for a system at constant pressure and temperature (G_{ini} and G_{fin} are the Gibbs free energies of the system in the initial and final states before and after the cluster formation). If M is the number of solute molecules in the system, G_{ini} is merely given by

$$G_{ini} = M\mu_{mother} \tag{19}$$

To find G_{fin}, one usually employs the Gibbs method [1] of introducing a surface which divides the system into a new phase of n molecules, a cluster of n molecules, and an old phase of the remaining $M - n$ solute molecules. Then G_{fin} is written down as

$$G_{fin} = (M-n)\mu_{mother} + n\mu_{crystal} + \Phi_n \tag{20}$$

where Φ_n is the total surface energy of the n-sized cluster (except for nucleation of bubbles when Φ_n contains also pressure-volume terms). Substituting

the above expressions (19) and (20) for G_{ini} and G_{fin} in Eq.(18) and recalling Eq.(1) yields

$$\Delta G = -n\Delta\mu + \Phi_n \qquad (21)$$

The ΔG function has a maximum of value ΔG^* at $r = r_c$, or $n = n^*$. The cluster of n^* molecules is the critical nucleus, r_c is the radius of curvature of critical nuclei and ΔG^* is the nucleation barrier. One of the major problems in the nucleation theory is to find ΔG^* which, physically, is the energy barrier of nucleation.

If the occurrence of foreign body in the system will reduce the interfacial (or surface) free energy, it will also lower the nucleation barrier. Let ΔG^*_{homo} be the homogeneous nucleation barrier, and ΔG^*_{heter} be the heterogeneous nucleation barrier (the nucleation barrier in the presence of the foreign body.) We can define here a factor describing the lowering of the nucleation barrier due to foreign body:

$$f = \Delta G^*_{heter} / \Delta G^*_{homo}. \qquad (22)$$

It will be shown in the following how ΔG^*_{heter} and f are derived [24,25].
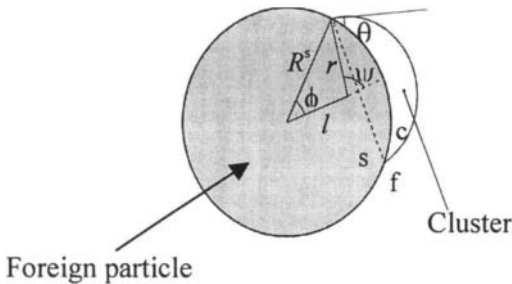
Figure 2. A generic picture of nucleation on the foreign particle.

As shown in Fig.2, we assume that nucleation occurs at a foreign body with a radius of R^s. The mother phase is represented by subscript f, the cluster of the crystalline phase by c and the foreign body by s. If we denote volume by V and surface area of the foreign body by S, then the free energy of formation of a cluster of radius r on a nucleating particle of radius R^s is given, according to Eq.(21), by,

$$\Delta G = -\Delta\mu V_c/\Omega + \gamma_{cf}S_{cf} + (\gamma_{sf} - \gamma_{sc})S_{sc} \qquad (23)$$

where γ_{ij} is the surface free energy between phases i and j and Ω is the volume per structural unit. Assume that the concept of contact angle can still be applied in this case. We have then

$$m = (\gamma_{sf} - \gamma_{sc})/\gamma_{cf} \approx \cos\theta, \quad (-1 \le m \le 1) \qquad (24)$$

Referring again to Fig.2, we have,

$$S_{sc} = 2\pi(R^s)^2(1 - \cos\phi), \ S_{cf} = 2\pi r^2(1 - \cos\psi),$$
and
$$V_c = \tfrac{1}{3}\pi r^3(2 - 3\cos\psi + \cos^3\psi) - \tfrac{1}{3}\pi(R^s)^3(2 - 3\cos\phi + \cos^3\phi), \qquad (25)$$
with
$$\cos\phi = (R^s - r\cos\theta)/l = (R^s - rm)/l, \qquad (26)$$

$$\cos\psi = -(r - R^s\cos\theta)/l = -(r - R^s m)/l, \qquad (27)$$
and
$$l = [(R^s)^2 + r^2 - 2R^s rm]^{1/2}. \qquad (28)$$

To evaluate the critical free energy ΔG^*_{heter}, we can substitute the expression (25) into (23) and require that

$$(\partial\Delta G/\partial r) = 0. \qquad (29)$$

Regarding the fact that the radius of curvature r_c of critical nuclei is only determined by γ_{cf} and the driving force $\Delta\mu$ [28,29], we then have

$$r_c = 2\Omega\gamma_{cf}/\Delta\mu = 2\Omega\gamma_{cf}/kT\ln(1 + \sigma) \qquad (30)$$

Now substituting the expressions (24)-(30) into Eq.(23) and writing,

$$R' = R^s/r_c = R^s\Delta\mu/\Omega\gamma_{cf} = R^s kT\ln(1 + \sigma)/\Omega\gamma_{cf}. \qquad (31)$$

The free energy of formation of critical nucleus is

$$\Delta G^*_{heter} = \Delta G^*_{homo} f \qquad (32)$$
with

$$\Delta G^*_{homo} = \frac{16\pi\gamma_{cf}^3\Omega^2}{3[kT\ln(1 + \sigma)]} \qquad (33)$$

$$f = f(m, R')$$

$$= \frac{1}{2} + \frac{1}{2}\left(\frac{1-mR}{w}\right)^2 + \frac{1}{2}R'^3\left[2 - 3\left(\frac{R'-m}{w}\right) + \left(\frac{R'-m}{w}\right)^2\right]$$

$$+ \frac{3}{2}mR'^2\left(\frac{R'-m}{w} - 1\right) \qquad (34)$$

and

$$w = [1 + (R')^2 - 2R'm]^{1/2}. \qquad (35)$$

Here R' is actually the dimensionless radius of curvature of the foreign body (substrate) in reference to the radius of critical nuclei r_c. In other words, it only makes sense if the curvature of a foreign body or a substrate is referred to the curvature of critical nuclei.

Substituting appropriate values of R^s, m, γ_{cf} and $\Delta\mu$ into (30)-(35), one can calculate $f(m,R')$ and ΔG^*_{heter} for any nucleation process. Note that the factor $f(m,R')$ varies from 1 to 0. Obviously, this factor plays an important role in determination of the heterogeneous nucleation barrier ΔG^*_{heter}. One can see from Eq.(10) that the influence of foreign particles on the nucleation barrier can be fully characterized by this factor.

In Fig.3a is shown $f(m, R')$ as a function of R' for a given m. When $R' \rightarrow 0$, $f(m,R') = 1$, implying that the foreign body as a nucleating substrate "vanish" completely. In practice, if foreign bodies are too small, c.f. clusters of several molecules, nucleation on these substrates will not be stable. Then, the foreign bodies play no role in lowering the nucleation barrier. If $R' >> 1$, the foreign body can be treated as a flat substrate compared to the critical nuclei. In this case, $f(m,R')$ is a solo function of m, and the curvature of the foreign body has no effect on the nucleation kinetics. Eq.(34) is then reduced to [10,11]

$$f(m,R') = f(m) = \frac{1}{4}(2 - 3m + m^3) \qquad (36)$$

The dependence of $f(m)$ on m is given in Fig. 3b.

Concerning the effect of m, as shown in Fig.3, heterogeneous nucleation occurs in the range of m between 1 and -1, depending on the interaction and the structural match between of the foreign body surface and the nucleating phase, and correspondingly $f(m,R')$ changes between 0 and 1. Evidently, the strong interaction between substrate molecules and structural units in the nucleating phase leads to $m \rightarrow 1$. Apart from this, the structural

match between the foreign body surface and the nucleating phase plays an important role, in particular for the nucleation of a crystalline phase. Let us look at the structural match between the nucleating phase and the substrate in this case. According to Aleksandrov [30], for a good substrate, the interfacial energy between the substrate and the nucleating phase can be quantitatively expressed, if the misorientation angle φ is low:

$$\gamma_{cs} \approx \gamma_{cs}(\alpha) + \frac{\varepsilon b \varphi}{4\pi(1-\nu)}\left(1 - \frac{\varphi - \varphi_{max}}{\varphi_{max}}\right) \qquad (37)$$

$\gamma_{cs}(\alpha)$ = minimal specific interfacial free energy at a given orientation α;

ε = elastic modulus;

ν = Poisson constant;
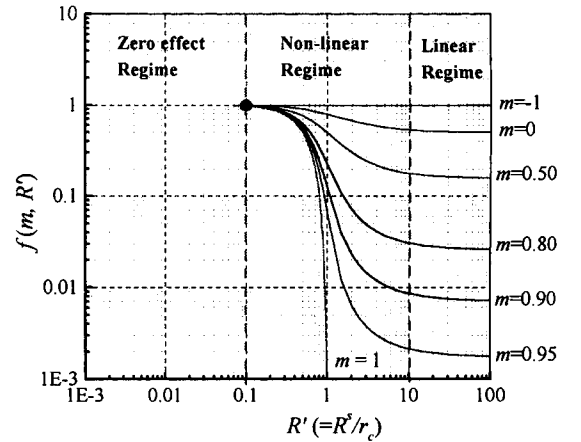
(a)
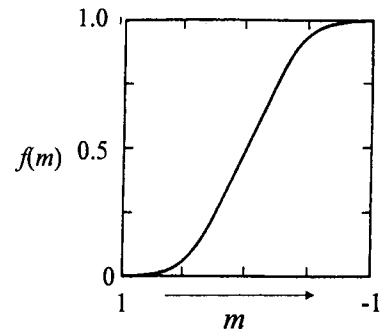
(b)

Figure 3. (a) Dependence of the interfacial correlation function $f(m,R')$ on m and R'; (b) Dependence of the interfacial correlation function $f(m,R')$ on m at $R' > 10$.

b = Burgers vector,

φ = misorientation angle from the given orientation.

Presuming the nature of the bonding between the substrate and the crystalline phase at the interface is similar to the bulk of crystalline phase, $\gamma_{cs}(\alpha)$ is dependent very much on the structural match between the substrate and the nucleating phase.

We notice that for a given crystalline phase and a substrate, an optimal structural match is the crystallographic orientation $\{hkl\}$, which the strongest *average* interaction or the lowest interfacial energy between the crystalline phase and the substrate between the two phases. This orientation corresponds to the (minimal) cusp at the γ-plot [31]. In our case, the plot is a $\gamma_{cs}(\alpha)$ plot (α: the angle to the reference orientation, c.f. Fig.4).
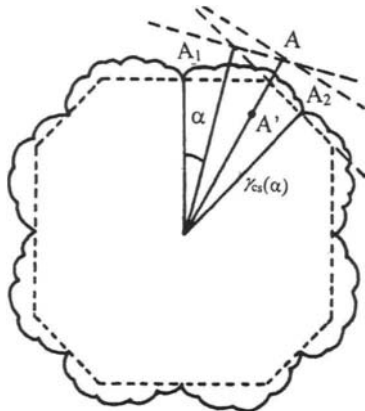
Figure 4. A $\gamma_{cs}(\alpha)$ plot in two dimension.

Combining Eqs.(24) and (36) yields

$$m \approx \frac{\gamma_{sf}}{\gamma_{cf}}\left\{1 - \frac{1}{\gamma_{sf}}\left[\gamma_{cs}(\alpha) + \frac{\varepsilon b \varphi}{4\pi(1-\nu)}\left(1 - \frac{\varphi - \varphi_{max}}{\varphi_{max}}\right)\right]\right\} \quad (38)$$

Obviously, an excellent structural match between the nucleating phase and the substrate $(\gamma_{cs}(\alpha)\,\varphi \to 0$, at $\varphi \to 0)$ leads to $m \to \gamma_{sf}/\gamma_{cf}$. In the case where $\gamma_{sf} \approx \gamma_{cf}$, one has then $m \to 1$. It follows from Eq.(34) that $f(m,R') \to 0$ (c.f. Fig.3), implying that ΔG_{heter}^{*} almost vanishes completely (c.f. Eq.(22).) This only occurs when the growth of crystals is well orientated in reference to the

structure of the substrate. In this case, the excellent epitaxial relation assumes.

As the structural match varies from a perfect to a poor match, m decreases from 1 to 0, -1. The extreme case will be $m \to$ -1, corresponding to the situation where the crystal-substrate correlation (interaction and structural match between nuclei and the substrate) does not exist. One has then $f(m,R')$ = 1 in this case (c.f. Fig. 3.) In this case, foreign bodies exert almost no influence on nucleation (the vanishing of the epitaxial effect), and nucleation is controlled by the homogeneous nucleation kinetics.
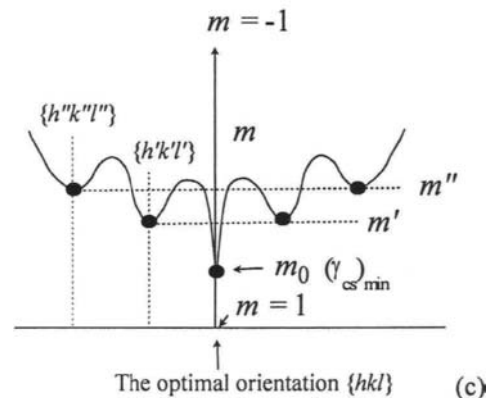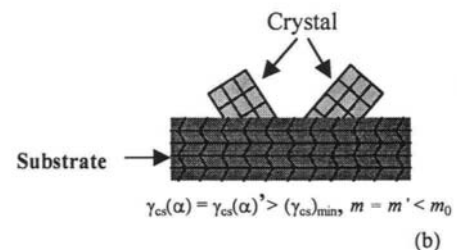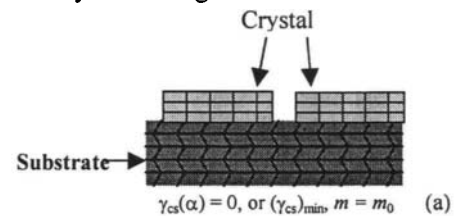
Figure 5. Schematic illustration of the structural match (a) and mismatch (b) between the crystalline phase and the substrate, and the implication for the structural matching parameter m (c).

As shown in Fig.4, due to the anisotropy of a crystalline phase, the deviations from the optimal structural match position to the secondary optimal

structural match (the second lowest $\gamma_{cs}(\alpha)$) will adopt a discrete value rather than a continuous change, which is the second lowest minimum of γ_{cs} in the orientation of $\{h'k'l'\}$. The similar principle holds for the further deviations. (See Figs.4,5.) Therefore, according to expressions (24) and (37), the deviations from the optimal structural match cause the change of m from m_0 to lower and discrete $m = m'$, m'', etc. (c.f. Fig.5.) This will be discussed further in Sec.4.

In many cases, one needs the expression ΔG^* as a function of n^*. Geometrically, in many cases of heterogeneous nucleation, [7,12]

$$f \cong V^* / V^*_{homo} = n^* / n^*_{homo} , \qquad (39)$$

where V^* and V^*_{homo}, n^* and n^*_{homo} are volumes and the number of the heterogeneously and homogeneously formed nucleus at the given supersaturation. It follows from Eqs.(22), (34) and (39) that

$$n^* = n^*_{homo} f_{n^*}\left(m, R^s, n^*_{homo}\right), \quad n^*_{homo} = \frac{32}{3(\Delta\mu)^3}\Omega^2 \gamma^3_{cf} \quad (40)$$

$$\Delta G^*_{heter}(\Delta\mu) = \frac{16}{3(\Delta\mu)^2}\Omega^2 \gamma^3_{cf} f\left(m, R^s, n^*_{homo}\right) \qquad (41)$$

with

$$f_{n^*}\left(m, R, n^*_{homo}\right) = \frac{1}{2} + \frac{1}{2}\left(\frac{1-mR}{w_{n^*}\left(3n^*_{homo}\Omega/4\pi\right)^{1/2}}\right)^2 + \frac{1}{2}\left(\frac{R^s}{\left(3n^*_{homo}\Omega/4\pi\right)^{1/2}}\right)^3$$

$$\times\left[2 - 3\left(\frac{R^s/\left(3n^*_{homo}\Omega/4\pi\right)^{1/2} - m}{w_{n^*}}\right) + \left(\frac{R^s/\left(3n^*_{homo}\Omega/4\pi\right)^{1/2} - m}{w_{n^*}}\right)^2\right]$$

$$+ \frac{3}{2}m\left(\frac{R^s}{\left(3n^*_{homo}\Omega/4\pi\right)^{1/2}}\right)^2\left(\frac{R^s/\left(3n^*_{homo}\Omega/4\pi\right)^{1/2} - m}{w_{n^*}} - 1\right)$$

$$\qquad (42)$$

and

$$w_{n^*} = \{1 + [R^s/\left(3n^*_{homo}\Omega/4\pi\right)^{1/2}]^2$$
$$- 2R^s m/\left(3n^*_{homo}\Omega/4\pi\right)^{1/2}\}^{1/2}. \qquad (43)$$

n^*_{homc} and r_c given in Eqs.(30) and (42) are often referred as the Gibbs-Thomson equations for the critical nucleus size.

3. NUCLEATION KINETICS

Let us repeat the picture of the heterogeneous nucleation model. On the substrate surface, some molecular processes occur due to transient visiting molecules which adsorb, form short lived unions, break-up, desorb etc.. An instantaneous census would show some distributions of subcritical nuclei (or clusters) with 1,2,3,.... molecules per cluster (c.f. Fig.6).
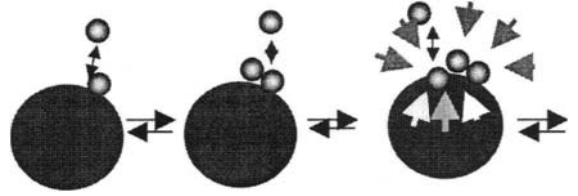
Figure 6. Schema of the process of nucleation at the surface of a foreign particle.

In any one with n molecules, the free energy changes for the n molecules to adsorb and form the n-mer with the size of r is given by Eq.(21). Nucleation begins with the formation of a cluster of size r_c with n^* molecules. r_c is given by Eq.(30).

The widely accepted kinetic model of nucleation (within the cluster approach) was used first by Farkas [3] in 1927. It is based on the Szilard scheme of successive "chain reaction" between monomer molecules and n-sized clusters:

Monomer $\rightleftharpoons$ Dimer... $\rightleftharpoons$ $(n-1)$ mer

$\rightleftharpoons$ n-mer $\rightleftharpoons$ $(n+1)$ mer... (44)
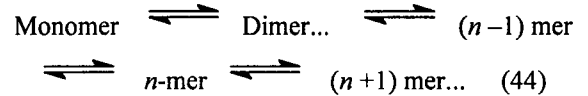
A master equation for the concentration $Z_n(t)$ of n-sized clusters at time t is written down in the form of continuity equation [12,17]

$$dZ_n/dt = J_{n-1} - J_n, \qquad (45)$$

where $J_n(t)$ is the flux through point n on the size axis. In this formulation J is thus the flux through the nucleus size n^*, i.e. $J \sim J_{n^*}(t)$, so that the nucleation rate can also be time-dependent.

3.1 Equilibrium, stationary and non-stationary states

The basic problem in the nucleation kinetics is to solve the master equation (44) in the unknown cluster size distribution $Z_n(t)$, since knowing $Z_n(t)$ allows the determination of the nucleation rate. There exist three physically distinct states of the

system which are of particular interest: *the equilibrium, the stationary (or steady)* and *the nonstationary states*. In the following, we will briefly discuss *the equilibrium* and *the nonstationary states*. Our concentration will focus more on the *steady* state.

3.1.1. The equilibrium state

In equilibrium $dZ_n/dt = 0$ and $J_n = 0$. Then $Z_n = C_n$ is the equilibrium cluster size distribution. (C_n is the equilibrium concentration of *n*-sized clusters.) According to the Boltzmann law, one can easily obtain:

$$(C_n/Z) = (C_1/\Sigma)^n \exp(-\Delta G_n/kT) \qquad (46)$$

(for all *n*; $n = 2,3,4,...$) with the effective total number of "molecules" per unit volume:

$$\Sigma = C_1 + \sum_{n=2} C_n . \qquad (47)$$

Here ΔG_n denotes the free energy barrier to form an *n*-mer. Since usually $C_1 >>> \Sigma C_n$ ($n > 1$), then $\Sigma \sim C_1$, thus

$$C_n \cong C_1 \exp(-\Delta G_n/kT). \qquad (48)$$

It follows that the $\Delta\mu$ dependence of the concentration $C^* = C_{n*}$ of nuclei is given by

$$C^*(\Delta\mu) \cong C_1 \exp[-\Delta G^*_{heter}/kT]. \qquad (49)$$

Let κ_n be the rate of molecule addition. That is,

$$\kappa_n = \beta_{kink} K_n \qquad (50)$$

where K_n is the collision rate of monomers with an *n*-sized cluster, and β_{kink} the conversion probability (c.f. Sec. 3.2.) Also let κ_n' be the rate at which the cluster lose molecules. Obviously, at the equilibrium state, one has the detailed balance between the growth and disintegration of clusters,

$$\kappa_n C_n - \kappa'_{n+1} C_{n+1} = 0. \qquad (51)$$

Since $\Delta G^*_{heter}(\Delta\mu)$ has a maximum at $n = n^*$, C_n displays a minimum at the critical nucleus size (Fig.7). The increasing, nonphysical branch of C_n at $n > n^*$ reflects the fact that the mother phase is saturated.
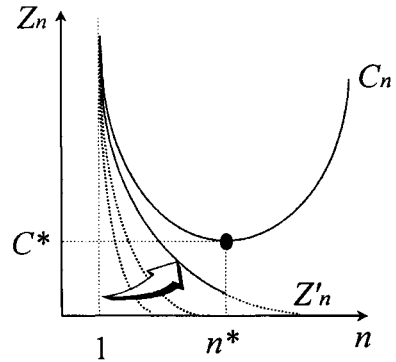
Figure 7. Equilibrium (C_n), stationary (Z'_n) and non-stationary (Z_n) cluster size distributions.

3.1.2. The stationary (or steady) state

In stationary state again $dZ_n/dt = 0$, but $J_n = $ constant $= J_{n*} = J$. Then, $Z_n \sim Z'_n$ is the steady-state cluster size distribution illustrated in Fig.7. The stationary nucleation rate for homogeneous nucleation is given by the Becker-Doering formula [5]

$$J = \frac{zK^*}{v_m} \exp\left(-\frac{\Delta G^*_{homo}}{kT}\right) \qquad (52)$$

with

$$z = Z'_{n*}/C_{n*} - Z'_{n*+1}/C_{n*+1} \qquad (53)$$

where z is the so-called Zeldovich factor [6], $K^* = K_{n*}$ is the frequency of monomer attachment to the critical nucleus, v_m denotes the average volume of structural units in the mother phase.

We notice that up to now, the extensive work carried out in the nucleation kinetics is mainly focused on homogenous nucleation. Concerning heterogeneous nucleation, the major concentration is focused on the effect on the nucleation barrier. Therefore, we will, in the following discussion, look at the kinetics of the steady state for heterogeneous nucleation [25,26].

Based on the definition of J_n one has

$$J_n = \kappa_n Z'_n \qquad (54)$$

At the steady state, *J'*, the formation rate of critical nuclei per unit volume-time around a foreign particle, is equal to the steady state growth of clusters at the surface of the particle. This can then be expressed as

$J' = J_{n*} = J_n$ = constant

= critical sized nuclei formed/unit volume – time

$\equiv \ldots = \kappa_n Z'_n - \kappa'_{n+1}Z'_{n+1}$= constant. (55)

Before further analyzing (51) or (55), we introduce the following two boundary conditions:

(1) $Z'_{n*} = 0$; (56)

(2) $\lim_{n \to 1}(Z'_n/C_n) = 1$. (57)

The first is because whenever a $(n^*\text{-}1)$-mer becomes an n^*-mer, it disappears from our population to begin a new stage - growth. But this event does not disturb the distribution of other clusters (mainly monomers). The second is because both C_n and Z'_n are large numbers when n is small. It follows that the perturbed concentration of monomers Z'_1 is almost equal to the equilibrium concentration.

Proceeding, rearranging Eq.(55) yields

$J' = \kappa_n C_n[(\kappa_n/C_n)-(\kappa_n+ 1/\kappa_n)(Z'_n+ 1/C_n)]$, (58)

and then using Eq.(51) we obtain

$J' = \kappa_nC_n[(Z'_n/C_n) - (Z'_n+ 1/C_n + 1)]$. (59)

From Eq.(59) a set of equations can be written:

$$\left.\begin{array}{ccccc} J'/\kappa_1 C_1 & = & (Z'_1/C_1) & - & (Z'_2/C_2) \\ J'/\kappa_2 C_2 & = & (Z'_2/C_2) & - & (Z'_3/C_3) \\ \vdots & = & \vdots & \vdots & \vdots \\ J'/\kappa_{n*-1}C_{n*-1} & = & (Z'_{n*-1}/C_{n*-1}) & - & 0 \end{array}\right\} \quad (60)$$

Adding Eq.set (60), and taking (56) into account

$$\sum_{n=1}^{n^*-1}(J'/\kappa_n C_n) = Z'_1/C_1 = 1 \quad (61)$$

and the growth rate J' is then given by,

$$J' = \left[\sum_{n=1}^{n^*-1}(\kappa_n C_n)^{-1}\right]^{-1} \quad (62)$$

To obtain J', the sum must be evaluated numerically. Nevertheless, the sum can be replaced by the integral and Eq.(62) can approximately be rewritten as

$$J' \approx \left[\int_1^{n^*}(\kappa_n C_n)^{-1}dn\right]^{-1}. \quad (63)$$

The evaluation of the integral in Eq.(63) is somewhat difficult. But it can be simplified by determining the dominant terms within the range from $n = 1$ to n^*, on which we can focus our attention. Certainly C_n decreases exponentially with n so that for large values of n the inverse C_n term is large. Therefore, the key issue is to find the variation of κ_n with n.

According to Eq.(50), to find κ_n, we need to derive the expression of K_n. By definition, K_n is proportional to the surface area of the cluster and the diffusivity of growth units D. For a cluster created on a spherical substrate as shown in Fig. 2, this area is given by

$S_{cf} = 4\pi r^2 f''(m,R')$ (64)

with

$f''(m,R') = \dfrac{1+(1- R'm)/w}{2}$ (65)

.

It follows then that

$K_n \cong DC_1 4\pi r^2 f''(m, R')$ (66)

Obviously, K_n (or κ_n) increases as the square root of n $(r \sim n^{1/2})$. For large values of n, κ_n is then only a weak function of n.

Returning to Eq.(63), if regions of high n contribute most and, in these regions, κ_n is not a strong function of n, one may remove κ_n from under the integral sign and, to approximate it as a constant equal to its value when $r = r_c$, i.e.,

$\kappa_n \sim \kappa_{n*} = 4\pi\beta_{kink}C_1(r_c)^2 f''(m,R')$ (67)

As long as we have concluded that the important region for n in the integral of Eq.(63) is where $n \sim n^*$, we can simplify the evaluation by rewriting it, as:

$(J')^{-1} \approx (\kappa_n)^{-1}\int_1^{n^*}(C_n)^{-1}dn$, (68)

and with Eq.(48)

$(J')^{-1} \approx (\kappa_n C_1)^{-1}\int_1^{n^*}\exp(\Delta G_n/kT)dn$ (69)

Expanding ΔG_n about n^* in a Taylor series, we find:

$$\Delta G_n = \Delta G^* + (\partial \Delta G_n / \partial n)_c (n - n^*)$$
$$+ \tfrac{1}{2}(\partial^2 \Delta G_n / \partial n^2)_c (n - n^*)^2 + ... \quad (70)$$

Define

$$y \equiv n - n^* \quad (71)$$

then

$$(J')^{-1} = (\kappa_{n*} C_1)^{-1} \left[\exp(\Delta G^* / kT) \right]$$
$$\times \int_{-n^*}^{0} \exp\left[(\Delta G_n') y / kT + (\Delta G_n'') y^2 / 2kT + ... \right] dy \quad (72)$$

$\Delta G_n'$ at $n = n^*$ is zero. Let Eq.(72) truncate after the second term. One has then after approximating the lower limit as $-\infty$,

$$J' = 2\kappa_{n*} \left[Q / (2\pi kT) \right]^{1/2} \quad (73)$$

$$Q = -\left(\partial^2 \Delta G_n / \partial n^2 \right)_{n=n^*} \quad (74)$$

where Q is positive so that the integral is an error function.

The last remaining step is to evaluate Q. For the heterogeneous nucleation, the second derivative of ΔG_n is very complex. Note that the derivatives are evaluated at $n = n^*$ so are not a function of n. In most cases where R' is not small, $f(m,x)$ turns out to be constant for a given foreign body. We can approximate the value of $Q = -(\partial^2 \Delta G_n / \partial n^2)_{n=n^*}$ by $f(m, R')(\partial^2 \Delta G_n^{homo} / \partial n^2)_{n=n^*}$ therefore Q can be expressed as

$$Q \approx \gamma_{cf} \left(\frac{2^5 \pi \Omega^2}{3^4} \right)^{1/3} (n^*)^{-4/3} f(m, R') \quad (75)$$

or, with Eq.(30),

$$Q \approx \gamma_{cf} \Omega^2 f(m, R') / (2\pi r_c^4) \quad (76)$$

By combining Eqs.(66)-(68) and (73), an expression for J' is obtained, as

$$J' = 4D\beta_{kink} f''(m, R') \Omega \left[\frac{\gamma_{cf}}{kT} f(m, R') \right]^{1/2}$$
$$\times (C_1)^2 \exp\left[-\frac{\Delta G_{homo}^*}{kT} f(m, R') \right] \quad (77)$$

The average nucleation rate in the solution depends on the density and size of foreign particles occurring in the system, and is given, according to $J = 4\pi a(R^s)^2 N^\circ J'$, by

$$J = 4\pi a (R^s)^2 N^\circ f''(m, R') \left[f(m, R') \right]^{1/2}$$
$$\times B \exp\left[-\frac{16\pi \gamma_{cf}^3 \Omega^2}{3kT [kT \ln(1 + \sigma)]^2} f(m, R') \right] \quad (78)$$

with

$$B = (C_1)^2 4D\beta_{kink} \Omega \left(\frac{\gamma_{cf}}{kT} \right)^{1/2} \quad (79)$$

(a is the dimension of growth unit.) In the equations, $f''(m,R')$, like $f(m,R')$, is also a function of m and the relative size of particles R', and gives a similar plot as given by Fig.3. Introducing the term $4\pi a(R^s)^2 N^\circ$ is based on the fact that the heterogeneous nucleation takes place only in the liquid layers adjacent to the nucleating particles. Evidently, only for this part of solutions, the nucleation rate will be effectively influenced by the foreign particles. Therefore, the relative effective volume fraction for heterogeneous nucleation is equal to the volume of the liquid is proportional to the density and surface area of nucleating particles occurring in the system, namely $4\pi a(R^s)^2 N^\circ$.

In the case of homogeneous nucleation, one has $f''(m,R') = f(m,R') = 1$, and $4\pi a(R^s)^2 N^\circ \to 1$. In this case, Eq.(43) is converted to

$$J = B \exp\left[-\frac{16\pi \gamma_{cf}^3 \Omega^2}{3kT [kT \ln(1 + \sigma)]^2} \right]. \quad (80)$$

Obviously, this expression is similar to that given by (52) for the homogeneous nucleation rate, meaning that Eq.(43) is applicable to describe both homogenous and heterogeneous nucleation.

3.1.3. The non-stationary (or steady) state

When nucleation is nonstationary, $dZ_n/dt \neq 0$ and the flux J_n is a function of both n and t. The nucleation rate is then time-dependent and this nonstationary nucleation rate $J_{nonst}(t) = J_{n*}(t)$. According to Kashchiev [32], this is approximately expressed as

$$J_{nonst}(t) = J \left[1 + 2 \sum_{i=1}^{\infty} (-1)^i \exp(k^2 t / \tau) \right]. \quad (81)$$

In this equation, J is the nucleation rate of the steady state given by (78). According to this formula, $J_{nonst}(t)$ at $t = 0$ and gradually increases to the stationary nucleation rate J. It turns out [32] that $J_{nonst}(t) \sim J$ for $t > 5\tau$, where

$$\tau = 4 / \pi^3 z^2 \kappa_{n^*}. \tag{82}$$

is the nucleation time-lag. Physically, τ is a measure of the time needed for the transformation of the initial cluster size distribution into the steady-state one (as illustrated by the arrow in Fig.7). The results in the theory of nonstationary nucleation have also been reported [33-35]. Since $z = 0.1$, [36] Eq.(40) shows that τ is largely determined by κ_{n^*}, in particular the diffusivity D. With typical values of D, this predicts time lags from microseconds for nucleation in vapors and less viscous solutions to days for nucleation in solids and glass-forming melts. [36]

3.2. Induction time in crystallization

One of the most common ways in nucleation study is to measure the induction time of nucleation t_{nucl} at different supersaturations. Due to the crystallization sequence, what we normally measure is the induction time t_i in crystallization, which is defined as the mean time elapsing before appearance of an observable amount of the new phase. Actually, t_i includes the time t_g for the growth of crystals to the observable size and t_{nucl}. As mentioned in Sec.3.1, there is certain time required to establish nucleation from time zero to the steady state. This is the transient period t_{nonst}, which is associated with nucleation of the non-stationary state. Denoting the induction time for nucleation of the steady state t_s, one has then

$$t_i = t_g + t_{nonst} + t_s. \tag{83}$$

Since the free energy barrier for 3D nucleation is much higher than that in 2D nucleation (c.f. Sec.5), the growth of crystals then much easier than nucleation in most cases. If crystals with a sufficiently small size can be detected by certain techniques, we then can have $t_g \ll t_{nucl}$ ($= t_{nonst} + t_s$). Nowadays, the laser light scattering method [26] promises the detection of particles from several nm to several tens of nm. This has already been very close to the critical size of nuclei in many cases. Under such a situation, we can even assume $t_g \rightarrow 0$.

Apart from this, t_{nonst}, according to the previous section, is determined to a large extent by diffusivity. If the mother phase is not too viscous, such as aqueous solutions, we normally have t_{nonst} = a few microseconds [36]. This implies that $t_{nonst} \ll t_s$. Therefore, we can approximate (83) by

$$t_i \cong t_{nucl} \cong t_s. \tag{84}$$

In other words, this implies that under the normal condition, the nucleation rate J is time-independent. By definition, one has

$$J = 1/(t_s V), \tag{85}$$

where V is the volume of the system. It follows then from Eq.(78)

$$\ln t_s = \frac{\rho f(m, R')}{[\ln(1+\sigma)]^2} - \ln\{Vf''(m, R')[f(m, R')]^{1/2} \beta_{kink}(C_1)^2 B'\}. \tag{86}$$

with

$$\rho = 16\pi\gamma_{cf}^3 \Omega^2 / 3(kT)^3. \tag{87}$$

where $B' = B|(R^s)^2 N^a|$. Notice that in applying this relationship to study the nucleation kinetics, V should be kept constant under a given condition. This can be an important step to gain a set of consistent data.

For a given system, since ρ, C_1, and B' are constant. Changes in the slope and/or the intercept of the $\ln(t_s) \sim 1/[\ln(1+\sigma)]^2$ plot will then correspond to the modifications in $f(m, R')$, $f''(m, R')$ and β_{kink}. We will discuss this in the following section.

On the other hand, if the mother phase is very viscous, such as a glass phase, t_{nonst} cannot be ignored. Then, Eqs.(84) and (86) cannot be applied directly.

4. INTERFACIAL EFFECTS OF FOREIGN BODY AND FLUID MOLECULE

4.1. Size effect of foreign body

As mentioned before, $f(m, R')$ in the exponential term of Eq.(76) describes the reduction of the nucleation barrier from a genuine homogeneous nucleation ΔG_{homo}^* to the actual heterogeneous nucleation ΔG_{heter}^*, due to the presence of foreign bodies (c.f. Eq.(22).) This factor, which is a
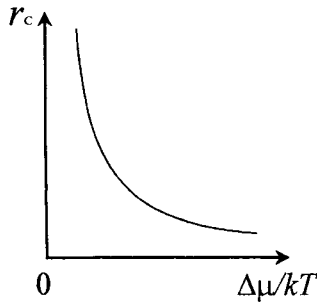
Figure 9. Dependence of the critical radius of nuclei on the thermodynamic driving force.

(a)

(b)
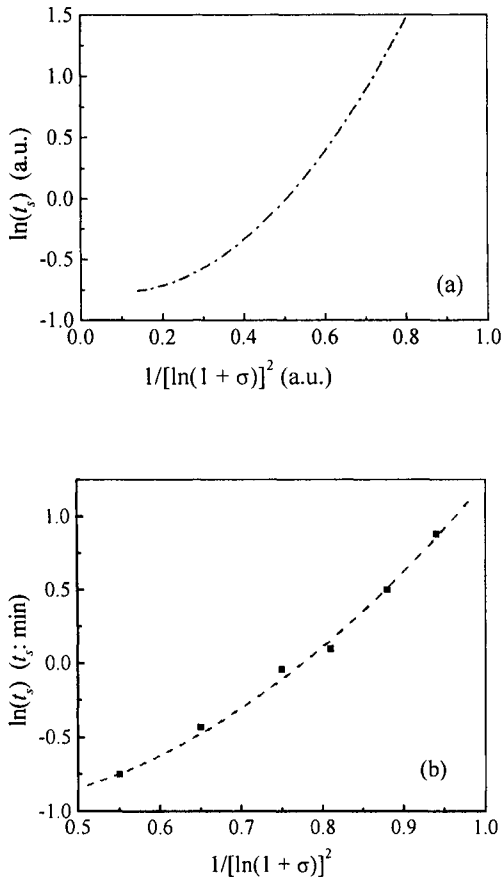
Figure 10. (a) The effect of particle size on the nucleation rate at different supersaturations in the non-linear regime. $m = 0.8$, $\gamma_{cf}/kT = 1.5$, $R^s = 5a$. (b) The nucleation of $CaCO_3$ in the nonlinear regime. (The details to be published.)

function of m and R', can be applied to the description of the correlation between the substrate and the nucleating phase.

According to Eq.(31), the relative radius of foreign body R' is determined by the radius of the curvature of foreign particles and the critical radius of nuclei. Since r_c is inversely proportional to $\Delta\mu/kT$ ($= \ln(1 + \sigma) \approx \sigma$), it will decrease parabolically with the increase of $\Delta\mu/kT$ (Fig.9.) For a given R^s (not too large and too small), R' turns out to be proportional to $\Delta\mu/kT$.

At high supersaturations, r_c can be much smaller than R^s. According to Eq.(31), this gives rise to a large R'. In the case where $R' \geq 10$, $f(m,R')$ becomes almost constant at different supersaturations. It follows from Eq.(86) that a linear $\ln(t_s) \sim 1/[\ln(1+\sigma)]^2$ plot will be obtained. Therefore, this regime is defined as the linear regime (see Fig.3a.)

Conversely, as supersaturation decreases, r_c will increase. If $R' \leq 10$, $f(m,R')$ will drastically increase with decreasing supersaturation (c.f. Fig.3a.) It follows from Eq.(86) that the slope of the $\ln(t_s) \sim 1/[\ln(1+\sigma)]^2$ plot will also increase drastically and non-linearly. (See Fig.10.) We refer this regime to the non-linear regime (Fig.3a.)

At very low supersaturations or very small R^s, we have $R' \leq 0.1$. Within such a regime, $f(m,R')$ reaches its maximum ($f(m,R') = 1$). This implies that from the point of view of nucleation barrier, foreign bodies plays no effect on the nucleation kinetics. Any nucleation occurring at the surface of foreign bodies having the radius much smaller than the critical nucleus radius r_c will be unstable. This is similar to the Gibbs-Thomson effect. Since the nucleating phase cannot "see" the foreign bodies, we define this regime as the "zero size" effect regime. In this regime, the $\ln(t_s) \sim 1/[\ln(1+\sigma)]^2$ plot reaches its maximal slope. Fig. 10a shows the correlation between $\ln(t_s)$ and $1/[\ln(1 + \sigma)]^2$ in the nonlinear regime. An example of $CaCO_3$ nucleation occurring in this regime is given in Fig.10b.

4.2. Interfacial correlation between the substrate and the nucleating phase

Let us first have a look at the case where the radius of foreign body R^s is much larger than r_c ($R' \geq 10$). In this case, the substrate can be regarded as essentially flat, and $f(m,R')$ is then dependent only on m (c.f. Fig.3b.) As discussed in Sec.2.2.,

heterogeneous nucleation occurs in the range of m between 1 and -1, and $f(m,R')$ between 0 and 1, depending on the nature of the substrate surface. Based on Eq.(86), plotting $\ln(t_s) \sim 1/[\ln(1+\sigma)]^2$ will give rise to a straight line whose slope is determined by ρ and $f(m,R')$. Obviously, for a given system (ρ, B' = const.), the slope of the straight line will change accordingly with $f(m,R')$. In this sense, the slope of the $\ln(t_s) \sim 1/[\ln(1+\sigma)]^2$ plot gives the relative $f(m,R')$ for the system. One can analyze the change of the correlation between the substrate and the crystalline phase in terms of the variation of the slope.

4.2.1. The effect of supersaturation on the interfacial correlation

Kinetically, the occurrence of foreign bodies will on one hand lower the nucleation barrier, resulting in an increase in the nucleation rate, on the other hand, exert also a negative impact on the surface integration. As shown in Fig.6, nucleation on a substrate will reduce the effective collision of structural units to the surface of clusters, where the structural units are incorporated into the crystal phase. This tends to slow down the nucleation kinetics, in contrary to the nucleation barrier lowering effect. This negative effect is described by $f''(m,R')$ and $f(m,R')$ appearing in the pre-exponential term of Eq.(78).

These two controversial effects play different roles in different regimes. At low supersaturations, the nucleation barrier is very high (c.f. Eqs.(32), (33)). The nucleation rate will be substantially enhanced if the nucleation barrier is suppressed effectively ($f(m,R') \to 0$). Therefore, the heterogeneous nucleation with a strong interaction and an optimal structural match between the substrate and the nucleating phase ($m \to 1$) will be kinetically favored. Conversely, at higher supersaturations, the exponential term associated with the nucleation barrier becomes less important. Instead, the issue of effective collisions, described by the pre-exponential factors $f(m,R')$ and $f''(m,R')$, becomes more dominant. Nucleation on substrates having larger $f(m,R')$ and $f''(m,R')$ ($m \to 0, -1$, the weak interaction and poor structural match between the substrate and the nucleating phase) becomes kinetically more favorable (c.f. Figs.5b,c.) From the point of view of statistic physics, this implies a less restriction from the substrate, and more orientational freedom (or a large entropy.)

It follows from Eq.(78) that if σ changes from low supersaturations to high supersaturations, nucleation will be governed by a sequence of progressive heterogeneous processes with the increasing $f(m,R')$. The above results are illustrated in Fig.11 by the plot of J (in arbitrary units) versus the relative supersaturation σ, based on Eq.(86). Obviously, homogeneous nucleation is the upper limit of the sequence of heterogonous nucleation.
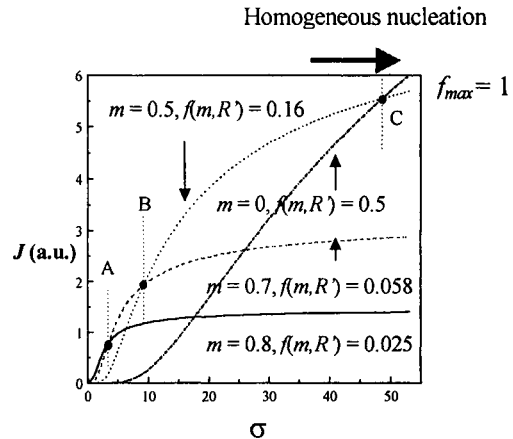
Figure 11. Dependence of the relative nucleation rate on the supersaturation. The effect of m on the nucleation rate. As shown, foreign bodies with a large m (or a small $f(m,R')$) will control the kinetics at low supersaturations while those with a small m (or a large $f(m,R')$) will control the kinetics at high supersaturations.
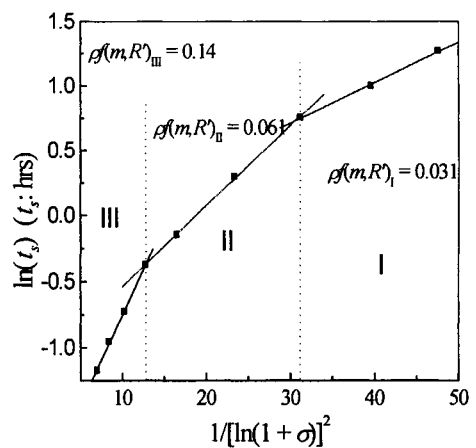
Figure 12. Dependence of $\ln\tau$ on $1/[\ln(1+\sigma)]^2$ for N-lauroyl-L-glutamic acid di-n-butylamide nucleating from iso-stearyl alcohol solutions. Within the range of supersaturations where the

experiments were carried out, three straight lines with different slopes intercept with each other, dividing the space into three regimes.

If σ changes progressively from a low to a high value, one should expect that a set of straight lines intercept with each other in the $\ln t_s \sim 1/[\ln(1 + \sigma)]^2$ plot. The slope of these lines will increase with σ. The measured induction time as a function of supersaturation for N-Lauroyl-L-glutamic acid di-n-butylamide nucleating from iso-stearyl alcohol solutions is given in Fig.12.

As shown, the interfacial correlation factor $f(m,R')$ subsequently increases from $0.031/\rho$ to $0.14/\rho$, with increasing supersaturation. This is in excellent agreement with our prediction given above

4.2.2. The epitaxial and counter-epitaxial effects of additives

Epitaxial growth promoted by substrates should result from the lowering of $f(m,R')$. This can be identified from the decrease of the slope of the $\ln(t_s) \sim 1/[\ln(1 + \sigma)]^2$ plot (c.f. Eq.(86), and Fig.13). Conversely, for the nucleation inhibition resulted from the weakening of the substrate/crystal phase interaction, $f(m,R')$ and the slope of the $\ln(t_s) \sim 1/[\ln(1 + \sigma)]^2$ plot should increase. This is the so-called "epitaxial effect".

If the modification of the surface of foreign bodies leads the interfacial correlation toward $m \rightarrow -1$, the effect is then the counter-epitaxial effect. Actually, both the epitaxial and the counter-epitaxial can in principle result from the adsorption of some additives on the surface of foreign bodies. The key point here is that the adsorption of additives at the surface of foreign bodies will significantly modify the interaction and structural match between foreign bodies and the nucleating phase. This brings about a change in the interfacial free energy between the nucleating phase and the substrate from γ_{sc} to $\gamma_{s'c}$.

In the case of nucleation promotion, the adsorption of additives will give rise to a stronger interaction and/or a better structural match between the substrate and the nucleating phase, which significantly reduces γ_{sc}. If the additives adsorb on the surface of clusters at the same time, according to Eq.(5), this leads to $m \rightarrow \gamma_{sf}/\gamma_{cf} \sim 1$ and $f(m,x) \rightarrow 0$, which can be identified from the lowering of the slope of $\ln(t_s) \sim 1/[\ln(1 + \sigma)]^2$, and the increase of the intercept due to the negative change in

$\ln\{f''(m,x)[f(m,x)]^{1/2}\}$. The transition of curve 0 to curve 1 in Fig.13 illustrates such a change. In practice, one of the examples is the application of surfactants to promote nucleation.[37]

Conversely, the adsorption of additives leading to the repulsion and the interfacial structure mismatch between the substrate and the nucleation crystalline phase will cause a substantial increase in γ_{sc}, and consequently. It follows that $m \rightarrow -1$ and $f(m,x) \rightarrow 1$ (c.f. Eq.(38)). This enhances the nucleation barrier and reduces the nucleation rate at a given supersaturation. The effect can be identified from the increase in the slope of $\ln(t_s) \sim 1/[\ln(1 + \sigma)]^2$, and a decrease of the intercept (from line 0 to line 2 in Fig.13.)
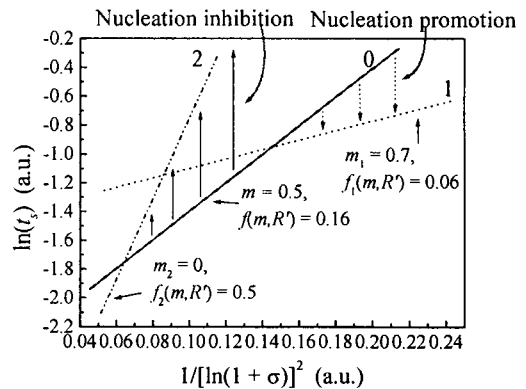
Figure 13. The effect of interaction parameter m on the nucleation rate at different supersaturations. Foreign particles with a large m (strong interaction and better structural match with the nucleating phase) will control the kinetics at low supersaturations while those with a small m will control the kinetics at high supersaturations.

4.3. Interfacial processes and non-epitaxial interfacial effects
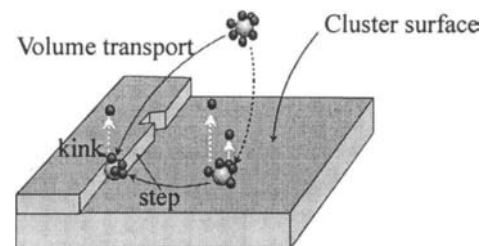
Figure 14. Schematic illustration of the surface integration occurring at the surface of cluster.

During the "growth" of clusters, structural units should be transported from the bulk of the mother phase to the surface of the clusters, and be incorporated into the clusters at the kink sites (see Fig.14.)

Before the incorporation of structural units (c.f. Fig.14), the desorption of adsorbed impurities and solvent molecules from the surface and the re-orientation and conformational adaptation of structural units to the crystal surface will occur first. [38]. Associated with this process, a free energy barrier $\Delta G^{\neq}$ needs to be overcome. Actually, the kink integration coefficient β_{kink} occurring in Eq.(86) is a factor describing this process, which is defined as [28]

$$\beta_{kink} = v \frac{a}{\lambda_o} \exp(-\Delta G^{\neq}/kT) \qquad (88)$$

(v denotes the vibration frequency of structural units in the neighborhood of the surface, λ_o is the average distance between two kinks at the surface and $\Delta G^{\neq}$ is the activation free energy for kink integration).
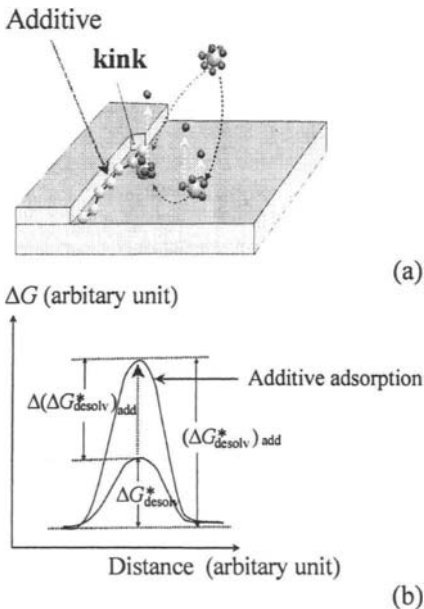
Figure 15. (a) Schematic illustration of adsorption of additives at a kink site and (b) the enhancement of the desolvation energy.

Although this process has been realized for long time, unfortunately it was never unambiguously defined. Recently, the effect on the kinetics of crystal growth has been identified [39-42]. This has a direct implication for our understanding on the nucleation process. It follows from recent computer simulations and theoretic analyses [39-42] that $\Delta G^{\neq}$ should include:

(a) the desolvation energy barrier ΔG^{*}_{desolv}, the energy required to remove solvent or impurity molecules adsorbed at the kink sites;

(b) the orientational and conformational entropy barrier ΔG^{*}_{order} due to the adaptation of molecules to the crystalline structure at the crystal surface.
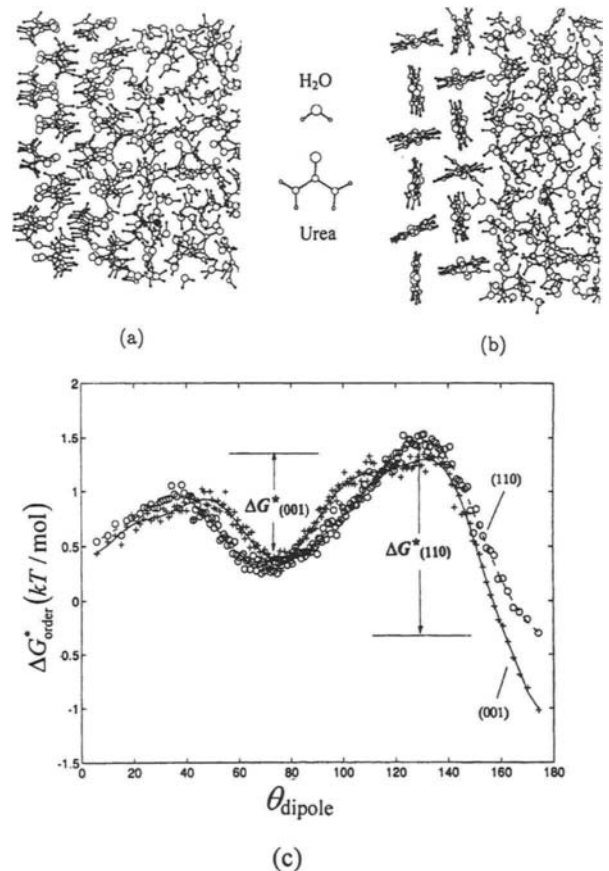
Figure 16. Snapshots of interfaces between crystalline urea and saturated aqueous solution. (a) The (001) interface; (b) The (110) interface. (c) ΔG^{*}_{order} associated with the two orientations.

The occurrence of ΔG^{*}_{desolv} is quite obvious. In many cases, a stronger adsorption of additives at the surface of clusters will enhance ΔG^{*}_{desolv}, therefore

substantially reduce β_{kink} and consequently suppress the nucleation rate (c.f. Fig.15.)

(a)

(b)   (c)
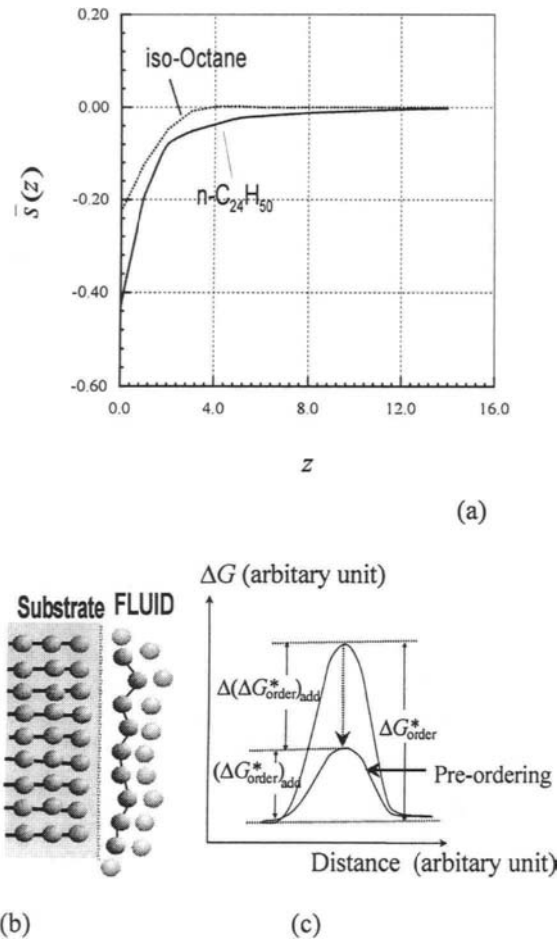
Figure 17. (a) The average order parameter $\bar{s}(z)$ of n-$C_{24}H_{50}$ molecules as a function of z, generated by the density functional theory calculations [40,41]. The density of segments in the bulk $\varphi_A = 0.1$; the torsion energy of the chains $\varepsilon^{tor} = 1.5kT$. The order parameter, $s = \frac{1}{2}(3 < \cos^2\alpha > - 1)$, α is the angle between a bond connecting two segments units and the normal. When the molecules are completely parallel to the surface, $s(z) = -\frac{1}{2}$. A random bond distribution will result in an order parameter $s(z) = 0$ [40,41]. (b) Illustration of the ordering of paraffin molecules at the solid-fluid interface. (c) The pre-ordering of liquid may lower $\Delta G^*_{order}/kT$, therefore affect the nucleation kinetics.

Here, we will focus on ΔG^*_{order} since this effect will play an important role in the nucleation of complicated and macromolecule molecules. The crystallization of these molecules becomes increasingly important due to the implications for life science.

The crystallization of complex molecules is very much different from the crystallization of simple molecules [38-42]. During the crystallization, complex fluid molecules having statistically random orientations and conformations will be converted into solid complex molecules with a given orientation and conformation required by the crystal structure. [39-42]. Fig. 16 shows the {001} (Fig.16a) and {110} (Fig.16b) interfacial structures between a urea crystal and its saturated solution, generated by molecular dynamic simulations [38,39,43]. This implies that in order to be incorporated into the crystal structure, fluid molecules at the kink sites of clusters should adapt to a specific orientation and conformation required by the crystal. This means that fluid molecules should surpass the conformational entropic barrier ΔG^*_{order} in order to be incorporated into the crystal. Fig.16c shows the calculated ΔG^*_{order}, $\Delta G^*_{(001)}$ and $\Delta G^*_{(110)}$ for the {001} and {110} orientations, respectively. The height of ΔG^* and the change in ΔG^* will exert a direct impact on the nucleation rate in terms of β_{kink} (c.f. Eq.(88).)

In many cases, the substrate will induce the pre-ordering at the crystal-fluid interface. Fig.17a shows the preordering of chain-like molecules at a flat substrate [38-42]. Evidently, it is much easy to acquire the crystalline order from the pre-ordered liquid molecules. Therefore, this will lower ΔG^*_{order}, consequently promote nucleation (c.f. Fig.17b.)

The change in ΔG^*_{order} due to the molecular ordering can also be identified from the $\ln t_s \sim 1/[\ln(1 + \sigma)]^2$ plot, based on Eqs.(88) and (89). In the case of nucleation promotion, the pre-ordering of molecules will shift the straight line downward. An example is given in Fig. 18, where line 0 moves downward to line 2. According to Eqs. (86) and (88), the difference in the intercepts in this case equals to the change in the entropic barrier $\Delta(\Delta G^*_{order}/kT)_{add} [= (\Delta G^*_{order}/kT)_{add} - \Delta G^*_{order}/kT \cdot]$. (See Fig.17c.)

We notice that the nucleation promotion caused by the pre-ordering of liquid molecules at the solid-fluid interface has never been recognized before. This effect has been experimentally identified for

the nucleation of paracetamol from aqueous solutions.
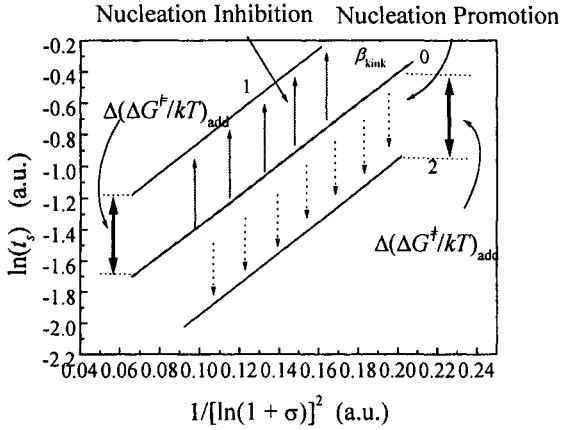
Figure 18. Illustration of the change in kink kinetics and the corresponding shift in the $\ln(t_s) \sim 1/[\ln(1 + \sigma)]^2$ plot.

5. HETEROGENEOUS TWO-DIMENSIONAL NUCLEATION GROWTH

When the growth of a crystal surface occurs under its critical roughening temperature, there will be a non-zero free energy, the so-called step free energy associated with the creation of a step of unit length at the surface (c.f Fig.14). Such a crystal face has the atomically smooth surface. Due to the step free energy, the creation of a new layer on the existing layer of the crystal surface requires overcoming a free energy barrier, so-called two-dimensional nucleation barrier. If the crystals are free of screw dislocations, they grow by the mechanism of 2D nucleation [29] and the growth rate R_g is largely determined by the 2D nucleation rate.

Similar to 3D nucleation, the occurrence of foreign particles at the crystal surface will lower the nucleation barrier, therefore promote the growth at relatively low supersaturations. Assuming that foreign particles have the shape of circular disk, the heterogeneous 2D nucleation barrier can be expressed analogously to heterogeneous 3D nucleation, as [44,45]

$$\left(\Delta G^{\bullet}_{heter}\right)_{2D} = \left(\Delta G^{\bullet}_{homo}\right)_{2D} f_{2D}\left(m_{step}, R'_{2D}\right), \qquad (89)$$

with

$$\left(\Delta G^{\bullet}_{homo}\right)_{2D} = \frac{\Omega \left(\gamma^{step}_{cf}\right)^2 \pi h}{kT \ln(1+\sigma)} \qquad (90)$$

$$f_{2D}(m_{2D}, R'_{2D}) = \tfrac{1}{\pi} \{ \text{arc } \cos[(R'_{2D}\, m_{2D} - 1)/w_{2D}]$$
$$+ ((R'_{2D})^2 - 2(R'_{2D})m_{2D}) \text{ arc } \cos[(R'_{2D} - m_{2D})/w_{2D}] - (1 - R'_{2D}\, m_{2D})[(w_{2D})^2 - (1 - R'_{2D}\, m_{2D})^2]^{1/2}/(w_{2D})^2 - (R'_{2D})^2(R'_{2D} - m_{2D}) \times [(w_{2D})^2 - (R'_{2D} - m_{2D})^2]^{1/2}/(w_{2D})^2\} \qquad (91)$$

and

$$w_{2D} = [1 + (R'_{2D})^2 - 2(R'_{2D})m_{2D}]^{1/2}, \qquad (92)$$

$$r_c^{2D} = \Omega \gamma^{step}_{cf} /\Delta\mu = \Omega \gamma^{step}_{cf} /kT\ln(1 + \sigma) \qquad (93)$$

$$R'_{2D} = R^s/r_c^{2D} = R^s kT\ln(1 + \sigma)/\Omega \gamma^{step}_{cf}. \qquad (94)$$

In Eqs. (89)-(94), "2D" occurring at the superscripts and the subscripts indicates the 2D nucleation quantities similar to 3D nucleation, γ^{step}_{ij} denotes the step free energy, and h is the height of the step. Here, the step correlation function $f_{2D}(m_{2D}, R'_{2D})$ has a similar shape as the surface correlation shown in Fig.3 [44,45].

Within the framework of 2D nucleation growth mechanisms, if the spreading rate U of the monomolecular steps on the crystal face is high enough, the crystal grows by the so-called mononuclear (MN) mechanism, since after its formation the very first nucleus can cover the crystal face before the appearance of other nuclei on it [29].

When spreading of the monomolecular steps on the crystal face is relatively slow, the so-called polynuclear (PN) or birth-and-spread mechanism is effective. Many nuclei then appear and grow simultaneously on the crystal face (Fig.19) [8,46].
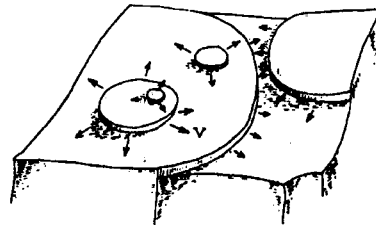
Figure 19. Schematic illustration of the birth and spread model.

This model allows for both nucleation of critical size clusters and subsequent growth at a

finite rate. Therefore we adapt the following three assumptions regarding the growth or spreading of the growing nuclei: a) There is no intergrowth between nuclei; b) The lateral spreading velocity U is a constant, independent of the island size; c) Nuclei can be born anywhere, around nucleating particles, which can occur on incomplete layers as well as on islands. As for homogeneous 2D nucleation growth [44], these assumptions lead to the following expression for the rate of heterogeneous 2D nucleation growth

$$R_g = h(J_{2D})^{1/3}(U)^{2/3} \qquad (95)$$

Here J_{2D} is the heterogeneous 2D nucleation rate given by [44],

$$J_{2D} =$$
$$\left\{ \frac{2D_s C_1^2}{\pi} \left[\frac{\Omega \ln(1+\sigma)}{h} \right]^{1/2} \exp\left(-\frac{\Omega\left(\gamma_{cf}^{step}\right)^2 \pi h}{(kT)^2 \ln(1+\sigma)} f_{2D}(m_{2D}, R_{2D}^{'}) \right) \right\}$$
$$\times \beta_{kink} \delta\left(m_{2D}, R_{2D}^{'}, R^s, N^o\right) \qquad (96)$$

with

$$\delta\left(m_{2D}, R_{2D}^{'}, R^s, N^o\right) = \left\{ \frac{\psi}{\pi} N^o \pi a R_{2D}^{'} \left[f_{2D}(m_{2D}, R_{2D}^{'}) \right]^{1/2} \right\}, \qquad (97)$$

where D_s denotes the surface diffusivity.

In the case of the growth of crystals from solutions or the melt, the step velocity has a form [28]

$$U = \pi \Omega \beta_{kink} C_1 \zeta \sigma, \qquad (98)$$

with

$$\zeta = \left[1 + (\pi \beta_{kink}/D\lambda_s)\ln(\lambda_s/\pi a) \right]^{-1}, \qquad (99)$$

where λ_s is the diffusion mean free path. The normal growth rate R_g is given, according to Eqs. (95), (96) and (98), as

$$R_g = A_1 \sigma^{2/3} [\ln(1+\sigma)]^{1/6}$$
$$\times \exp[-A_2 f_{2D}(m_{2D}, R_{2D}^{'})/T^2 \ln(1+\sigma)]$$
$$\times [\delta(m_{2D}, R_{2D}^{'}, R^s, N^o)]^{1/3} \qquad (100)$$

with

$$A_1 = (h\Omega)^{5/6} \beta_{kink} [2D_s \pi/av]^{1/3} [(C_1)^2 \zeta]^{2/3} \qquad (101)$$

and

$$A_2 = \frac{\Omega \gamma_{cf}^{step} \pi h}{3k^2}. \qquad (102)$$

At $\sigma \ll 1$, Eq.(100) can be rewritten as

$$R_g = A_1 \sigma^{5/6} \exp[-A_2 f_{2D}(m_{2D}, R_{2D}^{'})/T^2 \sigma]$$
$$\times [\delta(m_{2D}, R_{2D}^{'}, R^s, N^o)]^{1/3} \qquad (103)$$

Note that Eq.(103) includes also homogeneous 2D nucleation as a limited case. In the case of homogeneous 2D nucleation, $\psi = \pi$, $R^s \to a/\sqrt{2}$ and $N^o \to 1/a^2$. Therefore, $\delta(m_{2D}, R_{2D}^{'}, R^s, N^o) = f_{2D}(m_{2D}, R_{2D}^{'}) = 1$, and Eq.(103) describes the rate of homogeneous 2D nucleation growth. Obviously, $f_{2D}(m_{2D}, R_{2D}^{'})$ and $\delta(m_{2D}, R_{2D}^{'}, R^s, N^o)$ characterize the difference in the growth rate between homogeneous 2D nucleation and heterogeneous 2D nucleation. In this sense, this model covers both heterogeneous and homogeneous 2D nucleation growth.

Similar to 3D nucleation, heterogeneous 2D nucleation will govern the growth of dislocation-free crystals at lower supersaturations, while homogeneous 2D nucleation will control the growth of crystals at higher supersaturations. This has been confirmed by the experiments of the dislocation free growth of the {100} faces of $Ba(NO_3)_2$ crystals from aqueous solutions [44,45]. (See Fig.20.)
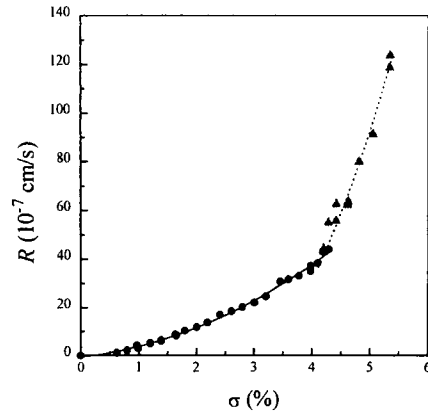
Figure 20. R_g vs. σ for the dislocation free growth of the {100} faces of $Ba(NO_3)_2$ crystals from aqueous solutions. • and ♦: experimental data. ——: fitting based on the heterogeneous 2D nucleation model; ----: fitting based on the homogeneous 2D nucleation model.

We notice that foreign particles occurring on the growing surface of crystals exert some very important impacts on the growth of crystals, which has not been realized before. It has been found that apart from the growth promotion, the adsorbed solid

particles may cause the pattern formation [47] and a special type of surface kinetic roughening [48].

6. GENERAL REMARKS AND CONCLUSIONS

In practice, foreign bodies occur almost everywhere, including our crystallization systems. These foreign bodies are ranged from dusts, the wall of crystallizers, macromolecules, fluid-fluid interfaces or bubbles, secondary nuclei or even the amorphous phase or metastable phase from the same crystalline system, etc. In other words, it is very difficult, or in some cases almost impossible to get rid of these entities. In many cases, people mixed up bulk nucleation with genuine homogeneous nucleation, and claimed that they measured "the homogeneous nucleation rate". In some other situations, the nucleation after filtration was often regarded "homogeneous nucleation". We have detected recently in our nucleation experiments that foreign particles still have some effects on the ice nucleation from ultra-clean deionized water (18.2MΩ, filtered by a 22nm filter), which was carried out in a containerless system (the results to be published). The key point is that the suppression of the nucleation rate by filtering the system does not guarantee the occurrence of homogeneous nucleation. Homogeneous nucleation can only be verified when the absolute value of the slope of the $\ln(t_s)$ (or $\ln J$) $\sim 1/[\ln(1 + \sigma)]^2$ plot reaches its maximum, which in many cases is difficult to achieve.

We also notice that there are many confusions, in the area of nucleation inhibition. One of the most common arguments associated with this subject [49,50] is that the adsorption of additives on the surface of growing nuclei lifts the nucleation barrier by the enhancement of solid-fluid interfacial tension, so that the nucleation rate is reduced correspondingly. This violates basic principles of surface science [51]. According to the Gibbs equation [51],

$$\Gamma_{add} = -\frac{1}{RT}\frac{d\gamma}{d\ln a_{add}} \qquad (104)$$

Here R denotes the gas constant and a_{add} the activity of the additives in the solutions, and Γ_{add} is defined as the surface excess of additives for the case where the surface excess of solvent is equal to zero in the chosen dividing surface. Obviously, the amount of additive adsorption can be described by Γ_{add}. The implication of Eq.(104) is that in the case of positive adsorption ($\Gamma_{add} > 0$), the addition of additives will lead to the decrease of the interfacial free energy. This means that the adsorption of additives on the surface of growing nuclei will lower the interface energy γ_{cf}.

In conclusion, heterogeneous nucleation is a generic event occurring in most nucleation processes. With the increase of supersaturation, instead of a single process, heterogeneous nucleation is a sequence of progressive processes which reveal a wide spectrum of heterogeneous characteristics described by $f(m,R')$ and $f''(m,R')$. At low supersaturations, nucleation will be governed by the process with an optimal structural match and a strong interaction between the substrate and the nucleating phase ($f(m,R') \to 0$); as supersaturation increases, the nucleation process will be controlled by the nucleation with the poor structural match and more oreintational freedom ($f(m,R') \to 0$). Obviously, homogeneous nucleation is the upper limit of this spectrum.

On the other hand, if the radius of the curvature of foreign particles is not large, the effect of foreign particles can be suppressed in terms of the "zero size" effect.

Additives adsorbed at the solid-fluid interface will cause the epitaxial and/or counter-epitaxial effects.

Substrates may promote nucleation without changing the nucleation barrier via inducing the pre-ordering of fluid molecules at the interface.

REFERENCE

1. J. W. Gibbs, *Collected Works. Vol. I: Thermodynamics* (Longmans and Green, New York, 1928).
2. M. Votmer and A. Weber, Z. Phys. Chem. **119** (1926) 277.
3. L Farkas, Z. phys. Chem. **125** (1927) 236.
4. R. Kaischew and I. Stranski, Z. Phys. Chem **B26** (1934) 317.
5. R. Becker and W. Doering, Ann. Phys. **24** (1935) 719.
6. J. B. Zeldovich, Acta Physicochim. URSS **18** (1943) 1.
7. J. P. Hirth and G. M. Pound, *Condensation and Evaporation* (Pergamon, Oxford, 1963).

8. A. E. Nielsen, *Kinetics of Precipitation* (Pergamon, Oxford, 1964).

9. F. F. Abraham, *Homogeneous Nucleation Theory* (Academic, New York, 1974).

10. B. Lewis and J. C. Andersen, *Nucleation and Growth of Thin Films* (Academic, New York, 1978).

11. O. Söhnel and J. Garside, *Precipitation, Basic Principles and Industrial Application* (Butterworth-Heinemann, Oxford, 1992).

12. A. C. Zettlemoyer, Ed., *Nucleation* (Dekker, New York, 1969).

13. S. Toschev, in: *Crystal Growth: An introduction,* Ed. P. Hamm (North-Holland, Amsterdam, 1973) p. 1.

14. I. Gutzow, Contemp. Phys. **21** (1980) 121; 243.

15. S. Stoyanov and D. Kashchiev, in: *Current Topics in Materials Science,* **7,** Ed. E. Kaldis (North-Holland, Amsterdam 1981) p. 69.

16. S. Kotake and I. L. Glass, Prog. Aerospace Sci. **19** (1981) 129.

17. D. Kashchiev, Cryst. Res. Technol. **19** (1984) 1413.

18. I. Gutzow, D. Kashchiev and 1. Avramov, J. Non-Cryst. Solids **73** (1985) 477.

19. K. F. Kelton, Solid State Phys. **45** (1991) 75.

20. D. W. Oxtoby, in: *Fundamentals of Homogeneous Fluids,* Ed. D. Henderson (Dekker, New York, 1992) p. 407.

21. A. Laaksonen, V. Talanquer, and D.W. Oxtoby, Annu. Rev. Phys. Chem. **46** (1995) 487.

22. C.L. Weakliem and H. Reiss, J. Chem. Phys. **101** (1994) 2398; X.C. Zeng and D.W. Oxtoby, J. Chem. Phys. **94** (1991) 4472.

23. R. McGraw, and A. Laaksonen, Phys. Rev. Lett. **76** (1996) 2754.

24. N.H. Fletcher, J. Chem. Phys. **29,** 572 (1958).

25. X.Y.Liu, J. Chem. Phys. **111** (1999) 1628.

26. X.Y.Liu, Langmuir **16,** (2000) 7337; X.Y.Liu, K.Tsukamoto and M. Sorai, *Langmuir* **16** (2000) 5499-5502.

27. R. Fowler and E.A. Giggenhein, *Statistical Thermoddynamics* (Cambrideg University, London, 1960).

28. A.A. Chernov, *Modern Crystallography III-Crystal Growth*, (Springer-Verlag, Berlin, 1984).

29. B.Mutaftschiev, in: *Handbook on Cryst. Growth,* Ed. D.T.J. Hurle (North-Holland, Amsterdam, 1993) p.187.

30. H.G. Scheider, Ed., *Epitaxie-Endotaxie* (VEB Deutscher Verlag für Grundstoff-industrie, Leipzig, 1969).

31. C. Herring, Phys. Rev. **82** (1951) 87.

32. D. Kashchiev, Surface Sci. 14 (1969) 209.

33. G. Shi, J. H. Seinfeld and K. Okuyarna, Phys. Rev. **A41** (1990) 2101.

34. D. T. Wu, J. Chem. Phys. **97** (1992) 2644.

35. V. A. Shneidman and M. C. Weinberg, J. Chem. Phys. **97** (1992) 3621; 3629.

36. D. Kashchiev, in: *Sci. and Tech. of Crystal Growth,* Eds. J.P. van der Eerden, O.S.L. Bruinsma (Kluwer Academic Publ., Dordrecht, 1995) p.53-66.

37. R. J. Davey, J. Garside, A.M. Hilton, D.McEwan, and J.W. Morrison, J. Crystal Growth **166** (1996) 970-975.

38. X.Y. Liu and P. Bennema, in: *Current Topics in Crystal Growth Research* **2** (1995) 451.

39. X.Y. Liu, E.S Boek, W.J. Briels and P. Bennema, Nature **374** (1995) 342-345.

40. X.Y. Liu and P. Bennema, J. Chem. Phys. **98** (1993) 5863.

41. X.Y. Liu, Phys. Rev.**B60** (1999) 2810-2817.

42. P. Bennema, and G.H. Gilmer, in: *Crystal Growth: an introduction,* Ed. by P. Hartman (North Holland, Amsterdam, 1973) p.263-326.

43. X.Y. Liu, E.S Boek, W.J. Briels and P. Bennema, J. Chem. Phys. **103**, (1995) 3747-3754.

44. X.Y. Liu, K. Maiwa and K. Tsukamoto, J. Chem. Phys. **106** (1997) 1870.

45. X.Y. Liu and K. Tsukamoto, J. Chem. Phys. **107** (1997) 10351.

46. W. B. Hillig, Acta Metallurg. 14 (1966) 1868.

47. X.Y.Liu, E.P.G. Van Den Berg, A.R.A. Zauner, P. Bennema, J. Phys. Chem.**B104** (2000) 11942-11949.

48. X.Y. Liu, to be published.

49. P.L. Davies, G.L. Fletcher and C.L. Hew, in: *Environmental Stress and Gene Regulation,* Ed. K.B. Storey (BIOS Scientific, Oxford, 1999) p.61.

50. B. A. Hendriksen and D.J.W. Grant, J. Crystal Growth **156** (1995) 252-260.

51. A.W. Adamson, *Phys. Chem. of Surfaces,* (John-Wilay & Jons, New York, 1990).

Challenges in crystal growth science and the microgravity tool

A.A. Chernov

Universities Space Research Association (USRA)/NASA/Marshall Space Flight Center,
4950 Corporate Dr., Suite 100, Huntsville, AL 35806, ph.: (256) 544-9196,
e-mail: alex.chernov@msfc.nasa.gov

ABSTRACT

Selected scientific topics where microgravity experiments help to solve general problems of crystal growth science are discussed. These are coupling between interface kinetics, diffusion and convection mass transport, dendritic growth, pattern formation, detached growth and crystal perfection. Specific attention is paid to biomacromolecular crystallization. Recent findings show that at least some of the crystals grown in space are purer than their terrestrial counterparts. This may open the way to faster move of the biocrystallization area from art to science and help to eliminate ever-narrowing bottleneck for development of structural biology.

1. Opportunities

First orbital flight of "Sputnik" in 1957 and first manned flight of Yuri Gagarin in 1961 opened the era of space flights. This breakthrough was prepared by dreams of K. Tsiolkovski in the 20th of the last century to fly to other planets and practical efforts to build reactive weapons during the World War II (Russian "Katyusha", German V-2 of Werner von Braun, etc.). Military goals were a major incentive to create space industry, which culminated in numerous great engineering achievements like numerous manned and automated earth-orbital flights and landing on moon, flights to Mars and Venus. These achievements triggered numerous civilian applications in telecommunications, materials, robotics and technologies. Out of the atmosphere astronomy studies of cosmic rays and earth magnetic field, Galley comet monitoring of earth surface and atmosphere from space, including weather and harvest forecast and global climate changes, like glacier retreat or expansion, are just some examples of new global scale studies that came with the cosmic era. Satellite communication and TV brought about dramatic changes in everyday life. Military applications remain vital activities in space. Supersonic transportation is unsolved problem, which also drives space – related engineering development.

Microgravity (μg) as a by-product of these developments and a crucial factor for human space flights, induced numerous biomedical studies. Essential reduction of buoyancy and solutal convection and corresponding striations in semiconductor crystals grown during the first experiments in μg [1, 2, 3] originally created an illusion that the convection induced striations may be fully eliminated in the melt grown silicon, III-V, II-VI, and other semiconductor and oxide single crystals. In reality, Marangoni convection driven by surface tension gradient on a free liquid interface was found to be strong enough to produce striations by itself [4]. It turns out that Marangoni convection flow rate, maximal near the cooled crystal and hot melt-crucible wall, may be even comparable to the buoyancy convection rates. This was one of the first lessons learned from μg experiments important for terrestrial crystal growth technology. Also, in Bridgeman growth striation amplitudes were indeed damped several times. In Bridgeman configuration, crystals grown in space were found to have only a tiny ridge and point contacts with the crucible wall and order(s) of magnitude lower dislocation density [3, 5].

Lack of buoyancy seemed also to be attractive to create foam materials. Open space is about infinite power vacuum pump so valuable for molecular beam epitaxy. So far, however, neither was explored essentially. The reasons are high cost and low speed of research and development. Space experiments require a lot of specific and more sophisticated hardware not only for experiments itself, but also for control and remote communication. Even when it is done, each improvement in the system is very slow: one should wait for the next flight for at least year-scale time, while, in conventional labs, an improvement may take a week. Unavoidable safety and justification requirements aggravate the difficulties. It is cheaper to build liquid helium (4.2K) cryopump, also of infinite power.

However, possibility in microgravity to reduce the usually complex convection heat and mass transport to pure diffusion makes it possible deeper insight into coupling and relative contributions of interface shape, molecular incorporation and transport phenomena to the overall crystallization kinetics and crystal morphology, trapping of dopants, creation of defects, pattern formation in ingots, ripening processes, colloidal crystallization and mass crystallization. Some of the first crystallizations of proteins in space resulted in a better quality and larger crystals, though, for unknown reasons. Crystal growth and material science in space received much of attention. Only till 1987, total numbers of publications approached 1800, as summarized in two overview books of Regel [6].

These developments resulted in the turn form the idealistic slogan of "factories in space" to serious studies of crystallization, solidification, and other material formation phenomena which understanding is of fundamental importance per se and help to improve terrestrial technologies. Lack of convection facilitates also precise measurements of heat and mass diffusivities, including thermodiffusion and processes in binary systems. This data is additional contribution to transition from art to science of growing crystals. So, microgravity as a scientific tool emerged. Its high cost was partly mitigated by development of space industry because of other incentives, independent of the materials science, so the space experiment opportunities were offered to the crystal growth community. After this, the practical issue was and partly is if and how µg tool may help in solving crystal growth and materials science problems.

2. What is missing in crystal growth science?

The simplest answer is "quantitative knowledge and predictive power". Major concepts have been generated and established during the last century: nucleation, interface structures and origin of the normal vs. layerwise growth modes, step sources and behavior of the layerwise growing faces, various kinds of epitaxy, morphological stability for rounded and faceted shapes, trapping of impurities and their influence on the growth mode, creation of defects like dislocations, stress, inclusions, striations, zonal structures, etc. These concepts capture basic phenomena qualitatively or even semiquanitatively. However, predictive power of these concepts is limited on one hand by insufficient elaboration of the concepts themselves and, on the other hand, by lack of basic constants often following from the same concepts. Further development is required in growth kinetics, coupled kinetic-transport stability analyses,

and in the numerous, practically crucial problems of defect formation. Where µg may help?

Molecular interactions and thus interfacial phenomena are not influenced by gravity. To study these interfacial processes, one should use vigorous stirring of solution or melt to assure that the crystallization driving force at the interface is identical to the reliably measurable or calculated supersaturation, undercooling and composition in the bulk fluid or gas. This approach works if molecular incorporation kinetics at the growing interface is much slower that the bulk mass and heat transport, like, e.g. in case of crystal growth from aqueous solution. However, even in this case, processes in the boundary layer may influence interface morphology and its stability [7, 8, 9, 10, 11, 12].

On the opposite, if transport rate is much faster or comparable to the interface incorporation rate, these processes are more strongly coupled and give rise to dependence of the fluid or gas temperature and composition at the interface on both the interface processes and transport rates. Elimination or at least essential reduction of liquid flows in microgravity may simplify the picture helping to discriminate the surface and bulk phenomena and to reveal consequences of their coupling. Morphological stability, dendritic growth, dissipative pattern formation in melts, alloys and solutions, defect formation in single crystals and ripening processes are, among others, the areas where microgravity environment help to obtain new quantitative knowledge supporting basic terrestrial materials technologies. Some of these issues are discussed below on the "pedestrian" level.

3. Surface processes vs. bulk diffusion and convection. Dendrites

Interfacial kinetics provides boundary conditions for the mass and heat transport equations. This is coupling between the surface phenomena and the convection dependent bulk transport in the fluid phase: a growing crystal generates temperature and concentration distribution, which influences the growth rate, shape and perfection of the crystal itself. Morphological instabilities and dendrites are famous examples of this feedback. Let us review in simple terms general framework of such feedback and how it is influenced by convection.

The boundary condition of heat and mass transfer come through the dependence of the interface rate V on the driving force $\Delta\mu/kT$ immediately at the interface:

$$V = V (\Delta\mu/kT). \qquad (1)$$

Here $\Delta\mu$ is the difference between chemical potentials of the crystallizing species in the mother medium and in the crystal. For the melt growth, $\Delta\mu/kT = \Delta S \Delta T$, where ΔS is entropy of fusion and ΔT is the supercooling at the interface:

$$\Delta T = T_m - T_m \Gamma_1 K_1 - T_m \Gamma_2 K_2 - T(\vec{r}, t) \qquad (2)$$

Here T_m is the melting point at the planar interface, K_1 and K_2 (cm^{-1}) are two major interface curvatures and $T(\vec{r}, t)$ is the temperature at the point r and time t under consideration.

$$\Gamma_i = \tilde{\gamma}_i / Q \quad \text{and} \quad \tilde{\gamma}_i = \gamma + \partial^2\gamma / \partial\varphi_i, \quad i = 1,2 \qquad (3)$$

are, respectively, the capillarity constants and interface stiffness with respect to the two independent angles, φ_1 and φ_2, determining the interface orientation. Q (J/cm^3) is the latent heat. For spherical crystal of a radius R, the full curvature, $K = K_1 + K_2 = 2/R$ and the capillarity lengths $\Gamma_1 = \Gamma_2$ (cm). The critical nuclei has the radius $R_c = 2T_m\Gamma/(T_m - T_\infty)$ which follows from eq. (2) at $\Delta T = 0$ and $T = T_\infty$.

The actual temperature at the interface, $T(\vec{r}, t)$, is determined by both the incorporation and transport kinetics and is the quantity through which the coupling with the external transport occurs. In the simplest case of rough interface and linear growth kinetics,

$$V = \beta^T \Delta T \qquad (4)$$

where β^T (cm/sK) is the kinetic coefficient. For simplicity, let us consider a spherical crystal of a radius R (Fig. 1)
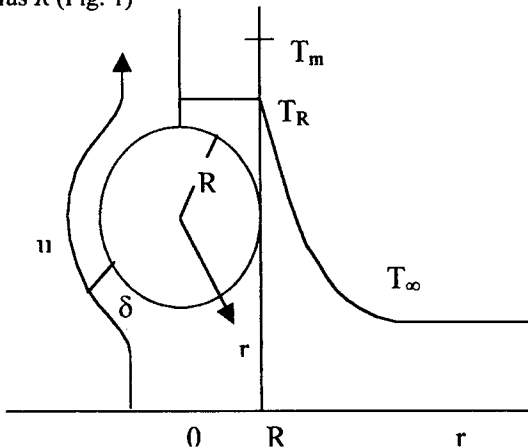
Fig. 1

in a pure melt supercooled from the melting point, T_m, down to the temperature T_∞ far away from the crystal. The latent heat at the growing crystal surface rises the temperature by $T_Q = Q/c$ where c (J/cm^{3}K) is heat capacity per unit of the solid volume. For metals, $T_Q \sim 10^2$K. For succinonitrile, $T_Q =23.3$K. If $T_m - T_\infty < T_Q$ the crystal surface and interior are heated to the temperature T_R close to the melting point, T_m. The rest of the latent heat is dissipated through the supercooled melt surrounding the crystal. The heat balance at the interface is thus

$$\kappa\frac{\partial T}{\partial r} = Q\beta^T (T_m - T_R) \qquad (5)$$

where κ (J/cm s K) is the heat conductivity. In the steady state Laplace approximation, eq. (5) can be rewritten by selecting typical size L where the temperature essentially changes between T_∞ and T_m. At the beginning of growth, $L \leq R$. At the latest stages, at the time $t \gg R^2 / 4a$, $L \simeq 2\sqrt{at}$, one has $R \ll L$, where a is thermal diffusivity of the melt, $a \sim 1.13 \cdot 10^{-3}$cm^2/s for succinonitrile. To find T_R, eq. (5) may be approximated as:

$$\frac{T_R - T_\infty}{L} = \beta^T T_Q(T_m - T_R). \qquad (6)$$

This equation gives supercooling $T_m - T_R$ at the growing interface, so that:

$$V = \beta^T (T_m - T_R) = \beta^T (T_m - T_\infty - 2T_m B\Gamma / R)$$
$$/(1 + B)$$
$$B = \beta^T T_Q L / a = \beta^T QL / \kappa. \qquad (7)$$

Protrusions on the shape perturbation, if applied to the sphere exceeding critical radius of stability, $>7 R_c$, grow and give rise to a dendrite. Directions of the dendrite branches correspond to the maximal perturbed rate,

$$\delta V = \delta\beta^T (T_m - T_R) - 2\beta^T BT_m\Gamma\delta(1/R). \qquad (8)$$

The first term is proportional to $\delta\beta^T$, and corresponds to the kinetic anisotropy, the second one is proportional to $\Gamma\delta$ $(1/R)$, i.e. to the stiffness anisotropy. If, in the eq. (7), the undercooling $(T_m - T_\infty)/T_m \leq 2B\Gamma/R$ i.e. is low enough, the surface stiffness prevails. At high undercooling, anisotropy of

β^T i.e. the first terms in eqs. (7) and (8) determines the dendrite branch direction [13].

Parameter B in eq. (7) determines relative importance of the bulk transport rate vs. average surface incorporation rate, since it presents the ratio of the latter, $\beta^T T_Q$, to the former, a/R or a/L. For $B >> 1$, the growth is controlled by bulk transport. For rough surfaces, $\beta^T \sim$ 10-100 cm/sK, $T_Q \sim$ 10-10^{2}K, so that $\beta^T T_Q \sim$ 10^2-10^4 cm/s. This strongly exceeds a/L at $a \sim$ 10^{-1} − 10^{-3} cm^2/s and $L > R \geq$ 10^{-4}cm. For succinonitrile, γ = 9 mJ/m^2, Q = 45.7 J/cm^3, $\Gamma \sim$ 2·10^{-8}cm, a = 1.13·10^{-3}cm^2/s, κ = 2.22 J/cm·s·K, T_Q = 23.3K. Thus, at R = 10^{-4}cm, B = 124 >> 1. At later growth stages, when $L \simeq 2\sqrt{at} > R$, the B constant is even larger so that the process is fully controlled by heat transport. Even for the critical nuclei at $T_m - T_\infty$ = 3K, R_c = 4.4·10^{-6}cm, one has $B \sim$ 5. For complex melts or solutions the typical surface incorporation rate equivalent to $\beta^T T_Q$ may be as low as 10^{-2} − 10^{-4} cm/s, and surface kinetics may be the rate determining step for mm-scale crystals.

If the crystal is fixed in the liquid, its higher temperature induces convection. This convection will disturb diffusive heat transfer from or mass transport to the crystal if the convection rate, u, is comparable or exceeds the heat transport, a/R, i.e., if

$$u > \sim a/R. \qquad (9)$$

For solution growth, heat diffusivity a should be replaced by mass diffusivity, $D << a$, (10^2- 10^4 times).

To estimate u, let us consider a spherical crystal of the radius R growing from the supercooled melt. We take $B >> 1$ in eq. (7), so that the growth is heat transport limited and the crystal surface is at $T \sim T_m$, i.e. is by $\Delta T = T_m - T_\infty$ warmer than the bulk of the melt. Since the melt is convectively flowing around the crystal its temperature tends to increase from T_∞ to T_m only within the boundary layer δ. The melt velocity at a distance δ from the crystal surface is $\sim \delta/R$, since typical length of liquid flow around the crystal is of the order of its radius, R. Then each element of the melt may be heated during the time $\sim \pi R^2/u\delta$. Therefore the effective thermal boundary layer

$$\delta \simeq 2(a\pi R^2/u\delta)^{1/2}, \text{ or } \delta \simeq (4\pi aR^2/u)^{1/3} \qquad (10)$$

Total buoyancy force F_b inducing convection is then force per unit volume, $g\rho k_T \Delta T$, times the volume of the boundary layer,

$$F_b = 4\pi R^2 \delta g\rho k_T \Delta T. \qquad (11)$$

In quasi-steady state convection (when the melt volume is much larger that that of the crystal) the buoyancy force is compensated by the viscous Stockes force, i.e. $F_b = 6\pi R\eta u$. This balance gives the typical convection rate

$$u = (2gk_T \Delta T/3v)^{3/4} R^{5/4}(4\pi a)^{1/4} \qquad (12)$$

At g = 10^{3}cm^2/s, k_T = 10^{-3}K^{-1}, ΔT = 3K, v = 10^{-2}cm^2/s, a = 10^{-3}cm^2/s, R = 10^{-1}cm, eq. (11) gives u $\sim$ 1cm/s. This is much faster than the heat transfer rate, $a/R \sim$ 10^{-2}cm/s, see eq. (9). From the eqs (9) and (12), convection may be insignificant only if

$$\Delta T << 3av/2 (4\pi)^{1/3} R^3 gk_T \qquad (13).$$

For the figures used above, eq. (13) means $\Delta T <<$ 6·10^{-3}K.

Because of the strong dependence of ΔT on R in eq. (13), the affordable supercooling is even smaller for larger crystals.

Therefore at fast interface kinetics, even slight convection brings about larger undercooling at the crystal surface and is resulted in larger growth rate as expected from undercooling calculated for stagnant melt. These circumstances made the use of microgravity inevitable to quantitatively analyze dendritic growth and to resolve long standing problem on the conditions determining the rate and shape of the dendrite tip and its side branching [14].

Dendrites and Gravity
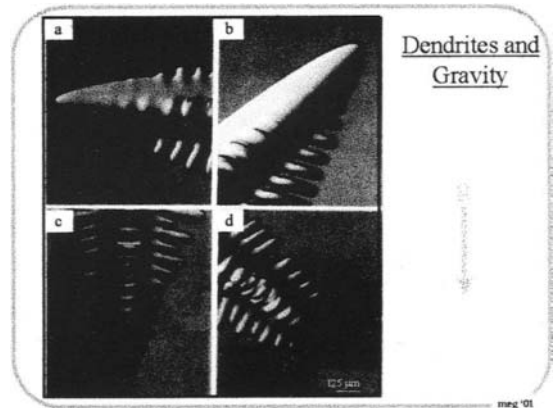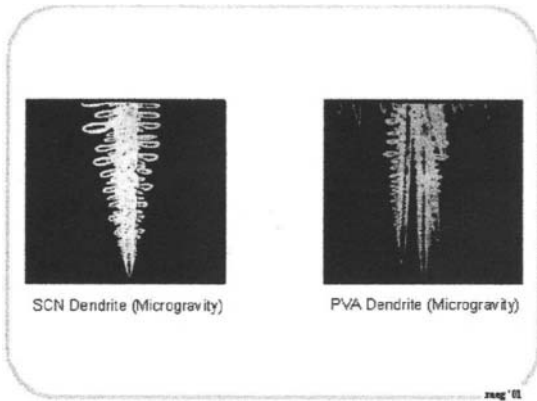
Fig. 2. Dendrites: of succinonitrile a-e and pivalic acide (f): a-d are taken under normal gravity, asymmetric shapes comes from convection flows; e, f are dendrites in microgravity. [M.E. Glicksman, M.B. Koss, L.T. Bushnell, J.C. LaCombe and E.A. Winsa, Materials and Fluids Under Low Gravity, Proceedings, Berlin, Germany (1995) p. 63-75.]

Fig. 2a-d shows the succinonitrile (SCN) dendrites growing at normal gravity and thus showing asymmetry.

Fig. 2e,f shows the SCN (e) and pivalic acid (f) dendrites in microgravity in the set up used in microgravity experiment.

(a)   (b)
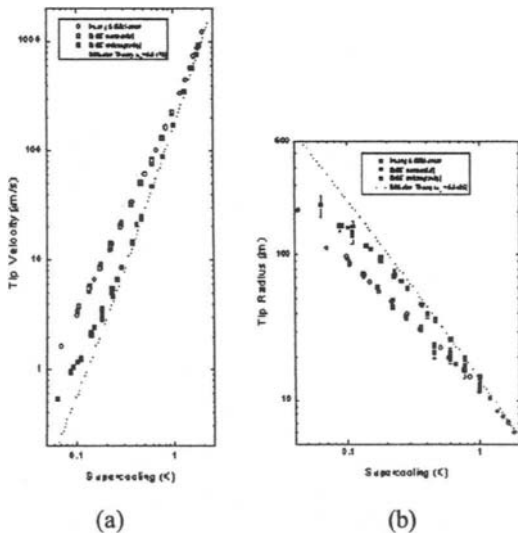
Figs. 3a, b [15] demonstrates difference in the tip growth rate and radius on earth and in space. An order of magnitude rate decrease in microgravity seen in Fig. 3a [15] demonstrate importance of convection and makes much more solid background for the existing quantitative theory of dendrite growth which is the key

process in metallurgical casting and ingot formation. (courtesy of M.E. Glicksman) [15]

4.   Pattern formation

Convection induced changes of both temperature and concentration at the growing flat interface and may essentially change conditions of morphological stability and pattern selection [16, 17, 18].

Classical examples of pattern formation are cellular or rod structures to which originally flat crystal-liquid (melt, alloy) interface transforms after it looses its stability [15, 19, 20, 21].
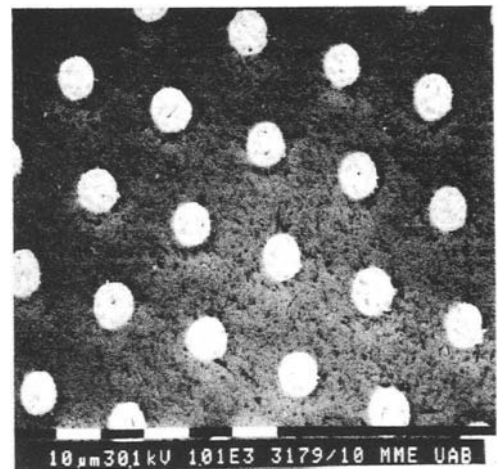
Fig. 4 (courtesy of J.B. Andrews) shows cross-section of Al-In rod (Al) eutectics made parallel to the growth interface [22]. In eutectic solidification, two components are alternatively distributed between mutually parallel lamella and rods so that both crystallizing components should diffuse tangentially along the average cellular front to reach the nearest lamella. The narrower the interlamella spacing and thus the diffusion path, λ, the lower supercooling ΔT is required to drive each of these lamella to grow at a given growth rate, V. On the other hand, under the condition of fast interface kinetics, reaching the same growth rate and undercooling requires wider lamella to minimize capillarity depression of the undercooling. Therefore, the total supercooling may be presented as [19]

$$\Delta T = A_1 V\lambda + A_2/\lambda \tag{14}$$

where the constant A_1 depends on the diffusivity, liquidus slopes on the phase diagram and volume fraction of the components and A_2 is proportional to capillarity constants [19, 20].

The supercooling at the steady interface should be minimal to achieve the same fixed growth rate V, i.e. this should be the easiest growing pattern. Minimization of eq. (14) with respect to λ results in classical condition [20]:

$$\lambda^2 V = const. \tag{15}$$

This condition works for numerous systems with rough interfaces, though quantitative discrepancy exists between predictions and experimental data, as is illustrated by the Table 1 for the Al-In alloy [21, 22].

Table 1. The products of the interrod spacing, λ, and growth rate, V, for the Al-In system [22].

Source	$\lambda^2 V$ (m^3/s)
Predicted Values	
Coriell et al.	4.0×10^{-17}
Stöcker & Ratke	1.8×10^{-15}
Experimental Results	
B. Andrews, et al.	9.0×10^{-16}
Grugel & Hellawell (1981)	4.5×10^{-16}
Vinet & Potard (1983)	5.3×10^{-16}
Kamio et al. (1991)	7.8×10^{-16}

The diffusion fields of the two eutectic components redistributing to form the rods or lamella of alternating composition penetrates into the liquid over the length comparable to the pattern period, $\lambda \sim$ 10-100μm. Thus the liquid flow perturbs this diffusion field and thus the pattern period and arrangement. The question is what is the influence of the flow on this structure. This problem is not solved and contribution of the flow is not exactly known despite of intensive efforts [23].

Even less is known what parameters control pattern formation when the growing surfaces of one or both phases are faceted. In this case, the supercooling ΔT in eq. (14) or the equivalent supersaturation at the interface should include a kinetic term, which may be essentially larger than the diffusion and capillarity ones. The same fundamental problem of how to discriminate contributions from the interface kinetics, diffusive and convection transport is not solved and is the one in which experiments in μg may help.

5. Ripening

In the classical Ostwald ripening in an ensemble of crystals in solution or melt (or droplet of one liquid in the other), larger particles grow at the expense of the smaller ones. The latter dissolve because surface energy tries to maintain higher equilibrium concentrations above their surfaces, than the concentration over the bigger particles. The first theory was developed by Lifshits and Slyozov [24] and Wagner [25]. The theory assumes pure diffusion for material transport between the particles, infinite interface kinetics and small volume fraction of the ripening particles (less that $\sim$ 10%). Then the average particle radius, $\overline{R}(t)$, should rise in time as

$$\overline{R}^3(t) - \overline{R}^3(0) = Kt \tag{16}$$

where $\overline{R}(0)$ and K are initial average particle size and the kinetic constant. The latter,

$$K = 8T_0 \Gamma D / 9m(C_S - C_L) \tag{17}$$

is expressed in terms of coarsening temperature, T_0, capillarity length Γ, diffusivity D, liquidus slope m at T_0 and liquidus and solidus concentrations, C_S and C_L. This problem has many practical applications in metallurgy, describes behavior of point defects and voids in crystals, ripening of solid solutions, etc, and coarsening in thin solid films as well since the theory was generalized to two-dimensional systems [26]. Besides the basic evolution of the average size, eq. (16), the theory predicts universal asymptotic particle size distribution, independent of the initial distribution. In other words, the relative radii $R/\overline{R}(t)$ should be distributed according the function independent of time. Detailed quantitative confrontation of the theory with experiment requires not only guaranteed pure diffusion and isotropic conditions but also lack of internal stress around the particles. Therefore, the matrix should be liquid. This leads to μg, though numerous tests have been successively made in solid solutions.

Experiments in μg on the solid Sn spheres in the Sn-Pb liquid [27, 28] showed general agreement with the prediction of eq. (16) and with the qualitative view of the asymptotic particle size distribution function, though the experimental distribution is less sharp at $R/\overline{R}(t) \to 0$. Also, this shape of this distribution was found to be not steady state - it changes with time.

These discrepancies are the larger the higher is the volume fraction of the solid particles.
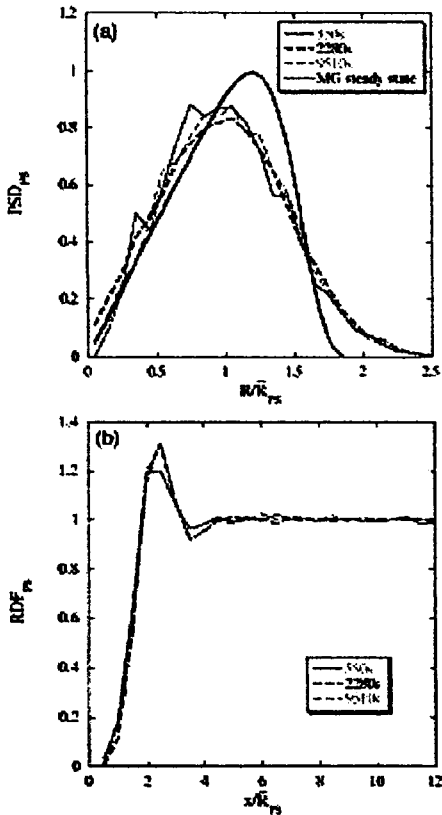

Fig. 5. The temporal evolution of (a) and the scaled radial (particle size) distribution function PSD and (b) the scaled RDF (radial distribution function) for volume fraction $\phi = 0.7$

Fig. 5 shows the normalized size and radial distribution functions as found from subsequent planar sample section for the maximal tested 70% volume fraction of the solid particles. The physics behind the deviation from the steady state comes from the fact that the theory assumes an average concentration of solution in the matrix between all of the ripening particles. In reality, however, each particle mainly "feels" the nearest particles around it. It looks like electrostatic "Debye screening" of the diffusion field of a selected particle by its nearest neighbors. Under these conditions it is not clear if the strict steady state distributions of size and interparticle distance may be achieved at any long time t, though deviation from the universal distribution function type is not dramatic. Similar to the pattern selection problem, again, numerous systems with slow interface kinetics are much less understood.

6. Detached growth

One of the first experiments on melt growth in space was growing of InSb and Ge in quartz ampoule by Bridgeman technique in cylindrical furnace [3]. A crystal originally grown in this ampoule on Earth was partly melted in space and the molten portion was re-grown in μg. This space re-grown portion showed that the crystal touched the crucible wall only along relatively rare small ridges or spikes while major part of the surface was smooth and did not indicate features of a contact with the wall [1, 4, 29, 30, 31, 32, 33]. Also, dislocation density in the microgravity re-grown portion was several times less than that in the terrestrial seed. Similar surface morphology and reduction of the dislocation density from $10^5 - 10^6 cm^{-2}$ to $10^2 cm^{-2}$ was confirmed by Bridgeman crystallization of Ge making use of quick melting by explosive burning of the special cartridge surrounding the ampoule and subsequent quick free recrystallization [5]. In the latter experiment, reduction of the dislocation density was also explained by lack of stress due to the difference in thermal expansion coefficients between the crystal and the ampoule. In addition, one-dimensional temperature field with small second radial derivatives inducing small thermal stress was estimated for this case. Finally, the growth front was estimated to move so fast that the dislocations from the send and that due to the thermal stress were unable to propagate quick enough to follow the front.

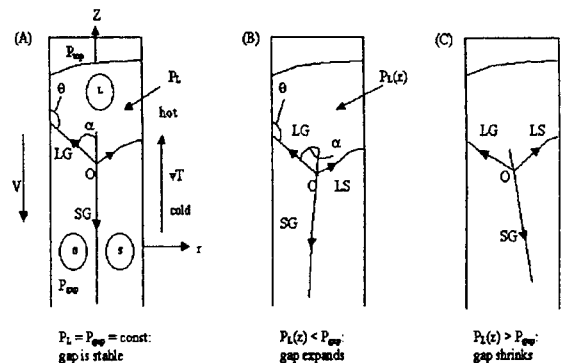Reasons for detachment are still under investigation. Physics behind it may be summarized as shown in Fig. 6.

Fig. 6

This drawing presents half of the longitudinal central (through the z-axis) cross-section of the ampoule with the melt above and the crystal below. Detached growth presumes coexistence of the liquid (L, melt), solid (S, crystalline) and gas (G). The gas phase fills

the gap between the solid and the crucible wall. This coexistence occurs along the nearly circular triple phase line seen in Fig. 6a as the point O.

Along this line, nearly conical liquid-gas (*LG*), cylindrical solid-gas (*SG*) and dome-like liquid-solid (*LS*) interfaces join. The *LS* is the crystallization front. The triade of vectors *LG*, *LS* and *SG* lie within these interfaces, the cross-section plane of the drawing and are normal to the triple phase line O. The vector lengths are the interfacial free energies between the phases, γ_{LG}, γ_{LS} and γ_{SG}. The angles between the interfaces are $(LG^\wedge GS)$, $(LG^\wedge LS)$ and $(LS^\wedge SG) = 2\pi - (LG^\wedge GS) - (LG^\wedge LS)$. Values of these angles follow from the condition that the total interfacial energy of all these interfaces should be minimal with respect to all variations of these angles, i.e. with respect to all movements of the point O along the z and r-axes within the plane of the drawing [34, 35]:

$$cos\,(LG^\wedge LS) = (\gamma^2_{LG} + \gamma^2_{LS} - \gamma^2_{SG})/2\,\gamma_{LG}\,\gamma_{LS}$$

$$cos\,(LG^\wedge SG) = (\gamma^2_{LG} + \gamma^2_{SG} - \gamma^2_{LS})/2\,\gamma_{LG}\,\gamma_{SG}$$

$$(18)$$

These conditions are similar to the equilibrium configuration along the line where three-grain boundaries meet. Any shift of the point O by a distance ξ along the z, or r, or both axes is (not shown in Fig. 6) associated with the change $\delta N_{z, r}$ in the number of particles transferred from liquid to solid, $\delta N_{z, r} \propto \xi^2$ since this change is proportional to the surface change induced by the movement of the point O in Fig.6. The same shift of the point O requires the change δF of the total surface energy. $\delta F \propto \xi$ since this change is associated with variations in the lengths of the lines meeting in the point O. Thus $\delta F/\delta N_{z,r} = A_{z, r}/\xi \rightarrow \infty$ as $\xi \rightarrow 0$, unless the so defined coefficients $A_{z, r} = 0$. These are two latter conditions (for z and r) that generate eqs. (18). The ratios $\delta F/\delta N_{z,r}$ are changes of the chemical potential difference between the melt and the crystal. Therefore the fact that $\delta F/\delta N_{z,r} \rightarrow \infty$ when $\xi \rightarrow 0$ means extreme stability of the mutual orientations of the phase boundaries near the triple line 0 satisfying eqs. (18) at the which $\delta F/\delta N_{z,r} = 0$. Therefore, at equilibrium along the line 0, the triade of vectors γ_{LG}, γ_{LS} and γ_{SG} should turn as a whole if, for any reason, one of the triade shoulders turns. One of the consequences is that the so-called growth angle α (Fig. 6a) between the extention of *SG* to the liquid phase and the interface *LG* is a constant for a given material, at a given composition and temperature at the point O. This angle [36, 37]

$$\alpha = \pi - (LG^\wedge SG) = arcos\,[(\gamma_{LG}^2 + \gamma_{SG}^2 - \gamma_{LS}^2)/2\gamma_{LG}\gamma_{SG}].$$
$$(19)$$

Typically, the *LS* free surface energy is about ten times less than the *SG* and *LG* free surface energies. Therefore, according to eq (19), the growth angle, α, is expected to be small which makes detached growth difficult as explained below.

Cristallographic anisotropy of γ_{SL} and γ_{SG} along the line projected as point O in Fig. 6a determines the side (*SG*) shape of the Czochralski pulled or zone grown crystals. Theory of this shaping follows from general surface behavior along the line O [34, 35, 36].

The principles outlined above allowed to predict wetting or dewetting with the change of gas pressure, P_{gap}, within the crystal - crucible gap, pressure above the melt, P_{top}, and the hydrostatic pressure P_L in the melt at the *LG* interface. Increase of the pressure P_{gap} within the gap (keeping P_{top} and P_L constant) induces rotation of the *LG* meniscus at the point O and thus of the whole γ-triade clockwise (Fig. 6b) so that during the growth the point O will move up towards the crucible axis z. As a result, the gap becomes wider and detached growth is maintained. Decrease in the gap pressure relative to the total pressure in the liquid along the meniscus will result is closing the gap. Evidently, the gravity induced hydrostatic component in the liquid tends to close the gap and cease detached growth unless compensated by rise of the pressure in the gap. Unfortunately, high P_{gap} induces bubbles in the melt and voids in crystal.

The detached growth is evidently (see Fig. 6a) impossible if the melt-crucible wetting angle, θ, is too small. Practically, one should have $\theta + \alpha \geq \pi$. For Ge, $\alpha = 11°$ [37], so that detachment requires $\theta > 169°$. The measured melt-crucible wetting angles vary form e.g. 70 - 90° for the CdTe/SiO$_2$ (quartz ampule) pair to 140 - 150° for GaAs/pBN and 173 $\pm$ 3° for Ge/pBN [38]. For Ge on quartz, wetting angles $\theta > 150°$. Their decrease in time was observed for SiC, graphite and pBN (presumably due to chemical modification of the surfaces). The Ge$_{1-x}$ S$_{ix}$ (x = 0.02 $-$ 0.13) melt wet pBN at $\theta = 165°$ without decrease in time. Indeed, detached growth under normal gravity conditions was realized for Ge in pBN mitride crucibles [39].

Since the crystal-crucible gap is typically narrow, within $10^{-1} - 10^{-3}$cm range, stability of detachment remains an issue. Among other factors, hysteresis of wetting angle during the growth process, and stability

of growth rate should be considered. This possibility is explained in Fig. 7 comparing equilibrium wetting angle (Fig. 7a) with the case of fast growth (Fig. 7b).
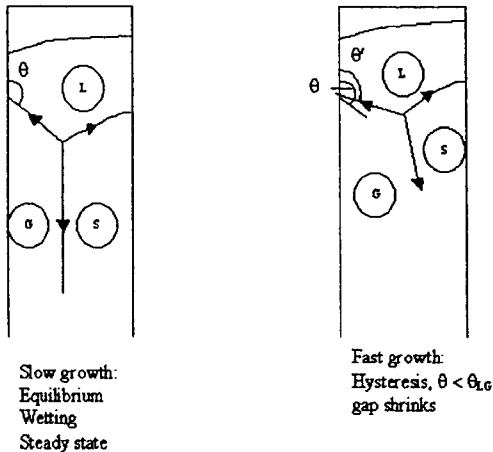

Slow growth:
Equilibrium
Wetting
Steady state

Fast growth:
Hysteresis, $\theta < \theta_{LG}$
gap shrinks

Fig. 7

During the growth, the melt should be continuously detached from the wall so that actual dynamic wetting angle θ' might be slightly less than the equilibrium one, as shown in Fig. 7b. Resulting slight rotation of the surface energy triad around the triple point O counterclockwise may stop dewetting. Changes in the growth rate and temperature might slightly affect growth angle. This may happen because of the change in the thickness of the thin melt film covering the crystal-vapor surface, for the azimuthal crystal surface orientations subject to surface melting. In particular, the growth angle may decrease as the supercooling at the growing interface increases [36] which decrease should work against dewetting. These phenomena associated with the growth dynamics might cause "spontaneous" dewetting observed on earth.

Fully detached growth not only reduces dislocation density. Lack of the stress at the surface may also reduce or eliminate twinning so typical for the II-VI compounds, since, probably; the twins are nucleated in the areas of stress concentration. These problems have not been solved so far.

7.   Biomacromolecular crystallization

Biomacromolecular crystals are rare exceptions (like hemoglobin C) in living nature. These crystals are needed to reveal, by x-ray diffraction, atomic structure of the biomolecules of which the crystals are grown in laboratory. There are hundreds of thousands various proteins, viruses, nucleic acids and their complexes. High structural resolution of atomic positions within these molecules requires high crystal perfection to collect as many Bragg's reflection as possible (of the order of 10^4 or more, especially at the periphery of reciprocal space) since molecular weight vary from 10 to 10^3 kDa or more. At present, recording of the Bragg's reflections from planes separated by 1Å is considered excellent. Resolution of 2 Å is good. Theoretical limit is about a half of the wavelength, ~ 0.5Å. To my feeling, at least a half of crystals grown in the recent structural genomic and drug design projects diffract worse than 3Å while many crystals show resolution 5-9Å and are useless for structural determination though might be good for storage of the proteins.

First attempts to crystallize proteins in space [40] and latter efforts (see, e.g., [41, 42, 43]) showed that the space-grown crystals are sometimes of higher quality than their terrestrial counterparts, i.e. diffract to higher resolution, visually better and are larger. Some references may be found in recent review on macromolecular crystallography in microgravity [44]. There is no rigorous statistics of success or failure of microgravity vs. terrestrial growth because scattering from crystal to crystal even in one experiment is high. Also, sometimes, not absolutely identical materials were used for space and earth cases, etc. Nevertheless, there is a feeling that about 20% or more of experimental runs in space resulted in larger crystals and, much more important, allow better structural resolution of molecules of which these crystals are built. The questions why and when space may help still have no answers, after about 20 years of the experiments.

Biomacromolecules are usually being crystallized by adding precipitants, like NaCl, ammonium acetate, $CdSO_4$ and many other salts or, not less often, polyetheleneglycols of various molecular weight and other chemicals [45]. Extreme difficulties in extracting and purification of various biomacromolecules from tissues, cells and production of these cells force researchers to work with milligrams or even micrograms of the substances to be crystallized. Out of ~ 10,000 crystals in the Protein Data Banks, the phase diagrams are known for several only.

This is the surface of the macromolecules, not its bulk that matters in crystallization. Therefore, from the crystallization standpoint, the expression/ purification problems are aggravated by the fact that the surfaces of these big, 1-20nm in diameter, molecules are very sensitive to pH and composition of solution. Last but not least, the biomolecules are often unstable in time.

These difficulties led to the predominately empirical approach to biocrystallization, i.e. to screening through hundreds of droplets in which solutions differ one from another by different combinations of macromolecule, precipitant, and buffer concentrations, pH, temperature and techniques (batch, sitting drop, hanging drop, size of wells, seeding, etc.). Because of the same complexity absolute majority of biocrystallographers and biochemists striving to solve new molecular structures as the #1 priority. These laboratories, though numerous, are unable to afford systematic work on the biomacromolecular crystallization science. Crystallization per se was and is not the area of major scientific interest in these laboratories. Fortunately, however, NASA, ESA, NASDA, Canadian Space Agency, being encouraged by the first empirical findings, funded fundamental studies of protein crystallization and a big progress was made during the last two decades. It was proven that biocrystallization follows the same laws as crystallization of small molecules, and the biocrystals have defects of the same type. Large size of biomolecules allowed much more successful application of static and dynamic light scattering [46, 47] and atomic force microscopy [48, 49] than the small molecules. Interferometry was productively applied to biocrystals growth as well [50, 51]. For further references see also the book [42], reviews [52, 53, 54] and Proceedings of eight International Conferences on biomacromolecular crystallization held since 1986 till 2000 [55, 56].

Despite of the deep similarity in crystallization processes of the biomacro- and conventional small molecules, there are several fundamental features of biomacromolecules that make a difference.

The size of macromolecules, 1 – 20nm, is essentially larger than the angstrom range of molecular or Debye screened ionic forces. This new scale and hydrophilicity of the biomolecular surfaces brings about essential amount of solution to the crystal, 40 – 70% of its volume. This solution, essentially water, fills the packing holes between molecules which are large as compared to the diameter of the water molecule, $\sim 2\text{\AA}$. The short-range intermolecular forces have difficulties to dock a large macromolecule into the lattice in the correct orientation because only a small torque is acting on the docking molecule. Nevertheless, exact orientation is required and indeed reached to build a lattice with orientational long-range order. This is at least one of the reasons, which makes kinetic coefficients for the steps, β_{st}, and faces $p\beta_{st}$, of biomolecular crystals ~ 100 times smaller than the ones for inorganic solution growth or even more as, seen in the Table 2. The vicinal slopes of the dislocation hillocks, $p \sim 10^{-2}$, are listed in this Table as well, since the product $p\beta_{st}$ is the kinetic coefficient characterizing the face growth kinetics and, being multiplied by $\omega(C-C_e)$, give the face growth rate (to the linear approximation). Here ω is specific biomolecular volume; C and C_e are the actual and equilibrium concentration of biomolecules. Since kinetic coefficients for biocrystals are small, surface incorporation kinetics is typically the limiting stage for biocrystallization. Therefore the protein concentration C at the step is not much lower than that in the bulk solution.

Table 2   Kinetic Coefficients β_{st} and vicinal slopes p [57, 58, 59, 60, 61, 62]

SUBSTANCE, FACE	β_{st}, cm/s	p	$p\beta_{st}$, cm/s
ADP, KDP, DKDP (100) NH₄H2PO₄, KH₂PO₄	$(5-12)\cdot 10^{-2}$	$3\cdot 10^{-4}$-$8\cdot 10^{-3}$	10^{-4}-10^{-3}
ADP (101)	0.4^{-1}	10^{-4}-$5\cdot 10^{-3}$	$4\cdot 10^{-5}$-$5\cdot^{-3}$
BaNO₃ (111)	$1.3\cdot 10^{-2}$	3-$15)\cdot 10^{-4}$	$4\cdot 10^{-6}$-$2\cdot 10^{-5}$
KAl (SO₄)₂·12H₂O (111) alums	$8\cdot 10^{-2}$	$(0.4$-$3.5)\cdot 10^{-3}$	$3\cdot 10^{-5}$-$3\cdot 10^{-4}$
Y₃Fe₅O₁₂ (110), (211)		$(0.3$-$3)\cdot 10^{-2}$	$(0.4$-$1)\cdot 10^{-3}$
(YSm)₃(FeGa)₅O₁₂(111)	$1.4\cdot 10^{-2}$		10^{-2}
(EuYb)₃Fe₅O₁₂(111)			$(0.1$-$3)\cdot 10^{-3}$
lysozyme (101) 14,300	$4.6\cdot 10^{-5}$	$(1.1$-$1.5)\cdot 10^{-2}$	$\cdot 10^{-7}$
canavalin 147,000	$9\cdot 10^{-4}$	$9\cdot 10^{-3}$	$9\cdot 10^{-6}$
thaumatin 22,000	$2\cdot 10^{-4}$		2D nucleation
catalase 25,000		$3.2\cdot 10^{-5}$	2D nucleation

Similar scaling explains low crystal-solution interfacial energies ~ 1 erg/cm^2 or even less. Another consequence of the new scaling missing in the world of small molecules is softness of the biomacromolecular crystals: their Young moduli are at least 100 times lower than those for inorganics. Like with surface energy, this difference, at least partly, comes from the ratio between cross-section area of a biomacromolecule and the area of intermolecular contact where real binding is realized. The scaling issues are discussed in [54]. As it was already mentioned, essential difficulty in biomacromolecular crystallization comes from extreme sensitivity of biomolecular surfaces to solution composition. Probably, about every area on the macromolecular surface is capable of making intermolecular contact – under relevant chemical conditions in solution. This versatility results in several crystalline modifications for each protein crystallized at specific pH, temperature, and precipitant.

Slow kinetics requires high supersaturations ($C/C_e \simeq 2^{-10}$) to grow at least a part of a millimeter size crystals in weeks or months. This high supersaturation essentially excludes the trail-and-error process of attachments and detachments to/from the lattice working to select proper molecules in proper orientation in inorganic systems. We are thus forced to assume that the orientation of biomolecules joining the growing crystal is selected in solution near its kink positions.

Biomacromolecular crystals are similar to colloidal crystals in terms of the two size scales present in both cases. Essential difference, however, comes from very specific chemical (mainly hydrogen) binding between biomacromolecules vs. repulsive or short range attractive, but still not specific interaction between structureless colloidal particles. Nevertheless, some similarity between dependence of solubility on the average intermolecular interaction strength has been demonstrated [63]. Monte Carlo simulation showed that the equilibrium between colloidal liquid and "vapor" (high or low colloidal particles density in solution, respectively) disappears when "the range of attractive part of the Yukawa potential is less than approximately 1/6 of the hardsphere diameter" [64]. Only the crystal – liquid (fluid of intermediate density) equilibrium remains. So, the "vapor" (diluted colloidal solution) - solid condensation will happen without intermediate liquid-liquid phase separation. The dense liquid may still appear as a metastable phase. This phenomenon may be of interest for biomacromolecular solutions in which, indeed, liquid-liquid phase separation sometimes occurs first, as a metastable phase [65]. Intermediate dense liquid

phase covering the solid-"vapor" interface, i.e. surface "melting", may essentially decrease the nucleation barrier since the overall "vapor"-solid surface energy decreases [66].

Very important condition for biomacromolecular crystallization is purity of solution with respect to irregular biomacromolecular species, like dimers, trimers, fraction of protein molecules and molecules with chemically modified surfaces. These microheterogeneous (homologous) impurities, being trapped by a crystal, induce stress, strain and mosaicity [67]. Ageing of solutions in which these impurities are spontaneously generated is an additional problem. Presence of "foreign" impurities much different from the crystallizing protein is probably of less influence on the crystal quality.

As it was mentioned above, a rationale behind improvements or deterioration of biomacromolecular crystals grown in space is still missing. To my view, this is the result of the overwhelming empirical "let us see;" approach just brought to space from traditional screening approach in terrestrial experiments.

However, it was found recently that micro-heterogeneous impurities have distribution coefficients (with respect to the crystallizing protein) exceeding unity. This distribution coefficients, K, is defined as

$$K = (C_{is}/C_{ps})/(C_{il}/C_{pl}), \qquad (20)$$

i.e. is defined with respect to the crystallizing protein. Here C stands for concentration, i for impurity, p for protein, s and l for the solid (crystal) and liquid (solution) phases, respectively. As it follows from the eq. (22) below, the impurity balance at the growing crystal interface is characterized by the distribution coefficient

$$k = C_{is}/C_{il} = K (C_{ps}/C_{pl}) >> K. \qquad (21)$$

This coefficient is defined with respect to the impurity concentrations per unit volume. The latter inequality in eq. (21) follows from the fact that the concentration of the crystallizing protein in solution is low as compared to the concentration of this protein in the crystal, typically about a percent or even less. So, $k \sim 100\ K$. In all the systems we studied so far, for homologous impurities, $K > 1$, while for the "foreign" proteins $K < 1$, though $k >> 1$, as it can be seen in the Table 3 for ferritin. Results for lysozyme are similar. Table 3 gives also an example of purification of ferritin crystal in space [68]. Namely, distribution coefficients in space are ~ 2.5 times lower than on the

ground for dimers while other impurities turned out to be below detection limit (10^{-3}) for the space grown crystals.

Table 3.: Distribution Coefficients of impurities in ferritin. Monomer: pI = 4.0, Charge: (–), Mr = 450 kDa; 24 subunits, Mr = 19kDa; Solution: ferritin –1mg/ml, NaAc – 0.2M, pH 5.0, $CdSO_4$ – 2.5% w/v; impurity: dimer 0.05 mg/ml (5%w/w), others –10% w/w. The volumetric distribution coefficients, k, are calculated by making use of measured K-values and of Eq. (21), initial lysozyme concentration, C_{pl}, for C_{ps}/C_{pl} = 457) [68]

Impurity, M_r (kDa), pI, Charge			Ground		Space	
			K	k	K	k
Dimer	900	4.2 –	4	1.8×10^3	1.5	690
Ribonuclease	14	9.6 +	0.1	46	$<10^{-3}$	
Insulin	5.7	5.2 +	0.02	9.1	$<10^{-3}$	
Cytochrome C	13	10.6 +	0.02	9.1	$<10^{-3}$	
Myoglobin	17	7.0 +	0.006	2.74	$<10^{-3}$	
Ovalbumin	43	4.6 –	$<10^{-3}$		$<10^{-3}$	

This purification in space may be qualitatively understood as follows, in agreement with the diffusion theory developed earlier [67, 68]. This approach is based on the impurity balance at the growing surface, $r = R$:

$$D_i \, \partial C_{il}/\partial r = (k - 1) \, V C_{il} \qquad (22)$$

where D_i is the diffusivity of impurity in solution, V is the crystal growth velocity and the crystal is approximated by a sphere of a radius R. If $k > 1$, impurity depletion zone develops around this crystal as it was visualized experimentally in gel [69]. The typical dimensionless parameter that controls the impurity distribution follows from eq. (22). It is D_i $(k - 1)$ V/R. The larger this parameter, the deeper is the depletion. Diffusivity of impurities is only slightly lower for dimers ($2 \cdot 10^{-7}$ cm^2/s for dimers vs. $3.2 \cdot 10^{-7}$ cm^2/s for monomers) and might be even larger for the other smaller size impurities. The major difference comes from the distribution coefficient, $k \gg 1$. In many cases this depletion may be several times deeper than the depletion with respect to the crystallizing protein.

In absence of solutal convection, the just nucleated crystal starts growing in the impurity rich solution. If $k > 1$, this crystal sucks impurity from the surrounding solution so that impurity depletion zone develops and, as a result, each next portions of the crystal grows from purer solution - at the expense of the impurity enriched core. Thus, if this core does not give rise to mosaicity, there is a chance for higher perfection of the main volume of the crystal and for better structural

resolution – because x-ray diffraction reflection intensities are proportional to the diffracting volume. If convection is present it brings the impurities to the growing crystal continuously so that detrimental effect of the impurities is stronger.

Our concept of the impurity depletion zone is supported not only by ours but also by the Canadian flight data with 10 proteins [70]. In those 5 of these proteins (cholesterol oxidase, duck I crystalline, fru-1.6 phase, H162N duck II crystalline, plasminogen activator inhibitor-1) in which solution light scattering indicated ageing, weaker mosaicity was found in the microgravity grown crystals. The quality of the other 5 (E-coli aldolase, lysozyme, thaumatin, rabbit muscle aldolase, xylanase) protein crystals which solutions did not show sign of ageing and thus, presumably, creation of the microheterogeneous impurities, did not show improvement in space as compared with their terrestrial counterparts.

The impurity depletion zone is fundamentally different from the depletion with respect to the crystallizing protein: the impurity depletion is controlled by the impurity distribution coefficient rather than by the growth kinetic coefficient. Namely, the latter is typically small as compared to the diffusion rate, D/R, so that the depletion with respect to the crystallizing protein should be weak (~ 10% for lysozyme) [71], several times weaker than the depletion with respect to the microheterogeneous impurity. More important in this respect is, however, the fact that decrease in supersaturation may easily be achieved under terrestrial condition just by choosing

lower protein concentration. On the contrary, impurity depletion zone, corresponding diffusion filtration and crystal purification cannot be achieved on earth other than by thorough purification of the initial solution and/or in gels. From our impurity depletion zone concept the following recommendation may follow:

- Positive effect of space on the biomacromolecular crystallization may be expected for the systems either insufficiently purified from microhetrogeneous (homologous) impurities or for the systems in which these impurities are spontaneously generated during the time comparable with preparation of the experiment and during the crystallization process itself. It is important that the ideal purification of proteins is hardly possible, at least at this point of development.

- To improve crystal quality, μg experiments should be made in sufficiently large volumes of solution: otherwise the growing crystals will trap major portion of the impurities anyhow, so that the quality of the space grown crystals will not differ from that of their terrestrial counterparts.

- The μg experiments require such a low level of gravity and vibrations that the impurity depletion zone is not essentially destroyed.

Of course, presence of the homologous impurities may be not the only factor, which deteriorate crystal quality. Therefore, broad analysis of imperfection in the biocrystals is needed to make this area the science rather than the art. The latter demand is supported by the present development of genomics and proteomics: success of the empirical screening approach in this great effort hardly exceeds 50%. As a result, the rest of the biomacromolecules that do not produce crystals allowing high structural resolution are piling up. Fast progress in the x-ray structural determination will make this determination fully automated in 5-7 years and the crystallization bottleneck will become even more serious obstacle on the way of structural biology. Rationale approaches should be more intensively developed now.

8. Conclusions

Application of μg to crystallization and solidification allowed realization that Marangoni convection may be essential even under terrestrial conditions and is superimposed on the buoyancy or solutal convection in terrestrial technologies. Discovery of detached growth in μg and further studies suggested a possibility of detached growth on earth. Though this detachment is not easy to achieve it may essentially reduce dislocation density and formations of twins in the melt and vapor grown crystals. Application of μg allowed to essentially support present theory of dendritic growth proving that the earlier discrepancy with experiment came from buoyancy convection in the undercooled melt. Experiments in μg allowed to build rigorous foundation for the Ostwald ripening theory, showed limit of its applicability and support "diffusion screening" concept. Experiments in μg allowed to essentially reduce striations. The fast interface kinetics was found to be able to imprint in to the growing crystal even slight composition inhomogeniety that exists e.g. in the II-VI semiconductor melt in presence of residual acceleration as low as 10^{-6} g during the space flight [72, 73]. Thus, application of μg demonstrates extreme sensitivity of the growing rough interface to the solidification condition. Pattern selection remains an issue in materials science and in the non-linear dynamics and is influenced by liquid flow. This is another area where μg studies may be essential. Crystallization and pattern selection processes related to the smooth, facetted surfaces remain much less elaborated than the one on the rough surfaces.

Experiments in μg suggested that convection may have detrimental effect on the biomacromolecular crystals - proteins, viruses, DNA, RNA and their complexes. The mechanism behind this influence and understanding of positive effect of microgravity is emerging only last years. The rationale for the biomolecular crystallization as a whole is not developed. However, fundamental studies funded by Space Agencies made it possible to make a big progress in biomacromolecular crystallization science and techniques. Well-established similarity between the crystallization of biological and small molecule and identification of specific differences open the way to the rationale approach to biomacromolecular crystallization. In particular, the discovered diffusional purification in space via the impurity depletion zone contributes to this rationale.

Crystal growth and materials science emerged during the 20-th century as an important area of knowledge between physics, chemistry, engineering and biology. Despite of tremendous success and decisive contributions to the technical revolution, crystal growth science did not yet achieved, as a whole, quantitative level and does not have enough predictive power. Coupled interface heat and mass transport phenomena, remain, among the others, one of the most difficult areas. Eliminating gravity driven processes,

experiments in space and related investigations under terrestrial conditions essentially contributed to solid fundamental background for solution of related problems. Focus on fundamental knowledge is the most relevant way for future progress.

9. Acknowledgement

My thanks go to S. Lehoczky, F. Szofran, D. Gillies, B. Andrews, M. Glicksman and M. Koss for

10. Reference:

1. H.C. Gatos, in Materials Processing in reduced gravity environment in space. Proc. of MRS meeting, Nov. 1981, Boston, Mass, Ed. G.E. Rindon, North Holland, Amsterdam 1982, p. 676.

2. A.F. Witt, Space Res. 19 (1979) 503

3. R.L. Parker, Current Topics in materials science, Ed. E. Kaldis, 2 (1977) 851.

4. A. Eyer, H. Leiste, R. Nitsche, 5th European Symposium "Materials Science in Microgravity", Paris (1984) 173.

5. A.A. Chernov, S.N. Maximovski, L.A. Vlasenko, E.N. Kholina, V.P. Martovitskii, V.L. Levtov, Sov. Phys. Crystallog. 29 (1984) 222.

6. L.L. Regel, Materials Science in Space (Kosmicheskoe materialovedenie; Itogi nauki itechniki, Series: Issledovamie kosmicheskovo prostranstka) vol. 21 (1984), vol 29 (1987), Moscow (in Russian).

7. A.A. Chernov, Progr in Cyst Growth and Charact of Mat 26(1993)121.

8. A.A. Chernov, Contemp Phys 30 (1989) 251.

9. A.A. Chernov, J.Cryst Gr 118 (1992) 333.

10. A.A. Chernov, Y.G. Kuznetsov, I.L. Smol'ski,V.N. Rozhanski, Sov Phys Crystallog. 31 (1986) 705.

11. S.R. Coriell, B.T. Murray, A.A. Chernov, J. Cryst Gr. 141 (1994) 219.

12. S.R. Coriell, A.A. Chernov, B.T. Murray, G.B. McFadden, J Cryst Gr 183 (1998) 669

13. E.A. Brenner, V.I. Mel'nikov, Adv. Phys. 40(1991) 53.

14. M.E. Glicksman, S.P. Marsh, Handbook of Crystal Growth, Ed. D.T.J. Hurle, vol. 1b, North Holland, Amsterdam, 1993.

15. Y.W. Lee, R.N. Smith, M. Glicksman, M.B. Koss, Annual Rev. of Heat Transfer VIII (1996) 59

16. B. Billia, R. Trivedi, Handbook of Crystal Growth, Ed. D.T. J. Hurle, vol. 1b, North Holland, Amsterdam, 1993.

17. S.R. Coriell, G.B. McFadden, Handbook of Crystal Growth, Ed. D.T. J. Hurle, vol. 1b, North Holland, Amsterdam, 1993.

18. S.H. Davis, Handbook of Crystal Growth, Ed. D.T. J. Hurle, vol. 1b, North Holland, Amsterdam, 1993.

19. H. Biloni, W.J. Boettinger, in Physical Metallurgy, Ed. R.W. Cahn, Ch. 8 (1996) 669.

20. K.A. Jackson, J.D. Hunt, Trans. AIME 236 (1966) 1129.

21. J.B. Andrews, L.J. Hayes, Y. Arikawa, S.R. Coriell, Solidification and Gravity 2000, Proc. of the 3rd International Conf. Miskolc, Hungary, April 25-29, 1999, Transtech. Pub. (1999) 247.

22. J.B. Andrews, private communication

23. S.H. Davis, Handbook of Crystal Growth, Ed. D.T.J. Hurle, vol. 1b, (1993) 860

24. I.M. Lifshits, V.V. Slyozov, J. Phys. Chem. Solids 19 (1961) 35.

25. C. Wagner, Electrochem 65 (1961) 581

26. S.A. Kukushkin, V.V. Slyozov, Dispersed systems on solid surfaces (Dispersnye systemy na poverkhnosti tverdykhtel) Nauka Pub. St. Petersburg (1996) Russian.

27. V.A. Snyder, J. Alkemper, P.W. Voorhees, Acta Mater. 49 (2001) 699.

28. V.A. Snyder, J. Alkemper, P.W. Voorhees, Acta Mater. 48 (2000) 2689.

29. J.T. Yue, F.W. Voltmer, J. Cryst. Gr. 29 (1975) 329.

30. T. Dufar, I. Paret-Harter, P. Dusserre, J. Cryst. Gr. 100 (1990) 171.

31. W.R. Wilcox, L.L. Regel, Microgravity Science and Technology VIII/1 (1995) 56.

32. D. Larson, J.I.D. Alexander, D. Gillies, F.M. Carlson, J. Wu, D. Black, Joint Launch + One Year Science Review for USML-1 and USMP-1 with the Microgravity Measurement Group, September 22-24, 1993, Huntsville, Alabama. NASA CP-3272, vol. 1, p. 129

33. D. Gillies, S.L. Lehoczky, F.R. Szofran, D.J. Larson, C.H. Su, Y-G. Sha, H. A. Alexander, SPIE vol. 2001 (1993) 10.

34. V.V. Voronkov, Sov. Phys. Crystallog. 23 (1963) 415.

35. V.V. Voronkov, Sov. Phys. Crystallog. 23 (1975) 573.

36. V.V. Voronkov, Sov. Phys. Crystallog. 23 (1978) 137.

37. W. Bardsley, F.C. Frank, G.W. Green, D.T.J. Hurle, J. Cryst. Gr. 10 (1971) 263.

38. F. Szofran, A. Croel, private communication.

39. P. Dold, F. Szofran, Submitted to J. Cryst. Gr. 2001.

40. W. Littke, C. John, Science 285 (1984) 203.

41. L.J. DeLucas, M.M. Long, K.M. Moore, W.M. Rosenblum, T.L. Bray, C. Smith, M. Carson, S.V.L. Narayana, M.D. Harrington, D. Carter, A.D. Clark, R.G. Nanni, J. Ding, A. Jacobo-Molina, G. Kramer, S.H. Hughes, E. Arnold, H.M. Einspahr, L.L. Clancy, G.S.J. Rao, P.F. Cook, B.G. Harris, S.H. Munson, B.C. Finzel, A. McPherson, P.C. Weber, F.A. Lewandowski, T.L. Nagabhushan, P.P. Trotta, P. Reichert, K.P. Navia, K.P. Wilson, J.A. Thomson, R.N. Richards, K.D. Bowersox, C.J. Meade, E.S. Baker, S.P. Bishop, D.J. Dunbar, J. Prahl, A. Sacco, Jr., C.E. Bugg, J. Cryst. Growth 135 (1994) 183-195

42. A. McPherson, Crystallization of Biological Macromolecules, Cold Spring Harbor Laboratory Press, New York, 1999.

43. K. M. Moore, M.M. Long, L.J. DeLucas, CP456, Space Technology and Application International Forum – 1999, Ed. M.S. El-Genk, 1999, 217.

44. C.E. Kundrot, R.A. Judge, M.L. Pusey, E.H. Snell, Crystal Growth and Design 1 (2001) 87.

45. A. Ducruix, R. Giege (eds) Crystallization of Nucleic Acids and Proteins. A Practical Approach. Oxford Univ. Press, NY 1992.

46. A. George, W.W. Wilson, Acta Crystal. D50 (1994) 361; A. Malkin, A. McPherson, Acta Crystal. D50 (1994) 385.

47. A. Malkin, A. McPherson, Acta Cryst., D50 (1994) 385.

48. S.D. Durbin, G. Feher, J. Crystal Gr. 110 (1991) 41.

49. A.J. Malkin, T.A. Land, Y.G. Kuznetsov, A. McPherson, J.J. De.Yoreo, Phys,. Rev. Lett. 75 (1995) 2778.

50. P.G. Vekilov, M. Ataka, T. Katsura, J. Cryst. Gr. 130 (1993) 1232.

51. Y.G. Kuznetsov, A. J. Malkin, A. Greenwood, A. McPherson, J. Struct. Biol. 114 (1995) 184.

52. A.A. Chernov, H. Komatsu, Science and Technology of Crystal Growth, Kluwer Ac. Pub. 1995, p. 329.

53. F. Rosenberger, P.G. Vekilov, M. Mushol, B.R. Thomas, J. Cryst. Gr. 168 (1996) 1.

54. A.A. Chernov, Physics Reports, 174 (1997) 354.

55. J. Cryst. Gr. 76 (186); 90 (1988); 110 (1991); 122 (1992); 168 (1996); 196 (1999) and Acta Cryst. D50 1994.

56. The ICCBM8 Proceedings are to appear in the J. Cryst. Gr. in 2001.

57. A.A. Chernov, J. Cryst. Gr. 118 (1992) 333.

58. P.G. Vekilov, Y.G. Kuznetsov, A.A. Chernov, J. Cryst. Gr. 121 (1992) 643.

59. P.G. Vekilov, M. Ataka, T. Kasura, J. Cryst. Gr,. 130 (1993) 317.

60. Y.G. Kuznetsov, A. J. Malkin, A. Greenwood, A. McPherson, J. Struct. Biol. 114 (1995) 184.

61. T.A. Land, J.J. DeYoreo, J.D. Lee, Surf. Sci. 84 (1997) 136.

62. A.J. Malkin, Y.G. Kuznetsov, W. Glantz, A. McPhersonm, J. Phys. Chem., 100 (1996) 11736.

63. D.F. Rosenbaum, C.F. Zukoski, J. Cryst. Gr. 169 (1996) 752.

64. M.H. Hage, D. Frenkel, J. Chem, Phys. 101 (1994) 4093.

65. M.L. Broide, C.R. Berland, J. Pande, O.O. Ogun, G.B. Benedek, Proc. Nat. Ac. Sci. USA 98 (1991) 5660-5664.

66. P. Reim, D. Frenkel, Science 277 (1997) 1975.

67. A.A. Chernov, J. Cryst. Gr 174 (1997) 354; 196 (1999) 524; Acta Cryst A54 (1998) 859.

68. B.R. Thomas, A.A. Chernov, P.G. Vekilov, D.C. Carter, J. Cryst. Gr. 211 (2000) 149.

69. A.A. Chernov, J.M. Garcia Ruiz, B.R. Thomas, J. Crystal Growth, in print.

70. T.-S. Yoon, S. Tetreault, H.E. Bosshard, R.M. Sweet, J. Sygusch, ICCBM8 Proceedings, J. Cryst Gr., in print, 2001.

71. F. Otalora, M.L. Novella, J.A. Gavira, B.R. Thomas, J.M Garcia-Ruiz, Acta Cryst. D57 (2001) 412-417.

72. D.C. Gillies, F.A. Reeves, L.B. Jeter, J.D. Sledd, J.M. Cole, S.L. Lehoczky, SPIE Proc. vol. 2809 (1996) 329.

73. A.V. Bune, D.G. Gillies, S.L. Lehoczky, SPIE Proc. vol. 3123 (1997) 241.

Surface step dynamics: basic concepts, theory and simulation

Makio Uwaha [a*]

[a] Department of Physics, Nagoya University,
Furo-cho, Chikusa-ku, Nagoya 464-8602, Japan

Step dynamics is crucial for the growth of a crystal at temperatures lower than the roughening transition. Recent advancement of microscopy such as various kinds of high resolution electron microscopes and scanning microscopes has made direct comparison of statistical theory and quantitative atomic scale experiment possible. Short review of basic ideas on the motion of steps and some recent theoretical development are presented.

1. BASIC IDEA OF STEP MODELS

Atomic steps on a facet or on a vicinal face of a crystal play a crucial role in the growth of a faceted crystal. Energetically the most favorable form of a step on a large facet is straight and the most favorable distribution of steps is uniform. This configuration is realized in equilibrium thanks to the step stiffness and the repulsive interaction between steps. During growth, however, they change their form and distribution, and as a result the surface morphology changes. In order to understand the behavior of steps at a mesoscopic level we introduce two kinds of models: a lattice model for Monte Carlo simulation and a continuum string model for both mathematical analysis and simulation of large scale phenomena (Fig. 1).

1.1. Monte Carlo lattice models

In the lattice model for Monte Carlo simulation the crystal surface is represented by a single valued function $z = h(x, y)$, where the spatial coordinate is limited to an integer multiple of the lattice constant a. This is the so-called solid-on-solid (SOS) surface. Since in the simulation the system size is very limited, $L \times H$, a periodic boundary condition is imposed to mimic an infinite surface

$$h(x + L, y) = h(x, y), \tag{1}$$
$$h(x, y + H) = h(x, y). \tag{2}$$

To study a vicinal surface with N steps running along the x-axis, the second condition is modified to a helical boundary condition

$$h(x, y + H) = h(x, y) + Na. \tag{3}$$

In this paper the surface we have in mind is a (001) face, or its vicinal, of a simple Kossel crystal in coexistence with the vapor. Although a variety of algorithms have been used in crystal growth simulations, there are two types according to the treatment of steps and adatoms.

Standard lattice model

In the standard lattice model the surface energy is expressed as

$$\mathcal{H}(\{h(i, j)\}) = J \sum_{\text{n.n.pairs}} |h(i', j') - h(i'', j'')|, \tag{4}$$

where we have put the lattice constant unity and the sum is taken over the nearest neighbor pairs. This is the so-called absolute SOS model. There are several variations of the model: the discrete Gaussian model in which $|h(i', j') - h(i'', j'')|$ is replaced by $(h(i', j') - h(i'', j''))^2$, and the restricted SOS (RSOS) model in which the height difference is limited to unity. In the simulation the height may be changed by the usual Monte Carlo algorithm.

To study growth from the vapor, however, we need to consider adsorption, evaporation and surface diffusion explicitly. A convenient simulation algorithm was introduced by Gilmer and Bennema [1]. The adsorption rate is usually assumed
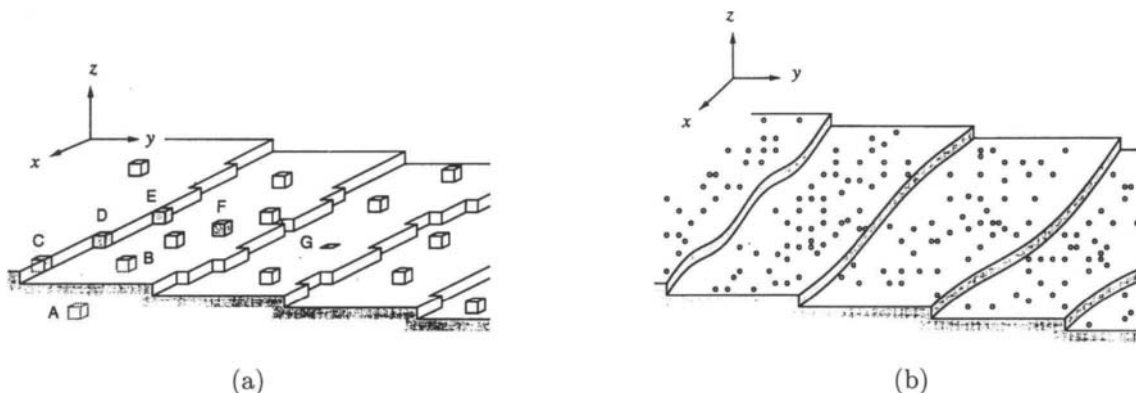
Figure 1. Vicinal face in the view of two models: (a) the lattice model, (b) the continuum step model. There are various kinds of atoms in (a): solid A, terrace B, step C, kink D, step (or adatom) E, adatom F, vacancy G.

to be a constant: $\tilde{f}$ atoms per unit time on a surface site. Putting this rate as the unit, we assign each process a relative transition probability p. For example, if the bond energy is ϕ_B the evaporation rate in equilibrium from a site with n lateral bonds is

$$\tilde{f}p_n = \tilde{f}\exp\frac{(2-n)\phi_B}{k_BT}, \tag{5}$$

where T is temperature of the crystal. The rate has been so chosen that the impingement and the evaporation are balanced on the kink site with $n = 2$. If the system is under supersaturation, the probability is changed to

$$p_n = \exp\frac{(2-n)\phi_B - \Delta\mu}{k_BT}. \tag{6}$$

The surface diffusion may be considered as a successive evaporation and impingement to the next site. The transition rate is assumed as

$$\tilde{f}p_n^D = \tilde{f}\exp\frac{\phi_B - E_D}{k_BT}\exp\frac{(2-n)\phi_B - \Delta\mu}{k_BT}, \tag{7}$$

where E_D is the energy barrier for diffusion. For an atom with the single vertical bond only ($n = 0$) the diffusion probability is larger than the evaporation probability (6) by the Boltzmann factor with the bond energy minus the diffusion barrier.

Using a model of this type, growth of a singular face and vicinal faces has been studied in detail [2, 3].

Restricted models

Although the above standard model can be used for general purposes, it is sometimes better to use a restricted model to study a particular aspect. For example, if we study the vapor growth of a vicinal face, several growth modes are distinguishable depending on the impingement rate of atoms from the vapor (f atoms per unit time per unit area): the two-dimensional nucleation mode for large f and the step flow mode for small f.

We give here an example of a Mote Carlo simulation model to study step dynamics in the step flow mode [4–6]. To elucidate the behavior of steps and the role of mobile atoms on the surface, we define an adsorbed atom (adatom) and a solid atom separately, and assume that the adsorbed atom and the solid atom can transform each other at a step position. Also the interaction between adatoms is neglected except the condition that two adatoms cannot occupy the same site, thus the adatoms are an ideal two-dimensional lattice gas. In this way two-dimensional nucleation on a terrace and formation of surface vacancy on a terrace are forbidden. Initially, the system contains n (≥ 1) steps running parallel to the x-axis,

and the height decreases on traversing the step in y direction. We call this a [10] step hereafter.

Gas atoms impinge on top of the crystal with a frequency f and are adsorbed on it. When an adatom among N_a adatoms in the system tries to make a random walk on the crystal surface to one of the four nearest neighbor sites, time increases $(4N_a)^{-1}$. Thanks to this choice the diffusion constant D_s takes the value unity. An adatom desorbs into the vapor with a lifetime τ. With the energy parameters of the previous lattice model, the diffusion coefficient is

$$D_s = a^2 \nu_0 \, e^{-E_D/k_B T}, \tag{8}$$

where ν_0 is the characteristic frequency of atomic motion, and the lifetime is

$$\tau = \frac{1}{\nu_0} e^{\phi_B/k_B T}. \tag{9}$$

In the present model we treat D_s and τ as independent parameters.

During diffusion when an adatom touches a step from the lower terrace, it tries to solidify with a probability

$$p_s = \frac{1}{1 + \exp\left(\Delta E_s - \phi\right)/k_B T}. \tag{10}$$

Here ΔE_s is the change of the step energy $E_s = \phi_B \times (\text{step perimeter}/a)$, and ϕ is the energy gain by the solidification. In this model supersaturation is controlled by changing the impingement rate f. To satisfy the detailed balance, the solid atoms at step sites can melt to become an adatom with a probability

$$p_m = \frac{1}{1 + \exp\left(\Delta E_s + \phi\right)/k_B T}. \tag{11}$$

From the parameters ϕ_B and ϕ one can calculate physical quantities such as the step stiffness $\tilde{\beta}_{[10]}$ as [7]

$$\frac{k_B T}{\tilde{\beta}_{[10]}} = \frac{2e^{-\phi_B/k_B T}}{(1 - e^{-\phi_B/k_B T})^2}, \tag{12}$$

and the equilibrium adatom density

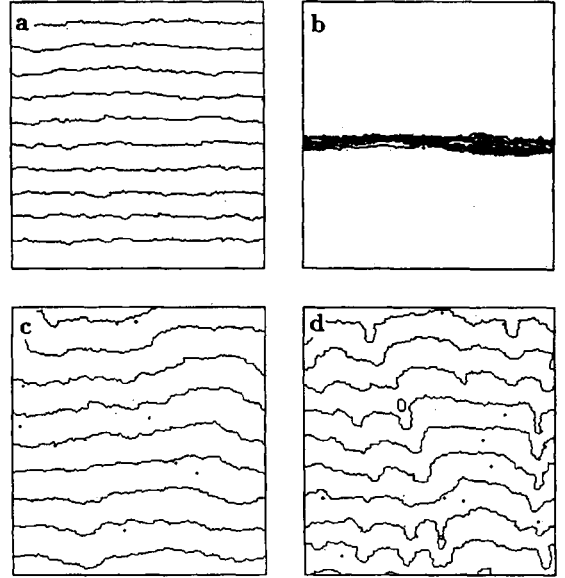$$c_{eq}^0 = \frac{1}{1 + e^{\phi/k_B T}}. \tag{13}$$

Figure 2. Time evolution of the [10] step profile at (a) $f = 3.0 \times 10^{-4}$, (b) $f = f_{eq} = 4.65 \times 10^{-4}$, (c) $f = f_c = 6.26 \times 10^{-4}$, and (d) $f = 8 \times 10^{-4}$ (from Ref. [6]).

Figure 2 shows time evolution of a single step at various impingement rate f (in this simulation adatom diffusion over a step is forbidden, see **2.1**). The system size is chosen to be $L \times H = 256 \times 128$, but the views in Fig. 2 are extended in y direction. The kink energy ϕ_B is taken as $\phi_B/k_B T = 2.0$, the energy gain $\phi/k_B T = 2.0$, and the lifetime of the adatom $\tau = 256$. Then the calculated physical quantities are $k_B T/\tilde{\beta}_{[10]} = 0.362$, $c_{eq}^0 = 0.119$, the equilibrium impinging rate $f_{eq} = c_{eq}^0/\tau = 4.65 \times 10^{-4}$ and the diffusion length $x_s = \sqrt{D_s \tau} = 16$. The step in (a) is receding downwards, and the steps in (c) and (d) are advancing upwards. As we will show in 2.1, the step is expected to become unstable at a critical impingement rate, which is calculated to be $f_{c,[10]} = 6.26 \times 10^{-4}$. With f smaller than the equilibrium value, the crystal is sublimating and the step is receding, as is shown in Fig. 2(a). At f_{eq} the step stops moving on average (Fig. 2(b)). For f larger than f_{eq} the

step grows forward. Until $f_{c,[10]}$ the step shows large fluctuations but remains stable (Fig. 2(c)), whereas for $f > f_{c,[10]}$ the step becomes unstable and develops deep grooves (Fig. 2(d)). The analysis of the step morphology for $f > f_{c,[10]}$ will be described in 2.1.

It is sometimes more convenient to use simpler models such as an SOS step model in which the position of step is assumed to be a single-valued function of x: $y = y_s(x)$. The solid-on-solid condition eliminates overhangs of a step and formation of surface vacancies.

1.2. Continuum step models

The step motion at length scales larger than the average kink distance can be described by a continuum step model. In the classic and the simplest form of the model [8] the surface of a crystal is divided into terraces by steps. Atoms at the surface are classified into adsorbed atoms(adatoms) and solid atoms, which form a stepped surface. On the terraces the density c of the adsorbed atoms obeys the diffusion equation:

$$\frac{\partial c}{\partial t} = D_s \nabla^2 c - \boldsymbol{v} \cdot \nabla c - \frac{c}{\tau} + f. \quad (14)$$

The first term on the right hand side is the surface diffusion with the diffusion coefficient D_s. The last term is the impingement rate from the ambient vapor phase, and the third term is the evaporation of adatoms with the lifetime τ. The second term is optional and represents drift of adatoms, which may exist under the influence of some external field, such as an electric field during direct current heating of a crystal [9]. Unless the adatom density is very high so that the step motion is fast, we may use the quasi-static approximation for the diffusion equation (14) by putting $\partial c/\partial t = 0$ [5,10],

$$\nabla^2 c - \frac{\boldsymbol{v}}{D_s} \cdot \nabla c - \frac{c - c_\infty}{x_s^2} = 0, \quad (15)$$

where $x_s = \sqrt{D_s \tau}$ is the surface diffusion length and $c_\infty = f\tau$ is the adatom density far from the steps.

Steps which bound the terraces are moving boundaries for the diffusion equation. The adatom current is the diffusion current given by

the gradient of the adatom density, supplemented with the drift current if it is present,

$$\boldsymbol{j} = -D_s \nabla c + c\boldsymbol{v}. \quad (16)$$

In general the adatoms flowing into a step either solidify or cross over the step onto the neighboring terrace. Thus the boundary conditions at the step for the front and the back side are given by [11]

$$-\hat{n} \cdot \boldsymbol{j}|_+ = K_+(c|_+ - c_s) + P(c|_+ - c|_-), \quad (17)$$

$$\hat{n} \cdot \boldsymbol{j}|_- = K_-(c|_- - c_s) + P(c|_- - c|_+), \quad (18)$$

where $\hat{n}$ is the normal vector of the step pointing to the downhill direction, the subscript $+$ $(-)$ indicates the front (the back) side of the step, $K_{+(-)}$ the kinetic coefficient and P the permeability (or transparency) coefficient of the step. In general the two kinetic coefficients K_+ and K_- are not equal because of a diffusion barrier. This asymmetry in the step kinetics is called the Ehrlich-Schwoebel effect [12,13]. The first term on the right represents solidification rate, which is proportional to the difference of the adatom density at either side of the step and that at equilibrium

$$
\begin{aligned}
c_s &= c_{eq}^0 \exp\left(-\frac{\Omega F_s}{k_B T}\right) \\
&= c_{eq}^0 \left(1 - \frac{\Omega F_s}{k_B T}\right),
\end{aligned} \quad (19)
$$

where c_{eq}^0 is the equilibrium adatom density for a straight step and F_s the force acting on the step. The force may be due to the step stiffness $F_s = -\tilde{\beta}\kappa$ ($\tilde{\beta}$: the step stiffness; κ: the step curvature) or due to the repulsive force from other steps. The second term in the right hand side of (17) or (18) expresses that the permeating current through the step is proportional to the difference of the adatom densities in the upper and the lower side of the step.

The flow of atoms may be most easily understood with the analogy to an electric circuit [14]. If there is no external field, the driving force of the current is the adatom density c. The density is related to the chemical potential as $\mu = k_B T \ln c$, which plays the role of an electric potential. The equivalent electric circuit for an isolated step is shown in Fig. 3(a). The difference between the

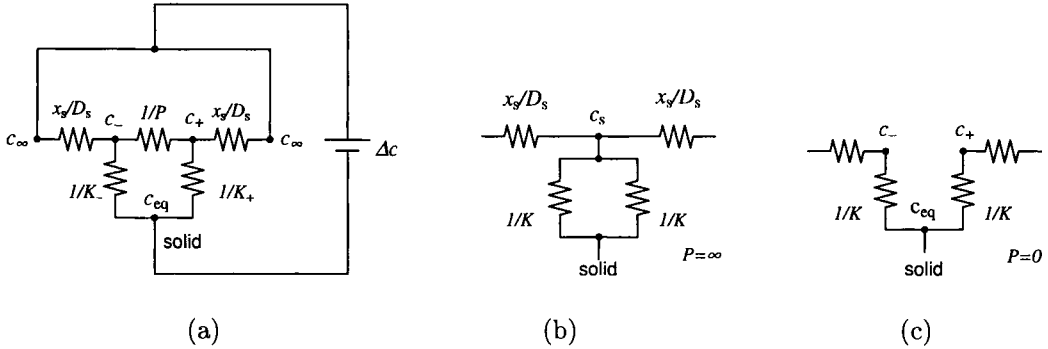(a)   (b)   (c)

Figure 3. Electric circuit equivalent to an isolated step without external field: (a) the general boundary condition, (b) the simplest model, (c) the impermeable step with asymmetry in step kinetics (from Ref. [5]).

adatom density far from the step c_∞ and that at equilibrium c_{eq} is the applied voltage to the circuit. The resistances x_s/D_s for the diffusion on the terrace and $1/K_\pm$ for the step kinetics are in series, and the resistances that in the upper terrace and in the lower terrace are in parallel. There is a connection permeating the step corresponding to the resistance $1/P$.

The original step model [8] does not distinguish the upper and lower side of the step as shown in Fig. 3(b), where $c_+ = c_- = c_s$. Another model frequently used takes account of the asymmetry of step kinetics as shown in Fig. 3(c), which excludes the over-step diffusion. The inclusion of the permeating path makes the diffusion field of neighboring terraces correlated and the mathematical treatment becomes very cumbersome so that we use the boundary conditions (b) or (c) in our analysis, i.e. our steps are perfectly permeable or impermeable.

With the boundary conditions (17) and (18) the diffusion equation (15) is solved, and from the solution, the local step velocity is determined by

$$V = \Omega K_-(c|_- - c_s) + \Omega K_+(c|_+ - c_s) \quad (20)$$

$$= \Omega \hat{n} \cdot (j|_- - j|_+), \quad (21)$$

where Ω is the atomic area. Then the new step position is calculated and we can trace the motion of the step.

For the continuum step model, it is computationally necessary and useful to make simplification of the above model. For multistep systems, instead of solving the diffusion equation and (21), one assumes a reasonable step velocity as a function of the distances to the neighboring steps [15,16]

$$\frac{\partial y_n(x)}{\partial t} = V_+(y_{n+1}(x) - y_n(x))$$

$$+V_-(y_n(x) - y_{n-1}(x)) + \gamma\frac{\partial^2 y_n(x)}{\partial x^2}, \quad (22)$$

where $y_n(x)$ is the position of the nth step, γ a constant proportional to the step stiffness. More drastic one is to neglect the displacement along the steps and use one-dimensional models. The one dimensional models are often used to study bunching instabilities when steps are not unstable for the fluctuation along the steps.

2. STEP DYNAMICS IN A VICINAL FACE

We first study morphological instabilities originated from the motion of steps in a vicinal face. A vicinal face consists of parallel equidistant straight steps (Fig. 4(a)). These equidistant steps are stabilized by a repulsive force between steps as well as the lattice periodicity in the atomic layer (the latter is irrelevant at temperatures above the roughening transition of this
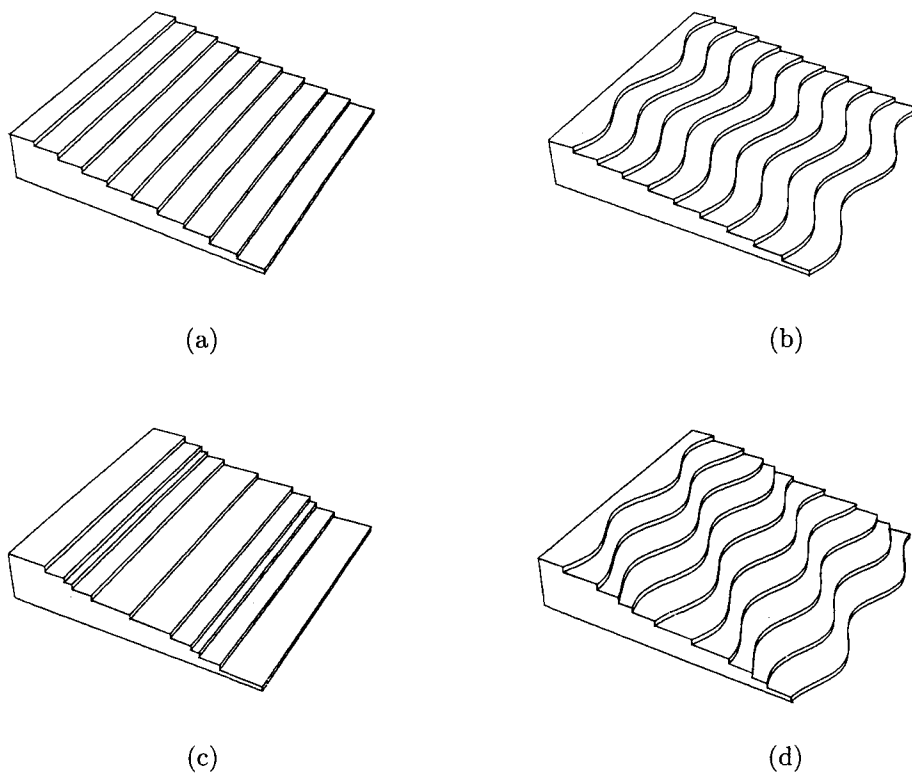
Figure 4. Vicinal face with sinusoidal modulation: (a) ideal vicinal, (b) in-phase wandering, (c) bunching and (d) bending.

vicinal face). The position of the nth step is expressed as $y_n(x) = nl +$ const., where l is the distance between steps. At a finite temperature steps can move freely in the y-direction and the crystal may grow or melt. In order to obtain a good quality crystal with a well-defined orientation of the surface, crystal growers want to achieve uniform motion of steps:

$$y_n(x,t) = nl + Vt. \tag{23}$$

In reality, however, steps fluctuate because of the thermal motion, which actually enables the steps to move by creating kinks. The surface now becomes rough due to the capillary motion of steps and, to make matters worse, instabilities of these steps sometimes occur. The morphological instabilities of steps are caused by a surface diffusion field of adatoms and lead to a chaotic or regular undulated state of the surface at a mesoscopic scale. We study the conditions under which these instabilities occur and the morphology of the surface in the instabilities.

In a linear regime modulation of a flat surface is a small amplitude sinusoidal wave as shown in Fig. 4. We may classify these instabilities according to the relative orientation between the steps and the wave vector $\boldsymbol{k}$:

wandering: If $\boldsymbol{k}$ and the steps are parallel, each step encounters sinusoidal modulation, which we call wandering (or meandering) of a step, and the modulation is all in-phase. The resultant surface morphology is ridges and troughs along the y-direction.

bending: If $\boldsymbol{k}$ and the steps are neither parallel nor perpendicular, the instability may be called bending of steps [17]. The form of each step is the same as that of the wandering, but the relative phase of the modulation is shifted.

bunching: If $\boldsymbol{k}$ and the steps are perpendicular, the instability is bunching of steps. The bunching of steps produces a stepped surface made up of giant steps.

In the following we study wandering and bunching of steps with the use of the models introduced in **1**.

2.1. Wandering of a single step

Wandering instability of steps is understood easily for a single isolated step. If the motion of the single step is controlled by diffusion of adatoms and the diffusion current of one of the terraces is dominant, a terrace bounded by the step is regarded as an edge of a two-dimensional crystal and a Mullins-Sekerka type instability [18] may occur as found by Bales and Zangwill [19]. For simplicity let us assume that the step absorbs atoms only from the lower terrace($K_- = 0$, one-sided model) and is impermeable($P = 0$). Then the problem is similar to the stability of a growing flat surface in a solution. Since a protruding part can grow more easily by absorbing atoms from the solution, the flat surface becomes unstable. The analogue is that the straight step starts to wander [19]. Using the continuum step model, one can calculate the linear growth rate ω_q of a fluctuation $\delta y_s(x,t) = \delta y_q e^{\omega_q t + iqx}$ superposed on the straight growing step $y_s(x,t) = y_0 + V_0 t$. The growth rare is [5]

$$\omega_q = -q^2 \Lambda_q \Omega D_s \Gamma + V_0 \left(\Lambda_q - \frac{1}{x_s} \right), \tag{24}$$

where

$$\Gamma = \frac{\Omega c_{eq}^0 \tilde{\beta}}{k_B T}, \tag{25}$$

$$\Lambda_q = \sqrt{q^2 + \frac{1}{x_s^2}}. \tag{26}$$

In (24) we have assumed that the kinetics is very fast $(K_+ \to \infty)$. The first term of (24) represents the stabilizing effect of the step stiffness and the second term represents the destabilizing effect of the surface diffusion. In the limit of long wavelength both terms are proportional to q^2, and its coefficient may be regarded as an effective step stiffness. When the impingement rate increases, the step velocity V_0 increases and the effective stiffness decreases. The instability occurs $(\omega_q > 0)$ if V_0 exceeds a critical value [19]

$$V_c = \Omega x_s(f_c - f_{eq}) = \frac{2\Omega D_s \Gamma}{x_s^2}. \tag{27}$$

Near the instability, because of the decrease of the effective stiffness, an anomalous increase of
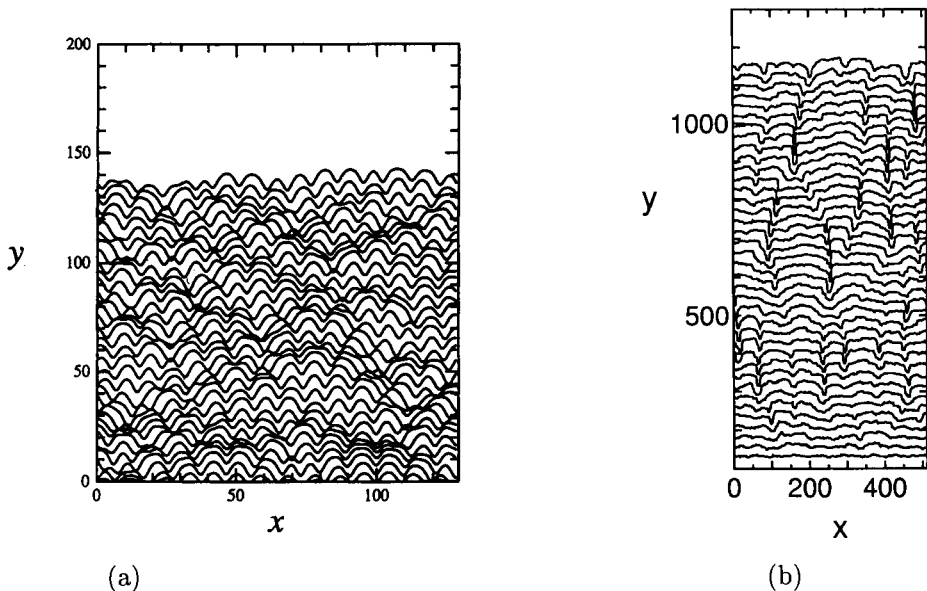
Figure 5. Time evolution of an isolated step which shows wandering instability. (a) a chaotic solution of the Kuramoto-Sivashinsky equation, (b) Monte Carlo simulation of chaotic step wandering.

the step fluctuation is expected [4]. On the other hand, suppression of the step fluctuation occurs in sublimation. Above the critical point the growth rate of fluctuation ω_q is positive for the wave number smaller than the critical value

$$q_c \approx \frac{1}{x_s} \sqrt{\frac{4(f - f_c)}{3(f_c - f_{eq})}}, \qquad (28)$$

which is obtained by expansion of (24) in q. The fastest growing mode $q_{max} = q_c/\sqrt{2}$ gives the characteristic wavelength of the instability. As the impingement rate f increases, the wavelength becomes shorter.

In the linear analysis the fluctuation grows exponentially as ω_q becomes positive, but in reality nonlinear effects become important. By a reductive perturbation method it has been shown that the step pattern near the instability is qualitatively described by a simple nonlinear equation (called Kuramoto-Sivashinsky equation) [20]:

$$\frac{\partial y_s}{\partial t} = -\frac{\partial^2 y_s}{\partial x^2} - \frac{\partial^4 y_s}{\partial x^4} + \frac{1}{2}\left(\frac{\partial y_s}{\partial x}\right)^2, \qquad (29)$$

where the length scale and time scales have been properly chosen (in this unit $q_c = 1$). This

equation produces a chaotic pattern of a step, which is shown as sequential stroboscopic patterns in Fig. 5(a). The initial pattern is a straight line with small amplitude random fluctuations at $y \approx 0$, and its time evolution is determined by integrating (29). A Monte Carlo simulation has been performed [5] for the growing step. The step pattern has been shown in Fig. 2, which clearly shows chaotic behavior (Fig. 2(d)). In Fig. 5(b) we show a simulation result with an SOS step in a large system ($L \times H = 512 \times 128$, $\tau = 256$, $f = 8 \times 10^{-4}$, an SOS step).

The chaotic motion may be suppressed by crystal anisotropy. The anisotropy results in orientation dependence in the step stiffness $\tilde{\beta}$ and in the kinetic coefficient K. Both anisotropies affect the step pattern, but in somewhat different way [21,22].

Wandering of a step is also expected in other systems if the diffusion field in front of the step is dominant for the step motion. For example, chaotic wandering is observed when a step leaves a step bunch into a large terrace [23]. In this case the upper terrace at the back of the step

is very narrow and adatoms are mainly supplied from the lower terrace although the Schwoebel effect may be absent. Another example of the asymmetry is due to drift of adatoms [24]. The drift can be induced by some external field such as an electric field in direct current heating of a Si specimen [25].

2.2. Wandering in a vicinal face

Wandering isntability also occurs in a vicinal face. The first effect of the neighboring steps is that the cutoff of the diffusion field is now the distance l between the steps instead of the diffusion length x_s, so that the destabilizing diffusion effect is weakened if the step distance is smaller than x_s [5,19]. The second effect is that the motion of neighboring steps is now correlated. Even without direct mechanical interaction between steps, the correlation appears through the diffusion field. Since the wandering fluctuations of neighboring steps work constructively, the inphase mode is the most unstable mode of step wandering [26].

If evaporation is negligible, i.e. $x_s \rightarrow \infty$, new features appear. For an isolated step ($l \rightarrow \infty$) the stabilizing term of (24) is proportional to q^3, from a similar reason as we will see in **3** ((47) corresponds to $\omega \sim -q^3$), and the second term $V_0 q$ always wins at small q: the steps are always unstable in growth. The instability occurs at a long wavelength. For a step in a vicinal face (small l), the wavelength of the instability is much longer than the terrace width $ql \ll 1$, the transport of adatoms along the steps is crucial for determining the step pattern. In this case, from a similar reason as we will see in (48) of **3** (diffusional transport along the interface), the stabilizing term is proportional to q^4 and the destabilizing term is proportional to q^2. A simple nonlinear equation for the step motion is derived with the use of a multiscale expansion by Pierre-Louis et al. [27,28]. Let us evaluate the adatom current j_s along the steps. For the initial equidistant straight steps, the adatom current is from the lower terrace to the step, and the component parallel to the step vanishes. For the one-sided model, all atoms fl coming from the vapor onto the terrace go to the upper step and
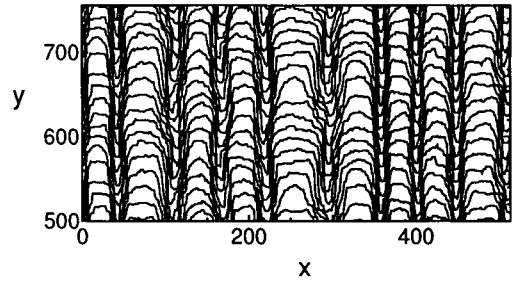

Figure 6. Vicinal face in growth under weak evaporation: $x_s = 64$, $l = 16$, $f = 6 \times 10^{-4}$.

$j = -\hat{n}f\zeta$ [29], where ζ is the length measured from the step edge in the direction of $\hat{n}$. When the step is deformed x component of the current appears from two origins. The first one is flux due to the tilt of the step,

$$
\begin{aligned}
j_{sx}^{(1)} &= \int_0^l j_x \, dy \\
&= \int_0^{l\cos\theta} j(-\sin\theta) \frac{dy}{d\zeta} d\zeta \\
&= f\tan\theta \frac{(l\cos\theta)^2}{2},
\end{aligned}
\tag{30}
$$

where $\tan\theta = \partial y_s/\partial x$ and $l\cos\theta$ is the step distance at the tilted part (note that $\zeta = (y - y_n)\cos\theta$). The second part is relaxation due to the chemical potential gradient: the current on the terrace along the step is

$$
j_s^{(2)} = l\cos\theta D_s \frac{\partial}{\partial s}\left(\frac{\mu}{k_B T}\right),
\tag{31}
$$

where s is the arc length of the step ($\partial x/\partial s = \cos\theta$) and the chemical potential is given by $\Omega\bar{\beta}\kappa$ (this is nothing but the force appeared in (19) multiplied by the atomic area Ω). Thus the advance rate $\partial y_s/\partial t$ measured from the uniform growth is given by the divergence of the current

$$
\begin{aligned}
\frac{\partial y_s}{\partial t} &= \frac{\partial}{\partial x}(j_{sx}^{(1)} + j_{sx}^{(2)}) \\
&= -\partial_x\left[\frac{\partial_x y_s}{1 + (\partial_x y_s)^2}\right. \\
&\left. + \frac{1}{1 + (\partial_x y_s)^2}\partial_x\left(\frac{\partial_{xx} y_s}{(1 + (\partial_x y_s)^2)^{3/2}}\right)\right],
\end{aligned}
\tag{32}
$$

where time and spatial coordinates are properly scaled and the subscript x denotes the spatial derivative in the second and the third lines. It has been shown that a zigzag step pattern will appear as a result of the instability and the amplitude w will increases as $w \propto t^{1/2}$ with a constant wavelength [27].

A Monte Carlo simulation similar to Fig. 2 and Fig. 5(b) for multiple steps with $l = 16$ and a long lifetime $\tau = 4096$ ($x_s = 64$) shows this regular in-phase step wandering (Fig. 6). The simulation was done in a system of the size $L \times H = 512 \times 128$, and a double system $L \times 2H$ is shown in the figure. Although the linearized equations of (29) and (32) are the same, nonlinear behaviors are thus completely different.

2.3. Bunching of steps

We first consider bunching instability of a vicinal face due to the Schwoebel effect. This is the simplest case though not an experimentally well studied case. The instability in the original BCF model has been studied [30] long time ago. The model shows an interesting symmetry in pairing behavior, which may not be realistic since the model neither incorporates the asymmetric kinetics nor the step interaction. Bennema and Gilmer studied the conditions for the stability of a step in a general form [31]. The step interaction is crucial for the development of the instability since noninteracting steps cannot avoid unrealistic collision of steps. A long-range elastic interaction between steps has been experimentally detected [32,33], and its role is very important in determining the morphology of vicinal faces. Here we first study a train of straight interacting steps with the asymmetry in kinetics taken into account.

To understand the role of the asymmetric kinetics, we first consider the one-sided model with infinitely fast step kinetics ($K_+ \to \infty$, $K_- = 0$) [34]. We suppose the step interaction is elastic and its potential can be written as [35] $U = g(y_n - y_m)^{-2}$ for steps located at y_n and y_m. Usually it is sufficient to take only the nearest neighbor interaction. If we choose proper units of time and distance as

$$\tilde{y} = \frac{y}{x_s}, \qquad \tilde{t} = \frac{t}{t_s}, \qquad t_s = \frac{\tau x_s^3 k_B T}{2A\Omega^2 c_{eq}^0},$$

the change of the position of nth step ($y_{n+1} > y_n$) is determined by

$$\frac{d\tilde{y}_n}{d\tilde{t}} = \left(\Delta \tilde{F} + \sum_{m=n\pm1} \frac{1}{(\tilde{y}_n - \tilde{y}_m)^3} \right) \times \tanh(\tilde{y}_{n+1} - \tilde{y}_n), \tag{33}$$

where the dimensionless driving force is defined as $\Delta \tilde{F} = \Omega(c_\infty - c_{eq}^0)t_s/\tau$. The first factor of (33) is the driving force consisting of the supersaturation term and the elastic force term. The other factor is the width of the lower terrace and is roughly proportional to the step distance if $y_{n+1} - y_n \ll x_s$.

If there are only two steps, from (33), their distance l changes as

$$\frac{d\tilde{l}}{d\tilde{t}} = \left(\Delta \tilde{F} + \frac{1}{\tilde{l}^3} \right) - \left(\Delta \tilde{F} - \frac{1}{\tilde{l}^3} \right) \tanh \tilde{l}. \tag{34}$$

The first term is from the right step with an infinitely large lower terrace and the second term is from the left step with a lower terrace of width $\tilde{l}$. In growth $\Delta \tilde{F} > 0$, and the distance always increases, but in sublimation with large negative $\Delta \tilde{F}$ the first term wins and the distance has a stable minimum. Thus in sublimation the interference of the two diffusion field effectively produces an attractive force which may win the elastic repulsion. In a vicinal face when the undersaturation is large enough ($|\Delta \tilde{F}|$ is large), neighboring steps form a stable bound pair. With increasing undersaturation bound states consisting of two or three pairs are formed [36].

The one-sided model is, of course, unrealistic. With moderate values of K (we assume $K_+ > K_-$) the pairing instability turns into a long wavelength instability of step density [37]. The elastic repulsion determines the surface stiffness parallel to the steps $\tilde{\alpha}_\parallel \sim g(a/l)$ [38]. Similarly to the wandering instability, the most unstable wave number k_{max} increases as $|\Delta \tilde{F}|$ is increased above the critical value. At first the linearly most unstable mode grows. But as shown in the numerical simulation (Fig. 7) it finally produces a regular array of bunches instead of chaotic motion [37]. This is consistent with the usual observation of periodic step bunches. By a nonlinear analysis it is found [39] that the difference arises from the
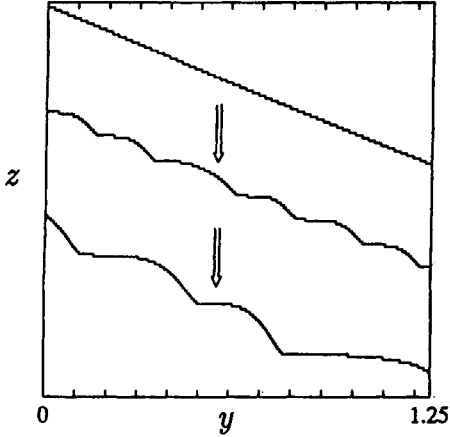
88

Figure 7. Evolution of the surface profile in the bunching instability. The step height is exaggerated.

symmetry of the system. Since the steps are moving to the left the dispersion of the linear mode has a term proportional to k^3 (a term proportional to k can be eliminated by a Galilean transformation) whereas the dispersion for a wandering step does not have odd power terms. Resulting nonlinear equation has a third derivative term and a form of the Benney equation [40]

$$\frac{\partial z}{\partial t} = -\frac{\partial^2 z}{\partial y^2} - \delta \frac{\partial^3 z}{\partial y^3} - \frac{\partial^4 z}{\partial y^4} + \frac{1}{2}\left(\frac{\partial z}{\partial y}\right)^2, \quad (35)$$

where z is the height of the surface from the initial uniform vicinal face and all variables are properly scaled. If the coefficient δ of the third derivative term is small, the equation is close to the Kuramoto-Sivashinsky equation and the solution is chaotic. On the other hand, if δ is large, as is the case near the critical point, the chaotic behavior is suppressed and one may obtain a periodic structure. Therefore origin of the difference between the chaotic behavior of a wandering step and the regular bunching is the difference of symmetry of the system.

Bunching of impermeable steps due to drift of adatoms

The physical origin of bunching is that the step

velocity depends on the size of the terrace of one side and its negative feedback to the fluctuation of the step distances [31]. A similar situation is realized by other causes than the asymmetry in step kinetics [41–45]. An important example is the bunching observed in direct current heating of Si(111) specimen [46–51]. Drift of adatomes induced by the external field may cause asymmetry of terraces in front and in the back of the step. This inevitably occurs if steps are impermeable.

To study the bunching of impermeable steps, we use the same model as that of Fig. 3(c) in one-dimension. The diffusion equation is given by (15) and the boundary conditions are (17) and (18) with $P =$. Here we summarize the result of the linear stability analysis [51]. For a small long wavelength perturbation $\delta y_m = \delta y_k e^{imkl+\omega_k t}$, the amplification rate ω_k is given by if $l \ll D_s/K \sim 1$

$$\frac{\tau \operatorname{Re} \omega_k}{\Omega c_{eq}^0} \approx \left(\frac{v\tau K}{2D_s} - \frac{\Omega}{k_B T}\frac{6g}{l^3}\right)(kl)^2$$
$$- \frac{\tau K\Omega}{k_B T}\frac{3g}{l^4}(kl)^4 \quad (36)$$

The coefficient of k^4 is determined by the repulsive interaction potential and negative. The vicinal face is stable for the short wavelength fluctuation. When the drift is in the step-down direction and its velocity exceeds the critical value,

$$v_c^B = \frac{12\Omega D_s g}{k_B T \tau K l^3}, \quad (37)$$

the coefficient of k^2 is positive and the vicinal face becomes unstable for the long wavelength fluctuation. If evaporation is negligible, $\tau \to \infty$, a vicinal face is unstable with any drift strength. The bunching starts as pairing of neighboring steps [52], and proceed as succesive pairing of step bunches [53].

In the above analysis we supposed that the steps are straight. However, the step wandering is also induced by the step-down drift [24]. Figure 8 shows snap shots of a step train with a small step distance in Monte Carlo simulation [54]. In the restricted Monte Carlo lattice model of 1.1 the kinetic coefficients can be assigned only approximately [54]. In Fig. 8 the system size is 128×256 and the number of steps is 32. Initially
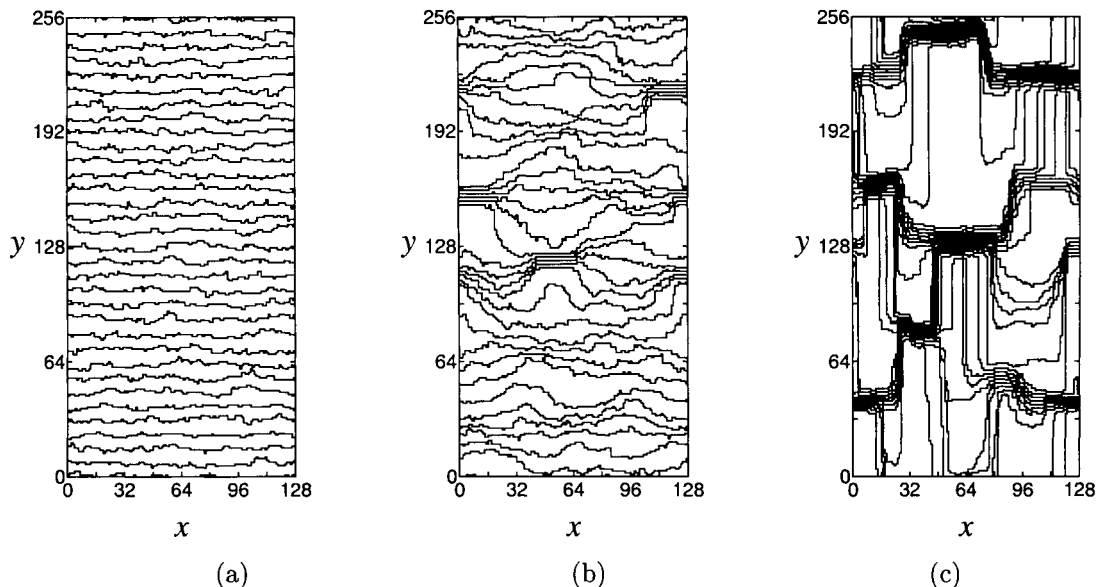
Figure 8. Train of impermeable steps with drift of adatoms: (a) stable with step-up drift, (b) bunching in the initial stage, and (c) in the late stage with step-down drift (from Ref. [54]).

the steps are equidistant with $l = 8$. The parameters are $x_s = 16$, $\tilde{\beta}/k_BT = 1.35$, $c_{eq}^0 = 0.10$ and $g/k_BT = 10$. Fig. 8(a) represents a snap shot of the step train with step-up drift ($v = -0.2$). As expected from the linear analysis, neither the wandering nor the bunching occurs. When the drift is in the step-down direction ($v = 0.2$), both bunching and wandering occur simultaneously. In the initial stage (Fig. 8(b)) the step wandering accompanied by bunching with a short length occurs. The bunches grow and connect with each other (Fig. 8(c)). This pattern may be compared to that of a two-dimensional continuum model [55].

Bunching of permeable steps due to drift of adatoms

The permeability of steps affect the behavior of steps [56] and may change the bunching instability. Recently Stoyanov [57] argued that a vicinal face consisting of perfectly permeable steps is unstable with the step-up drift in sublimation and showed that a large bunch is stable in the step-up drift. If the drift is in the step-up direction, the adatom density near the upper edge of the

bunch (we are considering a large bunch sloping down to the y-direction) is higher because atoms melted from the steps drift to the left. The high density region of adatoms decelerates the receding of the steps there and thus the slope of the bunch tends to become steeper. With step-down drift, to the contrary, steps in the lower edge are decelerated, and the slope of the bunch decreases so that the bunch will dissociate. If the crystal is growing with a high incident flux, the drift direction of adatoms for bunching is expected to be reversed. This is actually observed in experiment at 1160-1240°C in Si(111) [58].

For permeable steps (Fig. 3(b)), the adatom densities of neighboring terraces are coupled by the boundary condition (17). In the step flow model, we must solve simultaneous equations to determine the adatom densities and it is difficult to calculate the bunching of many steps. Therefore we use a continuum model [59], in which the drift of adatoms is readily taken into account [57]. Here we analyze the linear stability of a vicinal face for a long wavelength fluctuation and give an analytical expression for the condition of the instability.
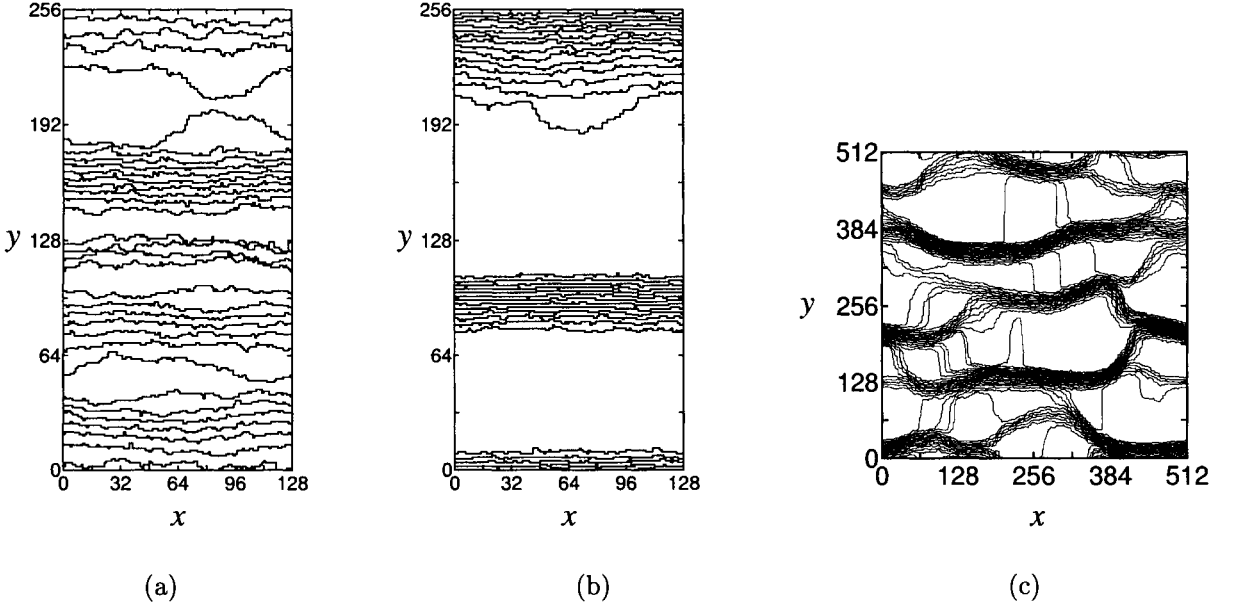
Figure 9. Train of permeable steps with drift of adatoms: (a) initial stage of bunching with step-up drift, (b) late stage, and (c) in the large system (from Ref. [54]).

We neglect the fluctuation along steps and assume that the steps are straight. When the step distance is small compared with the characteristic length of modulation, we can describe the surface profile with the density of steps $\rho(y)(=1/l)$. Time evolution equations of the adatom density and the step density are given by [57]

$$\frac{\partial c}{\partial t} = D_s \frac{\partial^2 c}{\partial y^2} - v\frac{\partial c}{\partial y} + F - \frac{1}{\tau}c$$
$$-2\rho K\left(c - c_{eq}(y)\right), \qquad (38)$$

$$\frac{\partial \rho}{\partial t} = -\frac{\partial}{\partial y}\left[2\rho\Omega K\left(c - c_{eq}(y)\right)\right], \qquad (39)$$

where $c_{eq}(y)$ is the local equilibrium density of adatoms. Equation (38) is the diffusion equation including the effect of solidification of adatoms at the steps. The decrease of the adatom density due to solidification is proportional to the local step density. Equation (39) is the continuity equation of the step density. The step current is ρV with the step velocity given by (20)

$$V = 2\Omega K\left(c - c_{eq}(y)\right). \qquad (40)$$

Since we consider modulation only in the y-

direction, the curvature of the steps vanishes and the equilibrium adatom density is determined by the step interaction. The interaction force is derived from the step energy ξ as $f = -\partial\xi/\partial y$, and therefore [60,34]

$$c_{eq} = c_{eq}^0 + \frac{\Omega c_{eq}^0}{k_B T}\frac{d\xi}{d\rho}\frac{\partial\rho}{\partial y}$$
$$= c_{eq}^0 + \gamma(\rho)\frac{\partial\rho}{\partial y}. \qquad (41)$$

The density dependent part of the step energy $\xi(\rho)$ is determined by the step interaction and

$$\frac{\partial\xi}{\partial\rho} = l^3\frac{d^2U}{dl^2} = 6g\rho, \qquad (42)$$

which is the surface stiffness in y-direction $\tilde{\alpha}_{\parallel}$ divided by a^2.

The linear stability of the vicinal face is determined by giving small perturbation to the step density, $\rho = \rho_0 + \delta\rho\,e^{iky+\omega_k t}$ and to the adatom density, $c(y) = c_0 + \delta c\,e^{iky+\omega_k t}$. Equations (38)-(41) determine the amplification rate ω_k for small k as

$$\omega_k = -ik\frac{V_0}{2\rho_0 K\tau} - k^2\frac{2\Omega\rho_0 K\gamma(\rho_0) - vV_0\tau}{2\rho_0\tau K} + \cdots,$$

where we have assumed $1 \ll \rho_0 K \tau$. The real part of ω_k represents the amplification of the fluctuation. The stability for the long wavelength fluctuation is determined by the coefficient of k^2. The first term in the numerator is the effect of the step repulsion and always stabilizes the vicinal face. The second term in the numerator can destabilize the vicinal face. The amplification rate becomes positive when $V_0 v$ exceeds the critical value,

$$(V_0 v)_c = \frac{2\Omega \rho_0 K \gamma(\rho_0)}{\tau} = \frac{12\Omega^2 c_{eq}^0 K \rho_0^2}{k_B T \tau} g, \qquad (44)$$

and the vicinal face is unstable. In sublimation, $V_0 < 0$, the instability occurs with the step-up drift. The drift direction to induce the bunching instability is opposite to that for the impermeable steps. Also it is reversed in growth (when $V_0 > 0$).

Figure 9 shows snap shots of a step train in Monte Carlo simulation without impingement of atoms. Since there is no impingement of adatoms, $V_0 \approx -\Omega c_{eq}^0/\rho_0 \tau$ and the receding steps become unstable when the step-up drift exceeds the critical velocity v_c^B,

$$v < v_c^B = -\frac{12\Omega K g}{k_B T l^3}. \qquad (45)$$

The parameters in Fig. 9 are the same as those of Fig. 8 except $\tilde{\beta}/k_B T = 0.54$. Then the critical drift velocity v_c^B is calculated as $v_c^B = -0.38$. Figures (a) and (b) represent snap shots with step-up drift ($v = -0.6$), and the step bunching occurs. In the initial stage of the bunching (a), the long wavelength fluctuation of the step distance appears. In the late stage (b), after the collision of small bunches, large bunches appear. When we carry out the simulation in a larger system (c), the bunches wander and sometimes collide with neighboring bunches. The pattern is similar to that observed in the experiment [47].

3. RELAXATION OF MESOSCOPIC STRUCTURES

3.1. Relaxation of small crystals with rough surfaces

When temperature is high and the surface atoms can move, a crystal relaxes to an equilibrium shape at the given temperature (and

the given pressure). We are not so long-lived enough to wait until a macroscopic crystal reaches the equilibrium shape, except some extraordinary materials such as ^{4}He at very low temperature[61]. Herring[62] estimated relaxation time τ for fine powder crystals and found that it increases as a power of the particle size. The exponent depends on the relaxation mechanism:

evaporation and condensation: The transport of material does not limit the relaxation rate but the surface kinetics does. The capillary force determines the surface velocity as $dR/dt \sim \tilde{\alpha}/R$, where $\tilde{\alpha}$ is the surface stiffness. By integration we have

$$\tau \sim R^2. \qquad (46)$$

volume diffusion: The diffusion current flowing into a curved surface is $j \sim 1/R^2$ so that the growth velocity is $dR/dt \sim \tilde{\alpha}/R^2$. Then

$$\tau = R^3. \qquad (47)$$

surface diffusion: The growth rate is determined by the surface current $dR/dt \sim -\nabla \cdot j_s$, the current is proportional to the chemical potential gradient $j_s \sim -\nabla \mu$, and the chemical potential of surface atoms is proportional to $\tilde{\alpha}/R$. Thus $dR/dt \sim \nabla^2(\tilde{\alpha}/R) \sim \tilde{\alpha}/R^3$, and

$$\tau \sim R^4. \qquad (48)$$

Mullins studied geometrically simple cases and derived analytical expressions for the relaxation time[60]. These laws are valid for a crystal surface above the roughening transition, where growth proceeds normal to the surface. The relaxation below the roughening temperature, where growth is only possible via lateral motion of steps, requires a different approach.

3.2. Relaxation of a faceted crystal

The relaxation of a facet and vicinal faces below the roughening temperature is controlled by the motion of steps [63,59,64–69]. In the following we study two simple examples: relaxation to an equilibrium shape when a round crystal is quenched

below the roughening transition and decay of a mesoscopic island on a facet.

Relaxation to an equilibrium shape

Suppose that a round crystal which is in the equilibrium form at a high temperature is quenched down below the roughening temperature. The new equilibrium shape at moderate temperature must have small facets bound by rounded edges. At the initial high temperature whole crystal surface is rough: steps, islands and kinks are everywhere and the step is not a well-defined entity on the surface. After the cooling, equilibration at a short length proceeds first: small islands disappear and well-defined steps emerge. The initial shape near one of the symmetry axes may be approximated by a paraboloid. Our question is how these steps move and how the facet appear on the surface [38]. Assuming circular symmetry around the symmetry axis, the initial shape is given by

$$Z(r) = Z_0 - \frac{r^2}{B}, \tag{49}$$

and the steps are contour of the paraboloid, whose radius is given by

$$R(z) = \sqrt{B(Z_0 - z)}. \tag{50}$$

When quenched, the crystal changes its shape to a new equilibrium shape [59] while the total mass of the crystal is conserved if evaporation and condensation are absent.

Evaporation and condensation

If evaporation and condensation are present at the steps in a steady uniform environment, the velocity of a step, which is determined by an equation similar to (20) with (19), is given by the supersaturation and the step stiffness $\Delta c = c - c_{eq}^0$

$$\begin{aligned}
\frac{dR_i}{dt} &= \Omega K \left(\Delta c - \frac{c_{eq}^0 \tilde{\beta} \kappa_i}{k_B T} \right) \\
&= \Omega K \Delta c \left(1 - \frac{R_c}{R_i} \right). \tag{51}
\end{aligned}$$

where $R_c = (\tilde{\beta}/k_B T)(c_{eq}^0/\Delta c)$ is the radius of a critical two-dimensional nucleus and $i = z/a$ is the step number of the step at height z (here and

after, the step numbering is opposite to that in the previous sections: a larger number implies a step at a higher location). In (51) the interaction between steps has been neglected for simplicity's sake. In the uniform environment steps whose radius is smaller than the critical radius shrink and steps whose radius is larger expand. The shrinkage and the expansion occur simultaneously with exchange of material to the environment so that the shape change proceeds smoothly.

The change of step radii $R_i(t)$ for steps interacting each other with elastic repulsion is shown in Fig. 10(a). (The interaction modifies the equilibrium deisity in (51) as (57).) The facet appear through gradual increase of the radius of curvature at the top. In the present case the total volume of the crystal is not conserved.

Uniform surface supersaturation

On the other hand, if evaporation is negligible and the total crystal mass is conserved, the relaxation proceeds via surface diffusion. Let us consider first a hypothetical situation in which surface diffusion is so fast that adatom density over the whole surface is uniform. (In reality such fast diffusion is possible only in a very small system.) The velocity of steps is determined by the same equation (51) with the step interaction, but now the atomic density in the environment is changing with time. It is given by the equilibrium density averaged over the whole step length. The surface atomic density is relatively low when outer steps are expanding at the beginning and becomes higher after they slow down. In Fig. 10(c) the change of radii of steps is plotted. Because of the temporal change of the surface supersaturation one of the steps shrinks rashly at the beginning and then expands again. In some cases (not shown here) steps annihilate so early that the final step number is actually fewer than that of the true equilibrium.

Surface diffusion

In reality the surface diffusion cannot be so fast and slow mass transport on the surface controls the morphological change. The relaxation of shape is very different: the step on the top shrinks because of strong step tension and disap-
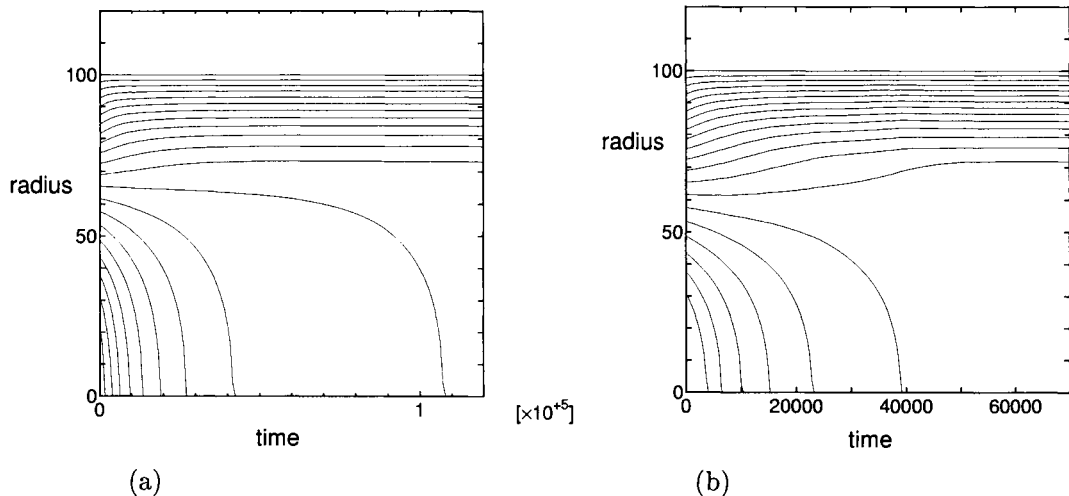
Figure 10. Relaxation to the equilibrium shape with a facet from a high temperature parabolic shape. (a) Relaxation limited by step kinetics in a uniform constant environment. (b) Conserved system with extremely fast surface diffusion.

pears while other steps do not move much. Only when the top step vanishes, does the second step start to shrink. The relaxation is a layer-by-layer annihilation of the top terraces. Macroscopically one first observes the appearance of a small facet on the top of a round crystal and then sees slow enlargement of the facet until the equilibrium size is reached. The growth of the facet radius R_f roughly obeys a power law

$$R_f \sim t^\nu, \tag{52}$$

where t is the annealing time. The growth exponent ν reflects the rate limiting mechanism, which is discussed in the decay problem in 3.3. Simulation result will be shown in 3.4.

3.3. Decay of a mesoscopic island on a facet

Another example of shape relaxation is decay of a three-dimensional island on a large facet (or on a substrate). In some cases one can prepare the initial shape by an artificial way, and the initial shape is not necessarily any equilibrium shape. For example, it may be a cone shape (see [70])

$$Z(r) = Z_0 - \frac{r}{A}, \tag{53}$$

and the step radius is given by

$$R(z) = A(Z_0 - z). \tag{54}$$

This shape relaxes to an equilibrium shape with a facet on the top of the island. However, since atoms can escape from the island onto the large facet (or to the substrate), the island disappears before it reaches the equilibrium shape.

Although this case looks different from that of 3.2, the initial relaxation processes, i.e. the appearance of a small facet and its expansion, are common. Here we study the process of expansion of the facet, whose behavior depends on the rate-limiting processes.

Israeli and Kandel [67,68] found a universal scaling relation in this relaxation process. As for the facet size they found that the exponent is $\nu = 1/4$ if the initial shape is a cone. Consider an island which consists of concentric step loops on a large facet as shown in Fig. 11. With the static approximation, the density of adatoms $c_i(r)$ on the ith terrace that is bounded by the ith and the $(i + 1)$th steps obeys the diffusion equation, which is (15) without the second and the third term,

$$\left(\frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r} \right) c_i(r) = 0. \tag{55}$$

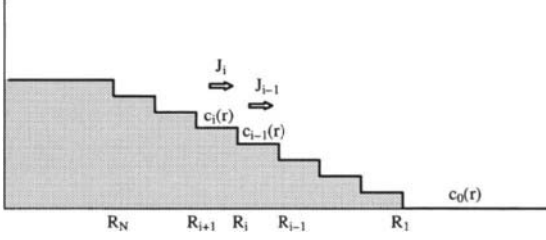The boundary conditions for (55) at the inner and the outer steps are the formulae for the adatom

Figure 11. Profile of an island consisting of circular steps (from Ref. [69]).

current

$$-j_i(R_{i+1}) = D_s \frac{\partial c_i}{\partial r}\bigg|_{R_{i+1}}$$
$$= K(c_i(R_{i+1}) - c_{\text{eq}}^{i+1}),$$
$$j_i(R_i) = -D_s \frac{\partial c_i}{\partial r}\bigg|_{R_i}$$
$$= K(c_i(R_i) - c_{\text{eq}}^i), \qquad (56)$$

where c_{eq}^i is the equilibrium adatom density at the ith step. We have assumed that the kinetic coefficient K of the step is the same for the upper and the lower terraces, and that the steps are impermeable: $P = 0$. The equilibrium adatom density for the ith step is given by (19)

$$c_{eq} = c_{eq}^0 \left[1 - \frac{\Omega}{k_B T}\left(-\tilde{\beta}\kappa + F_i - F_{i-1}\right)\right] \qquad (57)$$

with [71,67]

$$F_i = 4g \frac{R_{i+1}}{R_i + R_{i+1}} \frac{1}{(R_i - R_{i+1})^3}, \qquad (>0) \quad (58)$$

which is the strength of repulsive force acting on the steps at both ends of the ith terrace. If configuration of the steps, $\{R_i\}$, is given, the equilibrium density of each step (57) is determined, the boundary conditions (56) are fixed and the diffusion equation (55) can be solved.

From (56) the velocity of the ith step is [67]

$$\frac{dR_i}{dt} = \Omega\left(j_i(R_i) - j_{i-1}(R_i)\right)$$
$$= \Omega \frac{J_i - J_{i-1}}{2\pi R_i}, \qquad (59)$$

where

$$J_i = 2\pi \frac{c_{\text{eq}}^{i+1} - c_{\text{eq}}^i}{\frac{1}{D_s}\ln\frac{R_i}{R_{i+1}} + \frac{1}{K}\left(\frac{1}{R_i} + \frac{1}{R_{i+1}}\right)} \qquad (60)$$

is the total adatom flux: $J_i = 2\pi r j_i(r)$. If the step kinetics is fast, the first term in the denominator dominates and the process is limited by the slow surface diffusion between steps. If the opposite is the case, the process is limited by the slow step kinetics (attachment and detachment at the step). In the vicinal slope of the crystal, since $l_i \equiv R_i - R_{i+1} \ll R_i$, the first term of the denominator is $l/D_s R$ and the second term is $2/KR$, therefore the ratio $2D_s/Kl$ determines which is the case.

Kinetics-limited case

In the kinetics-limited case the adatom density c_i on the ith terrace is approximately constant over the terrace. Total adatom flux on the ith terrace is

$$J_i = 2\pi R_{i+1} K(c_{eq}^{i+1} - c_i)$$
$$= 2\pi R_i K(c_i - c_{eq}^i). \qquad (61)$$

For the first and the last steps we impose the following conditions. The top step receives no current from the top terrace

$$J_N = 0, \qquad (62)$$

which is common in the two problems. The boundary condition at the bottom is different. For the relaxation to equilibrium

$$J_0 = 0, \qquad (63)$$

and we assume that R_N is fixed. For the decay of an island, the adatom density on the substrate is at equilibrium and

$$J_0 = 2\pi R_1 K(c_{eq}^1 - c_{eq}^0) \quad \text{or} \quad c_N = c_{eq}^0. \qquad (64)$$

The diffusion effect may not be negligible for the motion of the top step and the bottom step since the top facet and the substrate are large. It is necessary to check if the idealized condition holds for the application to real systems. For example, in Si(111) the decay of a single layer island of a

nanometer size is most likely limited by the step kinetics [72]. The condition for the bottom step does not affect much to the initial behavior of the facet size, and the relaxation and the decay problems are similar.

Suppose that the envelope of the initial island shape is the cone given by (53). Small central steps shrink in a short time and a small facet appears on the top. Assuming that the slope does not change much, the radius of the top facet is given by

$$R_f(t) \approx R_0(z_f(t)), \tag{65}$$

where z_f is the height of the top facet. In the kinetics-limited case ($D_s/Kl \gg 1$ in (60)) the diffusion current is given by (61) and if we neglect the small difference of R in (61) the density on the ith terrace is expressed as

$$c_i = \frac{1}{2}(c_{eq}^i + c_{eq}^{i+1}). \tag{66}$$

From (61) and (66) in the continuum representation ($z = ia$ with the step height a) the surface current is

$$j(R) = \frac{J(R)}{2\pi R}$$
$$= \frac{1}{2}Ka\frac{\partial c_{eq}}{\partial z}. \tag{67}$$

If the step tension is much stronger than the step repulsion (this is certainly true in the initial stage), we can neglect $F_i - F_{i-1}$ in the equilibrium density (57). The current (67) becomes

$$j(R) = \frac{1}{2}K\Gamma a\frac{\partial}{\partial z}\left(\frac{1}{R}\right), \tag{68}$$

where Γ is given by (25). From the mass conservation the change of the facet height z_f is related to the current at the facet edge as

$$\pi R_f^2\frac{dz_f}{dt} = -2\pi R_f j(R_f)\Omega a, \tag{69}$$

or

$$\frac{dz_f}{dt} = -2\Omega a\frac{j(R_f)}{R_f}. \tag{70}$$

The change of the facet radius is given by

$$\frac{dR_f}{dt} = \frac{dz_f}{dt}\frac{dR}{dz}\Big|_{z=z_f}$$
$$= -\frac{dR}{dz}\Big|_{z_f} 2\Omega a\frac{j(R_f)}{R_f}. \tag{71}$$

With (68) we obtain

$$\frac{dR_f}{dt} = \left(\frac{dR}{dz}\Big|_{z_f}\right)^2 K\Gamma\Omega a^2\frac{1}{R_f^3} \tag{72}$$

Since the change of the profile is rather small during the enlargement of the facet, we estimate the change of the facet size using (65) as

$$R_f^4 \approx 4\Gamma K\Omega a^2 A^2 t. \tag{73}$$

Although the island shape near the top facet is modified by the atoms flowed out, the overall shape does not change much and the above estimate remains valid. The exponent $\nu = 1/4$ of this simple argument agrees with the more complete derivation of [67,68], in which the effect of interaction is also included.

The advantage of the present simple argument is that one can use (72) for other initial shapes. For a paraboloid, $R_0(z) = [B(z_0 - z)]^{1/2}$, the scaling is now

$$R_f^6 \approx \frac{3}{2}\Gamma K\Omega a^2 B^2 t. \tag{74}$$

Diffusion-limited case

In the opposite limit where step kinetics is fast, the rate limiting process is surface diffusion and the outgoing flux is obtained from (60) with $K \to \infty$ as

$$J_i(R_i) = 2\pi D_s\frac{c_{eq}^{i+1} - c_{eq}^i}{\ln\dfrac{R_i}{R_{i+1}}}$$
$$\approx 2\pi D_s\Gamma\frac{1}{R_i}. \tag{75}$$

In the second line the step interaction is neglected and the step distance is assumed to be small. Then the step velocity is

$$\frac{dR_i}{dt} = \Omega D_s\Gamma\frac{1}{R_i}\left(\frac{1}{R_i} - \frac{1}{R_{i-1}}\right). \tag{76}$$

The enlargement rate of the facet (71) can be used with (75) and we obtain

$$R_f^4 \approx 8\Gamma D_s\Omega aAt. \tag{77}$$

For the paraboloid a similar estimation gives

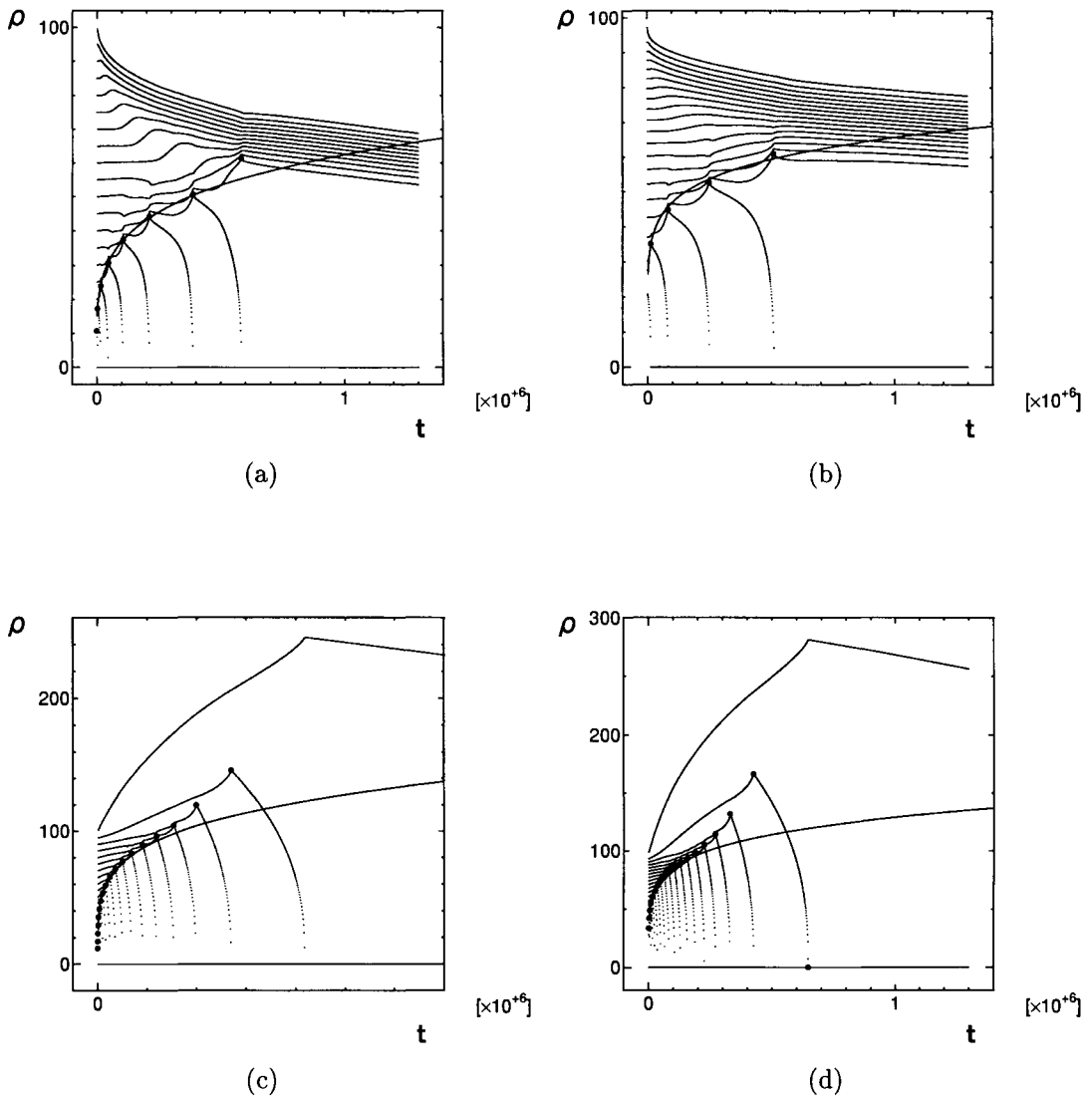$$R_f^5 \approx 5\Gamma D_s\Omega aBt. \tag{78}$$

Figure 12. Expansion of the top facet. The initial shape and the limiting process are (a) cone, step kinetics, (b) paraboloid, step kinetics, (c) cone, diffusion, (d) paraboloid, diffusion. The extra solid curve is the fitting for the facet size (from Ref. [69]).
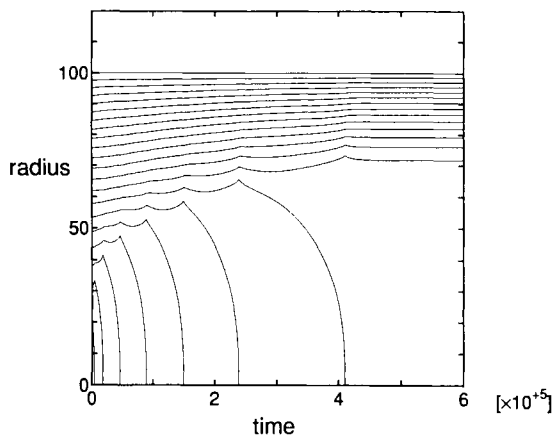
Figure 13. Relaxation to the equilibrium shape with a facet from a high temperature paraboloidal shape.

We note that for the cone the exponent $\nu = 1/4$ is independent of the rate limiting process, while it differs for the paraboloid, $\nu = 1/6$ and $\nu = 1/5$.

3.4. Simulation result and experiment

Finally we introduce some results of simulation and experiment related to the decay of three-dimensional islands and the facet relaxation. Our theoretical analysis does not take the step permeability into account. The step permeability is known to induce pairing of steps in the decay of islands [71], but we do not understand its mechanism and the role of the permeability yet.

Decay of a conical island on a facet

Several results of numerical integration of (59) for the island decay is shown in Fig. 12. The radius is scaled as $\rho = R_f(k_B T/\Omega\tilde\beta)$ and time is also scaled. A smaller value of g is used for the diffusion-limited case. In the kinetics-limited case a rough fitting for the cone is $R_f \sim t^{0.23}$ while (73) gives $R_f \sim t^{0.25}$. For the paraboloid $R_f \sim t^{0.14}$ while (74) gives $R_f \sim t^{0.17}$. In the diffusion-limited case a rough fitting for the cone is $R_f \sim t^{0.23}$ while (77) gives $R_f \sim t^{0.25}$. For the paraboloid $R_f \sim t^{0.18}$ while (78) gives $R_f \sim t^{0.20}$. The naive estimation of **3.3** agrees the simulation result.

In the kinetics-limited case the layer-by-layer

annihilation of terraces stops because bunching proceeds from the bottom steps as shown in (a) and (b) Then a mesa bound by the circular step bunch shrinks slowly [69].

Ichimiya et al. [70] observed decay of a small silicon island on a large Si(111) facet and found the exponent $\nu = 1/4$ in accordance with theory. The island was formed by an STM(scanning tunneling microscope) tip and made of only several thousand atoms. The observed initial shape is not a cone but a pyramid reflecting the crystal anisotropy. The observed shape of the pyramid has a small top facet from the beginning since the real starting shape cannot be observed. The height decrease from the extrapolated pyramid top was measured and found to change at $465°$C

$$Z_0 - z_f \approx Ct^\nu \tag{79}$$

with $C = 6.1 \pm 0.4$ Å/s$^{1/4}$ and $\nu = 1/3.9$. The exponent agrees well with the theoretical value $1/4$, and C is also in fair agreement with the calculated value $C = 3.1 \pm 0.3$ Å/s$^{1/4}$ from (73) and parameters values obtained in the decay experiment of a two-dimensional island [72]. It is rather surprising that the continuum step picture works at such a small scale.

Relaxation to the low temperature equilibrium form

Fig. 13 shows a result for the relaxation to a faceted equilibrium shape via surface diffusion. Here the total mass is conserved and the bottom step is fixed at $\rho_1 = 100$. The initial shape is a paraboloid and the final shape is the equilibrium shape with a facet, which is the same as the final shape of Fig. 10(b). In the calculation the parameter representing ratio of the kinetic coefficient to the diffusion coefficient is set as $(K/D_s)(\Omega\tilde\beta/k_B T) = 4$.

Thürmer et al. [73] observed the relaxation process of a Pb crystal on a Ru substrate. In the experiment the initial shape was an equilibrium shape with a small facet and the crystal was quenched to a lower temperature. The layer-by-layer annihilation of top terraces was clearly observed and the crystal relaxed to the new equilibrium shape at the low temperature. Qualitative features are common in the experiment and in the

simulation.

REFERENCES

1. G. H. Gilmer and P. Bennema, J. Appl. Phys. **43**, (1972) 1347.
2. Y. Arima and T. Irisawa, J. Cryst. Growth **104** (1990) 297.
3. T. Irisawa and Y. Arima, J. Cryst Growth **163** (1996) 22.
4. M. Uwaha and Y. Saito, Phys. Rev. Lett. **68** (1992) 224.
5. Y. Saito and M. Uwaha, Phys. Rev. B**49** (1994) 10677.
6. M. Uwaha, Y. Saito and M. Sato, J. Cryst. Growth **146** (1995) 164.
7. H. J. Leamy, G. H. Gilmer and K. A. Jackson, in *Surface Physics of Materials*, Vol.1, ed. J. M. Blakely, (Academic, New York, 1975) 121.
8. W. K. Burton, N. Cabrera and F. C. Frank, Phil. Trans. Royal Soc. London, **243** (1951) 299.
9. H. Yasunaga and A. Natori, Surf. Sci. Rep. **15** (1992) 205.
10. F. Liu and H. Metiu, Phys. Rev. E **42** (1994) 2601.
11. M. Ozdemir and A. Zangwill, Phys. Rev. B **45** (1992) 3718.
12. G. Ehrlich and F. G. Hudda, J. Chem. Phys. **44** (1966) 1039.
13. R. L. Schwoebel and E. J. Shipsey, J. Appl. Phys. **37** (1966) 3682.
14. H.-C. Jeong and E. D. Williams, Surf. Sci. Reports **34** (1999) 171.
15. D. Kandel and J. D. Weeks, Phys. Rev. B **52** (1995) 2154.
16. H.-C. Jeong and J. D. Weeks, Phys. Rev. B **57** (1998) 3939.
17. D.-J. Liu, J. Weeks and D. Kandel, Phys. Rev. Lett. **81** (1998) 2743.
18. W. W. Mullins and R. F. Sekerka, J. Appl. Phys. **35** (1964) 444.
19. G. S. Bales and A. Zangwill, Phys. Rev. B **41** (1990) 5500.
20. I. Bena, C. Misbah and A. Valance, Phys. Rev. B**47** (1993) 7408.
21. Y. Saito and M. Uwaha, J. Phys. Soc. Jpn. **65** (1996) 3576.
22. M. Sato and M. Uwaha, J. Crst. Growth **198/199** (1999) 38.
23. P. Finnie and Y. Homma, Phys. Rev. Lett. **82** (1999) 3500.
24. M. Sato and M. Uwaha, J. Phys. Soc. Jpn. **65** (1996) 2146.
25. M. Degawa, H. Nishimura, Y. Tanishiro, H. Minoda and K. Yagi, Jpn. J. Appl. Phys. **38** (1999) L308.
26. A. Pimpinelli, I. Elkinani, A. Karma, C. Misbah and J. Villian, J. Phys. Condens. Matter **6** (1994) 2661.
27. O. Pierre-Louis, C. Misbah, Y. Saito, J. Krug and P. Politi, Phys. Rev. Lett. **80** (1998) 4221.
28. F. Gillet, O. Pierre-Louis and C. Misbah, Eur. Phys. J. B **18** (2000) 519.
29. J. Villain, J. Phys. I **1** (1991) 19.
30. W. W. Mullins and J. P. Hirth, J. Phys. Chem. Solids **24** (1963) 1391.
31. P. Bennema and G. H. Gilmer, in *Crystal Growth: An Introduction*, ed. P. Hartman (North-Holland, Amsterdam, 1973) 263.
32. X.-S. Wang, J. L. Goldberg, N. C. Bartelt, T. L. Einstein and E. D. Williams, Phys. Rev. Lett. **65** (1990) 2430.
33. C. Alfonso, J. M. Bermond, J. C. Heyraud and J. M. Métois, Surf. Sci. **262** (1992) 371.
34. M. Uwaha, Phys. Rev. B **46** (1992) 4364.
35. V. I. Marchenko and A. Ya. Parshin, Zh. Eksp. Teor. Fiz. **79** (1980) 257 [Sov. Phys.-JETP, **52** (1980) 129].
36. M. Sato and M. Uwaha, J. Phys. Soc. Jpn. **66** (1997) 1054.
37. M. Sato and M. Uwaha, Phys. Rev. B **51** (1995) 11172.
38. P. Nozières, in *Solids far from equilibrium*, ed. C. Godrèche, (Cambridge, Cambridge, 1991) 1.
39. M. Sato and M. Uwaha, Europhys. Lett. **32** (1995) 639.
40. T. Kawahara, Phys. Rev. Lett. **51** (1983) 381.
41. N. Cabrera and D. A. Vermilyea, in *Growth and Perfection of Crystals*, ed. R. H. Doremus, B. W. Roberts and D. Turnbull (Wiley, New York, 1958) 393.
42. F. C. Frank, *ibid.*, 411.
43. A. A. Chernov, Usp. Fiz. Nauk **73** (1961) 277, [Sov. Phys.-USPEKHI **4** (1961) 116].

44. J. P. van der Eerden and H. Müller-Krumbhaar, Phys. Rev. Lett. **57** (1986) 2431.

45. D. Kandel and J. D. Weeks, Phys. Rev. B **49** (1994) 5554.

46. A. V. Latyshev, A. L. Aseev, A. B. Krasilnikov, S. I. Stenin, Surf. Sci. **213** (1989) 157.

47. Y. Homma, R. J. McClelland and H. Hibino, Jpn. J. Appl. Phys. **29** (1990) L2254.

48. S. Stoyanov, Jpn. J. Appl. Phys. **30** (1991) 1.

49. A. Natori, Jpn. J. Appl. Phys. **33** (1994) 3538.

50. C. Misbah and O. Pierre-Louis, Phys. Rev. E **53** (1996) R4318.

51. M. Sato and M. Uwaha, J. Phys. Soc. Jpn. **65** (1996) 1515.

52. D.-J. Liu and J. Weeks, Phys. Rev. B **57** (1998) 14891.

53. M. Sato and M. Uwaha, J. Phys. Soc. Jpn. **67** (1998) 3675.

54. M. Sato, M. Uwaha and Y. Saito, Phys. Rev. B **62** (2000) 8452.

55. M. Sato and M. Uwaha, Phys. Rev. E **60** (1999) 7120.

56. S. Stoyanov, Surf. Sci. **370** (1997) 345.

57. S. Stoyanov, Surf. Sci. **416** (1998) 200.

58. J.-J. Métois and S. Stoyanov, Surf. Sci. **440** (1999) 407.

59. M. Uwaha, J. Phys. Soc. Jpn. **57** (1988) 1681.

60. W. W. Mullins, in *Metal Surfaces*, ed. W. D. Robertson and N. A. Gjostein (Metall. Soc. AIME, Metals Park, 1963) 17.

61. S. Balibar and B. Castaing, Surf. Sci. Rep. **5** (1985) 87.

62. C. Herring, J. Appl. Phys. **21** (1950) 301.

63. J. Villain, Europhys. Lett. **2** (1986) 531.

64. A. Rettori and J. Villain, J. Phys. France **49** (1988) 257.

65. M. Uwaha and P. Nozières, in *Morphology and Growth Unit of Crystals*, ed. I. Sunagawa, (Terra Scientific, Tokyo, 1989) 17.

66. M. Ozdemir and A. Zangwill, Phys. Rev. B **42** (1990) 5013.

67. N. Israeli and D. Kandel, Phys. Rev. Lett. **80** (1998) 3300.

68. N. Israeli and D. Kandel, Phys. Rev. B **60** (1999) 5946.

69. M. Uwaha and K. Watanabe, J. Phys. Soc. Jpn. **69** (2000) 497.

70. A. Icimiya, K. Hayashi, E. D. Williams, T. L. Einstein, M. Uwaha and K. Watanabe, Phys. Rev. Lett. **84** (2000) 3662.

71. S. Tanaka, N. C. Bartelt, C. C. Umbach, R. M. Tromp and J. M. Blakely, Phys. Rev. Lett. **78** (1997) 3342.

72. A. Ichimiya, Y. Tanaka and K. Ishiyama, Phys. Rev. Lett. **76** (1996) 4721.

73. K. Thürmer, J. E. Reutt-Robey, E. D. Williams, M. Uwaha, A. Emundts and H. P. Bonzel, submitted.

Surface step dynamics: experimental observations

J.J. Métois, J.C. Heyraud and J.M. Bermond

CRMC2-CNRS , Campus Luminy, case 913, F-13288 Marseille Cedex 09, France

Under thermodynamical equilibrium conditions, the wandering of an isolated step and of step trains observed by reflection electron microscopy (REM), has been analyzed at 1170K on (111) vicinals. The step stiffness of an isolated step has been measured ($\beta_1 \sim 3 \times 10^{-11}$ J m^{-1}). From the analysis of terrace width distribution the existence of long-ranged 'energetic' step interactions decaying as $1/d^2$ can be established (d: normal distance to the step edge). The mechanism of adatom supply to a step can be determined from the time-correlation fonction of the step positions. It takes place by exchange of atoms at step edges with adatoms or vacancies on the terraces. A more macroscopic approach has been used also. The profiles of equilibrated silicon crystals have been studied in the vicinity of a (111) facet at 1170K. The equilibrium shapes of three-dimensional crystals have been produced by thermally equilibrating an array of small silicon columns on a silicon substrate. The samples have been observed *in situ* by transmission electron microscopy (TEM). The profiles have been measured along a (high symmetry) <110> zone, toward {110}. The theoretical prediction of a 3/2 power law for the profile equation has been checked involving repulsive $1/d^2$ step interactions (Prokowski-Talapov universality class behaviour). From the profile equation, the step interaction constant can be determined, it is in good agreement with previous experiments on the terrace width distributions.

1. INTRODUCTION

The thermodynamics of silicon surfaces has attracted a renewed interest during the past few years, due to the availability of modern observational techniques. The direct, *in situ*, imaging of steps opened a first line of approach.

Advantage has been taken of the atomic resolution of the scanning tunneling microscopy (STM) [1, 2] to determine the structure of the steps. However, on atomically resolved STM images, the field of view is limited to roughly 100 nm. To extract statistical data it is then necessary to study many selected areas on the specimen. On the other hand there are thermodynamically quantities that characterize either the steps themselves or their interactions, which can be defined by considering the steps on a more coarse-grained scale. Ignoring the detailed atomic step structure, the thermodynamics of a vicinal surface can be

formulated in terms of the step stiffness, of the step interactions and of the surface phase separation.

Hence, visualization techniques that do not achieve atomic resolution can be used to obtain thermodynamical information.

Such are, for instance, low energy electron microscopy (LEEM) [3] and reflection electron microscopy (REM) [4,5].

The *in situ* visualization of the motion of steps and of step trains over lengths of a few micrometers and possibly under conditions of thermodynamical equilibrium is their common feature.

One of the important applications of this emerging capability will be in understanding the kinetics of step motion on surfaces, which is important in growth and in the equilibration of surface structure. The simplest conditions under which step motion can be observed are equilibrium conditions in which the step position is fluctuating under thermal excitations.

Dynamical *in situ* REM experiments on the step meandering on Si (111) vicinal made it possible to measure the step stiffness, the step interactions and the surface mass transport [5].

However, a more macroscopic, thermodynamical, approach is possible: the equilibrium shape of a crystal may be used to determine the step-free energy and the step interactions via the measurement of the anisotropy of the surface free energy and of the equilibrium profile in the vicinity of a facet [6-9].

Unfortunately, the high chemical reactivity of the silicon raises some experimental difficulties. In particular, no suitable (inert) substrate has been found so far, unlike for metals. One promising way seems to be the use of micron-sized monocrystalline silicon protrusions on a silicon substrate. Here we find the competition between two mechanisms. On the one hand the sample surface tends to heal the protrusion out, to reach its *overall* stable shape. This process is driven by capillary effect, on a macroscopic scale. On the other hand, Nichols and Mullins [10] pointed out, that the driving force towards the equilibrium shape might establish the latter locally if the size and the geometry of the protrusion are favorable. The equilibrium shape of silicon, at 1323 K, is observed on small monocrystalline silicon columns, made by etching the surface of a silicon wafer. Using silicon itself as a substrate circumvents the chemical reactivity problem. The crystals are a few μm in size. By producing a bulbous crystal and annealing it, local thermodynamical equilibrium can be achieved in a time much smaller than that required for capillary induced changes to become appreciable [11].

2. EXPERIMENTAL

We used an UHV electron microscope (0.7 nm resolution) with a residual pressure in the 10^{-9} mbar range at the sample level [12]. The specimen holder (single tilt) allowed electrical heating of the specimen by passing a DC current through the latter. The images could be tape-recorded at video rate (30 frames/s).

The samples were cut from Si (111) wafers (18 x 2 x 0.3 mm^3). The samples were produced using an n-doped wafer (resistivity $\approx$ 1 Ωcm) or a p-doped wafer (resistivity $\approx$ 1-4 Ωcm). Prior to any experiment the sample was degreased by ultrasonic

treatment in desionized water, ethyl alcohol and acetone and so installed in the specimen holder that the azimuth direction of the incident electron beam was close to <110>. Flashing them at 1520K cleaned them.

The sample temperature was calibrated against the heating DC current by optical pyrometer (monochromatic, $\lambda = 0.65$ μm). For this measurement the sample holder was transferred into a separate vacuum system without dismantling the sample. The emissivity of silicon [13] and the absorption by the window of the vacuum chamber were taken into account. We estimate our measured temperatures to have an uncertainty of +/- 7.5 K at 1573 K

2.1. Sample preparation for the macroscopic approach.

Rectangular patterns of small columns were created on the polished side of the wafer by photolithography and chemical etching. The columns were arranged in a rectangular array and were 40 μm apart from each other. The patterns were so installed on the wafer that the column rows were parallel to <110>. The height of the columns was about 10 μm and their transversal dimension varied from 1 to 5 μm.
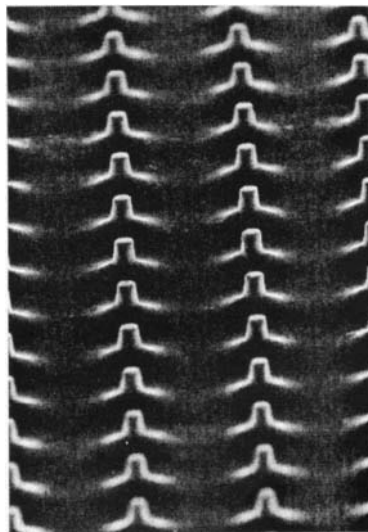

Figure 1. This is a part of a virgin sample as viewed by conventional SEM in grazing angle of around 30°. It shows an example of the column population prior to any thermal treatment. Note that the sample surface is flat between the columns. The columns are 40 μm apart and about 10 μm high.

No special attention was given to obtain a rigorous homogeneity in the transversal dimension of the columns. The columns that were produced were not strictly cylindrical. This was not deemed detrimental since the columns achieved a bulbous, fairly symmetrical, shape during the high temperature treatment that was applied at the beginning of each experiment. The cross section of the columns can be viewed in microscopy (figure 2)
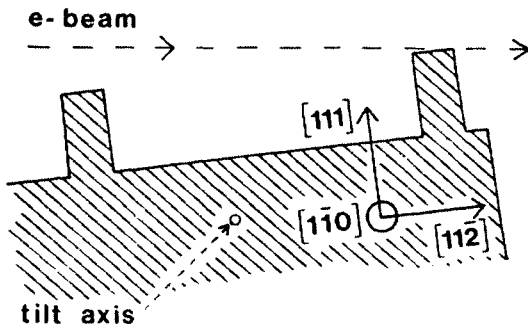

Figure 2. Schematic of the experimental arrangement in the TEM mode. It is a cross-sectional diagram showing the mutual positions of the sample and the electron beam. The direction of observation is along the electron beam. The upper-most column is being observed by looking at the sample under grazing incidence, perpendicularly to its longer side.

2.2. Sample preparation for the microscopic approach.

The samples were cut from Si (111) wafers. The microscope is used in the reflection mode (REM) the electron beam strikes the sample under grazing incidence (1° or less). A RHEED pattern is formed by the electrons reflected from a few top layers of the crystal. One diffracted beam is selected with an aperture and used to image the surface of the sample. The first consequence of the REM geometry is a severe foreshortening of the image but its main advantage is its sensitivity to the surface topography at the atomic scale. Although the surface was fairly disturbed by the cleaning treatment it was possible to find regions, on the (111) specimen, displaying a parallel system of steps with a fairly constant separation d. It has been shown that these steps are monatomic ones [14]. Regions of different d were selected on one sample and studied separately, Additionally some circular steps were also found. They appeared elliptical on the image. By measuring their elliptical shape we could calibrate the ratio of the magnification along the tilt axis to the

perpendicular one (around 50). The steps we studied on the (111) specimen were all very close to a <110> direction. Figure 3 shows an example of the REM images of an isolated step and of two step trains making slightly different angles with the tilt axis <112>.
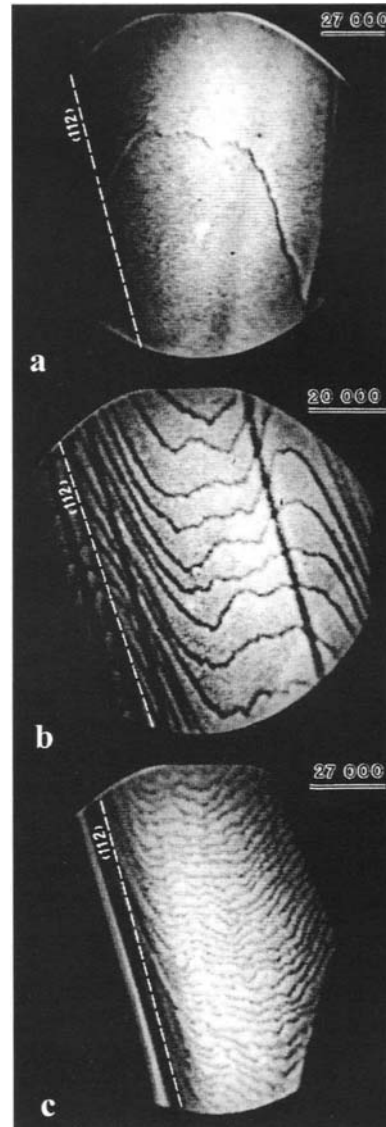
Figure 3. (a) An isolated step viewed by REM, it is pinned to the surface at points A and B. (b), (c) Two step trains with different densities. The dot-dashed line is the tilt axis<112> of the sample, which is normal to the electron beam. Due to the foreshortening of the REM images, the magnification noted at the top right of the images corresponds to the magnification along the tilt axis. In the perpendicular direction, the magnification is divided by about 50.

3. EQUILIBRIUM SHAPE (MACROSCOPIC APPROACH)

The first step to obtain the equilibrium shape of silicon consisted in transforming the straight columns into bulbous ones. For this purpose the sample was heated to 1623 K. Heating was interrupted when the desired shape was reached. On the one hand this process cleaned the sample surface. On the other hand a dramatic shape change took place, the column decreases while a neck is formed and a bulb is created at the top. Figure 4 shows a typical example of the shape evolution.
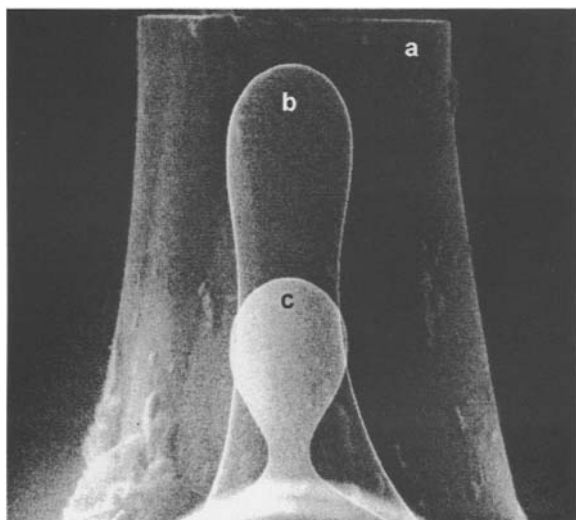
Figure 4. Typical stages in the morphological changes of the silicon columns upon heating at 1633 K. Superposition of three micrographs taken by conventional SEM. (a) Initial state. Height of the column ~ 8 μm, diameter ~ 5 μm. (b) After about 3 min at 1633 K. The height and diameter have decreased. The bulb is being formed. (c) After about 20 min at 1633 K a neck has been formed. The bulb diameter is about 1 μm. At this stage, the evaporation diffusion-process is stopped. Upon annealing at 1323 K facets appear on the surface.

In the second stage of an experiment the temperature was lowered down to 1323 K. Upon cooling, facets appear on the (previously smooth) bulb within 20 s and a stationary shape is reached. This shape will be termed stable in the following since it could be maintained for 1 h without appreciable decrease of the overall size of the columns. Fig. 5 shows one such shape.
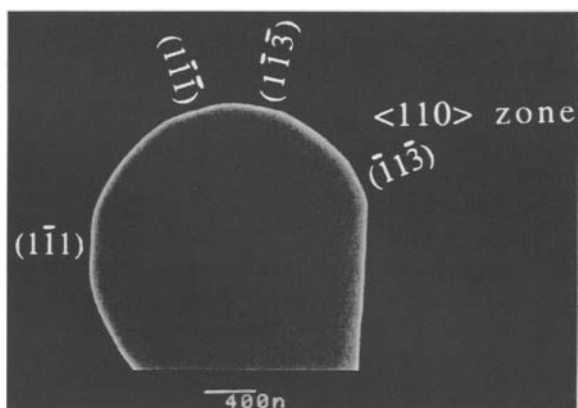
Figure 5. SEM micrograph of a clean silicon bulb after annealing at 1323 K (<110> zone). Facets are visible on the profile. Bulb diameter is ~1.6 μm.

The profiles were recorded and analyzed. Comparing their mutual angles to the values known from crystallography did indexing the facets. The measured geometrical angles always agreed with the theoretical ones within less than 2°. Such a slight discrepancy in the angular relationships is not unexpected, considering (a) the unavoidable imperfections in the alignment of the column rows with a crystallographic orientation on the wafer and (b) the fact that our single tilt prevented any fine in-situ adjustment of the sample position and of the sample edge with the tilt axis. We believe that this slight misalignment of the profiles does not preclude significant conclusions to be drawn from our measurements.

On the clean samples the only facets found at 1323 K were, along the <110> zone, well-characterized {111}, {113} and "curved" {110} and {100} facets. Indeed, the change in surface curvature is fairly small around {110} and even smaller around {100}. Therefore it is difficult to tell whether there exists {110} and {100} facets or whether the surface is merely rounded around these orientations. The special cases of {100} and {110} facets have been studies in details (see Refs. [15, 16]).

We must now give evidence that the observed profiles were representative of the equilibrium shape of silicon.

(i) The linear dimensions of the bulb did not decrease appreciably for an hour at 1323 K, whereas the profile reached a stable shape in less than one minute. This is evidence for an ovulation time at 1323 K much greater than the equilibration time of the bulb apex (much more than 1 h).

(ii) The stable shape of the profile was size independent.

(iii) Varying the temperature between 1323 K and 1573 K could produce reversible shape changes.

(iv) On the most bulbous columns, fairly good symmetry of the <110> profiles existed with respect to the bisectors of the angles between the two {111} facets and between the two {113} facets. These bisectors are crystallographic symmetry axis of the <110> zone but are not the macroscopic symmetry axis of the column (it is <111>). If the bulb shape was purely due to the capillary induced shape change, no symmetry should be found with respect to the previous mentioned bisectors.

So, our observations prompt us to conclude that the equilibrium shape of silicon was practically achieved on the apex of the columns. The extent of the equilibrated region can be estimated by checking the profile symmetries.

Whenever two symmetry axes can be found for a profile, the Wulff's point can be determined. It is located at the intersection of the axes. Then the relative anisotropy of the surface free energy (γ) can be measured by the inverse Wulff's construction. To conclude, it seems that the relative order of the surface free energy values is $\gamma_{111} > \gamma_{110} > \gamma_{113} > \gamma_{100}$ at the equilibration temperature of 1323 K.

We also have constructed the γ-plots for three profiles. The resulting plots are presented superimposed on the same figure (Figure 6).

Figure 7 shows one individual γ-plot. From figure 6, the absolute accuracy on the anisotropy can be estimated as roughly 1%. However, on an individual plot, the relative uncertainty is much smaller, as demonstrated by figure 7. Yet, figure 6 makes the general tendency apparent: The maximum anisotropy is about 4%. A cusp is found at <111> and <113> a broad minimum at <110> and <001>, in agreement with our observation of "curved" {110} and {100} facets and of a relatively small change in curvature of the profiles around these orientations.

Since all orientations are present on the equilibrium shape, the step free energy β_1 can be determined, in principle. Two methods can be used for this purpose [17]: Let γ_{hkl} be the facet surface free energy, a : the step height, r : the facet "radius" measured on the profile and h_{hkl} the Wulff's distance to the {hkl} facet.

Let also $(\partial/\partial\theta)(\gamma_{hkl}/\gamma_{111})$ be the slope of the (relative) γ-plot at the cusp. β_1 can be calculated using either one of the expressions:

$$\beta_1 = a\gamma_{hkl}\frac{r}{h_{hkl}} \tag{1}$$

$$\beta_1 = a\gamma_{111}\frac{\partial}{\partial\theta}\left(\frac{\gamma_{hkl}}{\gamma_{111}}\right) \tag{2}$$

We used formula (1) to determine β_1, from those profiles that were used to construct the γ-plots, formula (2) to calculate also β_1 from the individual γ-plots.

$\gamma/\gamma(111)$

Figure 6. Experimental γ-plots for the <110> zone. Superposition of the γ-plots measured on three profiles. Dots: experimental data. Solid curve: calculated mean value.
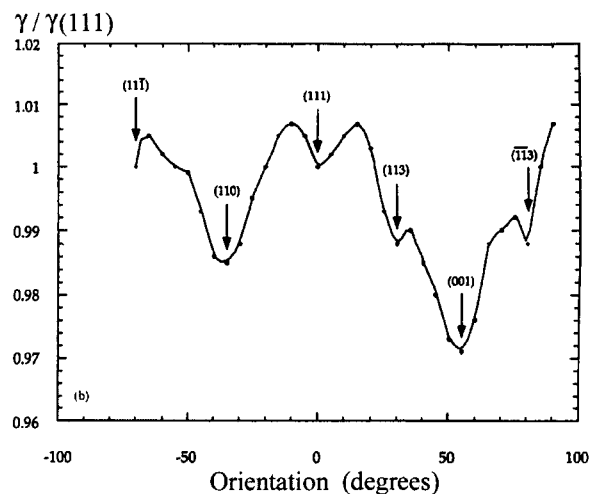
$\gamma/\gamma(111)$

Figure 7. One particular plot. Dots: experimental values. The line between the experimental points is a guideline for the eye.

Strictly speaking, <110> steps may have different structures. For example, on {111} vicinal misoriented toward {110}, the steps have a {111} riser, whereas, on {111} vicinal misoriented toward {001} they have a {001} riser. Considering the uncertainties on the slopes and the r, we disregarded the asymmetry around {111} and {113}. Taking γ_{111} = 1 J.m^{-2} and a = 0.313nm, we found, as averaged values, $\beta_1 \sim 3 \times 10^{-11}$ J.m^{-1}.

Yet, we remark that our results are not in disagreement with those found by Eaglesham et al. ($\beta_1 \sim 6 \times 10^{-11}$ J m^{-1}) [18]. They also agree with the values of the step stiffness measured in REM studies of the step fluctuations [see below].

Informations about the step interactions can also be obtained by observation of the equilibrium shape near a facet. Below the roughening temperature of a given facet, the equilibrium shape generally displays a rounded region surrounding the facet. Most remarkably, the crystal shape in this region [19] reflects the step-step interaction law. This is readily understood by considering the crystal profile along a mirror plane, e.g. the profile along a <110> plane of the cubic lattice. Then, for any vicinal orientation belonging to the rounded region surrounding {hkl} facet, the surface free energy may generally be represented [20] by a development:

$$\gamma(\theta)/\cos\theta = \gamma_{hkl} + \beta_1 n + \beta_2 n^2 + \beta_3 n^3 + ... \qquad (3)$$

Where θ is the misorientation angle from the facet and $n = |\tan\theta|/a$, the step density. The step interactions can be related to the β_2, β_3, ... terms, as explained below. Since the equilibrium shape is related to $\gamma(\theta)$ via the Wulff construction, the profile of the rounded regions depends on the nature of the interactions between steps.

Eq. (3) merely assumes that γ is a continuous analytic function of θ [20], it is essentially macroscopic. A physical meaning is assigned to the coefficients therein by considering the structure and the energetic of a vicinal surface. In the TLK model, γ_{hkl} is the specific free energy of the facet {hkl} and the terraces, and the steps possess a free energy β_1 per unit length and are supposed to interact through various mechanisms, for instance the entropic, elastic or dipolar repulsion's (see references in [21]). In principle, considering these interactions allows, β_2, β_3, etc., to be calculated. The

n^2 term in Eq. (3) is very often cancelled. Indeed, all the physically sensible interactions which have been considered theoretically so far (see references in [21]) result in an interaction decaying as the inverse square of the mean distance between steps. For repulsive interactions, this leads to $\beta_2 = 0$ in Eq. 1 [19], which can be limited to the cubic term in n^3 [21]. The profile must then obey a 3/2 power law [8]:

$$Z = \alpha Y^{3/2} \qquad (4)$$

Theory states that this is the signature of a Pokrowski-Talapov universality class behavior [19]. In other words, the 3/2 power law should be valid for any profile. Additionally, the expression between β_3, γ_{111} and α can be obtained theoretically

$$\beta_3/a^3\gamma_{111} = 4/27\alpha^2 h_{111} \qquad (5)$$

However, experiments aimed at checking the 3/2 power law did not yield unambiguous results.

We have determined the profile of equilibrated silicon single crystals in the vicinity of a {111} facet along the <110> zone, towards {110}. To measure the coordinates (Z, Y) of the profile points, it was necessary to make the {111} facet coincide with one pixel line on the computer display. Hence, we sometimes had to slightly rotate the image, through the computer program. In order to preserve the sharpness of the image, we only used a bilinear interpolation for this purpose. The position of the facet plane could be assessed with an uncertainty of 1 pixel. For a given position of the facet plane, we estimated the uncertainty on Z to be ±1 pixel. In the following, the corresponding error bars have been drawn on the figure 8 showing our experimental data.

In the first step of our analysis, we tried to decide whether our data were compatible with a power law. Due to the tangential junction, it is very difficult to locate the edge of the facet by a mere visual inspection of the image.

Consequently, the profile coordinates were fitted to the equation

$$Z = \alpha (Y - Y_0)^p \qquad (6)$$

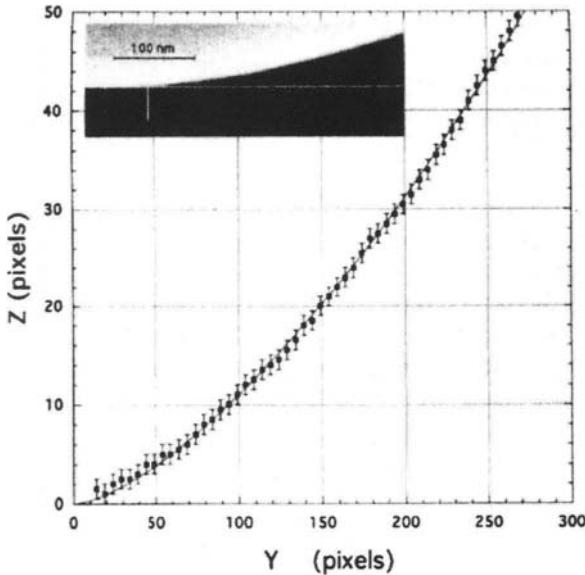with three unknown parameters α, p, and Y_0, which determines the position of the facet edge.

Figure 8. Example of the results of the data analysis. Height Z of the profile above the facet as a function of Y. Squares are experimental data. Solid line corresponds to the fit to Eq. 6. For the fit, the Y origin has been made to coincide with the facet edge, as obtained from the fit. The insert shows the recorded SEM image with vertical white bar at the position of the facet edge.

A least squares fit analysis was used to obtain α, p, and Y_0. Note that the length unit used to measure Z and Y is unimportant to check the power law itself. We measured Z and Y directly in pixels.

Every profile could be fitted to Eq. 6 up to (roughly) 17° away from {111}. Obviously a power law is valid, with an exponent whose value is close to p = 1.5 (with a 6% uncertainty). We checked that no significantly different value of p could be found by shifting the facet position by 1 or 2 pixels. We also made sure that the facet edge position, deduced from Y_0 was always reasonably situated.

So we must conclude that the equilibrium profiles can be described by a power law equation within the uncertainty of our measurements. Then, a quite reasonable value for the exponent is p=3/2.

For each profile the value of α obtained from Eq. 6 was used in Eq. 5 to calculate β_3. Since both terms in Eq. 3 are dimensionless, the length unit is again unimportant. So we measured h_{111} in pixels and scaled it to the magnification of the profile image. By so doing, we find $\beta_3/a^3 \sim 0.4$ J.m^{-2}.

4. THERMAL FLUCTUATION OF STEPS (MICROSCOPIC APPROACH)

The thermal fluctuation of an isolated step allows the measurement of the step stiffness. Let us first recall the definition of the step stiffness. Here we only consider a step which meanders around a mean line along which it is straight at T = 0 K. Let $\beta_1(\varphi)$ the specific step free energy of azimuth direction φ at T, the step stiffness is defined by: $\beta_1^* = \beta_1(\varphi = 0) + \partial^2\beta_1/\partial\varphi^2$

We now consider one isolated step, which is pinned at two points (A, B), distant from L, but which meanders in between. Let x be the abscissa of some point P between A and B and y (x, t) be the deviation of the step from AB at P. It can be shown [9] that the mean square fluctuation of y, averaged over time and over the whole length L, is

$$<\Delta y>= kTL/6\beta_1^* \qquad (7)$$

where k is the Boltzmann constant and T the temperature.

One such isolated step is shown by figure 3. We measured the deviation y from the AB line. Then the mean square deviation was calculated. By using the previous equation (7) the value of β_1^* was determined: $\beta_1^* = 7.10^{-11}$ J.m^{-1}.

The wandering of steps inside a train is limited by their interactions. Consequently the configuration step entropy is reduced, which raises the free energy of the step train. An interaction may arise because the steps cannot cross each other (and do not bunch into multiple height steps). This is the "entropic repulsion" [22]. Apart from this interaction already mentioned, various types of "energetic" interactions are conceivable. Elastic or electrical dipolar fields may reduce the step wandering further. The elastic interaction between steps is due to the lattice strain field around a step. For two parallel steps on an isotropic solid, the interaction energy per unit step length has been calculated [23]. For steps of similar signs this interaction is always repulsive. The dipolar electric interaction occurs via the electric field created by the change in charge distribution at the steps. This interaction may be repulsive or attractive. In a mean-field model, all these interactions (including the entropic one) lead to an additional d^{-2} term in the free energy per projected

unit area of the vicinal surface and only the cubic term exists in the development (1).

Since in situ experiments permit a direct statistical study of the step meandering in a train one may seek what characteristics of a step train give access to the step interactions. One such quantity is the distribution of the terrace widths.

The simplest model to treat the pure entropic interaction was devised by Gruber and Mullins [22]. One step is compelled to wander between two rigid walls, a distance 2<d> apart. Treating the step as a Markovian chain rigorously solved the problem. An analytical expression of the terrace width distribution is given by $\cos^2(\pi d/2<d>)$. The distribution is temperature independent, a characteristic of models with entropic interaction only. Joos et al. [24] elaborated this model by letting the whole train wander, the distribution is more peaked that the Gruber-Mullins'one.

The superimposed energetic interactions with the entropic one, will sharpen or broaden the distribution P(d) depending on their repulsive or attractive character and will make it temperature dependent. This expectation was verified by Joos et al. [24] They considered an interaction energy of the A/d^{-2} type and obtained exact solutions for two specific values of A (one repulsive and one attractive). The peak of the distribution was very well fitted by a Gaussian for the cases of repulsive interactions. Bartelt et al. [25] formulated the Gaussian approximation analytically in the case of dominant repulsive interactions decaying as A/d^{-2}. They found that the standard deviation of the Gaussian Δ is a linear function of d:

$$\delta = <d>(k^2T^2/48A\beta_1^{*})^{1/4} \qquad (8)$$

Therefore the constant A which characterizes the interaction may be experimentally determined if β_1^{*} is known.

Seven step trains with a nearly constant step separation ranging from 20 to 140 nm could be selected on the same sample. They were deemed long enough to be compared to a model of infinite steps. REM images were recorded at 1170K. Micrographs of the surface separated by a time interval of 0.25 seconds were extracted from the record. The distributions of terrace widths were measured from these data. All the distributions were well fitted by a Gaussian.
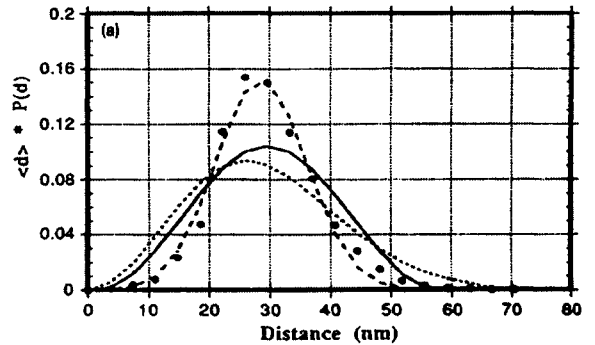
Figure 9. Example of a normalized experimental terrace width distribution <d> P(d) versus d and comparison of the data with various theoretical distributions. P(d) is the probability density of finding a terrace of width d. Long-dashed line: fit by a Gaussian.Short-dashed line: the "universal" law of Joos et al. The cos2 law of Gruber and Mullins is also shown for comparison (continuous line). Average step separation is <d> = 29.6 nm.

Figure 9 shows one such measured distribution and the Gaussian fit of the data.

The plot of the standard deviation δ versus <d> was compatible with a linear relationship (8) (Figure 10).
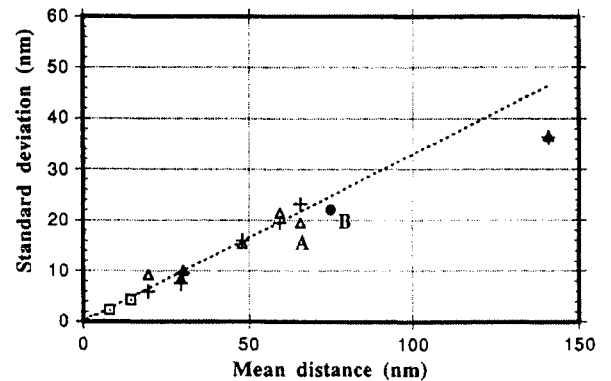

Figure 10. Standard deviation of the terrace width distribution versus the average step separation. (Δ) experimental values. (+) standard deviation of the Gaussian that fits the experimental distribution. ($\square$) experimental values by Wang et al. [2]. Short-dashed line: $\delta = 0.33$ <d> (see Eq. 8).

We have also plotted the experimental results by Wang et al. [2] for comparison.

We must conclude that our experiments establish the existence of repulsive step interactions varying as A/d^2, superimposed to the entropic one. We can use the analytical expressions (8) derived by Bartelt

et al to estimate the interaction constant A. We checked that A has the order of magnitude of an elastic interaction constant [23].

5. MECHANISM OF MASS TRANSPORT FOR STEP FLUCTUATION

There are two extreme cases of the many microscopic mechanisms which can cause step motion: (i) the exchange of atoms at the step edge with adatoms or vacancies on the terraces, (ii) the motion of atoms hopping from one step site to another on the same step. The two classes are clearly distinguishable by analysis of the time correlation function of the step position shown in figure 3b. Two Langevin's equations governing the mechanisms of mass transport express the balance between the minimization of the step curvature and the thermal noise of adatoms attaching /detaching for (i) or atoms hopping along the step edges for (ii) [26].

When the step is in thermal equilibrium, the correlation: $G(t - t') = <x^2> - <x(t) x(t')>$ of the fluctuation at any particular point on the step edge will be given, at early times, by $G(t) \propto t^{1/2}$ for the mechanism (i) and by $G(t) \propto t^{1/4}$ or the mechanism (ii).
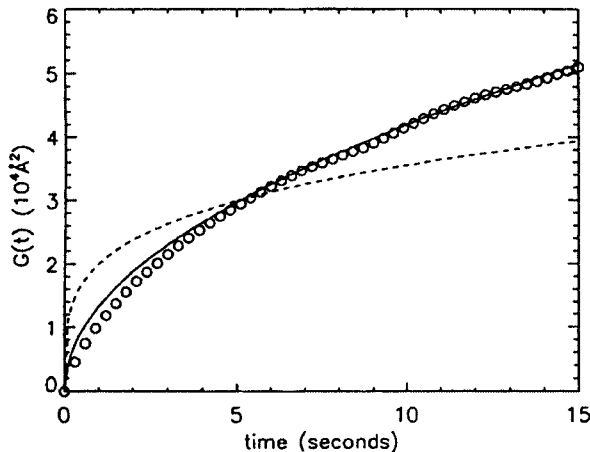
Figure 11. Correlation function G(t) deduced form averaging over around 800 data sets such as shown in figure 2b. We have only included the correlation functions for times shorter than 15 s, because for longer times, the statistical noise becomes large. The solid line is a fit to the square root time; the dashed line is a fit to fourth root time

Figure 11 plots G(t) from the data from figure 2b. The fit to the square root time dependence is

reasonably good, suggesting that the steps on Si(111) have characteristics of Brownian motion in which terrace adatoms or vacancies are allowed to attach and detach randomly from the step edge.

6. CONCLUSION

The surface steps considered as entities which can move by mass transport between them, is a very powerful approach to describe the dynamics of the surface below its roughening transition. The macroscopic analysis of the equilibrium shape of Si is a very fruitful way to get informations on the step free energy and their interactions. The equilibrium profile in the vicinity of (111) is well described by a 3/2-power law. It is a signature of a Pokrowski-Talapov phase transition, which is the consequence of a $1/d^2$ repulsive interaction between steps. Our macroscopic measurements are in qualitative agreement with the interaction law that was measured at microscopic scale by REM on the wandering steps on the vicinal. The step stiffness of an isolated step has been measured. From the analysis of terrace width distribution the existence of long ranged energetic step interactions decaying as Ad^{-2} can be established. The mechanism of adatom supply to a step has been determined from the time correlation of the step positions.

REFERENCES

1. B.S. Schwartzentruber, Y.W. MO, R. Kariotis, M.G. Lagally and M.B. Webb, Phys. Rev. Lett. 65 (1990) 1913.
2. X.S. Wang, J.L. Golberg, N.C. Bartelt, T.L. Einstein and E.D. Williams, Phys. Rev. Lett. 65 (1990) 2430.
3. R.M. Tromp and MC. Reuter, Phys. Rev. Lett. 68 (1992) 820.
4. A.V. Latyshev, A.L. Aseev, A.B. Krasilnikov and S.I. Stenin, Surf. Sci. 213 (1989) 157.
5. C. Alfonso, J.M. Bermond, J.C. Heyraud and J.J. Métois, Surf. Sci. 262 (1992) 371.
6. E.D. Williams, R.J. Phaneuf, J. Wei, NC. Bartelt and T.L. Einstein, Surf. Sci. 294 (1993) 219; 310 (1994) 451.
7. C. Herring, Phys. Rev. 82 (1951) 87.
8. J.J. Saenz and N. Garcia, Surf. Sci. 155 (1985) 24.

9. P. Nozières, in: Solids far from Equilibrium, Ed. C. Godrèche (Cambridge University Press, Cambridge, 1992) p. 1

10. F.A. Nichols and W.W. Mullins, J. Appl. Phys. 36 (1965) 1826.

11. T. Barsotti, J.M. Bermond and M. Drechsler, Surf. Sci. 146 (1984) 467.

12. J.J. Métois, S. Nitsche and J.C. Heyraud, Ultramicroscopy 27 (1989) 349.

13. F.G. Allen, J. Appl. Phys. 28 (1957) 1510.

14. R.J. Phaneuf, E.D. Williams and NC. Bartelt, Phys. Rev. B 38 (1988) 1984.

15. J.C. Heyraud, J.J. Métois, J.M. Bermond, Surf. Sci. 425 (1999) 48.

16. J.J. Métois, J.C. Heyraud, Surf. Sci. 446 (2000) L127.

17. R. Kern, Morphology of crystals, Materials Science of Minerals and Rocks, Ed. I. Sunagawa (Terrapubl. Tokyo) Vol.2 (1987).

18. D.J. Eaglesham, A.E. White, L.C. Feldman, N. Moriya and D.C. Johnson, Phys. Rev. Lett. 70 (1993) 1643.

19. C. Jayaprakash, C. Rottman and W.F. Sam. Phys. Rev. B30 (1984) 6549.

20. N. Cabrera. Surf. Sci. 2 (1964), p. 320.

21. E.D. Williams, Surf. Sci. 299-300 (1994) 502.

22. E.E. Gruber and W.W. Mullins, J. Phys. Chem. Solids 28 (1967) 875.

23. V.I. Marchenko and A.Y. Parshin, Sov. Phys. JETP 52 (1980) 129.

24. B. Joos, T.L. Einstein and NC. Bartelt, Phys. Rev. B 43 (1991) 8153.

25. N.C. Bartelt, R.L. Einstein and E.D. Williams, Surf. Sci. 240 (1990) L591.

26. N.C. Bartelt, J.L. Goldberg, T.L. Einstein, and E.D.Williams, Surf. Sci. 273 (1992) 252.

Elementary growth process in semiconductor epitaxy
-Molecular beam epitaxy as an example of epitaxy-

Tatau Nishinaga

Department of Materials Science and Engineering, Faculty of Science and Technology, Meijo Univerity, 1-501 Shiogamaguchi, Tenpakuku, Nagoya 468-8502, Japan
Email: nishinag@meijo-u.ac.jp

Elementary growth process of epitaxy is described taking molecular beam epitaxy as an example. First, we study the diffusion length of adatom incorporation and its dependence on arsenic pressure. Then, the surface diffusion between two surfaces(often they are facets), which we call inter-surface diffusion are described. Basing on these experiments the elementary growth process of epitaxy is discussed.

The understanding of the elementary growth process allows us to make use of this knowledge to fabricate microstructures. We will discuss how to control the top size of pyramids by adjusting the arsenic pressure. Finally, we will describe a new technique of GaAs MBE for growing a thin layer of GaAs single crystal over a SiO_2 film. This technique has been called LAIMCE(Low angle incidence microchannel epitaxy) and some new results are described.

1. INTRODUCTION

Epitaxial growth is one of the most important techniques to fabricate various electronic and optical devices such as high electron mobility transistor(HEMT), light emitting diode and laser diode. Modern devices require very sophisticated structures, which are composed of extremely thin films of III-V alloys with various compositions. This kind of structure cannot be made without epitaxial growth.

Recently the epitaxial growth is also used for the fabrication of semiconductor nano-structures such as quantum wires and quantum dots because it gives highly perfect structures with high density. So far many techniques of epitaxial growth have been proposed to

be employed for the fabrication of nano-structures. For instance, proposed are epitaxial growth on patterned substrate[1-5], selective area epitaxy[6-10], self-assembling growth [11-13] and etc. To fabricate sophisticated structures including nano-structures, one should understand the elementary growth process so that one can precisely control the growth.

In the present lecture, we discuss the recent understandings of elementary growth processes of epitaxy taking molecular beam epitaxy(MBE) as an example. The advantage of MBE for understanding epitaxy is its simplicity in the growth reactions. Environment is of ultra-high vacuum where almost no impurity gaseous species present. The growth elements are also simple for

instance in the case of GaAs MBE they are only Ga atoms and As molecules.

What is important for understanding the growth processes is that the experiments to obtain the growth information should be carried out at growth temperature in real time. We have been employing both conventional MBE and microprobe-RHEED/SEM MBE[14-16] by the latter of which we can conduct in-situ observation of RHEED intensity oscillation in a small area and changes in growth morphology[17-18] as well as the measurements of surface diffusion length of incorporation [14,15,19,20]. In the present lecture, we will discuss the elementary growth process of MBE basing on these experimental works.

2. INCORPORATION DIFFUSION LENGTH

In the theory of crystal growth, two kinds of diffusion length are so far employed. One is the average distance of a growing atom between arrival and evaporating points, λ_s, which is schematically shown in Figure 1. In MBE of GaAs, Ga atoms are difficult to evaporate at usual growth temperature and stay until they are incorporated into the crystal. Hence, the diffusion length of evaporation has almost no meaning in the case of GaAs MBE.

The other one is the incorporation diffusion length λ_{inc} which is defined as the distance between arrival and incorporating points. The concept of the incorporation diffusion length is not clear compared with the evaporation diffusion length because the former depends on the step density and the arsenic pressure in the case of GaAs MBE, while the latter is defined on an ideal surface where no step exists and hence is a simple function of the temperature and depends on adsorption and desorption energies. However, the incorporation diffusion length plays a key role in the growth of nano-structures.

The incorporation diffusion length can be determined experimentally by measuring the distribution of growth rate on a facet where the growth atoms diffuse laterally from or to a next facet. This technique was established by Hata and Isu[14, 15].

Figure 2 shows the schematic illustration of microprobe -RHEED/SEM MBE system used for the measurements of incorporation diffusion length. With this machine, one can focus electron beam on the growing surface with small area of the order of 10 nm in size so that it is possible to measure the local growth velocity with this spatial resolution. After preparing the non-planar (001) substrate with a groove which has (111)A or (111)B side
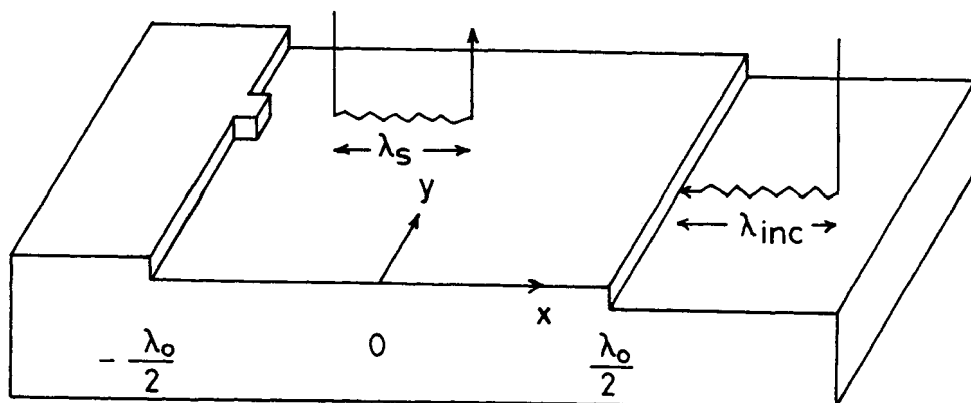
Figure 1 Definitions of two different surface diffusion lengths.
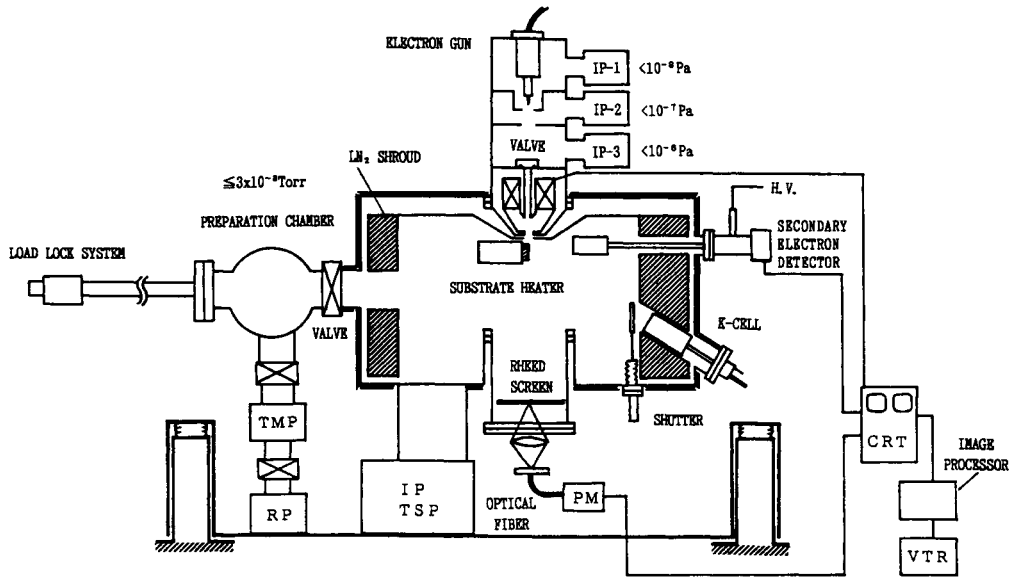
Figure 2 Microprobe-RHEED/SEM MBE employed for the measurement of incorporation diffusion length.

surface, we grow GaAs in the μ-RHEED/SEM MBE machine and measure the local growth rate on the top surface by RHEED intensity oscillation. Figure 3 shows the schematic growth rate as a function of the position from the corner between two facets. If the Ga atoms flow from the side surface to the top surface, the local growth rate shows a decreasing tendency as the distance is increased. On the other hand if the Ga atoms diffuse in opposite direction it shows the increasing tendency. Hence, one can know the direction of Ga lateral flow by measuring the distribution of the local growth rate. From the distribution of the growth rate, one can measure the diffusion length of incorporation basing on the fact that the change of the growth rate obeys exponential law as a function of the distance[14, 15].

In MBE of GaAs, Ga adatoms can diffuse in the distance on the order of 1 - 10 μm before the incorporation. It has been shown that on (001) surface the incorporation
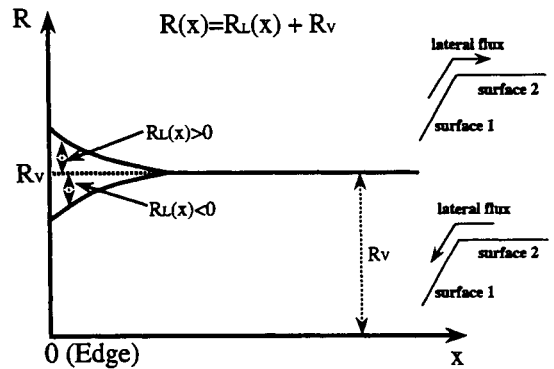
Figure 3 Schematic illustration of the growth rate, R(x) as a function of position x from the boundary between two facets. In the Figureure, R_L and R_V denote respectively the rates of growth which is caused by lateral flux and that by the direct flux. R_V is equal to the rate at the point far from the boundary.

113
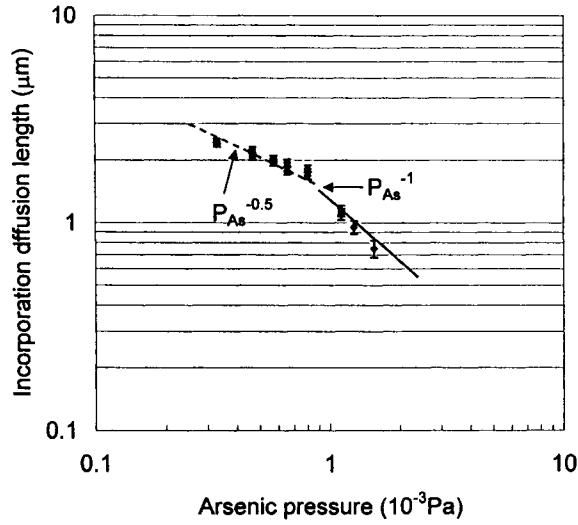
Figure 4 Arsenic pressure dependence of Ga incorporation diffusion length on (001)surface. Growth temperature and the side surface were respectively 580℃ and (110).

diffusion length of Ga depends on the arsenic pressure [20-23]. In our previous report, we showed the diffusion length is inversely proportional to the As_4 pressure. However, we found when the arsenic pressure is low, there is a region of the arsenic pressure where the diffusion length is inversely proportional to the square root of As_4 pressure. The dependence of λ_{inc} on arsenic pressure on (001) surface is given in Figure 4. Here, the arsenic pressure is recalculated so that the arsenic flux enter the (001) surface at right angle.

λ_{inc} is given by well known formula as,

$$\lambda_{inc} = \sqrt{D_s \tau_{inc}} \qquad (1)$$

where D_s and τ_{inc} denote respectively, the diffusion coefficient and the incorporation lifetime of Ga adatoms. Since the growth rate of GaAs, R_{GaAs} is proportional to the inverse of τ_{inc}, we get

$$R_{GaAs} = A_1 / \tau_{inc} \qquad (2)$$

where A_1 is a constant. From Figure 4, we get

$$\lambda_{inc} = A_2 P_{As4}^{-n} \qquad (3)$$

where A_2 is a constant and n takes the values of 0.5 and 1 when P_{As4} is lower than 8×10^{-4} Pa and higher than this value respectively. With the equations from (1) to (3) we get,

$$R_{GaAs} = A_3 P_{As4} \quad (P_{As4} \leqq 8 \times 10^{-4} \text{ Pa}) \qquad (4)$$

$$= A_3 P_{As4}^{2} \quad (P_{As4} \geqq 8 \times 10^{-4} \text{ Pa}) \qquad (5)$$

respectively for lower and higher As_4 pressures. The equation (4) shows that at lower As_4 pressure R_{GaAs} is proportional to P_{As4}, which means one As_4 molecule decomposes and gives active arsenic atoms or molecules for the growth. On the other hand, at high arsenic pressure, the equation (5) shows two As_4 molecules react to give the active arsenic atoms or molecules. There is no information to determine the kind of the active arsenic atoms or

molecules but As$_2$ is the most probable candidate[24-26].

Figure 5 shows the dependence of λ_{inc} on arsenic pressure on (110) surface. The arsenic pressure is recalculated so that the arsenic flux is incident on (110) at right angle. As seen in the figure, λ_{inc} is proportional to $P_{As4}^{-0.5}$ in low arsenic pressure region and to P_{As4}^{-1} in high pressure region, which means τ_{inc} is proportional to respectively P_{As4}^{-1} and P_{As4}^{-2}. Arsenic pressure dependency of λ_{inc} on (110) surface is exactly identical with that on (001).

3. INTERSURFACE DIFFUSION

When two facets are generated side by side during the growth, Ga atoms diffuse from one facet to the other depending on the growth conditions. We call this intersurface diffusion. If there is a difference in adatom concentration, surface diffusion occurs assuming there is no potential difference nor barrier between the facets for the adatoms to migrate across the boundary. Once
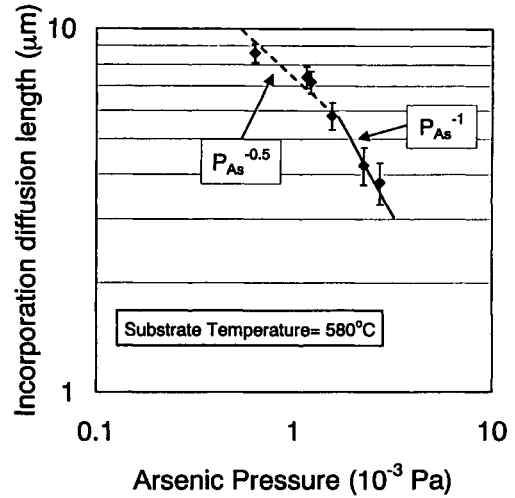
Figure 5 Arsenic pressure dependence of Ga incorporation diffusion length on (110) surface with the side surface of (001). The growth temperature was 580℃.
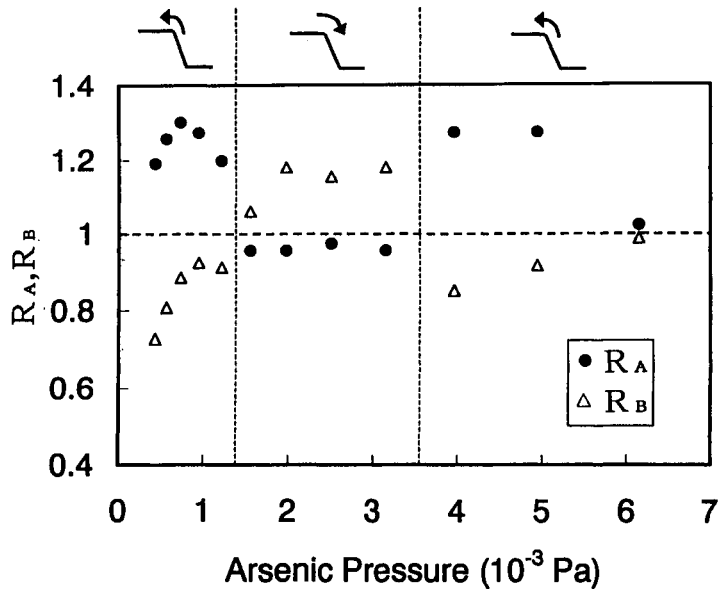
Figure 6 Normalized growth rate on (001) near the boundary and that on (111)B vs. arsenic pressure. R_A and R_B are $R_{corner}^{(001)} / R_{corner}^{(001)}$ and $R_{planar}^{(111)B} / R_{corner}^{(111)B}$ respectively.

intersurface diffusion occurs, one facet grows faster than the other, and the facet growing faster will disappear quickly[17, 18]. Thus, during the growth many facets appear and disappear and only a few facets remain in the final stage of the growth. For the fabrication of nano-structures, one should know relative growth rate among various facets during the growth. However, the growth rate of facets depends on the growth conditions and hence one should understand what is the factor which determines the growth rate.

As described in the previous section, by measuring the growth rate distribution on one facet near the boundary, one can know the direction and the length of the surface diffusion. We measured the distribution of the growth velocity on (001) surface of GaAs. As the side facet, we have chosen (111)B surface. To know the direction of the surface diffusion it is enough to measure the growth rate at 2 points. One is on the facet and close to the boundary and the other is the point on the facet very far from the boundary. We define each growth rate as R_{corner} and R_{planar} respectively. Figure 6 shows the results for a combination of (111)B-(001) facets. In the figure, closed circle and open triangle denote $R_{corner}^{(001)} / R_{planar}^{(001)}$ and $R_{corner}^{(111)B} / R_{planar}^{(111)B}$ respectively. Here, we define these normalized growth rates as R_A and R_B respectively. When the arsenic pressure is low, R_A is larger than unity. As easily understood with Figure 3, this indicates Ga adatoms diffuse from (111)B to (001). On the other hand, R_B is lower than unity which simultaneously indicates Ga adatoms diffuse from (111)B to (001) being consistent with the direction given by R_A. As the arsenic pressure is increased, R_A decreases and crosses the line of unity at the arsenic pressure of 1.4×10^{-3} Pa, which means the direction of the lateral flow is reversed, namely, from(001) to (111)B. At the same arsenic pressure, R_B also cross the line of unity from lower side to the higher side. This is

very important, since if this does not happen, one can not assume a pure two face intersurface diffusion[22]. As the arsenic pressure is increased, both R_A and R_B keep almost constant values but as the arsenic pressure is further increased, they again cross the line of unity which means the direction of the diffusion is again reversed. The direction of the diffusion for each arsenic pressure range is given on the top of the figure. Similar directional reversal was observed between (001) and (110) as shown in Figure 7.

4. ELEMENTARY GROWTH PROCESSES

Based on the experimental results given in the above, we discuss elementary growth processes involved in the MBE of GaAs.

4.1 Growth mode

In section 3, we have shown that Ga adatoms diffuse from one facet to the other. In Figure 3, the growth rate distribution on one face has been schematically given. With this distribution, as has been explained one can get experimentally the diffusion length of incorporation which has been shown in Figures 4 and 5 as a function of
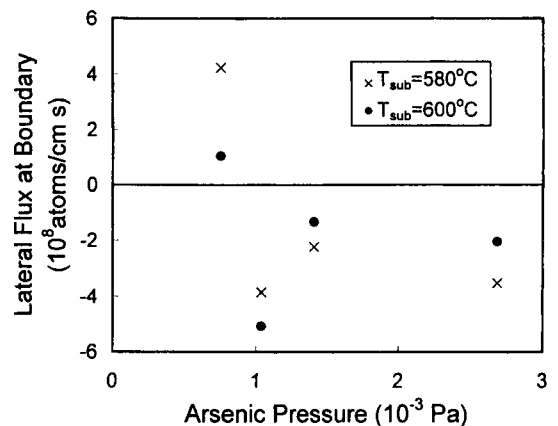

Figure 7 The arsenic pressure dependence of the lateral flux at the boundary between (001) and (110) for the growth temperatures of 580°C and 600°C. The positive flux means that the direction of the flux is toward (001) from (110).

arsenic pressure. Here, the experimental growth rate was measured by RHEED intensity oscillation so that the growth can be understood as being conducted in 2D nucleation mode. This means birth and spread of 2D nuclei are continually happening during the growth and Ga adatoms diffuse between and over the 2D nuclei. The incorporation lifetime of Ga, τ_{inc} which was defined in the section 2, should be the function of available number of incorporation sites. Hence, in this situation, the number changes periodically with the same periodicity but in opposite phase of RHEED intensity oscillation. Nevertheless, we can define the time average of τ_{inc} and λ_{inc} which we have measured in the section 3.

The average distance of the adatom incorporation is on the order of $1\mu m$ which means under such arsenic pressure Ga adatoms can diffuse crossing over many steps that are supplied from 2D nuclei. Hence, the sticking coefficient of Ga adatoms at the step edges should be much less than unity.

4.2 Change in the direction of intersurface diffusion

Intersurface diffusion occurs if there is a difference in Ga adatom concentrations between two facets. As we discussed, we have assumed that there is no potential difference nor barrier between these faces. But, there is no evidence for this assumption.

The adatom concentration of Ga, n_{Ga} is proportional to incident flux of Ga, J_{Ga} and τ_{inc}, as follows,

$$n_{Ga} = J_{Ga}\tau_{inc}. \qquad (6)$$

As we have discussed, τ_{inc} is inversely proportional to the available number of Ga sites, in other words, to the step density. The step density depends on the number of 2D nuclei and their sizes. Hence, τ_{inc} depends on the nucleation rate and the energy barrier for Ga adatoms to

enter and to leave the kink site of the step. There are almost no information for these energies so that we should be satisfied at this moment with qualitative discussions.

In section 3, we showed the experimental observation for the change in the diffusion direction. These experimental findings also give us information of elementary growth processes. Figure 6 has shown that the direction of the intersurface diffusion changes twice as the arsenic pressure is increased. At the second point of direction reversal (higher arsenic pressure side) it was found that the reconstruction of (111)B changes from ($\sqrt{19} \times \sqrt{19}$) to ($2\times2$) reconstructions as arsenic pressure is increased while the reconstruction of (001) is unchanged and keeps (2×4). It is known that (2×2) reconstruction consists of arsenic trimer which is rather difficult to be decomposed. Hence, once the (2×2) reconstruction is formed the generation of 2D nucleation might be more difficult which causes the increase of τ_{inc} and hence the increase of Ga adatom concentration.

As for the first point of directional reversal occurring at 1.4×10^{-3} Pa, we have explained in terms of difference in arsenic pressure dependence of τ_{inc} on (001) and (111)B facets. In our past paper, we extrapolated P_{As4}^{-4} dependency of τ_{inc} on (111)B and explained the reason for the directional reversal as the lines of τ_{inc} on (001) and (111)B crosses at one As_4 pressure[21]. But, as described in the section 3 recently we found there is a region in the lower arsenic pressure side where τ_{inc} shows the dependency of P_{As4}^{-2}. So that up to now, our previous conclusion for P_{As4}^{-4} dependency of τ_{inc} on (111)B, which is responsible for the direction reversal probably should be changed to P_{As4}^{-2} dependency. If this is the case, we should assure P_{As4}^{-1} dependency for (001) surface and P_{As4}^{-2} dependency for (111)B surface which allows the crossing of τ_{inc} on (111)B and τ_{inc} on (001) at the arsenic pressure of around 1.4×10^{-3} Pa.

5. FABRICATION OF MICROSTRUCTURES

In this section as one example of the micro-structure fabrication, we will describe the formation of the pyramids and the control of its top size by adjusting the direction of the intersurface diffusion basing on the knowledge on elementary growth process given in the former sections[27].

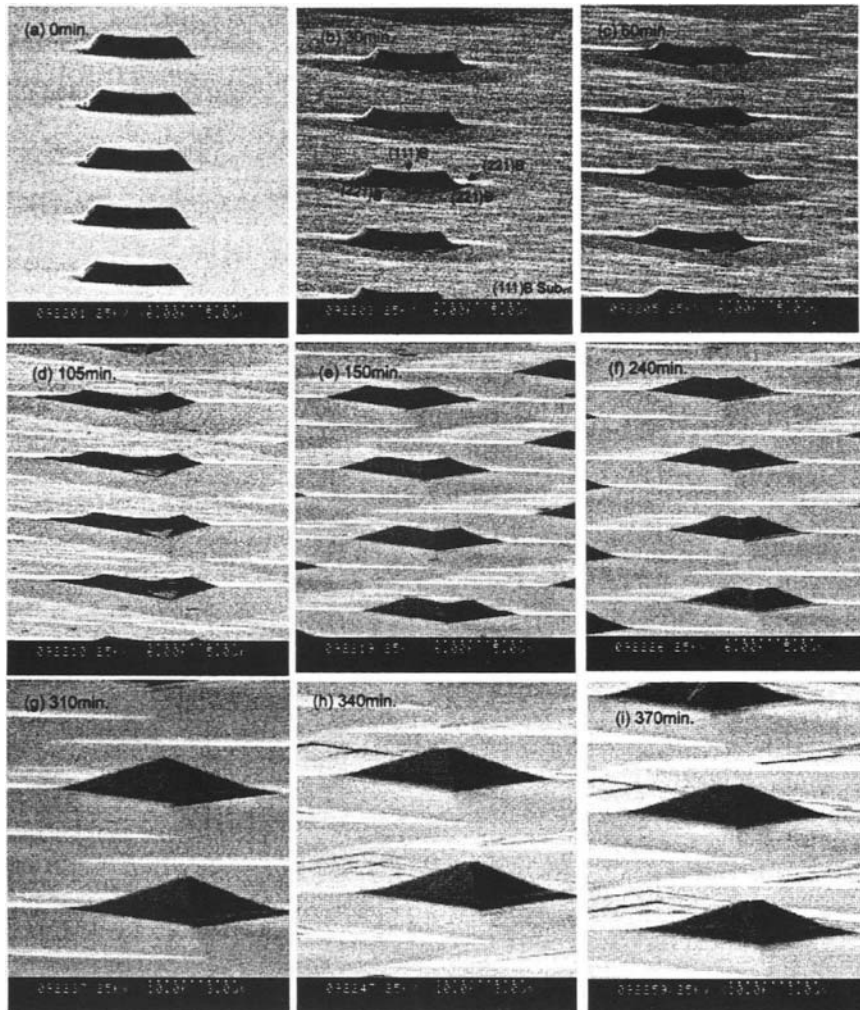## 5.1 Fabrication of pyramids and the control of the top size

Figure 8 Photographs of the real-time observations of the faceting and the shrinkage processes on the (111)B patterned substrate. (a) before growth: no special side facet appeared, (b) after 30min: {221}B appeared on the bottom of the mesa, (c) after 60min, (d) after 105min, (e) after 150min: the truncated pyramid with three complete {110} side facets, (f) after 240min: shrinkage of the top of the truncated pyramid happened, (g) after 310min: the sharp top pyramids were formed and (h) after 340min, (i) after 370min: by increasing the arsenic pressure, again truncated pyramid appeared[27].

The (111)B GaAs patterned substrates with mesa structures were employed. During the MBE those mesas were changed into the truncated pyramid. The height and the width of the mesa were about 2μm and 4μm respectively. The growth temperature were chosen as 580℃. The growth rate was kept at 0.5μm/h for the first 150 minuets and then decreased to 0.3μm/h to see the change of top size more in detail. The arsenic pressure was chosen at 1.1 x 10^{-3} Pa.

During the growth, in-situ SEM image was taken and the size of the truncated pyramid was measured from the images after the growth. Figure 8 shows the real-time photos of the pyramid formation, and Figure 8(a)-(g) show the pattern of the substrate before the growth(a), mesas in the beginning of the growth with {221}B facets on the foot of the mesa(b), {110} facets start to cover the mesa(c), {110} facets almost cover the sides(d), the truncated pyramids with three complete {110} side facets(e), the shrinkage of the pyramid top size(f) and the sharp top pyramids(g). To see the effect of arsenic pressure on the growth process, the arsenic pressure was increased to 4.7 x 10^{-3} Pa.

Figure 8 (h) and (i) show the result. As seen in the figure, the sharp top pyramids change their forms into truncated pyramids. This indicates the direction of surface diffusion was reversed when the arsenic pressure was increased. Hence, by increasing and decreasing the arsenic pressure, one can control the top size of the pyramid.

5.2 Theoretical calculation

In the following, we propose a simple model based on one-dimensional surface diffusion[27]. On the surfaces of the truncated pyramid, we must consider two-dimensional surface diffusion. But, we can solve this problem by approximating the truncated pyramid as a cone. Figure 9 shows the coordinates of the cone used in the calculation. In the figure, r, w, s and λ are defined as the distance from the center of the top surface, the half of the top width, the distance on the side wall from the top of the cone and $l\cos\theta$ where l is the side wall length, respectively.

In the following equations, Ga adatom concentration, diffusion coefficient and incorporation life time on the (111)B top, the bottom and the {110} side surfaces are
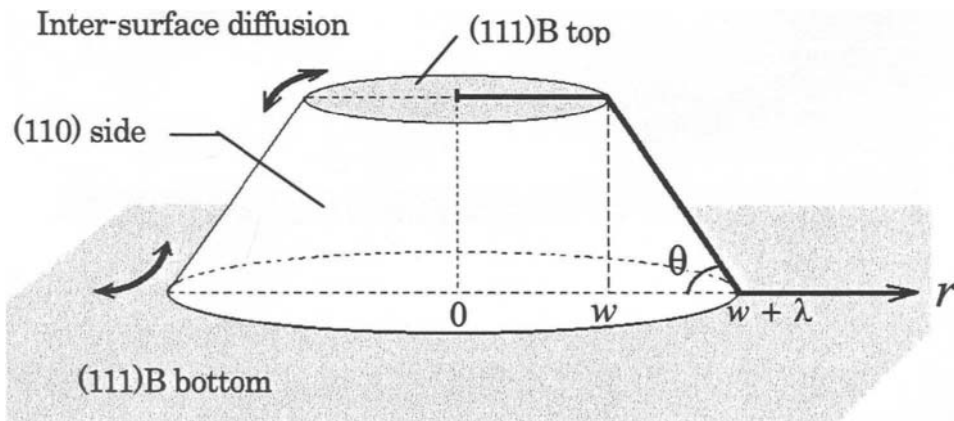
Figure 9 Coordinates used in the calculation. The truncated pyramid was approximated as a cone.

denoted by the suffices of 'top', 'bott' and 'side', respectively. The equations of the surface diffusion can be given as

$$D_s^{top} \frac{d^2 N^{top}}{dr^2} + \frac{D_s^{top}}{r} \frac{dN^{top}}{dr} - \frac{N^{top}}{\tau_{inc}^{top}} + J_{Ga} = 0$$

$$(0 \le r \le w),$$

$$D_s^{side} \frac{d^2 N^{side}}{ds^2} + \frac{D_s^{side}}{s} \frac{dN^{side}}{ds} - \frac{N^{side}}{\tau_{inc}^{side}} + J_{Ga} \cos\theta = 0$$

$$(w/\cos\theta \le s \le w/\cos\theta + l)$$

and

$$D_s^{bott} \frac{d^2 N^{bott}}{dr^2} + \frac{D_s^{bott}}{r} \frac{dN^{bott}}{dr} - \frac{N^{bott}}{\tau_{inc}^{bott}} + J_{Ga} = 0$$

$$(w + l\cos\theta \le r),$$

(7)

where N, τ, D_S, and θ are defined as the surface concentration of Ga adatom, its lifetime until incorporation into the crystal, surface diffusion coefficient and angle between (111)B substrate and {110} facet.

By denoting J_0 and Y_0 respectively as Bessel and Neumann functions, the Ga adatom concentration $N(r)$ or $N(s)$ on each surface can be given from above equations as[28]

$$N^{top}(r) = T_1 J_0\left(\frac{ir}{\lambda_{inc}^{top}}\right) + J_{Ga} \tau_{inc}^{top} \quad (0 \le r \le w),$$

$$N^{side}(s) = S_1 J_0\left(\frac{is}{\lambda_{inc}^{side}}\right) + S_2 \, \text{Re}\left[Y_0\left(-\frac{is}{\lambda_{inc}^{side}}\right)\right] + J_{Ga} \tau_{inc}^{side} \cos\theta$$

$$(w/\cos\theta \le s \le w/\cos\theta + l)$$

and

$$N^{bott}(r) = B_1 \text{Re}\left[Y_0\left(-\frac{ir}{\lambda_{inc}^{bott}}\right)\right] + J_{Ga} \tau_{inc}^{bott} \quad (w + l\cos\theta \le r),$$

(8)

under the boundary conditions of

$$D_s^{top} \frac{dN^{top}(r)}{dr} = 0 \quad (r = 0)$$

and

$$D_s^{bott} \frac{dN^{bott}(r)}{dr} = 0 \quad (r = \infty),$$

(9)

where T_1, S_1, S_2 and B_1 are all integral constants and eq.(9) means that there is no lateral flux of Ga adatoms at the center of the top surface because of a symmetry and at the distance far from the pyramid on the bottom surface.

In the present model, we assumed that there is no potential barrier for the surface diffusion across the boundary between (110) and (111)B, which gives the

Table 1 Parameters used in the calculation

	Experiment(dot)	Calculation(line)					
	P_{As4} [Pa]	$\lambda_{inc}^{top}, \lambda_{inc}^{bott}$ [μm]	τ^{top}, τ^{bott} [s]	D_S^{top}, D_S^{bott} [cm²/s]	λ_{inc}^{side} [μm]	D_S^{side} [cm²/s]	τ^{side} [s]
····●····	3.64×10^{-4}	9.9	0.098	1.0×10^{-5}	11.7	2.0×10^{-7}	6.9
—■—	6.38×10^{-4}	7.0	0.050	1.0×10^{-5}	8.3	2.0×10^{-7}	3.5
—◆—	8.64×10^{-4}	6.1	0.037	1.0×10^{-5}	7.2	2.0×10^{-7}	2.6
···▲···	1.16×10^{-3}	5.2	0.027	1.0×10^{-5}	6.2	2.0×10^{-7}	1.9
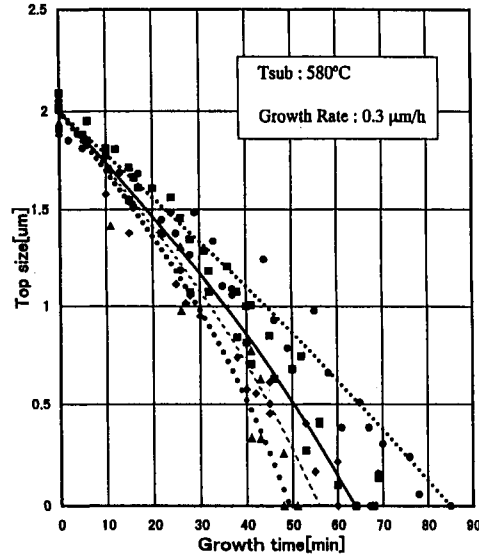
Figure 10 Time dependence of the top size of truncated pyramid for different As₄ pressures.

Marks and lines for each pressure are given in table 1.

boundary conditions as,

$$N^{top}(w) = N^{side}(w),$$

$$N^{side}(w/\cos\theta + 1) = N^{bott}(w + 1\cos\theta),$$

$$D_s^{top}\frac{dN^{top}(w)}{dr} = D_s^{side}\frac{dN^{side}(w/\cos\theta)}{ds}$$

and

$$D_s^{side}\frac{dN^{side}(w/\cos\theta + 1)}{ds} = D_s^{bott}\frac{dN^{bott}(w + 1\cos\theta)}{dr}$$

$$(10)$$

By using the eq.(8) with the boundary conditions of eq.(10), the Ga adatom concentration of each surface can be calculated. Also the growth rate $R(r)$ can be calculated

$$R^{top,bott}(r) = \frac{N^{top,bott}(r)}{\tau_{inc}^{top,bott}}$$

and

$$R^{side}(s) = \frac{N^{side}(s)}{\tau_{inc}^{side}}.$$

$$(11)$$

By repeating this procedure, we can obtain the top size of the truncated pyramid as a function of the growth time. Figure 10 shows the calculated results and the experimental data. Parameters used in the calculation were shown in table 1. In Figure 10, experimental top size means the length of one side of the top triangle, and the growth time of 0 was defined when top size takes the length of 2μm, because {110} side faceting was completed in all the experiments at this size. Incorporation diffusion length of (111)B and{110} surfaces were measured by Nishinaga et al.[21] and Yamashiki et al.[29] by microprobe-RHEED oscillation technique. From these experiments, surface diffusion length of (111)B and {110} were chosen respectively as 7.0μm for both λ_{inc}^{top} and λ_{inc}^{bott} at $P_{As4}=6.4\times10^{-4}$ Pa and 8.5μm for λ_{inc}^{side} at $P_{As4}=6.1\times10^{-4}$ Pa. Arsenic pressure dependence of the

(a) Ga adatom density[cm⁻³]

(b) Lateral flow of Ga adatom[cm-1·s-1]
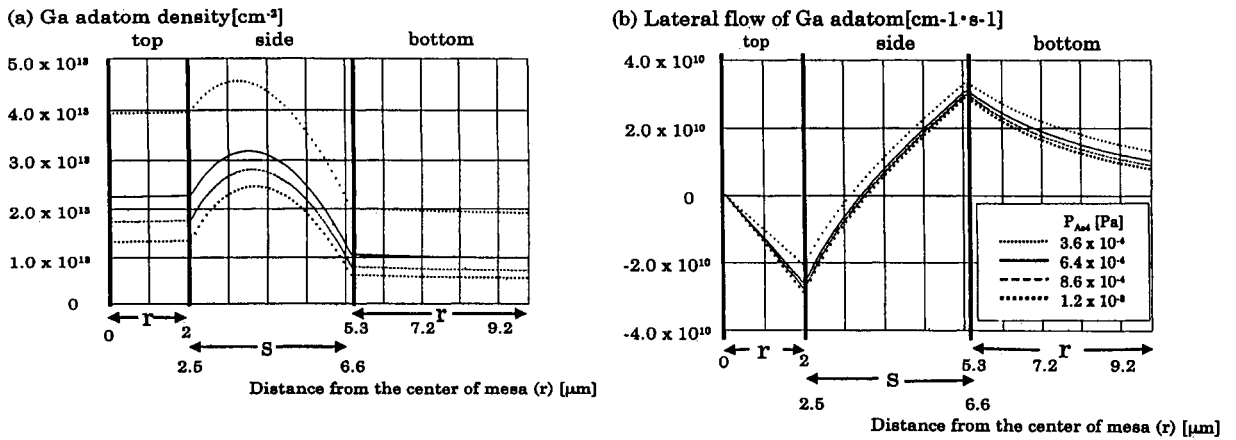
Figure 11 (a) Ga adatom density and (b) lateral flow of Ga adatoms, calculated from eq.(3), when the top size was 2μm.

The horizontal axis between r = 2(s=2.5) ~ r = 6.1(s = 6.6)μm is taken along the side of the cone.

surface diffusion length, $\lambda_{inc} \propto P_{As_4}^{-0.5}$ was employed for both surfaces[30].

On {110} facet, we employed diffusion coefficient of 1.4 x 10⁻⁶ cm²/s for D_S^{side}. The reason for the choice of this value is as follows. Yamashiki et al. [31] studied the intersurface diffusion of Ga between (001) and (110) facets and found the surface diffusion coefficient on (110) surface is nearly ten times larger than that on (001) surface. Although there is no reliable experimental data for the surface diffusion coefficient on (001), there is one theoretical work which gave the value of $D_S^{(001)}$ = 2.0 x 10⁻⁸ at 600°C in the direction which has the lowest diffusivity[32]. These two considerations gave the above value. Ga adatom lifetime τ_{inc}^{side} is given by eq.(1) and is shown in the table 1.

On (111)B surface, diffusion coefficients D_S^{top} and D_S^{bott} were determined so that calculation agrees with experiment. Ga adatom lifetimes τ_{inc}^{top} and τ_{inc}^{bott} were also obtained from eq. (1). Figure 10 shows good agreement between the calculation and the experiment when D_S^{top} and D_S^{bott} were chosen as 1.0 x 10⁻⁵ cm²/s. From this calculation, the lifetime and the diffusion coefficient of (111)B substrate were estimated respectively as 1/70 as small and 50 times as large as that of {110} side wall. The value of $D_S^{(111)B}$ = 1.0 x 10⁻⁵ cm²/s which shows the best fitting seems very large compared to the value on (001). Since the surface atomistic configuration is quite different to each other, we cannot eliminate the possibility for (111)B surface to have such high diffusion coefficient. However, as mentioned before, we assumed

there is no potential barrier between (110) and (111)B surfaces. The large value of $D_S^{(111)B}$ may include the error which comes from this assumption.

Figure 11(a) and (b) show respectively Ga adatom density and the lateral flow of Ga adatoms calculated from eq.(8) when top size was 2μm. As shown in Figure 11(a), Ga adatom density takes the maximum in the side wall and the maximum values become relatively smaller when the arsenic pressure is increased. Due to the geometrical discontinuities, the lateral flow shows discontinuities in its gradient at the top and bottom boundaries as shown in Figure 11(b). It is seen in the figure, as arsenic pressure is increased, the lateral flow toward the bottom is decreased while the flow toward the top is increased. On the other, hand when the arsenic pressure is decreased, Ga adatom can diffuse longer, so that the flow toward the bottom is increased and this makes the amount of the flow toward the top small. Basing on this model, we can explain why the mesa shrinks faster when arsenic pressure is higher as shown in Figure 10. This also suggests that the inter-surface diffusion model is valid for understanding the present experiments.

6. MICROCHANNEL EPITAXY BY LOW ANGLE INCIDENCE BEAM MBE

It was found that by sending Ga and As$_4$ fluxes with
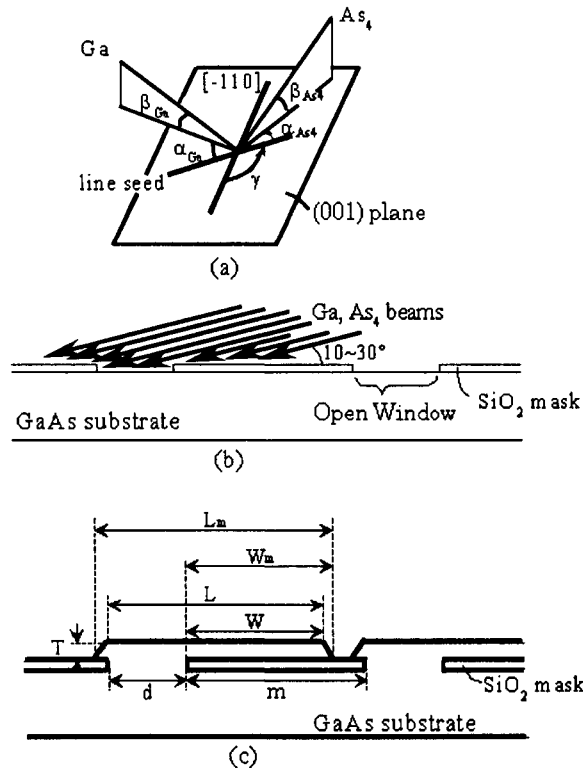
Figure 12 Schematic illustration of the low angle incidence beam microchannel epitaxy(LAIMCE). (a)angles employed in LAIMCE, (b) cross section of the substrate before growth and incident fluxes and (c) the cross section after the growth.

Table 1 Results of experiments with d + m = 3.3 μm, β_{Ga} and β_{As4} of $11°$ and $23°$ respectively.

epilayers alignment	LAIMCE parameters			figure of merit	
	α_{Ga}	α_{As4}	γ	W_m (μm)	W_m/T ratio
[-110]	$34°$	$11°$	$0°$	0.62	7.3
[110]	$56°$	$79°$	$90°$	poor morphology	
[010]	$79°$	$56°$	$-45°$	0.73	8.4
[010]+$10°$	$69°$	$46°$	$-35°$	1.45	22.3
[100] -$10°$	$1°$	$22°$	$35°$	1.29	19.8

low angle with respect to substrate in MBE, one can grow thin layer of GaAs over insulator film[33, 34]. This technique is called as LAIMCE(low angle incidence beam microchannel epitaxy). The schematic illustration of this technique is given in Figure 12. The definitions of various angles are shown in the figure (a) and the cross sections of the substrate and Ga/As$_4$ fluxes before and after the growth are given in (b) and (c) respectively. It was also found by sending the fluxes with low angles that the selectivity is improved largely so that the selective area epitaxy becomes easier even by MBE under normal growth conditions. In Figure 12, d shows the window (or microchannel) width, while W does that of the lateral grown part of the epilayer and T denotes its thickness. In some cases the value of W is difficult to be defined since the fronts of the lateral growth have various shapes. Therefore, we have defined L_m and W_m, respectively, as the maximum bottom width and the maximum width of the lateral grown part of the epilayer. Figures of merit of the epilayers are W_m and the W_m/T ratio. For device fabrication, large W_m is required and a greater value of the W_m/T ratio means the larger dislocation free area when this technique is applied to heteroepitaxy systems.

The growth conditions employed in the experiments were a substrate temperature, an As$_4$ pressure, and a growth rate of 610℃, 3.8×10^{-5} Torr and 1.08 μm/h, respectively. The LAIMCE angles β_{Ga} and β_{As4} were $11°$

and $23°$, respectively. The 1.08 μm/h growth rate was calculated from the period of the intensity oscillations of the reflection high-energy electron diffraction (RHEED) specular beam and corresponds to the growth rate of GaAs on GaAs (001) when Ga beam is perpendicular to the substrate. With the angle of Ga flux employed in LAIMCE, the horizontal and vertical growth rate were calculated as 0.21 μm/h and 1.06 μm/h respectively from simple geometrical consideration. The horizontal growth rate of 1.06 μm/h is calculated for an open window line seed perpendicular to the Ga flux ($\alpha_{Ga} = 90°$ in Figure 12a). The morphology and the shape of the grown layers were investigated by AFM picture and cross-section analysis.

We have employed an ULVAC MBC-300 MBE system for the growth. After organic chemical cleaning and chemical etching by 25%NH$_4$OH: 40%H$_2$O$_2$: H$_2$O= 4:1:20, a SiO$_2$ film was deposited on the GaAs (001) substrate by spinning of an organic solution (OCD, Tokyo Ohka Kogyo Co. Ltd.) and baking at 500℃. The oxide thickness was estimated by AFM to be 56 nm. The typical values of d and m were respectively 1 and 2.3 μm.
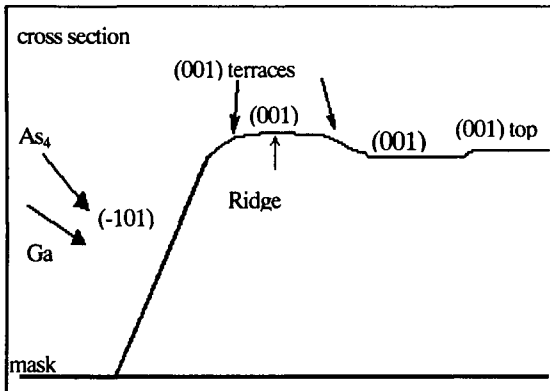
The results are given in Table 1[35]. When the microchannel is aligned in low index orientations such as [-110], [110] and [010], the fronts of lateral growth are covered by facets such as (111)A, (111)B and (-101) respectively. Once, the front is covered by a facet, the velocity of lateral growth is reduced and one can not get

124

Section Analysis

(a)
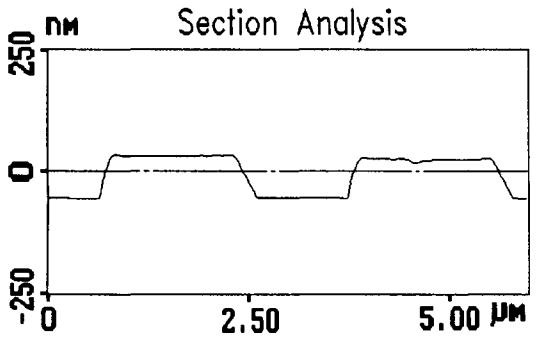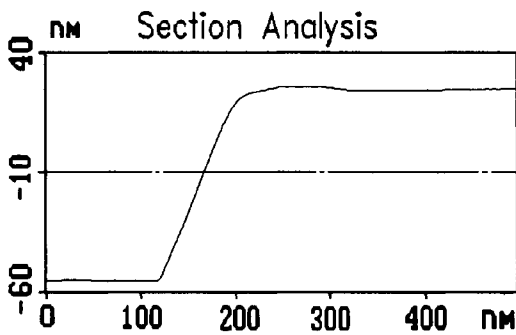
Section Analysis

(b)

(c)

Figure 13 AFM cross section analysis of the epilayer grown on the microchannel aligned along [010] direction. $d = 1$ μm, $d + m = 3.3$ μm. (a) cross section of the epilayer; (b) cross section of the left Ga rich sidewall of the epilayer shown in (a); (c) schematic of (b). $T = 87$ nm, $W_m = 730$ nm, $W_m/T = 8.4$.

high value of W_m/T ratio. Figure 13 shows the AFM cross sectional analysis for the microchannel orientation of [010]. It is seen in (c) that the front is covered by (-101) facet. As the direction of the microchannel is misoriented from [010], the side surface of the MCE layer becomes atomically rough and it serves as a sink of Ga adatoms that arrive on the MCE top surface. The intersurface diffusion of Ga adatoms from top surface to the side surface enhances the lateral growth and makes the value of W_m/T ratio high.

[010]+10° and [100]-10° are equivalent in crystallographic orientation. It is clearly seen from Table 1 that a difference in α_{Ga} and α_{As4} results in a difference in figures of merits between [100]-10° alignment and [010]+10° alignment. Since the [100]-10° direction is equivalent crystallographically to the [010]+10°, differences in W_m and W_m/T ratio between two alignments should be derived from the difference in the mass transport
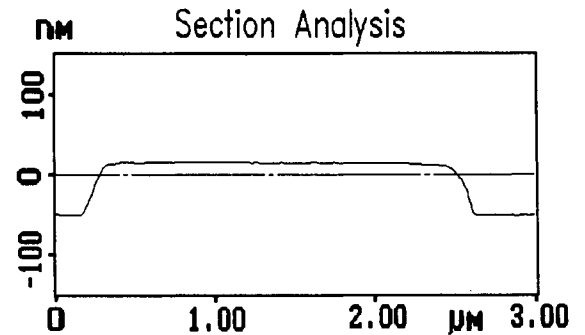
Section Analysis

Figure 14 AFM cross section analysis of the epilayer grown on the open window aligned along [100]-10° direction. $d = 1$ μm, $d + m = 3.3$ μm. $T = 65$ nm, $W_m = 1.29$ μm, $W_m/T = 19.8$.

Figure 15 3D-AFM image of the epilayer grown from microchannel aligned along [010]+10° direction.
$d = 1\ \mu$m, $d + m = 3.3\ \mu$m., $T = 65$ nm, $W_m = 1.45\ \mu$m, $W_m/T = 22.3$.

to the sidewalls. Table 1 shows that the value of W_m is 12% bigger for the [010]+10° alignment than that for the [100]-10° alignment although the direct Ga beam incidence is, from the geometrical calculation, up to 50 times larger on the sidewall of the epilayer of [010]+10° alignment, than that of [100]-10° alignment. The Ga atoms directly impinging on the sidewall do not contribute to lateral growth as much as concluded by the geometrical calculation. Therefore, intersurface diffusion is dominant in the growth process of LAIMCE.

Figure 14 shows AFM cross-section analysis of the epilayers grown from the microchannel aligned along [100]-10°. The formation of the round shaped sidewall and (001) terraces around the edge indicates that the sidewall and the edge of the epilayer are atomically rough. The

atomically rough sidewall enhances lateral growth and this explains the discrepancy in the W_m and W_m/T ratio between the epilayers with faceted sidewalls and the round shaped sidewall.

Figure 15 shows the 3D-AFM image of the epilayer grown from the microchannel aligned along [010]+10° direction. It is seen that the side surface is round and atomically rough. In this case, the thickness of the epilayer is 65 nm and W_m is 1.45 μm, which gives the value of W_m/T as 22.3. This ratio is extremely large even compared with LPE-MCE, which gives typically the ratio of 17.5[36]

7. ELIMINATION OF THE GROWTH ON (111)B FACET ADJACENT TO (001) FACET

126

Under the growth condition that the intersurface diffusion occurs from (111)B to (001), one can eliminate the growth on (111)B by decreasing the Ga flux so that the concentration of Ga adatoms is kept below the critical concentration for the 2D-nucleation[37-39]. To keep the Ga adatom concentration below the critical value there are several methods. The first one is directly to decrease the Ga flux on (111)B by adjusting the incident angle of the flux and the second is to rotate the substrate with (111)B side and (001) top configuration. The third one is to keep the width of (111)B facet small so that the accumulation of Ga adatoms is not enough to start the nucleation. By switching the growth of (111)B on and off while keeping (001) growth continuously on one can fabricate the special structures for devices[39].

8. CONCLUSIONS

Elementary growth process of epitaxy is described taking molecular beam epitaxy of GaAs as an example. First, two different definitions of the surface diffusion length were explained. Among them the diffusion length of adatom incorporation is much more important in MBE of GaAs and we studied its dependence on arsenic pressure in detail. Then, the surface diffusion between two surfaces, which we call inter-surface diffusion are described. It was shown that the direction of intersurface diffusion between (111)B and (001) is changed twice as arsenic pressure is increased. This was explained in terms of the different arsenic pressure dependence of Ga incorporation lifetime on (111)B and (001) and the change of surface reconstruction of (111)B. Basing on these experiments the elementary growth process of epitaxy is discussed.

The understanding of the elementary growth process allows us to make use of this knowledge to fabricate microstructures. It was confirmed that by changing arsenic

pressure, the top shape of pyramid can be controlled. Namely, by changing the arsenic pressure one gets flat top pyramid from sharp top pyramid or vise versa. We demonstrated how to control the top size of pyramids by adjusting the arsenic pressure.

Finally, we described the new technique of GaAs MBE to grow a thin layer of GaAs single crystal over a SiO_2 film. This technique has been called by us as LAIMCE(Low angle incidence microchannel epitaxy). By optimizing the growth parameters it was possible to grow GaAs layers of 65 nm thick and 1.45 μm wide on SiO_2 layer, which gives the width to the thickness ratio of as large as 22.3.

REFERENCES

1. E.Kapon, D.M.Hwang and R.Bhat, Phys. Rev. Lett. 63(1989)430.

2. X.Q.Shen, M.Tanaka, K.Wada and T.Nishinsga, J. Cryst.Growth 135(1994) 85.

3. S.Koshiba, H. Noge, H.Akiyama, T.Inoshita, Y. Nakamura, A.Shimizu, Y.Nagamune, M.Tsuchiya, H.Kano, H.Sakaki and K.Wada, Appl. Phys. Lett. 64(1994)363.

4. Y.Nakamura, S.Koshiba, M.Tsuchiya and H.Sakaki, Appl. Phys. Lett. 59(1991)700.

5. A.Madhukar, K.C.Rajkumar and P.Chen, Appl. Phys. Lett.62(1993)1547.

6. S.Tsukamoto, Y.Nagamune, M.Nishioka and Y.Arakawa, Appl. Phys. Lett. 62(1993)49.

7. S.Tsukamoto, Y.Nagamune, M.Nishioka and Y.Arakawa, J. Appl. Phys. 71(1992)533.

8. C.S.Tsai, J.A.Lebens, C.C. Ahn, A. Nouhi and K.J.Vahala, Appl. Phys. Lett.60(1992)240.

9. T.Fukui and S.Ando, Electron. Lett. 35(1989)410.

10. T.Fukui, S.Ando and Y.Tokura, T.Toriyama, Appl. Phys. Lett. 58(1991)2018.

11. M.Tabuchi, S.Noda and A.Sakaki, in: S.Namba, C.Hamaguchi, T.Ando(eds.), Science and Technology of Mesoscopic Structures, Springer, Tokyo, 1992, p.379.

12. D.Leonard, M.Krishnamurthy, C.M.Reaves, S.P.Denbaars and P.M.Petroff, Appl. Phys. Lett.,63(1993)3203.

13. T.Nishinaga, I.Ichimura and T.Suzuki, in I. Ohdomari, M.Oshima and Hiraki(Eds.), Control of Semiconductor Interfaces, Elsevier, Amsterdam, 1994, p.63.

14. M.Hata, T.Izu, A.Watanabe and Y.Katayama, Appl. Phys. Lett.56(1990)2542.

15. M.Hata, T.Isu, A.Watanabe and Y.Katayama, J. Vac. Sci. Technol. B8(1990)692.

16. T.Suzuki and T.Nishinaga, J.Cryst. Growth 142 (1994)49.

17. X.Q.Shen, H.W.Ren, M.Tanaka and T.Nishinaga, J.Cryst. Growth 169(1996)607.

18. X.Q.Shen, H.W.Ren and T.Nishinaga, J.Cryst. Growth 177(1997)175.

19. X.Q.Shen and T.Nishinaga, Jpn. J. Appl. Phys. 32(1993)L1117.

20. X.Q.Shen, D.Kishimoto and T.Nishinaga, Jpn. J. Appl. Phys. 33(1994)11.

21. T.Nishinaga, X.Q.Shen and D.Kishimoto, J. Cryst. Growth 163(1996)66.

22. A.Yamashiki and T.Nishinaga, Cryst. Res. Technol. 32(1997)1049.

23. T.Nishinaga, A.Yamashiki and X.Q.Shen, Thin Solid Films 306(1997)187.

24. C.T.Foxon and B.A.Joyce, Surf. Sci. 44(1974)69.

25. C.T.Foxon and B.A.Joyce, Surf. Sci. 50(1975) 434.

26. C.T.Foxon and B.A.Joyce, Surf. Sci. 64(1977) 293.

27. S.Kousai, A.Yamashiki, T.Ogura and T.Nishinaga, J. Cryst. Growth 198/199(1999)1119.

28. H.Takarabe, Bachelor Thesis, Department of Electronic Engineering, University of Tokyo (1996).

29. A.Yamashiki and T.Nishinaga, J. Cryst. Growth 198/199(1999)1125.

30. T.Nishinaga and A.Yamashiki, Thin Solid Films 343-344(1999)495.

31. A.Yamashiki, Doctor Thesis, Department of Electronic Engineering, University of Tokyo(1999).

32. T. Ohno, K. Shiraishi and T. Ito, Mat. Res. Soc. Symp. Proc. Vol.326(1994) 27.

33. G.Bacchin and T.Nishinaga, J. Cryst. Growth 208

(2000) 1.

34. G.Bacchin, A.Umeno and T.Nishinaga, Appl. Surf. Sci., 159(2000)270.

35. A.Umeno, G.Bacchin and T.Nishinaga, J. Cryst. Growth 220(2000)355.

36. Y.S.Chang, S.Naritsuka and T.Nishinaga, J. Cryst. Growth, 192(1998)18.

37. D.Kishimoto, T.Noda, Y.Nakamura, H.Sakaki and

T.Nishinaga, J. Cryst. Growth 209(2000)591.

38. D.Kishimoto, T.Nishinaga, S.Naritsuka, T.Noda, Y.Nakamura and H.Sakaki, J. Cryst. Growth, 212(2000) 373.

39. D.Kishimoto, T.Ogura, A.Yamashiki, T.Nishinaga, S.Naritsuka and H.Sakaki, J. Cryst. Growth, 216(2000)1.

Atomistic simulation of epitaxial growth processes

Tomonori Ito

Department of Physics Engineering, Mie University,
1515, Kamihama, Tsu-shi, Mie 514-8507, Japan

Recent progress in computer-aided simulation of epitaxial growth is reviewed in the area of semiconductor materials. Reliable predictions can now be made for a wide range of problems, such as static surface reconstructions and adatom migration potentials, using *ab initio* calculation methods such as pseudopotential method. In order to apply the *ab initio* calculation methods to epitaxial growth, a simple system energy formula is proposed. Using this energy formula, quantum mechanical simulation of epitaxial growth is carried out with Monte Carlo method. Predictability of the simulation is examined by investigating epitaxial growth processes of GaAs such as migration and adsorption behavior on GaAs surfaces. An overview of these issues is provided and the latest achievements are presented to illustrate the capability of the atomistic simulation of epitaxial growth by comparing experimental results.

1. INTRODUCTION

During the last decade, steady improvements in epitaxial growth techniques, such as molecular beam epitaxy [MBE] and metal-organic vapor phase epitaxy [MOVPE], have provided scientists and engineers with opportunities to fabricate substantially improved devices and to prepare novel device structures. This leads semiconductor technology to nanotechnology which has made it possible to create artificial materials, whose physical properties are on the atomic-scale such as superlattices, quantum dots, layers of molecules and microclusters fabricated by various epitaxial growth techniques. Furthermore, we can create material phases with new physical properties that are not found in nature today. These situations have inspired a number of theorists in the field of computer-aided materials science, since a fundamental understanding of these techniques should eventually lead to better epitaxially grown thin films and growth techniques. As a physical process, the growth of thin films is controlled by various complex kinetic processes that involve adsorption of atoms or molecules onto surface, their subsequent diffusion across the surface, dissociation of molecules and evaporation, etc. Despite the remarkable advances in the development of experimental techniques for epitaxial growth and the vigorous theoretical attempts

to understand epitaxial growth, the kinetic processes are not well understood at the fundamental level.

Computer-aided materials design is now a popular goal among researchers in computational materials science. This reflects recent developments in the use of electronic theory to explain many experimental results and to predict the physical properties of materials without actually synthesizing them. In our previous paper, we reviewed the development of these computational methods and discussed a wide range of problems [1]. One such problem is the behavior of various atomic species in semiconductors, including the stability and band structure of superlattices, lattice defects, alloy systems and epitaxial growth. The relationship among these issues including input parameters, computational methods and various material properties is schematically shown in Fig. 1. Pseudopotential methods used in our study are typical examples of computational methods for electronic structure and total energy in this figure. This figure reveals that these issues are not independent of each other but are closely inter-related. Therefore, these various issues have to be systematically investigated in considering future prospects in computational materials science and computer-aided materials design as its goal. Among them, computational methods at finite temperatures such as molecular dynamics and Monte Carlo methods are particularly
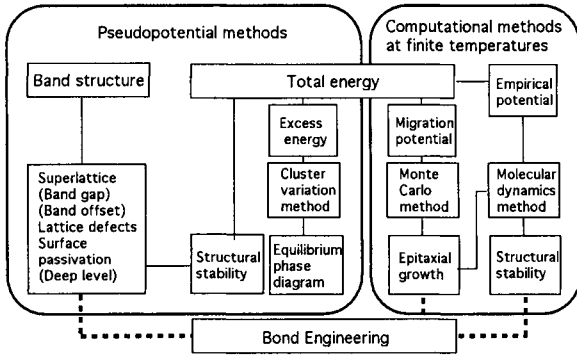
Figure 1. Relationship among computational methods and various material properties

important for predicting material properties, since materials are processed in the fabrication of thin films at high temperatures.

In this article, we will provide an overview of the theoretical-computational methods and our latest achievements for atomistic simulation of epitaxial growth processes of semiconductors such as GaAs based on quantum mechanics. Section 2 briefly reviews various computational methods such as pseudopotential methods and empirical interatomic potentials. Molecular dynamics and Monte Carlo methods are also shown as computational methods at finite temperatures. Applicability of these methods to epitaxial growth processes is exemplified by adatom migration on a GaAs(001) surface. It should be noted that Monte Carlo method combining with *ab initio* calculations is a powerful tool to investigate dynamic change in atomic configurations on the surface at finite temperatures. Section 3 gives a quantum mechanical approach for understanding surface reconstructions of GaAs(001) and elemental growth processes on the surface such as migration potentials, adsorption and desorption behavior. A simple energy formalism and newly developed electron counting Monte Carlo simulation are introduced in section 4. This approach is used to investigate migration potentials and initial growth processes on various GaAs surfaces. Our simple approach can give good estimates of the migration potentials of Ga adatoms on the surfaces with various orientations and undulations such as steps or As-dimer kinks. Finally, future prospects in computational materials science for semiconductors are discussed by introducing a recent achievement to epitaxial growth in section 5.

2. COMPUTATIONAL METHODS

Numerical computational methods can be categorized in two ways: (1) as static or dynamic calculation methods and (2) as *ab initio* [non-empirical or first-principles], semi-empirical and empirical methods. Static calculation is used to evaluate the static energy in a system without any dynamical contributions such as atomic diffusion, while dynamic calculation methods such as the molecular dynamics method can deal with atomic movements as functions of time and temperature. Alternatively, one can use the criterion of whether the method uses parameters adjusted to reference data or not, to classify methods into *ab initio* [non-empirical or first-principles], semi-empirical, or empirical methods. The number of input parameters required in the calculation becomes smaller in the order of empirical, semi-empirical and *ab initio* calculation methods, while the size of system which can be treated typically becomes smaller as the calculation approaches *ab initio* but this also depends critically on the properties.

According to these classifications, a non-empirical or *ab initio* dynamic method with only a few fundamental parameters should be ideal as a tool for computational materials science, since it can simultaneously predict various thermodynamic properties and electronic properties as functions of time and temperature. For example, Car and Parrinello have reported an *ab initio* molecular dynamics method that incorporates dynamic treatments non-empirically and gives satisfying results [2]. However, the *ab initio* molecular dynamics method requires very time-consuming numerical computations, which limits most of its applications to simulated annealing. Therefore, considering the capabilities of present computers and relative availability of the computational methods, static and dynamic methods based on *ab initio*, semi-empirical and empirical methods should be properly used to investigate individual problems in the materials science as complementary computational tools.

In this section, we outline the various computational methods commonly used in the field of materials science in semiconductors such as *ab initio* calculation methods based on electronic theory, pseudopotential methods, empirical interatomic potentials applicable to dynamic calculations and dynamical methods at finite temperatures such as molecular dynamics and Monte Carlo methods. A prototypical

application of these computational methods to atomistic simulation of epitaxial growth is exemplified by adatom migration of Ga on a GaAs(001) surface, where Monte Carlo simulation is combined with pseudopotential calculations.

2.1. *Ab initio* calculation methods

The development of the *ab initio* calculation is closely related to that of energy band structure computational methods. These are the orthogonalized plane waves [OPW], the pseudopotential, the tight-binding or the linear combination of atomic orbitals [LCAO], the Green function [KKR], linearized augmented plane waves [LAPW] and the linear muffin-tin orbitals [LMTO] methods [3-10]. All of these *ab initio* methods are based on the local density functional formalism laid down by Hohenberg, Kohn and Sham, which provides a powerful technique for electronic structure and total energy calculations [11,12]. The LAPW method was extensively applied by Hamann and coworkers [13-17]. They investigated the electronic and structural properties of various materials such as BaPbBiO, Cr, Mo, W, WC and As. Zunger *et al.* also used the LAPW method to evaluate thermodynamic properties for various semiconducting alloys and to predict the phase stability of those alloy systems[18,19]. The self-consistent FLAPW method is described in detail in a series of papers by Freeman *et al.* [20-22]. The self-consistent LMTO method has also been successfully used to study the electronic and ground-state properties of semiconductors. Lambrecht and Andersen have demonstrated that ground-state properties of C, Si and Ge can be accurately calculated using this method [23]. Christensen *et al.* have extensively applied the LMTO method to heterostructures and lattice defect formation in semiconductors [24,25].

In addition to these calculation techniques, the pseudopotential approach is a promising candidate as a tool for computational materials science because of its ability to calculate energy band structures and total energy. Ihm *et al.* developed a method for calculating the total energy within pseudopotential formalism combined with an *ab initio* pseudopotential [26]. Using this momentum space formalism for the total energy, Yin and Cohen showed that the crystal stability and pressure-induced phase transformations in Si, Ge and C can be accurately described [27]. Using these pseudopotential calculation methods, various properties of the combination of

elemental or compound semiconductor materials such as superlattices, quantum dots, lattice defects and alloy systems have been examined by many researchers [28-46]. Zunger *et al.* successfully developed electronic-structure theory of semiconductor quantum dots within a pseudopotential framework and applied it to Si, InP, InAs/GaAs, CdSe, GaAs/AlAs and InP/GaP nanostructures [47,48]. Further successful applications of these calculation methods to surface-related properties such as reconstruction and adatom migration potential were also demonstrated [49-52]. Moreover, GaN as promising material for optical devices was highlighted not only by experimentalists but also by a number of theorists. Lattice defects, surface and interface-related properties were extensively investigated by the *ab initio* calculations. Various calculation techniques were used to predict material properties [53-64].

In this article, we mainly focus on the *ab initio* pseudopotential method. The total crystal energy for semiconductors based on the *ab initio* pseudopotential method [26] is given by

$$E = E_i + E_{kin} + E_{e-i} + E_H + E_{xc}, \qquad (1)$$

where E_i, E_{kin}, E_{e-i}, E_H and E_{xc} are the electrostatic energy, electron kinetic energy, electron-ion interaction energy, Hartree energy, and exchange-correlation energy, respectively. Using the *ab initio* pseudopotential method, extensive properties and a complete tabulation of the pseudopotentials were shown by Bachelet, Hamann and Schlüter [65]. Moreover, the fully separable *ab initio* pseudopotentials proposed by Kleinman and Bylander are often used to reduce computing time [66]. The kinetic energy cut-off value in the calculation using the separable pseudopotentials has to be carefully chosen so as not to produce ghost bands, according to the criterion proposed by Gonze *et al.* [67,68]. The *ab initio* pseudopotential method is used in the calculations of both total energy and band structure.

2.2. Empirical interatomic potentials

Along with the development of the *ab initio* computational methods, empirical interatomic potentials for semiconductors have also been proposed by many authors for application to dynamic treatments such as molecular dynamics method and for use in the simulation of complex systems with a large number of atoms. This is because application of the *ab*

initio methods to the complex systems is currently limited by the large computational effort. Although the theory of two-body interatomic potentials for rare gas systems, ionic systems and metals is fairly well established, the theory of covalent solids is less developed because of the complexity of stabilizing the open tetrahedral structure of semiconductors, which requires three-body potentials instead of simple two-body pairwise potentials such as the Lennard-Jones and Morse potentials. The oldest empirical three-body potential for diamond or zinc blende structured semiconductors is the valence-force potential such as the Keating model which has been used with considerable success for studying phonons, elastic properties and excess energies of semiconductor alloy systems [69,70]. However, it is perturbative in nature and cannot properly be applied to systems with large distortion such as various crystal structures, defect structures, surfaces and melting.

To treat these more general structures, Pearson *et al.* treated the pairwise interaction by a Lennard-Jones term and the three-body term by the Axilrod-Teller potential for van der Waals interactions of three bodies [71]. Furthermore, Stillinger and Weber proposed an empirical interatomic potential, incorporating a three-body potential to model the melting of silicon [72]. Neither of these potentials, however, attempt to describe accurately the properties of non-tetrahedral forms of Si. Biswas and Hamann proposed a more general form for the three-body interaction and determined the eighteen parameters in that model by a least-squares fit to a large amount of data for calculated energies of real and hypothetical Si structures [73]. The resulting potential works well on bulk elastic properties and the high-density polymorphs, which make up the database.

Another approach closer to that outlined by Abell [74], who noted that cohesive energies can be modeled by pairwise interactions moderated by the local environment in attempting to explain the universal cohesive energy curves for various materials, has been implemented for Si by Tersoff [75,76]. This potential does a good job of describing elastic properties of bulk Si and a reasonable job of determining the energies of many defect structures and surface reconstructions, as well as low coordination number geometries. Kelires and Tersoff applied this potential to the Si-Ge binary system in conjunction with the Monte Carlo technique and successfully described lateral ordering in thin films [77]. Further attempts to

use Si empirical potentials for various properties, such as microcluster formation [78-83], surface reconstruction [84-87] and migration potential on Si surface [88-93], were made by many researchers. However, very few studies have applied empirical interatomic potentials to compound semiconductors.

Based on Abell's idea, Khor and Das Sarma noted that two basic relationships exist for elemental semiconductors [94]: that between the equilibrium lattice spacings r_e of various phases and the coordination number Z, and that between the corresponding cohesive energies D_e and Z. Ito *et al.* assumed similar rules exist for other materials and developed empirical potentials for cohesive energy calculations for the Si-Ge binary system, compound semiconductors and their binary systems [1,95,96]. Structural stability and epitaxial relationship to various thin films including metals were also clarified using this empirical interatomic potential [97-101]. The universal empirical potential is given by:

$$V_{ij} = A\exp[-\beta(r_{ij}-R_i)^\gamma][\exp(-\theta r_{ii}) -B_0\exp(-\lambda r_{ij})G(\eta)/Z_i^\alpha], \qquad (2)$$

where r_{ij} is the distance between the atoms, $Z_i =\Sigma_j\exp[-\beta(r_{ij}-R_i)^\gamma]$ gives the effective coordination number of atom i, R_i is the minimum distance between neighbors and $G(\eta)$ is the bond bending term. The potential parameters A, B_0, θ, λ, α, β, γ and η were determined using the cohesive energy, the bulk modulus, shear stiffness and the relative stability between zinc blende and rocksalt structures obtained by *ab initio* calculations and experiments [1].

2.3. Methods at finite temperatures

In semiconductor technology, semiconducting materials are processed at high temperatures in the fabrication of thin films and devices. Thus, their fabrication involves the formation of thin films, alloys, defects and related material properties at high temperatures, since atomic motion or atomic exchange should be taken into account. The usual way to incorporate atomic motion and atomic exchange is to use molecular dynamics [MD] or Monte Carlo [MC] methods. Molecular dynamics simulation yields the motion of a given number of atoms governed by their mutual interatomic interactions. The first molecular dynamics calculations were made by Alder and Wainwright for systems of hard-core particles [102-103] and by Rahman for systems with

continuous interaction potentials [104]. Since then molecular dynamics has been used extensively in the field of condensed matter physics, statistical mechanics and chemical physics to probe the subtle relationships that exist between the observable properties of matter and the motion of individual atoms or molecules [105].

In molecular dynamics, the classical equations of motion for an assembly of N interacting particles are solved.

$$m_i dv_i/dt = F_i = -\nabla_{ri} V(r_1, r_2, ..., r_N), \qquad (3)$$

where r_i and v_I are the positions and velocities of the particles. In Eq. (3), F is the total force acting on the i-th particle and $V(r_1, r_2, ..., r_N)$ is the average potential on the i-th particle due to all the other particles. Once the potential $V(r_1, r_2, ..., r_N)$ has been specified, the equations of motion can be numerically integrated. The output of the computation is a trajectory in phase space at discrete times t_n separated by a time increment $\Delta t = t_{n+1} - t_n$. From this trajectory, various thermodynamic quantities can be calculated.

A Monte Carlo method generally involves the use of random sampling techniques to estimate averages or to evaluate integrals. The Monte Carlo method introduced by Metropolis et al. is a very efficient and important sampling technique [106]. For a constant density simulation, a system of N particles is placed in an arbitrary initial configuration in a volume V, e.g., a lattice of a chosen crystal packing and of uniform density equal to the experimental density at temperature T; here N, V, and T are fixed. Configurations are generated according to the following rules: (i) select particles at random; (ii) select random displacements or random exchanges of particles; (iii) calculate the change in potential energy ΔU after displacing or exchanging the chosen particle; (iv) if ΔU is negative, accept the new configuration; (v) otherwise, select a random number h uniformly distributed over the interval (0,1); (vi) if $\exp(-\Delta U/kT) < h$, accept the old configuration; (vii) otherwise, use the new configuration and the new potential energy as the current properties of the system. This procedure is repeated for a suitable number of configurations in order to approach equilibrium configurations. The quantities that can be evaluated by the Metropolis Monte Carlo method are those that can be expressed as canonical averages of functions of pressure, energy and the radial distribution function.

The Metropolis Monte Carlo method, however, does not generate a true dynamical history of an atomic system in contrast to the molecular dynamics method, which yields the atomic motion. Thus, the Metropolis Monte Carlo method is suitable for achieving the equilibrium state but not for investigating non-equilibrium behavior such as adatom migration on the surface. The molecular dynamics method, however, limits the simulation duration to $t < 10^{-9}$ (s) because of its time-consuming numerical computations.

A kinetic Monte Carlo method based on the lattice-gas model is introduced to make up for these deficiencies in the Metropolis Monte Carlo and molecular dynamics methods. In the kinetic Monte Carlo simulation for atomic diffusion, individual atomic movements are simulated by the following procedure: (i) One first lists all the participating kinetic processes, and ascribes to each a corresponding rate of occurrence; (ii) The kinetic rates are assumed to be of Arrhenius form such as $R = R_0 \exp(-\Delta E/kT)$, where R_0 is the diffusion prefactor in hops per second and ΔE is the activation barrier or migration potential for the diffusion; (iii) Once these rates have been calculated, the simulation proceeds to activate the kinetic events corresponding to these rates and to follow the individual atomic movements. The kinetic Monte Carlo method enables the simulation to finish within a realistic time $t \approx 1$ (s) for epitaxial growth.

2.4. Application to epitaxial growth simulation

As discussed in the previous subsection, the molecular dynamics [MD] method can simulate kinetic processes rigorously without any approximations if the interatomic potential used in the simulation is reliable. However, a drawback of the MD approach is that the system size and time scale are much too small compared with actual epitaxial growth. On the other hand, the Monte Carlo [MC] approach can simulate fairly large systems and include certain aspects of the dynamics in the actual epitaxial growth time scale, but the necessary physical parameters need to be determined by other means, such as by experiment or by MD or ab initio calculations. Molecular dynamics simulation studies of epitaxial growth have been mainly limited to simple systems such as the hypothetical Lennard-Jones crystal and elemental semiconductors such as silicon because of their simple expression of interatomic

potentials. Schneider *et al.* performed the first full MD study of epitaxial growth based on a spherically symmetric Lennard-Jones type interatomic potential [107]. They extended their calculations to epitaxial growth of silicon on a Si(111) substrate using the interatomic potential proposed by Stillinger and Weber [108]. Taylor and Dodson simulated strained layer growth of a two-dimensional Lennard-Jones crystal lattice [109]. Two-component systems were investigated by Hara *et al.* [110].

Similar systematic epitaxial growth simulations using a Lennard-Jones potential were reported in a series of papers by Paik and Das Sarma [111-115]. Further MD simulations on a reconstructed Si surface were done by Ethier and Lewis for $Si_{1-x}Ge_x$ on Si(001)-2×1 surface [116], and Si and Ge on Si(001) incorporating surface steps by Roland and Gilmer [88-90] using the Stillinger and Weber potential. Moreover, Kitabatake and Greene investigated projectile incorporation and defect production resulting from 50-eV Si irradiation of (2×1)-terminated Si surface using Tersoff potential [117]. For the epitaxial growth of compound semiconductors, there have been very few MD simulations because of the complexity of the growth process involving different species such as cations and anions, and the difficulty of constructing of the interatomic potentials. Choi *et al.* investigated multilayer interfaces in the GaAs/AlAs system [118] and simulated GaAs cluster formation on GaAs, AlAs, and Si surface by evaluating cluster excess energies using their interatomic potentials in a static energy calculation [119]. Monte Carlo [MC] simulation has often been used in epitaxial growth simulations for compound semiconductors, particularly for GaAs.

The pioneering MC work investigating epitaxial growth was undertaken by Weeks, Gilmer and Jackson [120], and by Weeks and Gilmer [121]. They used the solid-on-solid [SOS] model, in which the crystal is described by an array of columns of atoms, with the requirement that every atom must possess a nearest neighbor immediately below it. Many aspects of crystal growth were simulated, including evaporation, migration, screw dislocations and the roughening transition at high temperatures. Based on the SOS model, Clarke and Vvedensky examined the kinetics of molecular beam epitaxial [MBE] growth by monitoring surface growth by calculating the evolution of the surface step density, which successfully mirrors the measurements of the temporal evolution

of specular reflection high energy electron diffraction [RHEED] intensities [122]. Using a similar approach, Shitara *et al.* investigated the transition from growth by nucleation of two-dimensional islands to step advancement on misoriented GaAs(001) surfaces during epitaxial growth in conjunction with RHEED experiments [123-125].

Atomistic simulation [specifically investigating the atomistic epitaxial growth of compound semiconductors] has mainly been performed by Madhukar and Ghaisas [126-128], and Shingh *et al.* [129,130]]. Early work focused on the temperature dependence of the growth front on GaAs(001). Das Sarma *et al.* performed similar atomistic MC simulation for $Si_{1-x}Ge_x$ and GaAs [131-133]. Lu and Metiu investigated the formation of an AlAs/GaAs tilted superlattice to simulate MBE growth without taking into account microscopic surface reconstruction [134]. Furthermore, Rockett, Farrell *et al.*, McCoy and Maksym, and Zhang *et al.* carried out MC simulations of MBE growth on Si(001) and GaAs(001), incorporating the effects of reconstruction in an explicitly atomistic fashion [135-138]. More recently, Grein *et al.* employed hybrid approach to the epitaxial growth of ZnSe on GaAs(001) based on MD to describe the initial kinetic behavior of deposited adatoms and MC displacements to account for subsequent equilibration [139]. In those works, however, there was some ambiguity in determining the adatom diffusion rate given by Arrhenius expressions because of the lack of experimental and theoretical data about the activation barrier.

2.5. *Ab initio*-based simulation

Considering these results from early studies and the advantages and disadvantages of each computational method, we believe that the fundamental processes, such as jump frequency and migration potential of adatoms, should be estimated by MD calculation or static calculation based on the *ab initio* calculation method. The MC method is suitable for investigating adatom migration on the surface and in epitaxial growth. In this subsection, we present our numerical study of kinetic processes that occur on a crystal surface combining *ab initio* pseudopotential calculations for the migration potentials with a kinetic Monte Carlo method. We limit the application of our *ab initio*-based MC approach to cation adatom migration across the GaAs(001) surface in the range of adatom coverage $\theta_{Ga} \leq 0.5$, because of the com-

plexity of As incorporation during the MBE growth that will be discussed in section 3.

The migration of adatoms on a semiconductor surface is one of the fundamental problems in semiconductor surface science. In particular, greater adatom migration is crucial in order to obtain high quality samples when using epitaxial growth techniques. This is because if an adatom migrates in a wide area, the surface becomes flat during the growth, and moreover, the synthesized interfaces become sharp. MBE is one approach that has successfully achieved high adatom migration [140]. Following this achievement, numerous experiments were carried out to clarify the mechanism of adatom migration during epitaxial growth. Neave *et al.* studied RHEED oscillations during MBE growth on a vicinal GaAs(001) substrate, where they estimated the activation barrier for the migration of a Ga adatom and discussed growth mode transition from two-dimensional growth to step flow growth [141]. Although RHEED oscillation measurements gave useful macroscopic information about adatom migration, they cannot individually provide intrinsic adatom migration or its relationship with microscopic surface reconstruction. To fill in this gap in the experiments, theoretical calculations are necessary to obtain microscopic information about intrinsic adatom migration.

Brocks *et al.* calculated potential energy surface of Si adatom migration on a Si(001) surface from the first-principles [40]. Their calculated activation barrier agrees well with the STM experiment by Mo *et al.* [142]. Adatom diffusion by orchestrated exchange on Ge(111) surface was proposed by Kaxiras and Erlebacher [143]. Oshiyama successfully applied *ab initio* total-energy calculations to step structures and the appearances of {311} facets on Si(001) surface, estimating the activation energy for diffusion of an adatom near a step edge [144]. A number of *ab initio* studies for Initial growth process of Si were reported [145-147]. More recently, novel migration mechanisms for Si adatoms on H-terminated Si(001) surface were also investigated by Nara, Sasaki and Ohno [148] and Jeong and Oshiyama [149]. Furthermore, Shiraishi successfully carried out *ab initio* pseudopotential calculations of the potential energy surface for Ga adatom migration on a GaAs (001) surface [52,150]. He found that the Ga adatom migration is expected to be very anisotropic. Kley, Ruggerone and Scheffler investigated diffusion
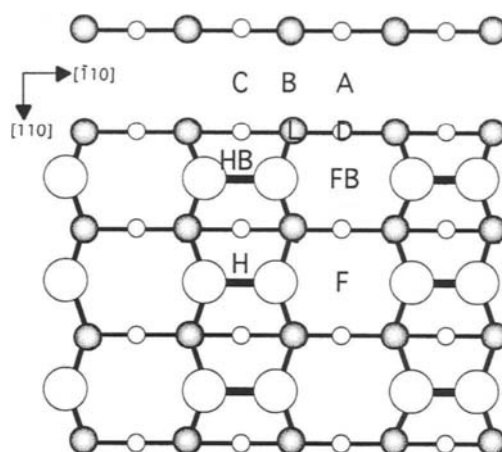
Figure 2. (4×4) surface unit cell for GaAs(001)-(2×4)β1 used in the calculation: adsorption sites are indicated by letters

channels of adatoms on GaAs(001)-(2×4)β2 surface to elucidate the strong diffusion anisotropy and the difference in diffusion barrier of Ga and Al adatoms [151].

However, the results do not treat over the entire coverage of adatoms on the surface but correspond only to high coverage, since the surface unit cells used in the studies were not large enough. To calculate the migration potential surface for various adatom coverages, Shiraishi *et al.* used the *ab initio* pseudopotential method on the basis of local density functional formalism [LDF] [152]. The surface migration potential $E(x,y)$ was calculated to optimize the z coordinate of an adatom and the positions of substrate atoms according to the Hellmann-Feynmann forces at the fixed (x,y) coordinates of an adatom. This potential $E(x,y)$ determines the activation energy of cation adatom diffusion on the GaAs(001) surface, which is essential in performing MC simulation. A (4×4) surface unit cell, which contains two (2×4) GaAs surface super structures, is used in the calculation as shown in Fig. 2, where the surfaces was investigated by increasing the number of adatoms in the surface unit cell. In this process, previously adsorbed adatoms are set at the most stable sites, and the migration potential of the adatom that arrives next is then calculated.

Figures 3(a) and 3(b) show the calculated migration potential surface of Ga adatoms at Ga coverage
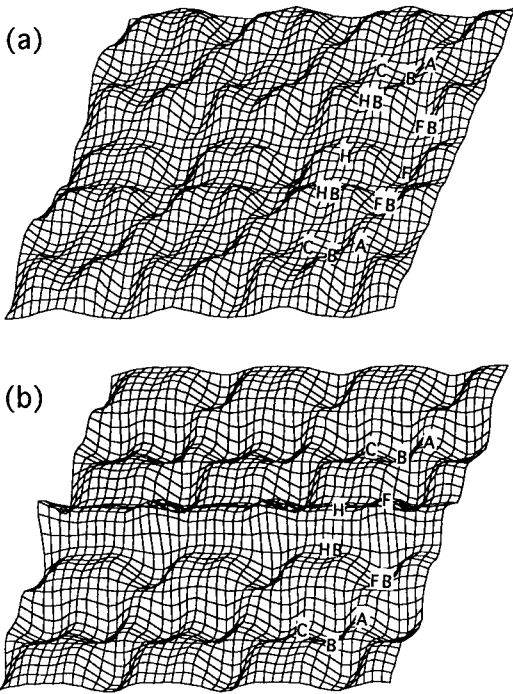
(a)

(b)

Figure 3. Calculated migration potentials for different Ga coverages: (a) θ_{Ga}=0.0625 and (b) θ_{Ga}=0.1875.

θ_{Ga} of 0.0625 and 0.1875, respectively. The most stable adsorption site at θ_{Ga} =0.0625 is found to be a long-bridge site [F site] which is located between two As-dimer rows. Missing dimer region sites [A, B, and C sites] in this figure are unstable. This shows that Ga adatoms tend to adsorb on the dimer region in the initial stage of epitaxial growth. The favorable migration path is along the [110] direction on the dimer region. As the number of adsorbed Ga adatoms increases [Ga coverage increases], the most stable site changes from the F site in the dimer region to the B site in the missing dimer region. This tendency can be recognized in Fig. 3(b), where the Ga bridge site in the missing dimer region [B site] becomes the most favorable at θ_{Ga} =0.1875. The B site is about 0.3 eV more stable than the second most favorable site [FB site]. From this figure, we can see that the most favorable migration path is along the missing dimer row. Once two long-bridge sites [F sites] are occupied, impinging Ga atoms tend to adsorb in the missing dimer rows and migrate through the missing dimer rows. These results imply that the coverage dependence of the migration potential is crucial in

investigating adatom migration on a GaAs(001) surface. Similar results were obtained for Al adatoms [153].

If we try to extend these results to the dynamical behavior of adatoms at high temperatures, the study is not amenable to a *ab initio* theoretical analysis at the present time due to the complexity of the migration process. Therefore, it is desirable to devise a numerical method that lets us gain some insight into the important factors involved in MBE growth. To this end, a direct kinetic MC approach is employed. A similar technique has been used by many researchers [126-133]. In contrast to the previous work, Ito *et al.* devised an MC approach that incorporates site-dependent activation barrier data obtained by the *ab initio* calculations.

In the kinetic MC simulation, one lists all the participating kinetic processes, and ascribes to each a corresponding rate of occurrence. These rates, assumed to be of Arrhenius form, are obtained from an energetics calculation obtained by the above procedure. The coverage of the first Ga layer can be recorded as a function of time, providing us with a growth profile. The following assumption were made in our MC simulation: (i) The adsorption of the atoms occurs at specific lattice sites with local energy minima including zinc blende lattice sites; (ii) At the temperature employed in our simulation, evaporation from the surface is negligible; (iii) Downward interlayer hop from the second layer to the first layer for Ga adatoms is included, whereas the diffusive processes in the upward z direction are ignored. An intralayer hop to an adjacent vacant site within the square sub-lattices and an interlayer hop downward to any one of the four vacant next nearest neighbor sites are considered; (iv) The growth profile is recorded for a lattice size 20×20 square sub-lattices. Periodic boundary conditions are imposed on the xy planes; (v) The activation barrier data, corresponding to local adatom coverage in a 5×5 cell around an adatom encountering kinetic event, are employed every moment during the simulation; (vi) The coverage-dependent activation barrier data at θ_{Ga}=0.0625, 0.125, 0.1875, 0.25, 0.375 and 0.5 are used as typical values in the range of θ_{Ga}<0.125, 0.125≤θ_{Ga}<0.1875, 0.1875≤θ_{Ga}<0.25, 0.25≤θ_{Ga}<0.375, 0.375≤θ_{Ga}<0.5, and θ_{Ga} ≥0.5, respectively. Corrections to the values are qualitatively made by changing potentials according to the ratio between the number of atoms in the dimer and missing dimer regions.

The kinetic rates are assumed to be of Arrhenius form:

$$R=R_0\exp(-\Delta E/kT), \qquad (4)$$

where R_0 is the diffusion prefactor in hops per second and ΔE is the site-dependent and coverage-dependent activation barrier, which is determined by the *ab initio* total energy calculations as shown in Fig. 3. Once all the rates have been calculated in this fashion, the incremental time unit for the simulation run must be determined. We note that the reciprocal of each rate gives us the basic time interval for each process. In addition, an important time scale in the simulation is given by the cation deposition rate of the order of $\mu m/h$, which is consistent with experimental situations.

Using the migration potentials shown in Fig. 3, the MC simulation for Ga adatoms was performed at 600 °C with a growth rate of 2 (ML/s) [153]. The ratio n/N of the number of Ga adatoms n in the dimer region [solid line] and missing dimer region [dotted line] to the total number of surface lattice sites N is shown in Fig. 4 as a function of surface coverage θ_{Ga}. This figure implies that Ga adatoms

favorably exist in the dimer region at the initial stage of the growth i.e., $\theta_{Ga}<0.1$, in which adatom migration is qualitatively governed by the migration potential as shown in Fig. 3(a). The number of Ga adatoms in the missing dimer region is found to linearly increase with increasing coverage and to approach that in the dimer region, which reflects the migration potential as shown in Fig. 3(b). At higher adatom coverage $\theta_{Ga}>0.3$, again, the number of Ga adatoms on As-dimers linearly increases, whereas the number of Ga adatoms in the missing dimer region almost stays constant. This is because anisotropic migration is suppressed at higher adatom coverage resulting from the migration potential with lower activation barrier. Similar results were obtained for Al adatom migration [153].

According to these results, MC approach combining with migration potentials obtained by *ab initio* calculations is a powerful tool to investigate kinetic processes on the complicated surface structures. In order to simulate more realistic growth, however, *ab initio* calculations require very time-consuming computations because they need migration potentials at various adatom coverage in a large surface unit cell with steps and/or kinks. Thus simple energy formalism is desirable not only for estimating migration potentials from the viewpoint of electronic theory but also for materials design in nano-technology.

Figure 4. Ratio n/N of the number of Ga adatoms n in the dimer region [solid line] and missing dimer region [dotted line] to the total number of surface lattice sites N.

3. ROLE OF ELECTRONS ON THE SURFACE

In order to apply *ab initio*-based MC simulation to epitaxial growth, elemental growth processes such as migration potentials, adsorption and desorption behavior of adatoms have to be simplified because of the capabilities of present computers. In this section, a simple interpretation for surface reconstruction, migration potential, and adsorption and desorption energies can be successfully shown by considering the number of electrons remaining in surface dangling bonds. Electrons in dangling bonds on the surface play an important role for understanding various surface-related properties.

3.1. Surface reconstruction

The atoms in compound semiconductors are sp^3 hybridized. In bulk materials, two hybridized orbitals, one from each type of atom, combine to form bonding and antibonding orbitals. Therefore, the bulk

138

properties can be understood by considering the nature of simple interatomic bonds. At the surface, however, some hybrid orbitals cannot form bonds, and partially filled sp^3 dangling bonds remain. The dangling-bond energy states, which can be estimated from the s and p atomic levels, dominate the surface properties for compound semiconductors. Farrell, Harbison, and Peterson proposed a simple criterion for use in predicting static surface reconstructions of GaAs(001) and the adsorption sequence of Ga atoms, where all As dangling bonds are kept filled and all Ga dangling bonds empty [136]. This leads to an electron counting model which can specify the stability of a compound-semiconductor surface based on a simple analysis of the dangling-bond energy levels.

Figure 5 shows the energy levels of the sp^3 dangling bond states of GaAs [154]. The dangling bond energy level of Ga is located in the conduction band, and should, therefore, be empty, while that for As is in the valence band and should be filled. In order to achieve this, electrons transfer from the dangling bonds of Ga to those of As. The electron counting model requires that a surface structure be found in which the number of available electrons in the surface layer will exactly fill all the dangling bond states in the valence band, leaving those in the conduction band empty. If this condition is satisfied, then the surface will be semiconducting, whereas partially filled dangling bonds may lead to a metallic surface. Pashley applied the electron counting model to a compound-semiconductor surface under the following assumptions.[203] (i) The lowest-energy structure is obtained by filling all dangling bonds on the anion atom [with V_n electrons] and keeping empty those on
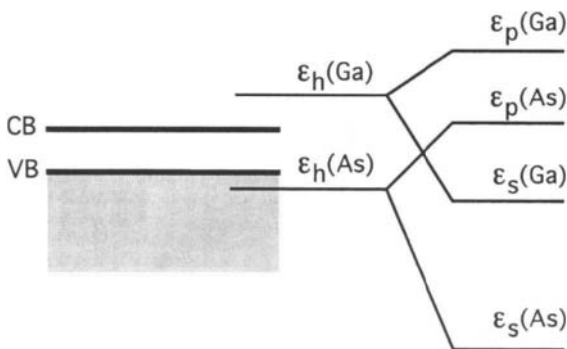
the cation atom [with V_p electrons]; (ii) The surface forms a $(2 \times N)$ reconstruction; (iii) The $2x$ periodicity arises from the formation of surface dimers; (iv) The Nx periodicity arises from the absence of surface dimers, leaving D dimers per unit cell where $D \leq N$. He counted up the number of electrons required to satisfy these conditions and equated that with the number of electrons available in order to determine the relationship between N and D, which gives $6D+8D=2V_nD+V_pN$. In the case of GaAs, $V_n=5$ and $V_p=3$, giving the relationship $4D=3N$. The smallest unit cell that satisfies this condition is a (2×4) unit cell consistent with stable $(2 \times 4)\beta 1$ or $(2 \times 4)\beta 2$ structures.

In the case of a Se-rich ZnSe(001) surface, $V_n=6$ and $V_p=2$. Thus, making the same assumptions as for GaAs(001), they found a (2×1) reconstruction to be consistent with the experimental results. Furthermore, the electron counting model was successfully applied to Ga- and Zn-rich surfaces, and to step and island structures on GaAs(001) and ZnSe(001) surfaces [154]. Although the electron counting model does not necessarily predict a unique surface structure, it gives a guiding principle for the surface structure from the quantum mechanical viewpoint, since any successful structural model must satisfy the electron counting model.

3.2. Migration potential

As seen in Sec. 2.5, as the number of surface Ga adatoms on GaAs(001)-$(2 \times 4)\beta 1$ surface increases, the most stable sites change from long-bridge sites in the As-dimer region [F sites in Fig. 2] to Ga-bridge sites in the missing dimer row [B sites in Fig. 2]. This also gives the sequence of energetically stable structures as a function of the number of Ga adatoms as shown in Fig. 6. Shiraishi and Ito considered the mechanism of this sequence from the viewpoint of strain-energy and band-energy contribution using the electron counting model [155].

The original surface consists of a full monolayer of Ga terminated with three As dimers, satisfying the electron counting model, i. e, the number of electrons in the Ga dangling bonds $\Delta Z=0$. With the first adsorption, ΔZ increases because a Ga adatom has nonbonding electrons. Thus, the first stable structure is mainly determined by the strain-energy contribution, which has been confirmed by energy calculation using our empirical interatomic potentials. With the second Ga adsorption onto F sites, surface As dimers are broken and strong Ga-As bonds form. This re-

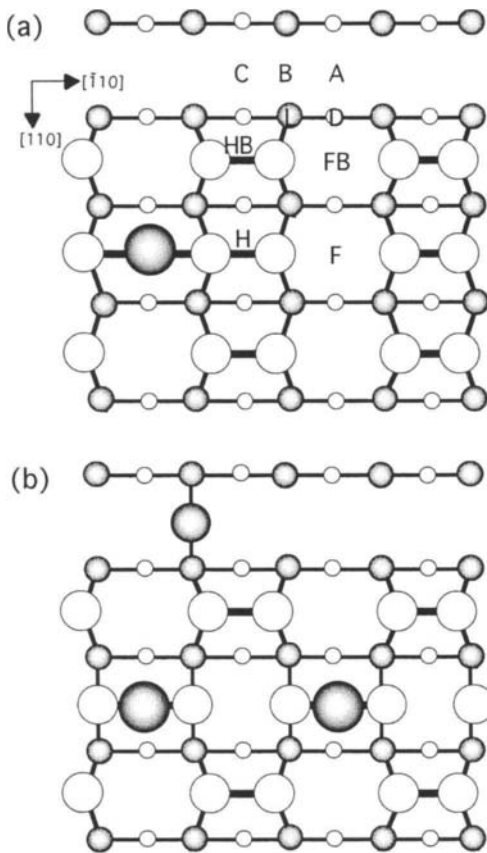Figure 5. Energy levels of the sp^3 dangling bond states of GaAs

Figure 6. Adsorption sequence obtained by *ab initio* calculations, after (a) the first Ga and (b) the third Ga adsorption, in which letters F and B denote the second adsorption site in (a) and the fourth adsorption site in (b), respectively. Gray and white circles denote Ga adatoms and surface As atoms, respectively.

duces the strain energy and ΔZ. The third and fourth adsorptions onto B sites tend to satisfy the electron counting model. The fourth adsorption causes the stabilization of band energy, thereby perfectly restoring the electron counting model. Reflecting these qualitative trends, the energy band structure changes from metallic to semiconducting at the fourth adsorption [155]. Furthermore, Farrell *et al.* proposed a sequence of electronically stable intermediate structures predicted by the electron counting [136,156]. By considering their results together with ours, the electron counting model give useful information not only about surface structure but also about epitaxial growth processes.

3.3. Self-surfactant effect

Surface reconstructions on GaAs surface have been the object of significant scientific and technological interest, stemming not only from the rich variety of behavior exhibited by this surface but also from the widespread use of epitaxially grown GaAs. Through numerous experimental studies, GaAs(001) surface has been found to exhibit, depending on the surface stoichiometry, a variety of structures ranging from the As-rich $c(4\times4)$ structure to the Ga-rich $c(8\times2)$ reconstruction [157-159]. With increasing annealing temperature in ultra high vacuum [UHV], the order of the appearance of the primary reconstructions is $c(4\times4)$, $(2\times4)/c(2\times8)$, (1×6), (4×6) and $(4\times2)/c(8\times2)$. Among these, As-rich (2×4) and $c(4\times4)$ surfaces are technologically important since molecular beam epitaxy [MBE] and metal-organic vapor phase epitaxy [MOVPE] growths usually begin and end with them, respectively [136,160].

The atomic structure of these surfaces has therefore been intensively investigated from both experimental and theoretical viewpoints [159,161-164]. However, in regards to the atomic structures of the (2×4) and $c(4\times4)$ surfaces, there are some puzzling questions as to how nucleation occurs at the initial growth stage and how layer-by-layer growth with stoichiometry is maintained on the GaAs(001) surfaces despite the deficiency of surface As atoms on (2×4) and the excess of surface As atoms on $c(4\times4)$ surface as shown in Fig. 7. A still unclear is the relationship between structural changes in GaAs surface,
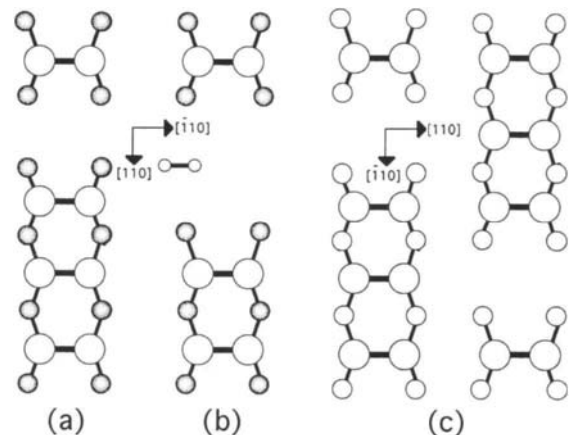
Figure 7. Schematic of (a) GaAs(001)-(2×4)β1, (2×4) β2 and $c(4\times4)$ surface structures.

such as the $c(4\times4)$ to (2×4), and temperature or the ratio of As and Ga. In order to answer these questions, theoretical investigations of adsorption behavior on the GaAs(001) surface are desirable from the microscopic viewpoint.

Using *ab initio* pseudopotential method, Shiraishi and Ito investigated the microscopic mechanisms of As incorporation into missing dimer rows on GaAs(001)-$(2\times4)\beta1$ surface and desorption of As addimers from GaAs(001)-$c(4\times4)$ surface without antisite formation [165,166]. Figures 7(a), 7(b) and 7(c) show the $(2\times4)\beta1$, $(2\times4)\beta2$ and $c(4\times4)$ surface structures. The $(2\times4)\beta1$ and $(2\times4)\beta2$ contain missing dimer rows and the As coverage is smaller than 1.0. On the other hand, the $c(4\times4)$ structure contains excess As ad-dimers and the As coverage is larger than 1.0. The calculated results of adsorption and desorption energies of As-dimers are shown in Fig. 8(a) on the $(2\times4)\beta1$ and 8(b) on the $c(4\times4)$ as a function of Ga adatom coverage, respectively. The As_2 adsorption energy drastically increase as the Ga adatom coverage increases. Similar results were obtained in the calculation for $(2\times4)\beta2$ surface. This large increase in As_2 adsorption energy indicates that As atoms tend to be favorably incorporated into the missing dimer row after the Ga atoms have adsorbed on the GaAs(001)-(2×4) surface. In other words, preadsorbed Ga atoms promote As incorporation into the missing dimer row without forming antisites during thin-film growth. The As_2 desorption energy decreases as the Ga adatom coverage increases. This dramatic decrease indicates that the chemical bonds between the second-layer As and the excess As ad-dimers are greatly weakened by the effects of preadsorbed Ga atoms. Thus preadsorbed Ga atoms promote the exchange of excess As dimers by Ga atoms without forming antisites during epitaxial growth.

These results imply that preadsorbed Ga atoms control the excess and deficiency of the As atoms on the surfaces in epitaxial growth processes. In this sense, preadsorbed Ga atoms act as "self-surfactant atoms". After controlling the excess and deficiency of As atoms on the surface, Ga atoms continue to adsorb and migrate on the surface, maintaining layer-by-layer growth. Thus Ga and As atoms are "mutual self-surfactant atoms" in the epitaxial growth processes. This is crucial mechanism by which layer-by-layer growth is maintained without antisite formation during GaAs epitaxial growth and this "self-
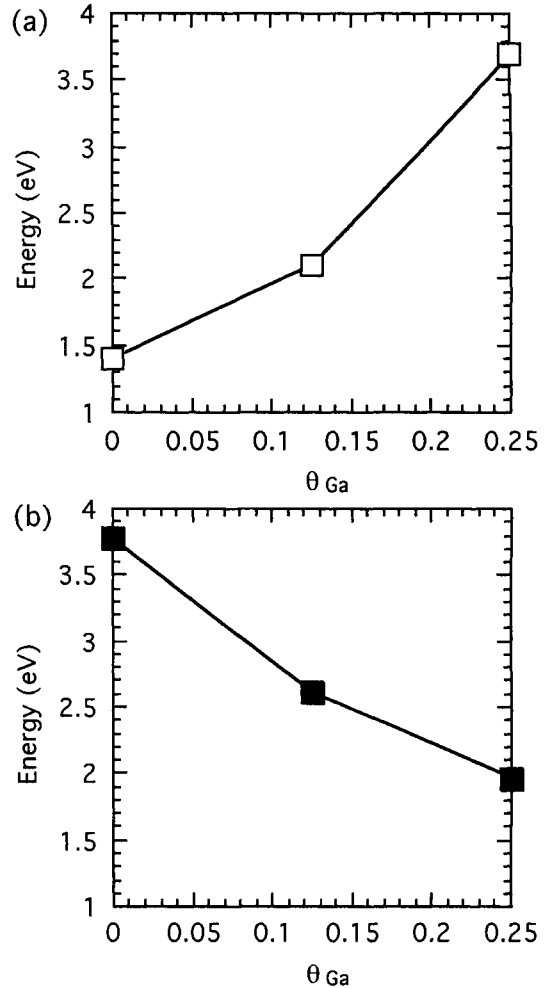
Figure 8. Calculated (a) As_2 adsorption energy and (b) desorption energy as a function of Ga adatom coverage θ_{Ga}.

surfactant effect" is universally applicable to other multicomponent semiconductor epitaxial growth. Furthermore, the self-surfactant effect can be also explained well by the electron counting model. It was found that the energies shown in Fig. 8 linearly depend on the number of electrons in the surface dangling bonds. This implies that adsorption and desorption on the GaAs surface occur to reduce the number of electrons in the dangling bonds and stabilize epitaxially grown GaAs surface. According to these findings, reducing the number of electrons in the surface dangling bonds and restoring the electron counting model is important for understanding epitaxial growth of GaAs.

4. EPITAXIAL GROWTH SIMULATION

In continuing to develop epitaxial growth simulation, we have to be in the process of applying the *ab initio* method with MC simulation to the dynamic change in atomic configurations undergone during epitaxial growth. In this section, adsorption behavior on GaAs(001)-(2×4)β2 surface is investigated by *ab initio* pseudopotential calculations and Monte Carlo simulation. Calculation of adsorption and desorption energies of As dimers on GaAs(001)-(2×4)β2 surface imply that Ga adatoms impinging on the surfaces play an important role to restore the electron counting model. Motivated by these findings, electron counting Monte Carlo [ECMC] method is developed using a simple energy formula based on the empirical interatomic potentials and the electron counting model. Using this energy formula, migration potentials for Ga adatoms and epitaxial growth processes of GaAs on GaAs(001) surface is systematically investigated.

4.1. Simple energy formula

As discussed in section 3.1, Farrell, Harbison and Peterson proposed a simple criterion for predicting static surface reconstructions of GaAs(001) and growth mechanisms where all As dangling bonds are filled and all Ga dangling bonds are empty [136]. This provides an electron counting model that can specify the stability of a compound semiconductor surface based on a simple analysis of the dangling bond energy levels. Furthermore, we pointed out that strain energy, in addition to the number of electrons in electron counting model, is crucial for interpreting the strong dependence of migration potentials on Ga adatom coverage.

Based on these facts, in order to investigate the migration potential of Ga adatoms on the GaAs surfaces, Ito and Shiraishi proposed system energy, E, which is defined as the summation of two terms as [167]

$$E = E_{bond} + \Delta E_{bond}. \tag{5}$$

$$E_{bond} = 1/2 \Sigma V_{ij}, \tag{6}$$

$$\Delta E_{bond} = C|\Delta Z|. \tag{7}$$

Here, E_{bond} is the cohesive energy estimated by empirical interatomic potentials V_{ij} and ΔE_{bond} due to the

charge redistribution in the dangling bonds is extracted from the results obtained by our *ab initio* calculations [152,155,168,169]. The universal empirical interatomic potentials [1,94-961 are used for the calculation of the E_{bond} of the system.

The energy contribution of charge redistribution from the Ga dangling bonds to As dangling bonds on the surface is estimated according to the electron counting model based on the data in Fig. 9. This figure summarizes our previous calculations of the migration potential for Ga adatoms at adatom coverage $\theta = 0.25$ on the GaAs(001)-(2×4)β1 surface [168,169]. Figure 9 shows that the data linearly depend on the number of electrons ΔZ remaining in the Ga dangling bonds with the slope of $C = 0.4$ eV/electron, because the dangling-bond energy level of Ga is located in the conduction band and, therefore, gives higher energy with an increase in the number of electrons in the Ga dangling bonds on the surface. The value of 0.4 eV is reasonable since the charge transfer from Ga dangling bonds to As dangling bonds on the surface is energetically costly in terms of electrostatic interactions between charged subunits [164]. These contributions give a small value compared to the energy-band gap value of 1.4 eV for bulk-GaAs.

A linear relationship with ΔZ was generally found in various properties, such as adsorption and desorption energies [165,166]. Furthermore, the
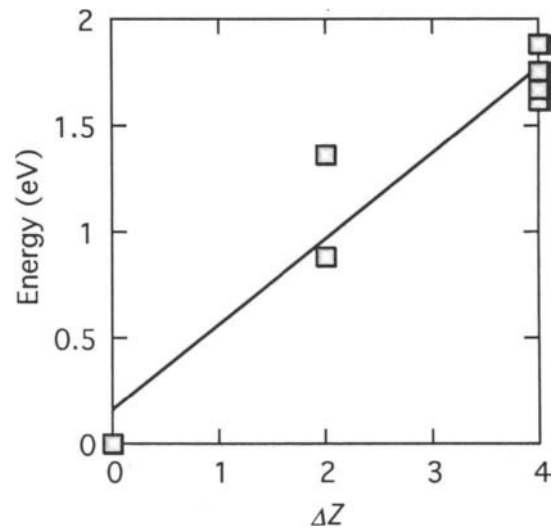
Figure 9. Relationship between energy with respect to ΔE_{bond} and the number of electrons ΔZ.

applicability of this assumption in ΔE_{bond} to the surfaces with As-dimer kinks or steps were confirmed by our preliminary *ab initio* calculations, where the step structure hardly affects the energy due to the charge redistribution. Zhang and Zunger also showed that the formation energy of steps of GaAs can be described by the linear combination of structural motifs including energies with respect to the charge redistribution, which are extracted from the *ab initio* calculations of flat GaAs(001) surfaces [170,171]. However, the versatility of coefficient value of 0.4 eV is not fully justified at the present time, since it was not directly determined by the *ab initio* calculations for As-dimer kinks or steps. In this regard, these equations are used to investigate qualitative trends among the migration potentials by calculating the energy, E, at various lattice points on GaAs(001)-(2×4)β2 surfaces with As-dimer kinks and steps.

4.2. Migration potential

Figure 10 shows the calculated migration potential for Ga adatoms on the flat GaAs(001)-(2×4)β2 surface, which clearly implies that energetically stable sites are located along the missing dimer rows [168]. The stability of missing dimer sites comes from the dangling bond characteristics of the Ga adatom, since a Ga adatom in the missing dimer row can form interatomic bond with another Ga atom to suppress the number of electrons in the Ga dangling bonds. This suppression diminishes the contribution of ΔE_{bond} in Eq. (7). These implications are consistent with the results obtained by *ab initio* total energy calculations, which are also shown in this figure [172,173]. It should be noted that the agreement between two results is good, notwithstanding the fact that the contribution of ΔE_{bond} is not directly determined by the *ab initio* results for (2×4)β2 but for the (2×4)β1.

Figures 11(a) and 11(b) respectively show the calculated migration potential for Ga adatoms on the GaAs(001)-(2×4)β2 surface near *A*- and *B*-type steps. A similarity to the results to Fig. 10 is found with respect to the stable lattice sites along the missing dimer row because of the smaller number of electrons in the Ga dangling bond. A difference between the *A*- and *B*-type step explicitly appears in the migration potential near the step edges. The *A*-type step edge does not affect the migration potential significantly, whereas the migration potential energy has the lowest value at the lattice sites on the lower terrace near the

B-type step edge [denoted G in Fig. 11(b)]. This is because the Ga adatom located at G site in the missing dimer row is weakly stretched by an As-dimer and As atom at the regular fcc sublattice and can form interatomic bond with another Ga atom to suppress the number of electrons in the Ga dangling bonds. Similar findings were obtained near As-dimer kinks [167,174,175]. Based on scanning tunneling microscopy [STM], Pashley *et al.* reported that *A*-type steps are relatively straight, while *B*-type steps result in rugged step edges on the GaAs(001) vicinal surfaces grown by MBE [176]. More elaborate STM study by Tsukamoto and Koguchi directly shows that Ga adatoms are found near not *A*- but *B*-type step edges [177]. These experimental findings are consistent with the calculated results shown in Fig. 11, where the *B*-type step edges provide stable adsorption sites along the missing dimer row while none of
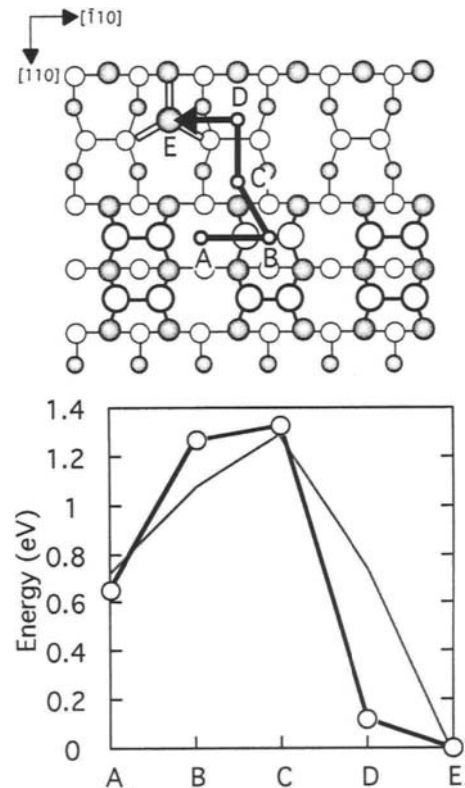
Figure 10. Calculated migration potential for Ga adatom on the GaAs(001)-(2×4)β2 surface. Gray and white circles denote Ga and As atoms, respectively
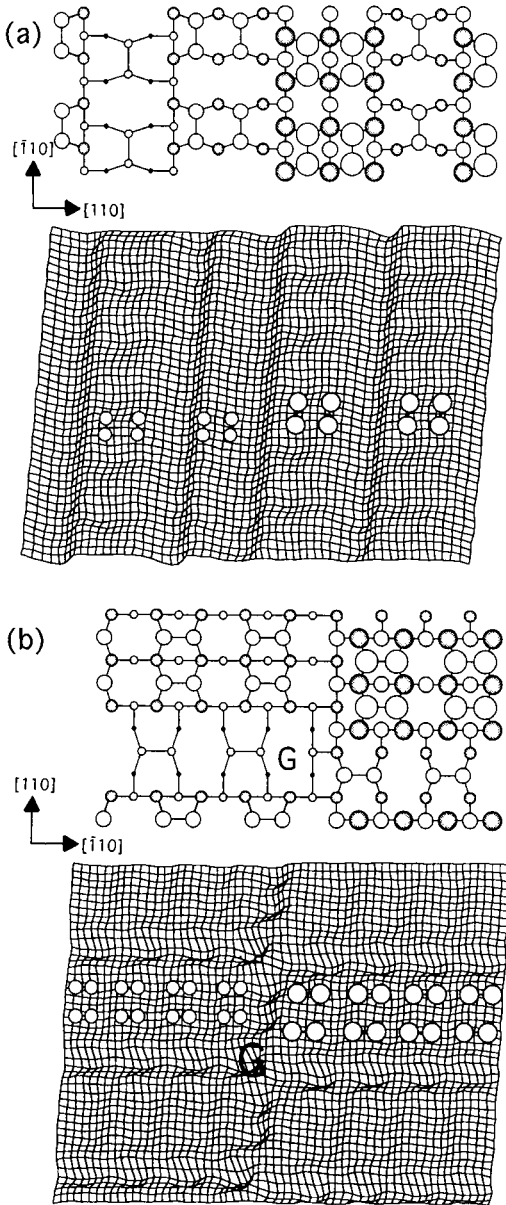
Figure 11. Calculated migration potentials for a Ga adatom on the GaAs(001)-(2×4)β2 surface near (a) A- and (b) B-type steps. Gray and white circles denote Ga and As atoms, respectively.

preferential lattice sites appears near the A-type step edges. Considering these facts, the simple energy formula is feasible for investigating epitaxial growth processes.

4.3. Electron counting Monte Carlo simulation

According to the results for adsorption behavior on the GaAs(001) surfaces, Ito and Shiraishi newly developed an electron counting Monte Carlo (ECMC) method to simulate GaAs epitaxial growth [178]. In the ECMC simulation, they make the following assumptions on a (4×4) surface unit cell, except for the calculations on a (16×4) surface unit cell for GaAs(001)-(2×4)β2 surface with As-dimer kinks or steps. (i) Adsorption of the Ga atoms occurs at specific lattice sites with zinc blende symmetry. (ii) On GaAs(001)-(2×4)β2 surface, Ga and As atoms are alternately impinged on the surface in four successive impingements of each atomic species. This assumption is based on the results of *ab initio* calculations as shown in Fig. 8(a) and limits As-coverage θ_{As} to the reasonable range of $0.5 \leq \theta_{As} \leq 0.75$ [136]. (iii) Dimerization of the surface atoms and nearest-neighbor exchange between adatoms are also included in each Monte Carlo [MC] step in the equilibration procedure. (iv) A growth profile is recorded for a lattice size of 4×4 or 16×4 sub-lattices. Periodic boundary conditions are imposed on the *xy* plane.

Equilibrium atomic arrangements in the (4×4) or (16×4) surface unit cell are obtained by lowering the energy, E, of the system in Eq. (5). Here, for simplicity, E_{bond} [in eV/atom] is described in terms of interatomic bond energies instead of by Eq. (6), typically 3.26 for Ga-As, 2.10 for Ga-Ga, 2.80 for As-As, 1.64 for Ga-Ga dimers, and 2.58 for As-As dimers. These are approximate values extracted from the calculations using empirical interatomic potentials[178]. The electronic contribution due to excess electrons ΔZ remaining in the dangling bonds is given as ΔE_{bond} [in eV/atom]ß=0.4|ΔZ|, as shown in Eq. (7). Atomic arrangements are generated according to the following rules: (i) Select adatoms on the surface at random. (ii) Select events as described in assumptions (3) and (4) at random. (iii) Calculate the change in energy ΔE after exchanging the chosen adatoms; (iv) If ΔE is negative, accept the new atomic arrangement; otherwise, (v) select a random number h uniformly distributed over the interval (0,1). (vi) If $\exp(-\Delta E/kT)<h$, accept the old atomic arrangement; otherwise, (vii) use the new atomic arrangement and the new energy as the current properties of the system. This procedure is repeated for a suitable number of configurations in order to approach equilibrium atomic arrangements

According to our kinetic Monte Carlo simulation based on the activation barrier data obtained by *ab initio* calculations, Ga adatoms randomly impinging on the GaAs(001)(2×4)β1 surface migrate to reside very quickly in the most stable lattice sites with zinc blende symmetry at 600 °C. This occurs in the range of Ga adatom coverage $\theta_{Ga} \leq 0.25$, because of the low activation barrier, $\Delta E \approx 0.7$-0.8 eV, for Ga adatoms on the surface. This implies that resultant atomic configurations at high temperatures can be specified by considering adsorptions with the lowest energy. Therefore, the ECMC simulation without incorporating adatom migration reflects real initial MBE growth processes at high temperatures.

4.4. Epitaxial growth processes

The typical epitaxial growth process on GaAs(001)-(2×4)β2 surface obtained by ECMC simulation is shown in Figs. 12(a) to 12(h) [178-180]. The GaAs(001) surface changes its structure from (2×4)β2 to another electronically stable structure, the (2×4)α shown in Fig. 12(c), in the first Ga impinging cycle. This agrees well with *ab initio* total energy calculations, which showed the (2×4)β2 surface favors Ga adatoms located at the lattice sites along the missing dimer row. Subsequent As adsorption produces the (2×4)β1 structure, which also satisfies the electron counting model, as shown in Fig. 12(d). The adsorption sequence on the (2×4)β1 surface is shown in Figs. 12(e) to 12(h). Impinged Ga adatoms occupy the lattice sites on the As-dimers at low Ga adatom coverage of $\theta_{Ga} \leq 2/16$ [Fig. 12(e)].

As the coverage increases to $\theta_{Ga} \geq 3/16$, Ga adatoms tend to reside in the lattice sites in the missing dimer region [Fig. 12(f)]. As shown in Fig. 12(g), once Ga adatoms stably reside in the lattice sites along the missing dimer row, As atoms occupy other vacant lattice sites or replace Ga adatoms in the missing dimer region to avoid the formation of energetically unfavorable Ga-Ga bonds. Finally, the As-missing dimer row is completely occupied by As adatoms and layer-by-layer growth consistent with the electron counting model continues simultaneously with the dimerization of Ga atoms on the surface as shown in Fig. 19(h).

The ECMC simulation of the number of electrons ΔZ (closed circles) in the Ga dangling bonds as a function of the number of adatoms N_{ad} on the GaAs(001)-(2×4)β2 surface is shown in Fig. 13. The simulation implies that the structural change

from initial (2×4)β2 to (2×4)β1 via (2×4)α is electronically favorable, since there are many electronically stable states, which are specified by the number of electrons $\Delta Z=0$, with small fluctuations in the range of $-1 \leq \Delta Z \leq 1$. On the other hand, the subsequent structural change from (2×4)β1 occurs with large fluctuations, such as $-4 \leq \Delta Z \leq 6$. The surface at $N_{ad}=8$ corresponds to (2×4)β1 consisting of a full monolay-
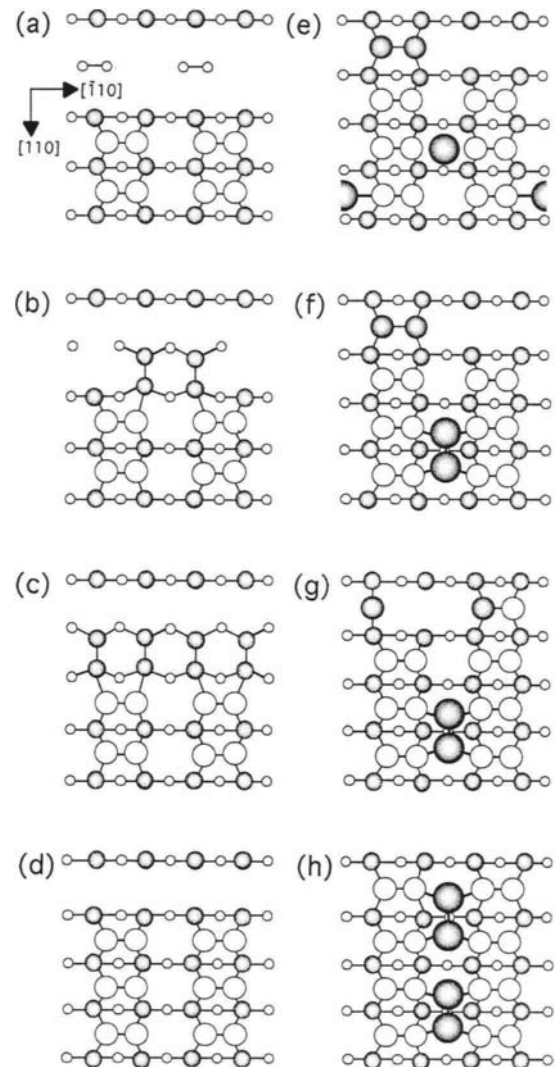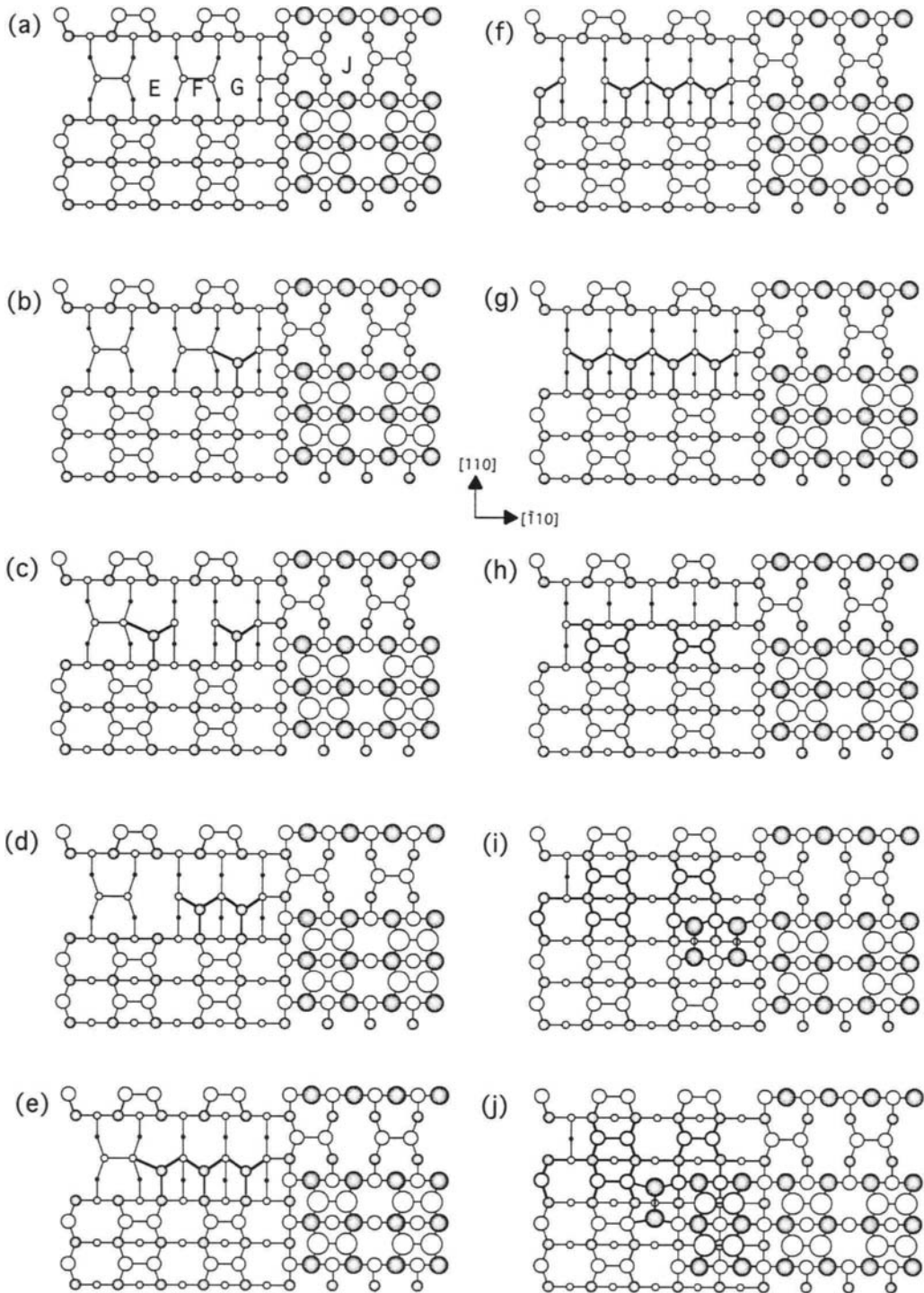

Figure 12. Growth process of GaAs on GaAs(001) - (2×4)β2 surface obtained by ECMC simulation. Gray and white circles denote Ga and As atoms, respectively.

er of Ga terminated with three As dimers, which satisfies the electron counting model. With the ninth adsorption in the As-dimer region, ΔZ increases because a Ga adatom has nonbonding electrons. The electron counting model is restored by the twelfth adsorption onto the lattice sites in the missing dimer region. The qualitative trends obtained by ECMC simulation are consistent with *ab initio* results. Furthermore, STM observations by Avery *et al.* verify the basic feature of these predictions [181].

The ECMC simulation for adsorption sequences near the B-type step edges is shown in Fig. 14 [179]. It is found that Ga adatoms preferentially occupy the lattice sites along the missing dimer rows near the B-type step edges. The first Ga adatom impinging on the surface resides in the most stable lattice site, i. e., G, as shown in Fig. 14(b). Figures 14(c) and 14(d) imply, however, that the second Ga adatom does not occupy lattice site J on the upper terrace, which is the second stable lattice site as shown in Fig. 14(a), but instead occupies lattice site F adjacent to G on the lower terrace via lattice site E. This is because occupying the F site decreases the number of electrons in the Ga dangling bonds ΔZ from one to zero, whereas ΔZ increases from one to two when a Ga adatom locates at the J site. A similar situation can be seen in Figs. 14(e) to 14(g). A Ga adatom impinging on the surface resides in the lattice site between the As dimer and As atom with smaller strain energy and the subsequent Ga adatom breaks As dimer to form a Ga-dimer array along the missing dimer row. Further As adsorptions form a $(2\times4)\beta1$-like structure near the B-type step edges as shown in Fig. 14(h). Subsequent Ga and As adsorptions respectively produce Ga dimers at the step edges and fill up the missing dimer row as shown in Fig. 14(i). The lattice sites near the B-type step edges on the As-dimer regions are also favorable for Ga adatoms, since Ga adatoms are not strongly stretched by As dimers and As atoms without dimerization. As a result, the nucleation occurs near the B-type step edges on the lower terrace as shown in Fig. 14(j).

Based on reflection high-energy diffraction [RHEED] observations during migration enhanced epitaxy [MEE] on misoriented GaAs(001) substrate, Yamaguchi and Horikoshi pointed out that the difference in critical temperature for step-growth can be explained by the different characteristics between A- and B-type steps; the B-type steps, unlike A-type
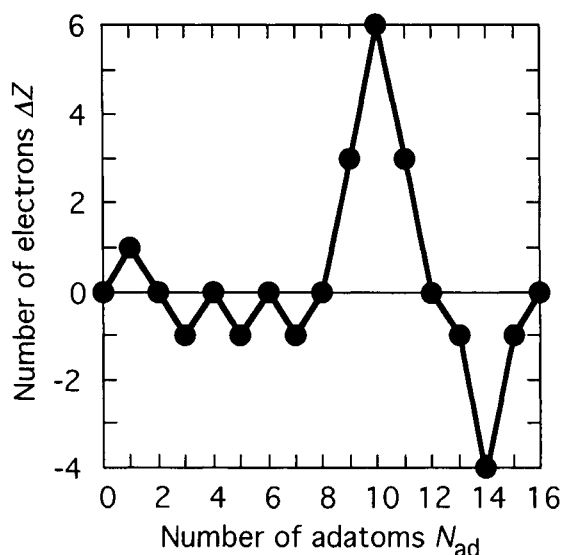
Figure 13. Change in the number of electrons ΔZ in the dangling bonds on GaAs(001) -(2×4)β2 surface as a function of the number of adatoms N_{ad} during ECMC simulation.

steps, provide active sites for Ga adsorption [182-184]. Our calculated migration potentials near the A- and B-type step edges and ECMC simulation near the B-type step edges are consistent with their experimental results. Moreover, Tsukamoto and Koguchi studied in situ STM observations of Ga adatoms on the MBE growth front on the GaAs(001)-(2×4)β2 surface[177]. Their observations support our predictions for atomic arrangements near B-type step edges, where Ga adatoms are self-organized in a missing dimer row.

In summary, Ga adatoms favorably reside in the lattice sites along the missing dimer row on the GaAs(001)-(2×4)β2 surface. The lattice site between an As atom and As dimer along the missing dimer row is favorable for Ga adatoms near B-type step edges, while A-type steps without such configurations do not provide active sites for Ga adsorption. Further ECMC simulation gives reasonable predictions for MBE growth processes on the GaAs(001)-c(4×4) surface in addition to that near As-dimer kinks on the GaAs(001)-(2×4)β2 surface [180,185]. This implies that the ECMC simulation can give predictions of atomic arrangements during epitaxial growth in advance of experiments and is feasible for atomistic simulation of epitaxial growth processes.

146

Figure 14. Typical growth process near B-type step edges on GaAs(001)-(2×4)β2 surface obtained by ECMC simulation. Gray and white circles denote Ga and As atoms, respectively.

5. SUMMARY

In this article we have briefly shown the power and applicability of computational methods in understanding epitaxial growth of GaAs. Similar approach has been successfully applied to investigate the adsorption behavior in more complicated system such as (111)-oriented and non-planar surfaces [186]. These successful applications show that the calculations can explain experimental results and be used to make useful predictions. More recently, Seino *et al.* investigated the microscopic migration of As adatom on a Ga-terminated GaAs(001) surface using *ab initio* calculations and SOS MC simulations [187]. They showed that the kinetics of both As and Ga adatoms are crucial during MBE growth. Incorporating electronic contribution similar to our approach, Itoh *et al.* performed MC simulations with large surface unit cell to simulate the nucleation and growth of islands on GaAs(001) and InAs(001) in homoepitaxy [188,189]. Itoh and Ohno found that the density of double As dimers evolves synchronously with the observed specular RHEED intensities in growth and after its interruption [190,191].

Success with the computer-aided approach will lead more realistic simulation of more complicated systems such as simulation with various surface orientations and heterosystems combining with various materials having different structures and different valencies. Taguchi *et al.* systematically investigated the growth process of GaAs on GaAs(111)A surface using *ab initio* calculations [192-194]. They pointed out that As acts as self-surfactant element, which maintains GaAs lattice. Sano and Nakayama performed the MC simulation of ZnSe/GaAs(001) heteroepitaxy incorporating the heterovalent effects to investigate the defect formation at the interface [195] . Based on the atomic arrangements obtained by the MC simulation, they calculated reflectance difference [RD] spectra of ZnSe(001) surface and clarify the origin of the spectra in relation to the surface reconstructions. Moreover, Kangawa *et al.* proposed a new theoretical approach for studying adsorption-desorption behavior of atoms on semiconductor surfaces [196]. They successfully calculated free energy of gas phase to discuss temperature and beam equivalent pressure dependence of adsorption-desorption behavior of Ga on GaAs surfaces.

These attempts will become increasingly important for discovering new characteristics. To approach the next stage in the development of materials science, atomistic simulation of epitaxial growth enables us to interpret growth process and predict dynamic change in atomic arrangements and resultant material properties. The results presented here are just an example of epitaxial growth simulations, where adsorption behavior on various complicated surfaces including non-planar surfaces can be investigated to realize desired structures in response to various purposes. Recent progress in fabrication techniques will make it possible to create various new metastable materials by controlling the atomic arrangements, such that many atomic species with different valencies can be selected. Furthermore, development of quantum mechanical simulations will make it possible to give guiding principles for identifying atomic configurations that would rise to a desired new property and for controlling adsorption and migration behavior of atoms to realize those identified configurations during epitaxial growth. We believe that atomistic simulation of epitaxial growth based on quantum mechanics will lead to great advances in materials science and technology.

REFERENCES

1. T. Ito, J. Appl. Phys. **77** (1995) 4845.
2. R. Car and M. Parrinello, 1985. Phys. Rev. Lett. **55** (1985) 2471.
3. C. Herring, Phys. Rev. **57** (1940) 1169.
4. J. C. Phillips and L. Kleinman, Phys. Rev. **116** (1959) 287.
5. J. C. Slater and G. F. Koster. Phys. Rev. **94** (1954) 1498.
6. J. Korringa, Physica **13** (1947) 392.
7. W. Kohn and N. Rostoker, Phys. Rev. **94** (1954) 1111.
8. M. Posternak, H. Krakauer, A. J. Freeman, and D. D. Koeling, Phys. Rev. B **21** (1980) 5601.

148

9. P. J. Feibelman and D. R. Hamann, Phys. Rev. B **28** (1983) 3092.
10. O. K. Andersen, Phys. Rev. B **12** (1975).3060.
11. P. Hohenberg and W. Kohn, Phys. Rev. **136** (1964) B864.
12. W. Kohn and L. J. Sham, Phys. Rev. **140** (1965) A1133.
13. D. R. Hamann, Phys. Rev. Lett. **42** (1979) 662.
14. L. F. Mattheis and D. R. Hamann, Phys. Rev. B **29** (1984) 537.
15. L.F. Mattheis and D. R. Hamann, Phys. Rev. B **29** (1984) 5372.
16. L.F. Mattheis and D. R. Hamann, Phys. Rev. B **30** (1984) 1731.
17. L. F. Mattheis, D. R. Hamann and W. Weber, Phys. Rev. B **34** (1986) 2190.
18. S.-H. Wei, A. A. Mbaye, L. G. Ferreira and A. Zunger, Phys. Rev. B **36** (1987) 4163.
19. S.-H. Wei, L. G. Ferreira and A. Zunger,. Phys. Rev. B **41** (1990) 8240.
20. H. J. F. Jansen and A. J. Freeman, Phys. Rev. B **30** (1984) 561.
21. C. L. Fu, A. J. Freeman and T. Oguchi, Phys. Rev. Lett. **54** (1985) 2700.
22. C. L. Fu and A. J. Freeman. Phys. Rev. B **35** (1987) 925.
23. W. R. L. Lambrecht, and O. K. Andersen, Phys. Rev. B **34** (1986) 2439.
24. U. Schmid, N. E. Christensen, M. Cardona and M. Alouani, *Proc. 20th Int. Conf. Physics of Semiconductors*, World Scientific Publishing, Singapore, 1990, pp. 865,.
25. Y. M. Gu, L. Fritsche and N. E. Christensen, Phys. Stat. Solidi **159** (1990) 617.
26. J. Ihm, A. Zunger and M. L. Cohen, J. Phys. C **12** (1979) 4409.
27. M. T. Yin and M. L. Cohen, Phys. Rev. B **26** (1982) 5668.
28. W. Andreoni and R. Car. Phys. Rev. B **21** (1980) 3334.
29. T. Nakayama and H. Kamimura, J. Phys. Soc. Jpn. **54** (1985) 4726.
30. S. Ciraci and I. P. Batra, Phys. Rev. B **36** (1987) 1225.
31. D. M. Bylander and L. Kleinman, Phys. Rev. B **34** (1986).5280.
32. D. M. Bylander and L. Kleinman, Phys. Rev. B **36** (1987) 3229.
33. C. G. Van de Walle and R. M. Martin, Phys. Rev. B **35** (1987) 8154.
34. A.Munoz, N. Chetty and R. M. Martin, Phys. Rev. B **41** (1990) 2976.
35. D. M. Bylanderand L. Kleinman, Phys. Rev. B **41** (1990) 3509.
36. G. A. Baraff, E. O. Kaneand and M..Schlüter, Phys. Rev. B **21** (1980) 3563.
37. J. E. Northrup and M. L. Cohen, Phys. Rev. B **23** (1981) 2563.
38. J. E. Northrup, M. L. Cohen, J. R. Chelikowsky, J. Spence and A. Olsen, Phys. Rev. B **24** (1981) 4623.
39. G. A. Baraff and M. Schlüter, Phys. Rev. B **30** (1984) 3460.
40. Y. B-Yam, and J. D. Joannopoulos, Phys. Rev. B **30** (1984) 1844.
41. Y. B-Yam, and J. D. Joannopoulos, Phys. Rev. B **30** (1984) 2216.
42. R. Car, P. J. Kelly, A. Oshiyama and S. T. Pantelides, Phys. Rev. Lett. **52** (1984) 814.
43. R. Car, A. Oshiyama and S. T., Pantelides, Phys. Rev. Lett. **54** (1985) 360.
44. A. Zunger, *Solid State Physics* Vol. 39, Academic Press, New York, 1986, pp. 275.
45. E. Tarnow, P. Dallot, P. D. Bristowe and J. D. Joannopoulos, Phys. Rev. B **42** (1992) 3644.
46. G. A. Baraff and M. Schlüter, Phys. Rev. B **33** (1986) 7346.
47. L. W. Wang and A. Zunger, J. Chem. Phys. **94** (1994).2158.
48. L. W. Wang and A. Zunger, J. Chem. Phys. **100** (1994) 2394.
49. K. C. Pandey, Phys. Rev. Lett. **47** (1981) 1913.
50. J. E. Northrup and M. L. Cohen, Phys. Rev. Lett. **49** (1982) 1349.
51. G. Brocks, P. J. Kelly and R. Car, Phys. Rev. Lett. **66** (1991) 1728.
52. K. Shiraishi, Appl. Phys. Lett. **60** (1992) 1363.
53. J. Neugeberger and C. G. Van de Walle, Phys. Rev. B **50** (1994) 8067.
54. J. Neugeberger and C. G. Van de Walle, Appl. Phys. Lett. **69** (1996) 503.
55. T. Matilla and R. M. Nieminen, Phys. Rev. B **55** (1997) 9571.
56. C. H. Park and D. J. Chadi, Phys. Rev. B **55** (1997) 12995.
57. D. J. Chadi, Appl. Phys. Lett. **71** (1997) 2970.
58. J. Elsner, R. T. Jones, P. K. Stich, V. D. Porezag, M. Elstner, Th. Frauenheim, M. I. Heggie, S. Öberg and P. R. Briddon, Phys. Rev. Lett. **79** (1997) 3672.

59. R. B. Capaz, H. Lim and J. D. Joannopoulos, Phys. Rev. B **51** (1995) 17755.

60. J. E. Jaffe, R. Pandey and P. Zapol, Phys. Rev. B **53** (1996) R4209.

61. J. E. Northrup and J. Neugebauer, Phys. Rev. B **53** (19969 R10477.

62. M. Städele, J. A. Majewski and P. Vogl, Phys. Rev. B **56** (1997) 6911.

63. J. E. Northrup, R. Di Felice and J. Neugebauer, Phys. Rev. B **56** (1997) R4325.

64. A. R. Smith, R. M. Feenstra, D. W. Greve, J. Neugebauer and J. E. Northrup, Phys. Rev. Lett. **79** (1997) 3934.

65. G. B. Bachelet, D. R. Hamann and M. Schlüter, Phys. Rev. B **26** (1982) 4199.

66. L. Kleinman and D. M. Bylander, Phys. Rev. Lett. **48** (1982) 1425.

67. X. Gonze, R. Stumpfand M. Scheffler, Phys. Rev. B **44** (1991) 8503.

68. T. Ohno, Phys. Rev. Lett. **70** (1993) 631.

69. P. N. Keating, Phys. Rev.**145** (1966) 637.

70. R. M. Martin, Phys. Rev. B **1** (1970) 4005.

71. E. M., Pearson, T. Takai, T. Halicioglu and W. A. Tiller, J. Cryst. Growth **70** (1984) 33.

72. F. H. Stillinger and T. A. Weber, Phys. Rev. B **31** (1985) 5262.

73. P. Biswas and D. R. Hamann, Phys. Rev. B **36** (1987) 6434.

74. G. C. Abell, Phys. Rev. B **31** (1984) 6184.

75. J. Tersoff, Phys. Rev. Lett. **56** (1986) 632.

76. J. Tersoff, Phys. Rev. B **37** (1988) 6991.

77. P. C. Kelires and J. Tersoff, Phys. Rev. Lett. **63** (1989) 1164.

78. E. Blaisten-Baroja and D. Levesque, Phys. Rev. B **34** (1986) 3910.

79. B. P. Feuston, R. K. Kalia and P. Vashishta, Phys. Rev. B **35** (1987) 6222.

80. J. R. Chelikowsky and J. C. Phillips, Phys. Rev. B **41** (1990) 5735.

81. W. Andreoni and G. Pastore, Phys. Rev. B **41** (1990) 10243.

82. B. C. Bolding and H. C. Andersen, Phys. Rev. B **41** (1990) 10568.

83. X. G. Gong, Phys. Rev. B **47** (1993) 2329.

84. K. E. Khor, and S. Das Sarma, Phys. Rev. B **36** (1987) 7733.

85. X. P. Li, G. Chen, P. B. Allen and J. Q. Broughton, Phys. Rev. B **38** (1988)3331.

86. H. Balamane, T. Halicioglu and W. A. Tiller, Phys. Rev. B **40** (1989) 9999.

87. J. H. Wilson, J. D. Todd and P. A. Sutton, J. Phys. Condens. Matter **2** (1990) 10259.

88. C. Roland, and G. H. Gilmer, Phys. Rev. B **46** (1992) 13428.

89. C. Roland, and G. H. Gilmer, Phys. Rev. B **46** (1992) 13437.

90. C. Roland, and G. H. Gilmer, Phys. Rev. B **47** (1993) 16286.

91. D. Srivastava and B. J. Garrison, Phys. Rev. B **46** (1992) 1472.

92. D. Srivastava and B. J. Garrison, Phys. Rev. B **47** (1993) 4464.

93. H. Balamane, T. Halicioglu and W. A. Tiller, Phys. Rev. B **46** (1992) 2250.

94. K. E. Khor and S. Das Sarma, Phys. Rev. B **38** (1988) 13191.

95. T. to, K. E. Khorand S. Das Sarma, Phys. Rev. B **40** (1989) 9715.

96. T. Ito, K. E. Khor and S. Das Sarma, Phys. Rev. B **41** (1990) 3893.

97. T. Ito, Jpn. J. Appl. Phys. **30** (1991) L1349.

98. T. Ito, Jpn. J. Appl. Phys. **31** (1992) L920.

99. T. Ito, Jpn. J. Appl. Phys. **32** (1993) L379.

100. T. Ito, Jpn. J. Appl. Phys. **35** (1996) 3376.

101. T. Ito, Jpn. J. Appl. Phys. **35** (1996) L1035.

102. B. Alder and T. E. J. Wainwright, Chem. Phys. **27** (1957) 1208.

103. B. Alder and T. E. J. Wainwright, J. Chem. Phys. **31** (1959) 459.

104. A. Rahman, Phys. Rev. A **136** (1964) 405.

105. see for example: F. F. Abraham, Advances in Physics **35** (1986) 1.

106. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller,and E. Teller, J. Chem. Phys. **32** (1953) 1087.

107. M. Schneider, A. Rahman and I. K. Schuller Phys. Rev. Lett. **55** (1985) 604.

108. M. Schneider, I. K. Schuller and A. Rahman, Phys. Rev. Lett. B **36** (1987) 1340.

109. P. A. Taylor and B. W. Dodson, Phys. Rev. B **36** (1987) 1355.

110. K. Hara, M. Ikeda,O. Ohtsuki, K. Terakura, M. Mikami, Y. Tago, and T. Oguchi, Phys. Rev. B **39** (1989) 9476.

111. S. M. Paik and S. Das Sarma, Phys. Rev. B **39** (1989) 1224.

112. S. M. Paik and S. Das Sarma, Phys. Rev. B **39** (1989) 9793.

113. S. M. Paik and S. Das Sarma,. Surf. Sci. **208** (1989) L53.

114. S. M. Paik and S. Das Sarma,. Surf. Sci. **208** (1989)L61.

115. S. M. Paik and S. Das Sarma,. Surf. Sci. **219** (1989) L607.

116. S. Ethier and L. J. Lewis, J. Mater. Res. **7** (1992) 2817.

117. M. Kitabatake and J. E. Greene, J. Appl. Phys. **73** (1993) 3183.

118. D. K. Choi, S,. M. Koch, T. Takai, T. Halicioglu and W. A. Tiller,. J. Vac. Sci. Technol. B **6** (1988) 1140.

119. D. K. Choi, T. Takai, S. Erkoc, T. Halicioglu and W. A. Tiller, J. Cryst. Growth **85** (1987) 9.

120. J. D. Weeks, G. H. Gilmer and K. A. Jackson, J. Chem. Phys. **65** (1976) 712.

121. J. D. Weeks and G. H. Gilmer, Adv. Chem. Phys. **40** (1979) 157.

122. S. Clarke and D. D. Vvedensky, J. Appl. Phys. **63** (1988) 227.

123. D. D. Vvedensky and S. Clarke, Surf. Sci. **225** (1990) 373.

124. T. Shitara, D. D. Vvedensky, M. R. Wilby, J. Zhang, J. H. Neave and B. A. Joyce, Phys. Rev. B **46** (1992) 6815.

125. T. Shitara, D. D. Vvedensky, M. R. Wilby, J. Zhang, J. H. Neave and B. A. Joyce, Phys. Rev. B **46** (1992) 6825.

126. A. Madhukar, Surf. Sci. **132** (1993) 345.

127. S. V. Ghaisas, and A. Madhukar, Phys. Rev. Lett. **56** (1986) 1066.

128. A. Madhukar and S. V. Ghaisas, CRC Crit. Rev. Solid State Matr. Sci. **14** (1988) 1.

129. J. S. Singh and K. K. Bajaj, Superlatt. & Micro struct. **2** (1986) 185.

130. J. Singh, S. Dudley and K. K. Bajaj, J. Vac. Sci. Technol. B **4** (1986) 878.

131. A. Kobayashi, and S. Das Sarma, Phys. Rev. B **37** (1988) 1039.

132. A. Kobayashi, S. M. Paik and S. Das Sarma, J. Vac. Sci. Technol. B **6** (1988) 1145.

133. T. Kawamura, A. Kobayashi and S. Das Sarma, Phys. Rev. B **39** (1989) 12723.

134. Y.-T. Lu and H. Metiu, Surf. Sci. **245** (1991) 150.

135. A. Rockett, Surf. Sci. **227** (1990) 208.

136. H. H. Farrell, J. P. Harbinson and L. D. Peterson, J. Vac. Sci. Technol. B **5** (1987) 1482.

137. J. M. McCoy and P. A. Maksym, J. Cryst. Growth **111** (1991) 178.

138. Z. Zhang, Y.-T. Lu and H. Metiu, Phys. Rev. B **46** (1992) 1917.

139. C. H. Grein, J. P. Faurie, V. Bousquet, E. Tourié, R. Benedek,and T. de la Rubia, J. Cryst. Growth **178** (1997) 258.

140. Y. Horikoshi, M. Kawashima and H. Yamaguchi, Jpn. J. Appl. Phys. **25** (1986) L868.

141. J. H. Neave, P. J. Dobson, B. A. Joyce,and J. Zhang, Appl. Phys. Lett. **47** (1985) 100.

142. Y. W.,Mo, J. Kleiner, M. B. Webb, and M. G. Lagally, Phys. Rev. Lett. **66** (1991) 1998.

143. E. Kaxiras and J. Erlebacher, Phys. Rev. Lett. **72** (1994) 1714.

144. A. Oshiyama, Phys. Rev. Lett. **74** (1995) 130.

145. G. Brocks and P. J. Kelly, Phys. Rev. Lett. **76** (1996) 2362.

146. T. Yamasaki, T. Uda and K. Terakura, Phys. Rev. Lett. **76** (1996) 2949.

147. E. Kim, C. W. Oh and Y. H. Lee,. Phys. Rev. Lett. **79** (1997) 4621.

148. J. Nara, T. Sasaki, and T. Ohno, Phys. Rev. Lett. **79** (1997) 4421.

149. S. Jeong, and A. Oshiyama, Phys. Rev. Lett. **79** (1997) 4425.

150. K. Shiraishi, Appl. Surf. Sci. **60/61** (1992) 210.

151. **A. Kley, P. Ruggerone and M. Scheffler, Phys. Rev. Lett. 79 (1997) 5278.**

152. K. Shiraishi, T. Ito and T.Ohno, Solid State Electron. **37** (1994) 601.

153. T. Ito, K. Shiraishi and T. Ohno, Appl. Surf. Sci. **82/83** (1994) 208.

154. M. D. Pashley, Phys. Rev. B **40** (1989) 10481.

155. K. Shiraishi and T. Ito, J. Cryst. Growth **150** (1995) 158.

156. J. P. Harbison and H. H. Farrell, J. Vac. Sci. Technol. B **6** (1988) 733.

157. P. K. Larsen, J. F. van der Veen, A. Mzur, J. Pollmann, J. H. Neave and B. A. Joyce, Phys. Rev. B **26** (1982) 3222.

158. P. K. Larsen, J. H. Neave, J. F. van der Veen, P. J. Dobson and B. A. Joyce, Phys. Rev. B **27** (1983) 4996.

159. D. K. Biegelsen, R. D. Bringans, J. E. Northrup and L. -E. Swartz, Phys. Rev. B **41** (1990) 5701.

160. F. J. Lamelas, P. H. Fuoss, P. Imperatori, D. W. Kisker, G. B. Stephenson and S. Brennan, Appl. Phys. Lett. **60** (1992) 2610.

161. A. P. Payne, P. H. Fuoss, D. W. Kisker, G. B. Stephenson and S. Brennan, Phys. Rev. B **49** (1994) 14427.

162. T. Hashizume, Q. K. Xie, J. Zhou, A. Ichimiya

and T.Sakurai, Phys. Rev. Lett. **73** (1994) 2208.

163. **J. E. Northrup, and S. Froyen, Phys. Rev. Lett. 71 (1993) 2276.**

164. **J. E. Northrup, and S. Froyen, Phys. Rev. B 50 (1994) 2015.**

165. K. Shiraishi, and T. Ito, Surf. Sci. **357/358** (1996) 451.

166. K. Shiraishi and T. Ito, Phys. Rev. B **57** (1998) 6301.

167. T. Ito and K. Shiraishi, Jpn. J. Appl. Phys. **35** (1996) L949.

168. K. Shiraishi, T. Ito and T. Ohno, Sakaki, *Nanostructures and Quantum Effects*, Springer-Verlag, Berlin, 1994, pp.294.

169. K. Shiraishi, Thin Solid Films **272** (1995) 345.

170. S. B. Zhang, and A. Zunger, Phys. Rev. B **53** (1996) 1343.

171. S. B. Zhang and A. Zunger, J. Cryst. Growth **163** (1996) 113.

172. K. Shiraishi, Y. Y. Suzuki, H. Kageshima and T. Ito, Appl. Surf. Sci. **130/132** (1998) 431.

173. T. Ito, and K. Shiraishi, Jpn. J. Appl. Phys. **37** ,(1998) L488.

174. T. Ito and K. Shiraishi, Appl. Surf. Sci. **121/122** (1997) 171.

175. K. Shiraishi, and T. Ito, Appl. Surf. Sci. **121/122** (1997) 98.

176. M. D. Pashley, K. W. Haberern and J. M. Gaines, Appl. Phys. Lett. **58** (1991) 406.

177. S. Tsukamoto and N. Koguchi, J. Cryst. Growth **201/202** (1999) 118.

178. T. Ito and K. Shiraishi, Surf. Sci. **357/358** (1996) 486.

179. T. Ito and K. Shiraishi, Surf. Sci. **386** (1997) 241.

180. T. Ito and K. Shiraishi, Jpn. J. Appl. Phys. **37** (1998) 4234.

181. A. R. Avery, H. T. Dobbs, D. M. Holmes, B. A. Joyce and D. D. Vvedensky, Phys. Rev. Lett. **79** (1997) 3938.

182. H. Yamaguchi and Y. Horikoshi, Jpn. J. Appl. Phys. **28** (1989) L1456.

183. Y. Horikoshi, H. Yamaguchi, F. Briones and M. Kawashima, J. Cryst.Growth **105** (1990) 326.

184. H. Yamaguchi, Y. Horikoshi, Jpn. J. Appl. Phys. **30** (1991) 802.

185. T. Ito and K. Shiraishi, Jpn. J. Appl. Phys. **35** (1996) L1016.

186. T. Ito and K. Shiraishi, Jpn. J. Appl. Phys. **37** (1998) L488.

187. K. Seino, A. Ishii and T. Kawamura, Jpn. J. Appl. Phys. **39** (2000) 4285.

188. M. Itoh, G. R. Bell, A. R. Avery, T. S. Jones, B. A Joyce and D. D. Vvedensky, Phys. Rev. Lett. **81** (1998) 633.

189. M. Itoh, G. R. Bell, B. A. Joyce and D. D. Vvedensky, Prog. Theor. Phys. Suppl. **138** (2000) 90.

190. M. Itoh and T. Ohno, Phys. Rev. B **62** (2000) 7219.

191. M. Itoh and T. Ohno, Phys. Rev. B **63** (2001) 125301.

192. A. Taguchi, K. Shiraishi and T. Ito, Phys. Rev. B **60** (1999) 11509.

193. A. Taguchi, K. Shiraishi and T. Ito, Phys. Rev. B **61** (2000) 12670.

194. A. Taguchi, K. Shiraishi and T. Ito, Jpn. J. Appl. Phys.**39** (2000) 4270.

195. K. Sano and T. Nakayama, Jpn. J. Appl. Phys. **39** (2000) 4289.

196. Y. Kangawa, T. Ito, A. Taguchi, K. Shiraishi and T. Ohach, Surf. Sci. (in press).

This Page Intentionally Left Blank

PART II

Materials Design and Functionality of Advanced Materials

This Page Intentionally Left Blank

Si bulk crystal growth: What and how?

Koichi KAKIMOTO

Research Institute for Applied Mechanics, Kyushu University,
816-8580, 6-1, Kasuga-Koen, Kasuga, Japan

This paper covers recent developments in silicon technology related to the magnetic field-applied crystal growth method. The mechanisms of suppression and enhancement of melt flow under the conditions of static magnetic fields such as vertical, transverse and cusp-shaped fields are discussed. Moreover, the mechanisms under dynamic magnetic field: rotating magnetic fields and current induced magnetic fields are also introduced.

1. INTRODUCTION

Over the past fifty years, single crystals of silicon have become increasingly important materials in computer and information technology fields. Attempts to produce pure silicon (i.e., defect-free single crystal of silicon) were motivated by the desire to obtain ultra large-scale integrated circuits (ULSIs) in which micro-voids of about 10 nm diameter [1] are formed during crystal growth. Research over the past decade on crystal growth of silicon has focused on analysis of the formation of such micro-voids during crystal growth using mass transfer and reaction equations and a temperature field in the crystals, obtained from global modeling.

Micro-voids are formed by agglomeration of vacancies that are introduced at a solid-liquid interface of silicon. In most past studies, it has been difficult to reduce the total number of such micro-voids because the vacancy flux in silicon crystals must be controlled to reduce the probability of the agglomeration. One of the

key points for controlling the vacancy flux in crystals, especially that near a solid-liquid interface, is control of the convection of melt, by which the shape of the solid-liquid interface can be controlled.

From the above point of view, efforts have been made to control the periodic and/or turbulent flow of melt inside a crucible of large diameter. The crystal growth industries have mainly focused on quantitative prediction of a solid-liquid interface, point defect distribution, oxygen concentration, and dislocation free growth. Steady (DC) [2-35] and dynamic (AC) [36-41] electromagnetic fields are opening up new fields to meet an increasing demand for large-diameter crystals. Figure 1 shows a schematic diagram of rotating magnetic fields in a growth system.

This paper presents results of experimental and numerical investigations of convection of melt under conditions of DC and AC electromagnetic fields, such as vertical, transverse, cusp-shaped, and rotating fields, and electric current in the system of Czochralski

156

(CZ) crystal growth.

2.MAGNETIC FIELD-APPLIED CRYSTAL GROWTH

2.1 Steady electromagnetic hydrodynamics

Electromagnetic hydrodynamics has a long history in the field of steel manufacturing. Since molten silicon, like molten steel, has many free electrons, electromagnetic hydrodynamics can be used to analyze convection in metallic melt of molten silicon.

Electric current (J) in the melt and Lorentz force (F), which is induced by the current, in the case of a steady electromagnetic field are shown in eqs. (1) and (2), respectively, where σ_e, E, B and v are electric conductivity of the melt, electric field, magnetic flux density and velocity of the melt, respectively.

$$J = \sigma_e \left(E + v \times B \right) \qquad (1)$$

$$F = J \times B. \qquad (2)$$

Due to the continuity condition of an electric current in the melt, eq. (3), which is a Poisson-type equation, should be satisfied.

$$\nabla \cdot J = 0. \qquad (3)$$

Figure 2 shows a schematic diagram of a magneto-hydrodynamic system of metallic melt. If we apply this kind of static magnetic field to electrically conducting melt, magnetic flux density of the order of 0.1 T is necessary to effectively suppress convection of the melt in an actual crystal growth system.

2.2 Dynamic electromagnetic hydrodynamics

When dynamic magnetic fields are used, electric fields in the melt can be expressed by eqs. (4) and (5). The Faraday's equation; eq.

(5) can be expressed by vector potential in eq. (6).

$$E = -\nabla\Phi - \frac{\partial A}{\partial t}. \qquad (4)$$

$$\nabla \times E = -\frac{\partial B}{\partial t}. \qquad (5)$$

$$\nabla \times A = B. \qquad (6)$$

By combining eqs. (1) and (4), the following equation for electric current in melt can be obtained.

$$J = \sigma_e \left(-\nabla\Phi - \frac{\partial A}{\partial t} + v \times B \right). \qquad (7)$$

Then the following equation for Lorentz force can be obtained.

$$F = \sigma_e \left(-\nabla\Phi - \frac{\partial A}{\partial t} + v \times B \right) \times B. \qquad (8)$$

The Lorentz force contains three terms such as electric fields, velocity fields and time dependence of vector potential, which corresponds to the second term on the right-hand side of eq. (8). The time dependence of vector potential plays an important role in convection of melt through Lorentz force due to the time dependence of magnetic field density by rotating magnetic fields. The following equation can be derived from using eqs. (3) and (7).

$$\nabla \cdot \left(-\nabla\Phi - \frac{\partial A}{\partial t} + v \times B \right) = 0. \qquad (9)$$

Therefore, the Poisson equation shown in eq. (10) can express electric potential in the melt.

$$\nabla^2\Phi = -\frac{\partial}{\partial t}(\nabla \cdot A) + \nabla \cdot v \times B. \qquad (10)$$
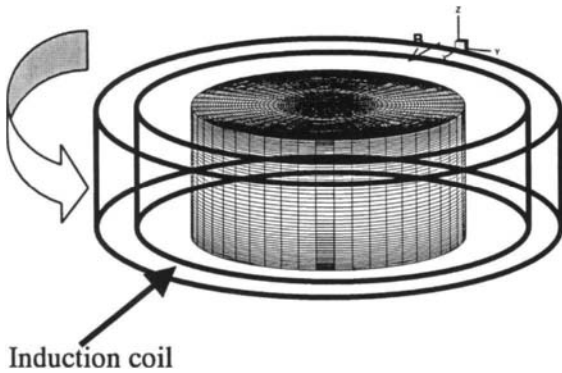
Figure 1. Schematic diagram of rotating magnetic fields.

Let us consider the mechanisms by which static and dynamic magnetic fields suppress and enhance the motion of metallic melts such as silicon and gallium arsenide (GaAs) semiconductors.

Figure 2 shows moving metallic melt inside parallel walls that are electrically insulated. Static magnetic fields are applied in the z-direction. Due to the coupling of magnetic field density and motion of the melt, the melt is subjected to Lorentz force in the x-direction, and the electrons therefore move in the x-direction. However, since the walls are electrically insulated, electrons accumulate near the wall. Consequently, electric potential is formed as is schematically shown in Fig. 2.

This phenomenon resembles the Hall effect in semiconductor solids, although the carriers of the melt and the solid are different. Since only the scattering electrons exists in the solid, we can obtain drift velocity of the free electron in the solids from the experiment of Hall effect.

If we can measure the electric potential difference, we have a possibility to estimate flow velocity from the electric potential

difference; however, it is very difficult to measure the electric potential difference in an actual system of crystal growth of semiconductors because semiconductor melts are chemically reactive, and their temperatures are high.
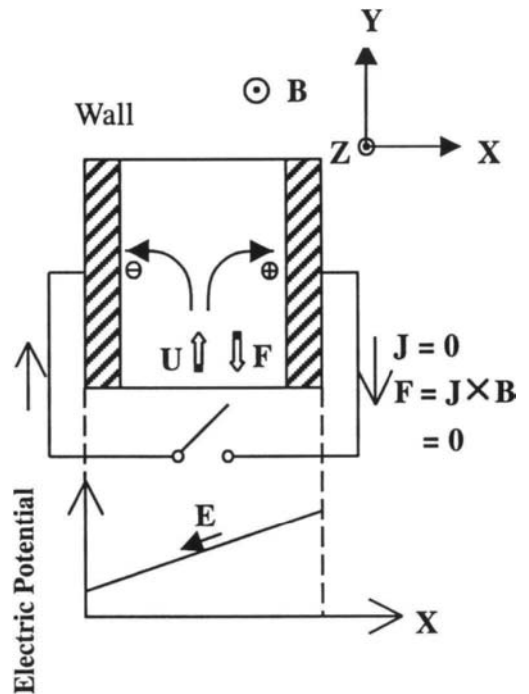
Figure 2. Electric current in metallic melt with electrically insulated walls under the condition of a magnetic field.

When the wall is electrically conductive, as shown in Fig. 3, electric current of the melt flows through the wall. Consequently, Lorentz force works effectively in the opposite direction to that of the melt motion. The force thus reduces the velocity of the melt.

The difference of electric potential becomes almost zero due to a large electric current that flows in the electrically conductive walls.
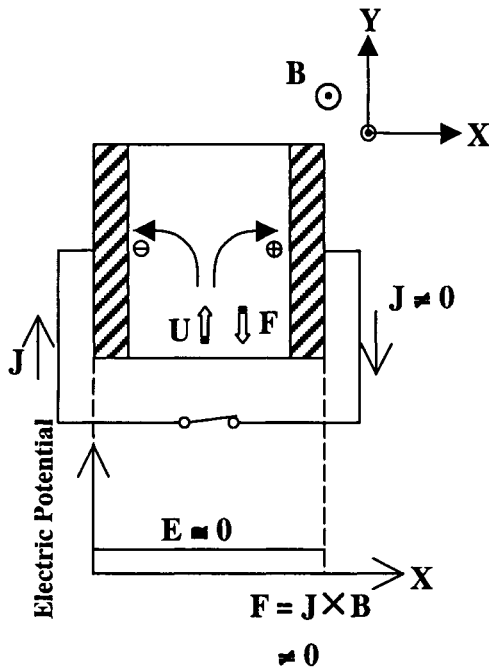
Figure 3. Electric current in metallic melts with electrically conductive walls under the condition of a magnetic field.

3. ACTUAL SYSTEM

3.1 Vertical magnetic fields

The vertical magnetic field-applied Czochralski (VMCZ) method was one of the methods used in early magnetic field-applied crystal growth. Figure 4 shows a schematic diagram of a VMCZ with a solenoid coil. Magnetic fields are applied in the z-direction; therefore, motion of the melt in the radial and/or azimuthal directions reacts with the magnetic field, while melt motion in the vertical direction does not react with the field.

Figure 5 schematically shows how electric currents flow and Lorentz forces work. Electric current is induced in the azimuthal direction by the radial motion of the melt under the vertical magnetic field. Therefore, Lorenz force works in the opposite direction. This

force suppresses velocity of the melt in the radial direction. Radial current, which is induced by a coupling with azimuthal velocity and magnetic field, cannot flow through a crucible wall due to electrically insulation of the wall. Thus, the Lorentz force cannot work, and the melt therefore flows freely in the azimuthal direction.
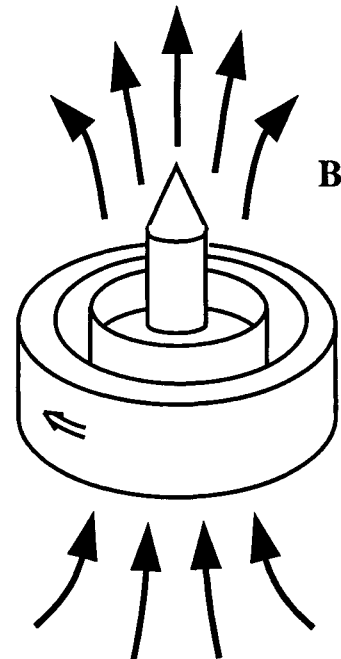
Figure 4. Schematic diagram of a VMCZ. Static current is applied to the cylindrical coil.

Figure 6 shows the experimental results of melt velocity in a meridional plane obtained by a visualization technique using an X-ray radiography method [22]. The dots show experimental data, and the lines obtained by numerical calculation using a three-dimensional configuration of the melt, show calculated results. This figure clearly shows that is a reduction in melt motion in a meridional plane. It was clarified from visualization that the motion in the azimuthal direction was not

suppressed.
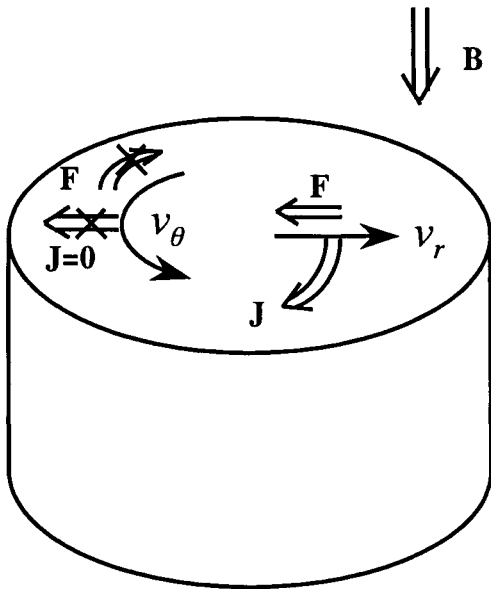
Figure 5. A schematic diagram of electric currents and Lorentz forces under vertical magnetic fields.

3.2 Transverse magnetic fields

The transverse magnetic field-applied CZ (TMCZ) method is only one case of MCZ to produce commercially available crystals in the present stage. Figure 7 shows a schematic diagram of the TMCZ system.

The TMCZ system has a non-axisymmetric configuration, and temperature and velocity fields therefore have two-folded symmetry. Although this system has such asymmetry, it has been used for actual production of charge-coupled devices (CCDs), since the system enables crystals with a low oxygen concentration to be produced. The CCDs should have a homogeneous and low oxygen concentration for reduction of inhomogeniety of image cells in the devices.

Since the system has asymmetry as shown

in Fig. 7, the flow and temperature fields in the system become three-dimensional. Figure 8 shows a schematic diagram of electric current and Lorentz force at the initial stage of application of transverse magnetic fields to the melt.
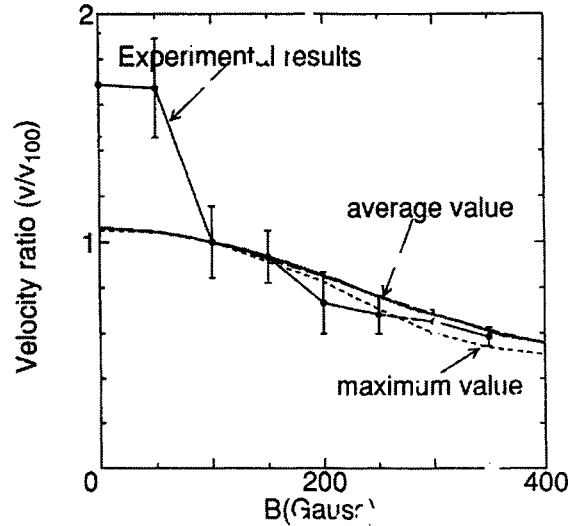
Figure 6. Relationship between magnetic fields and flow velocity in a meridional plane under vertical magnetic fields.
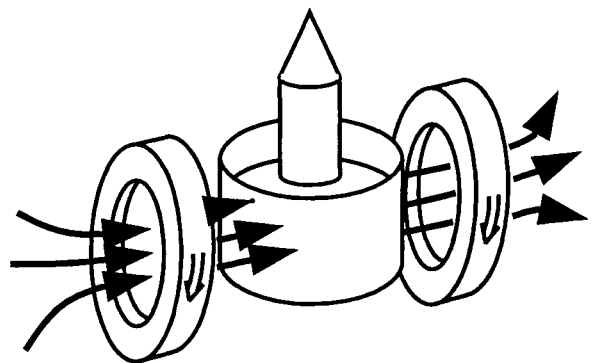
Figure 7. A schematic diagram of the TMCZ system.

An electric current induced by the magnetic field can flow in a direction parallel to

crucible wall at positions A and B in the figure. Therefore, Lorentz force works to the melt, and the melt motion is effectively suppressed.
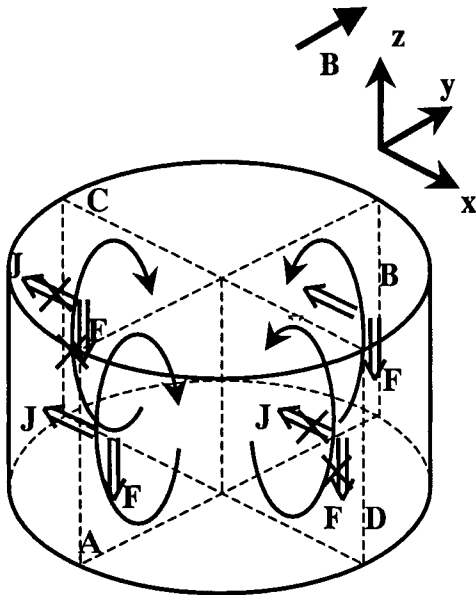
Figure 8. A schematic diagram of electric current and Lorentz force at the initial stage of application of transverse magnetic fields to the melt.

The situation regarding current flow and Lorentz force at positions C and D is the opposite to that of the above case. The electric current near the wall at positions C and D cannot flow into or from the electrically insulated wall. Therefore, the Lorentz force vanishes at these positions. Consequently, the melt motion cannot effectively be suppressed as shown in Fig. 9.

Only a downward flow remains in the y-z plane, while natural convection still exists in the x-z plane. Consequently, two roll cells remain in the melt in the case of a TMCZ system as shown in Fig. 10.

$$J = \sigma(-\nabla\psi + U \times B)$$
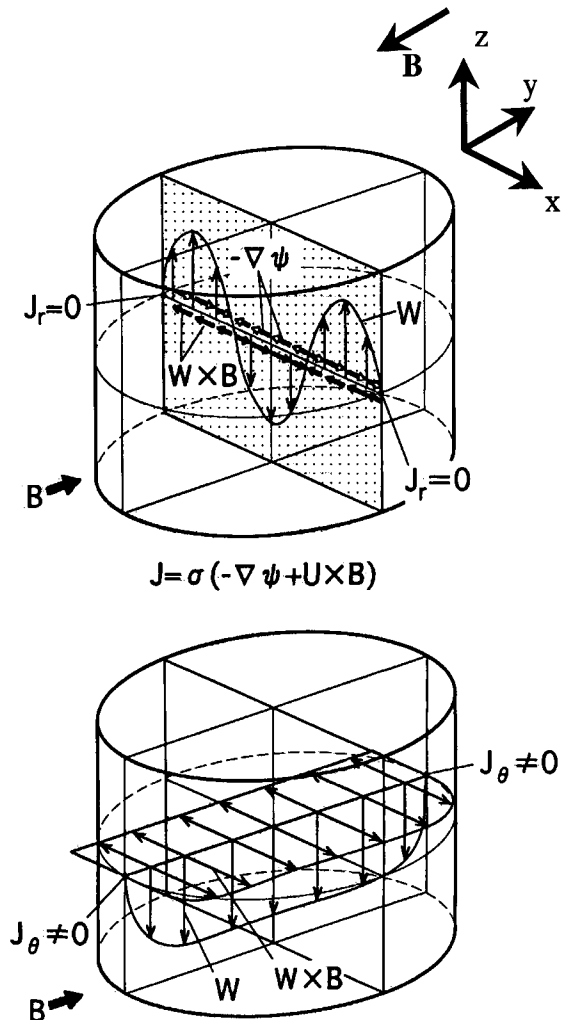
Figure 9. A schematic diagram of electric current and Lorenz force in the melt under transverse magnetic fields.

Due to the above electric boundary conditions, two main rolls aligned in parallel in the x direction remain. This means that the heat transfer in the melt in the x direction is larger than that in the y direction; therefore, an asymmetric temperature profile is formed in the melt.

161
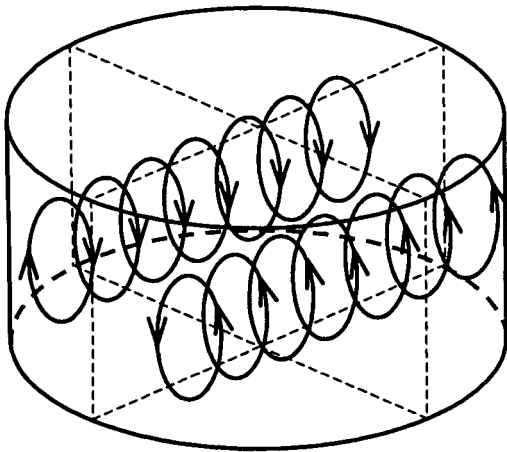
Figure 10. A schematic diagram of two roll cells in the melt under the condition of TMCZ.

Figure 12. An iso-surface of temperature in the melt.

Figure 11. Temperature contours at the top of the melt with magnetic field applied in the x direction.

Figure 11 shows temperature contours at the top of the melt when a magnetic field is applied in the x direction. An elliptic temperature distribution, which is due to inhomogeneous heat transfer in the melt can be seen.

Figure 12 shows an iso-surface of temperature in the melt under the condition of TMCZ. This result is due to the asymmetric velocity distribution shown in Fig. 13. Figures 13 (a) and (b) show velocity profiles in the x-z and the y-z planes, respectively. The figures indicate that two roles in a y-z plane enhance heat transfer from the crucible wall to the inside of the melt, while the downward flow in an x-z plane suppresses the transfer.

Calculation of such temperature and velocity distributions in the melt by time-dependent calculation shows that these asymmetric profiles are fixed in a laboratory frame except for a layer close to the crucible wall. Therefore, thin boundary layers of velocity, temperature, and oxygen near the crucible wall are formed. This phenomenon is a characteristic of a transverse magnetic field, which is static and non-axisymmetric. If an axisymmetric magnetic field such as a vertical or a cusp-shaped field is used, the melt rotates with the same angular velocity of the crucible.

162

(a)

(b)
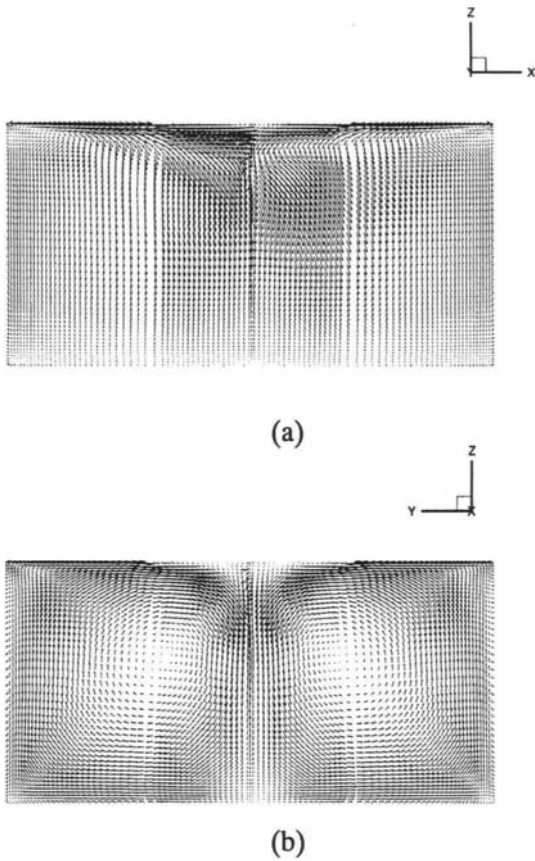
Figure 13. Velocity profiles in the x-z (a) and the y-z (b) planes.
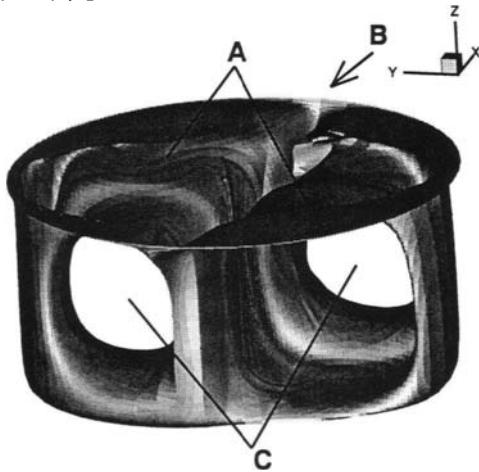
Figure 14. Calculated iso-surface of oxygen concentration in the melt under the condition of TMCZ.

Such asymmetric velocity profiles affect oxygen distribution in the melt. Figure 14 shows a calculated iso-surface of oxygen concentration in the melt. A and C in the figure show positions in which the oxygen concentration is large. This means that oxygen is distributed inhomogeneously, due to the velocity profiles shown in Figs.13 (a) and (b).

3.3 Cusp-shaped magnetic fields

The cusp-shaped magnetic field-applied CZ method has recently been used for actual crystal growth, especially for large-diameter crystals. Two solenoids, in which currents flow in opposite directions, can produce axisymmetric magnetic fields as illustrated in Fig. 15.
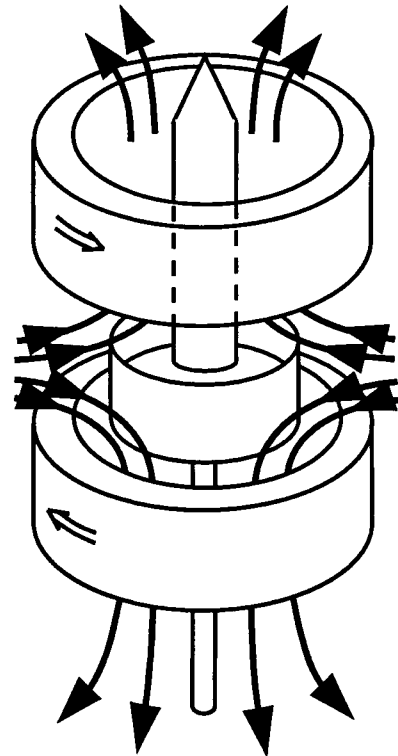
Figure 15. A schematic diagram of a cusp-shaped magnetic field-applied CZ system.

In this system, the diameters of the two solenoids, distance between them, and relative position between the solenoids and the melt can be freely selected. Therefore, an appropriate condition for the magnetic fields should be selected.

Since the diffusivity of oxygen in the melt is much smaller than that of temperature, therefore, we can recognize a fine structure, which is originated by the three-dimensional structure of velocity profile. Since the cusp-shaped magnetic fields are inhomogeneously distributed in the melt, electric current and Lorentz force distribution in the melt becomes complex, as shown in Fig. 16.

insulated, the electric current cannot pass though the crucible wall.
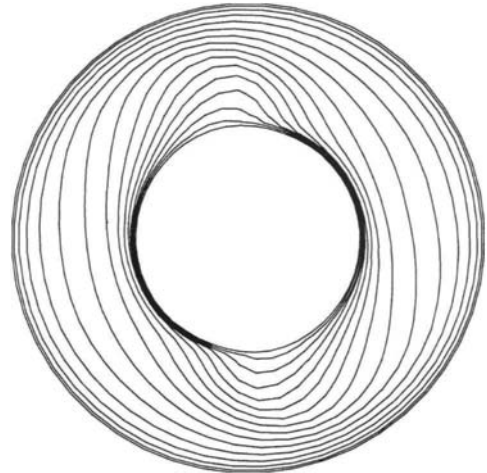
Figure 17. Calculated temperature field at the top of the melt under cusp-shaped magnetic fields.

Figure 16. A schematic diagram of electric current and Lorentz force distribution in the melt in the case of a cusp-shaped magnetic field-applied CZ system.
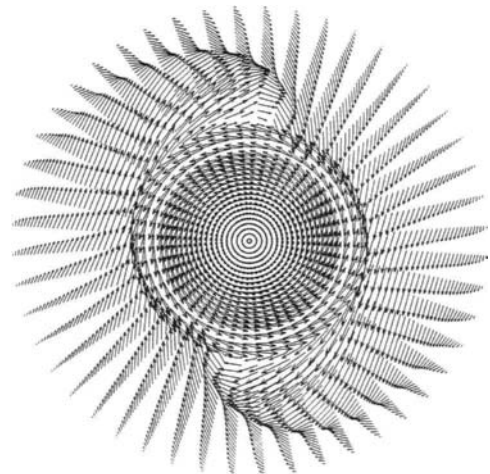
Figure 18. Calculated velocity field at the top of the melt under cusp-shaped magnetic fields.

Radial and vertical velocities induce electric currents along the crucible wall, which is electrically insulated, while the azimuthal component of velocity of the melt induces an electric current perpendicular to the crucible wall. Since the crucible wall is electrically

Figures 17 and 18 show calculated

temperature and velocity fields at the top of the melt under cusp-shaped magnetic fields. This is the same situation as that in the case of the VMCZ. Therefore, velocity in the azimuthal direction cannot be effectively suppressed; the melt rotates with the same angular velocity as that of the crucible.

The profiles show two-folded symmetry. Such asymmetric profiles rotate with the same angular velocity as that of crucible rotation. Moreover, some instability of flow still exists even in the case of an MCZ system.

3.4 Rotating magnetic fields

We have freedom to select the type of magnetic field from static and/or dynamic fields. If a static magnetic field is used, the purpose is to suppress the melt flow by using static fields. This technique is used to reduce velocity of the melt even under the condition of a crucible with a large diameter. Therefore, heat and impurity are transferred mainly by conduction. This means that the profiles of temperature and impurity become inhomogeneous due to small transfer rates.

The rotating magnetic field, which is a dynamic field, enhances rotation and circulation of melt flow. Since this method enables almost complete mixing of the melt, heat and impurity transfers from crucible to crystals become large. Figure 19 shows the velocity profile at a specific period in an x-z plane. A schematic diagram of the velocity profile is shown in Fig. 20. This figure shows that the flow is coming in from a crucible wall to the center of the melt. This makes it possible to increase oxygen concentration in the melt; thus, oxygen concentration in the crystal becomes large.

Figure 21 shows the distribution of electric potentials in the melt. This distribution is based on the distribution of flow velocities

shown in Fig. 19.
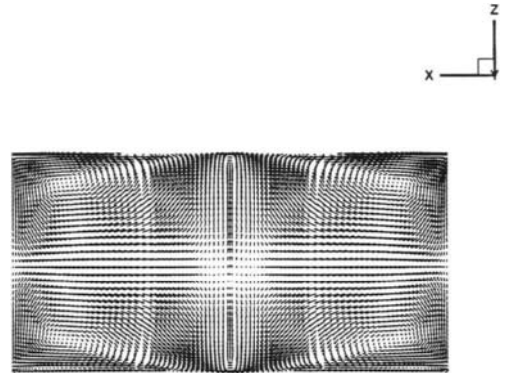
Figure 19. Velocity profile at a specific period in an x-z plane.
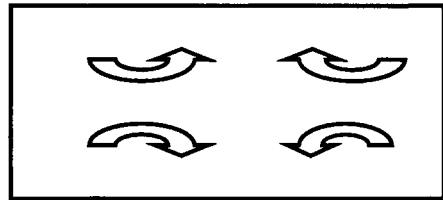
Figure 20. A schematic diagram of the velocity profile in the melt under a rotating magnetic field.

41	7.03695E-05
31	2.48036E-05
21	-2.07623E-05
11	-6.63282E-05
1	-0.000111894
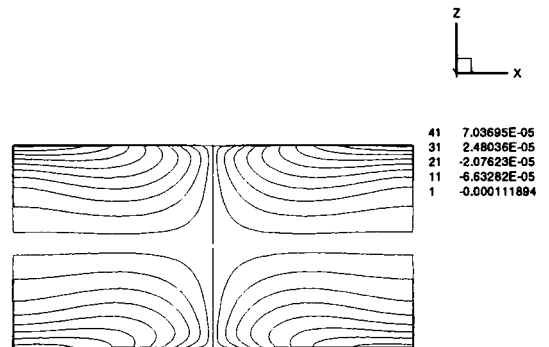
Figure 21 Distribution of electric potentials in the melt.

Figure 22 shows the temperature distribution at the top of the melt. It can be

seen that the profile is not axisymmetric; thus there exists some instability of flow in the melt under a rotating magnetic field.
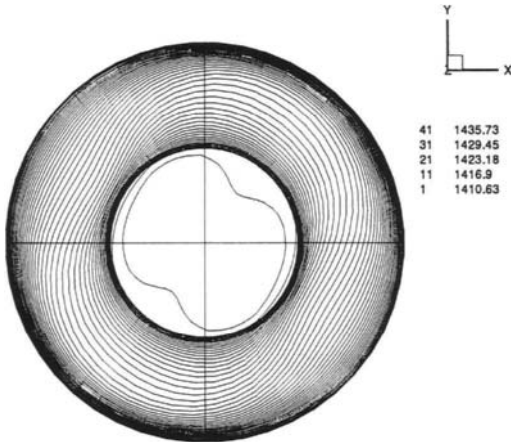
Figure 22. Temperature distribution at the top of the melt of silicon under a rotating magnetic field.

4. SUMMARY

We can select an appropriate magnetic field from various types of magnetic fields to control melt flow. However, there is a need to understand the phenomena occurring in the melt under various magnetic fields. The most appropriate magnetic field can be selected by taking into account how the melt flow couples not only the magnetic fields but temperature distribution in the furnace. If we select one type of magnetic field, computer simulation technique helps us to predict temperature, velocity and impurity concentration more quantitatively, since the technique is developing rapidly. Although current computer power is still not sufficient to be able to reproduce all of the phenomena occurring in a furnace in the case of a magnetic field-applied system, we are going to approach more quantitative prediction of the phenomena by using new algorithm and hardware.

REFERENCES

1. M. Itsumi, H. Akiya, and T. Ueki, The composition of octahedron structures that act as an origin of defects in thermal SiO2 on Czochralski silicon, J. Appl. Phys., **78**, 5984-5988, 1995.
2. H. Yamagishi, M. Kuramoto, and Y. Shiraishi, Solid State Phenom., **57-8** (1997) 37.
3. Y. C. Won, K. Kakimoto, H. Ozoe, Numerical Heat Transfer, A**36** (1999) 551.
4. Kyung-Woo Yi, Masahito Watanabe, Minoru Eguchi, Koichi Kakimoto and Taketoshi Hibiya, Jpn. J. Appl. Phys., **33**, (1994) L487.
5. M. G. Williams, J. S. Walker and W. E. Langlois, J. Crystal Growth, **100** (1990) 233.
6. A. E. Organ and N. Riley, J. Crystal Growth, **82** (1987) 465.
7. J. S. Walker and M. G. Williams, J. Crystal Growth, **137** (1994) 32.
8. J. Baumgaartl, M. Gewald, R. Rupp, J. Stierlen and G. Meuller, Proceedings of the VIIth European Symposium on Materials and Fluid Sciences in Microgravity, Oxford, UK, pp.10 (1989).
9. L. N. Hjellming and J. S. Walke, J. Crystal Growth, **87** (1988) 18.
10. S. Kobayashi, J. Crystal Growth, **85** (1987) 69.
11. M. Akamatsu, K. Kakimoto, H. Ozoe, Transport phenomena in thermal science and process engineering **3**, (1997) 637.
12. K-W Yi, K. Kakimoto, M. Eguchi, M. Watanabe, T. Shyo and T. Hibiya, J. Crystal Growth, **144**, (1994) 20.

13. K. Kakimoto and H. Ozoe, J Crystal Growth, **212** (2000) 429.

14. R. A. Brown, T. A. Kinney, P.A. Sackinger and D. E. Bornside, J. Crystal Growth, **97** (1989) 99.

15. H. Hirata and N. Inoue, Jpn. J. Appl. Phys. **23** (1984) L527.

16. H. Hirata and K. Hoshikawa, J. Crystal Growth **96** (1989) 747.

17. H. Hirata and K. Hoshikawa, J. Crystal Growth **98** (1989) 777.

18. H. Hirata and K. Hoshikawa, J. Crystal Growth, **113** (1991) 164.

19. K. Hoshi, T. Suzuki, Y. Okubo and N. Isawa, Extended Abstracts Electrochem. Soc. Spring Meeting (The Electrochem. Soc., Pennington, 1980) vol.80-1, p.811.

20. K. Hoshikawa, Jpn. J. Appl. Phys. 21 (1982) L545.

21. K. Hoshikawa, H. Kohda and H. Hirata, Jpn. J. Appl. Phys. 23 (1984) L37.

22. K. Kakimoto, H. Watanabe, M. Eguchi and T. Hibiya, J. Crystal Growth 88 (1988) 365.

23. Kakimoto K., Proceedings of 2nd Workshop on High Magnetic Fields, (1995) Florida, ed. by M. Schneider.

24. K. Kakimoto, Prog. Crystal Growth and Charact., **30** (1995) 191.

25. K. Kakimoto, Y. W. Yi and M. Eguchi, J. Crystal Growth, **163** (1996) 238.

26. K. Kakimoto, and M. Eguchi, J. Crystal Growth, **116** (1996) 1257.

27. K. M. Kim and W. E. Langlois, J. Electrochem. Soc. 133 (1986) 2586.

28. A. E. Organ and Riley, J. Crystal Growth **82** (1987) 465.

29. Z. A. Salnick, J. Crystal Growth 121 (1992) 775.

30. T. Suzuki, N. Isawa, Y. Okubo and K. Hoshi, Semiconductor Silicon 1981, eds. H. R. Huff, R. J. Kriegler and Y. Takeishi (The Electrochem. Soc., Pennington, 1981) p.90.

31. R. N. Thomas, H. M. Hobgood, P. S. Ravishankar and T. T. Braggins, Solid State Technol (April 1990) 163.

32. M. Watanabe, M. Eguchi, K. Kakimoto and T. Hibiya, J. Crystal Growth **128** (1993) 288.

33. M. Watanabe, M. Eguchi, K. Kakimoto and T. Hibiya, J. Crystal Growth **131** (1995) 285.

34. A. F. Witt, C. J. Herman and H.C.Gatos, J. Mater. Sci. **5** (1970) 882.

35. K.-W.Yi, M. Watanabe, M. Eguchi, K. Kakimoto and T. Hibiya, Jpn. J. Appl. Phys. **33** (1994) L487.

36. Y. Gelfgat, J. Krumins, B. Q. Li, J. Crystal Growth, **210** (2000) 788.

37. Y. Gelfgat, Jpriede, Magneto-hydrodynamics, **31** (1995) 102.

38. R. U. Barz, G. Gerbeth, Y. Gelfgat, J. Crystal Growth, **180** (1997) 410.

39. T. Kaiser and K. W. Benz, Phys. Fluids, 10 (1998) 1104.

40. F.-U. Brucker and K. Schwerdtfeger, J. Crystal Growth, **139** (1994) 351.

41. J. Virbis, Th. Wetzel, A. Muiznieks, B. Hanna, E. Dornberger, E. Tomzig, A. Muhlbauer, W. v. Ammon, Proceeding of the Third International Workshop on Modeling in Crystal Growth, (2000) 31.

Optimization of Melt Growth Processes by Experimental Analysis and Computer Modeling

G. Müller[a,b] and B. Fischer[b]

[a]Crystal Growth Laboratory, Dept. of Materials Science (WW6), University Erlangen-Nürnberg, Martensstr. 7, 91058 Erlangen, Germany

[b]Fraunhofer Institute of Integrated Circuits IIS-B, Schottkystr. 10, 91058 Erlangen, Germany

Strategies and tools are introduced how to optimize melt growth processes. Both experimental analysis and modeling by computer simulation play an important role. The success and power of this procedure is demonstrated by actual examples of the field of Czochralski growth of Si and Vertical Gradient Freeze (VGF) of GaAs.

1. INTRODUCTION

The most frequently used and most important method of producing bulk single crystals of a given material is by solidification of its own melt – called "crystal growth from the melt" or just "melt growth". Its technical and economic importance is due to the fact that large single crystals (several kilograms up to 300 kg) can be grown very efficiently and with a high yield at a relatively high growth rate (several millimeters per hour up to millimeters per minute). This growth rate is not limited by the transport of crystal species but in principle only by the removal of the heat of crystallization.

The control of the heat transport, therefore, plays an important role in the process optimization as will be shown later. This holds also with respect to the temperature gradients, which are causing thermal stress in the growing crystal. It may not exceed a certain value (specific for each material) in order to avoid unwanted crystal imperfections, e.g. dislocations.

Several techniques of melt growth will be described in chapter 3.

The main application of melt-grown crystals is in electronic industry, which consumes huge amounts of semiconductor single crystals (nearly 10000 t of silicon and about 600 t of GaAs per year). Also hundreds of tons of optical crystals are produced by melt growth, like ruby, sapphire, garnets, molybdates, zirconia and fluorides.

Melt growth cannot be used in cases where the material to be grown has no congruent melting point and where the melting point T_m or the vapor pressure p_m are too high (typical limits are $T_m < 2000\,°C$, $p_m < 100$ bar). In these cases crystals have to be grown from solutions or from the vapor phase.

2. STRATEGY OF PROCESS OPTIMIZATION IN CRYSTAL GROWTH

At first one has to answer the question what is meant by the "*optimization of a crystal growth process*"?

The primary goal in crystal growth is the fabrication of a single crystalline material with *certain physical and chemical properties* defined by the purpose of application of the crystal, either for R+D or fabrication of any devices. Physical and chemical properties of a crystal are defined by the structure of the crystal. There are several important ideas about structure and crystal growth, which need to be clarified. Thermodynamics dictate the *ideal structure* of any crystalline material; this is given by nature through the phase diagram. The *ideal (crystallographic) structure* can normally not be changed or manipulated by the crystal growth process, except in some special cases like hetero-epitaxy of strained layers, where even the crystallographic structure can be changed by lowering the symmetry.

However, the process of actually growing a crystal never results in this ideal structure. The *real*

structure, i.e. that obtained under the conditions existing during the crystal growth process, always exhibits defects. Crystal defects are defined as the difference between the *real structure* and the *ideal crystallographic structure* of a crystal. Crystal defects in many cases define the relevant electronic, optical and mechanical properties of a crystal.

Producing a *real crystal* with certain properties means, therefore, *defect engineering*, which is a good term to characterize an important task of crystal growth. Defect engineering is accomplished during melt growth by appropriate modification of growth conditions. It should be mentioned that part of this defect engineering can also be the post growth treatment of a crystal, i.e. the cooling down or annealing procedures. But also in these cases the success of the post growth procedures is strongly depending on the as-grown properties of the crystal.

Now one can use these correlations according to figure 1 to define an optimized crystal growth process as one which allows to control the formation of certain crystal defects, as well as avoiding others which are unwanted, by a precise control of a well defined set of crystal growth parameters.
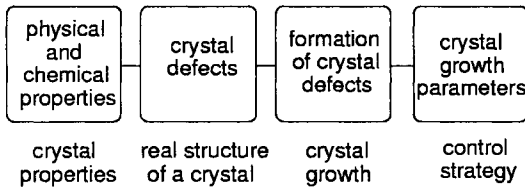
Figure 1. Schematic relation of crystal properties and crystal growth processing.

In the following chapter 3 some important techniques of melt growth will be briefly explained. Later on it will be discussed what *different kinds of parameters*, like geometry, temperature fields, gas atmosphere (pressure, flow), mechanical movements (crystal/crucible rotation), magnetic fields, and even the direction of gravity etc., have a certain influence on the crystal properties. As a result it comes out clearly that a typical melt growth process like crystal pulling (Czochralski method) is a very complex procedure, as the interaction of various highly non-linear coupled physico-chemical processes has to be controlled by a large set of parameters.

One can define *optimization of a crystal growth process* by searching for a process window in a n-fold parameter space, as sketched in figure 2. This process window Ω is defined by a set of parameters

x_1^s, x_2^s,... x_n^s, which provide growth conditions for the crystal with the wanted physical and chemical properties according to figure 1. It is important that the n-fold region Ω, which corresponds to the *optimized crystal growth process*, has certain extensions in all n directions of the parameter space, as sketched by the n-fold box in figure 2. Only a minimum size of the n edges of the box guarantees that unavoidable small variations δx_1, δx_2,... δx_n of the parameters (arrows in figure 2) keep the system still under stable growth conditions, which provide the specified crystal properties. Typical unavoidable variations δx of the parameters in a crystal growth run or from run to run can have various origins like tolerances in mechanical parts, aging of thermocouples, chemical reactions of heaters and shielding materials, or fluctuations in the whole system.
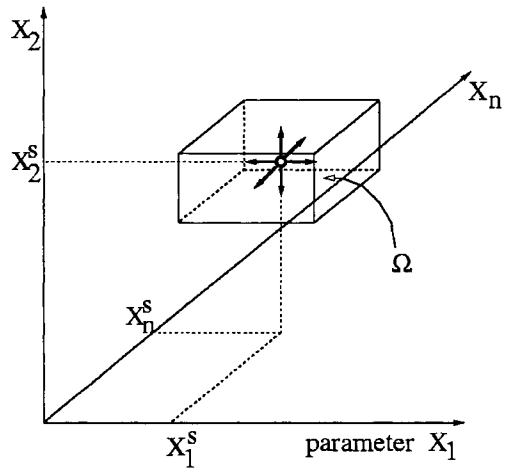
Figure 2. n-fold parameter space for crystal growth. Ω is the process window with optimum growth conditions, defined by the parameters x_1^s, x_2^s,... x_n^s.

Each crystal supplier must guarantee a certain set of crystal properties, which are written down in a data sheet called *crystal spec*, from "specification". Crystals which are "out of spec" are waste. Finding the box Ω in figure 2 is thus essential for the economic success of a crystal growth company.

The goal of this article is to give strategies how crystal growth processes can be optimized, i.e. how the box Ω in figure 2 can be found, by using the possibilities of experimentally analyzing a crystal growth configuration and by rigorously taking advantage of computer modeling. The field of applica-

tion of this strategy is nearly the whole field of crystal growth, bulk growth as well as epitaxy. But for two reasons this article is focussing mainly on melt growth of semiconductors. Firstly, due to the limited space it is necessary to concentrate on one subject. Secondly, the authors are most experienced in that subject.

3. IMPORTANT FEATURES OF INDUS-TRIAL MELT GROWTH PROCESSES

3.1. Fundamentals

Processes of melt growth are based on the liquid-solid phase transition of an element or a compound. The use of equilibrium thermodynamics and especially phase diagrams is very helpful in finding the appropriate growth conditions. Melt growth processes are characterized by relatively low driving forces (Gibbs free energy), as homogeneous nucleation has to be avoided in the supercooled melt region. All melt growth techniques make use of an oriented seed crystal typically of the same material (heterogeneous nucleation).

On a microscopic scale, crystal growth from the melt occurs mainly under conditions of an atomically rough interface. This means that the shape and position of the solid-liquid interface is nearly identical to the melting point isotherm. Only in certain cases, where the growth interface is formed by a highly packed dense atomic face, e.g. (111) in semiconductors, the crystal can grow with an atomically smooth interface deviating from the curved shape of the isotherm T_m. Such planar sections of the interface are called facets.

In any case, melt growth is not limited by growth kinetics but by the transport rate of heat. For that reason the mechanisms of heat transport are treated in detail in chapter 4. For further details on the fundamentals of melt growth, the reader is referred to text books like Rosenberger [1], earlier Schools on Crystal Growth, e.g. ISSCG10 [2], and the Handbook of Crystal Growth [3].

3.2. Melt growth techniques

The *Czochralski (Cz) method* is the dominant method for the production of bulk single crystals of a wide range of electronic and optical materials (figure 3). The feed material is melted in a cylindrically or bowl-like shaped crucible with a free liquid surface at the top. The heating of the melt is provided by resistance or radio-frequency heaters. It is important to adjust the power of the heater(s) carefully so that a certain portion of the dipped seed is remelted and a melt meniscus is formed. The pulling rod, i.e. the seed, is slowly lifted (often under rotation) and the melt crystallizes at the interface of the seed by forming a new crystal portion. The shape of this crystal, especially the diameter, is controlled by adjusting the heating power, pulling rate, and rotation rate of the pulling rod, i.e. the crystal.
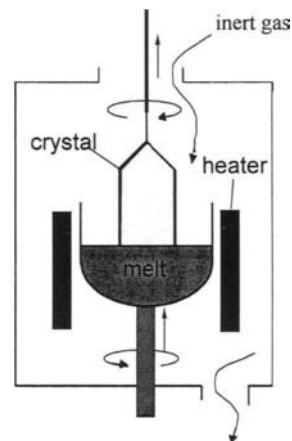
Figure 3. Schematic principle of the Czochralski (Cz) method.

For economic reasons it is important to grow the cylindrical part with a constant diameter and a length as large as possible. For this purpose an automatic diameter control is necessary, either by control of the meniscus shape (e.g. for silicon) or by weighing the crystal (e.g. for GaAs, InP) or the melt (for oxides). As the gas atmosphere has to be controlled during the growth process, the whole assembly is maintained in a vacuum-tight vessel, which is filled with a gas (inert gas for semiconductors, oxygen or air for oxides).

Materials with a high partial pressure of one or more components can also be grown by the Cz principle by using a liquid encapsulating of the melt surface (*Liquid Encapsulated Czochralski (LEC) method*). The main disadvantage of the LEC method is the high thermal stress in the growing crystal caused by large temperature non-linearities and large axial temperature gradients (about 100 Kcm^{-1}) owing to the low thermal conductivity of the liquid encapsulant. This problem can be overcome by other growth methods described below. This fact is illus-

trated by a plot of axial temperature profiles in GaAs crystals grown by various methods in figure 4.

The *Vapor-pressure-controlled LEC (VLEC or VCz) method* is a variant of the LEC method, which was introduced to reduce the thermal stress in the growing crystal by an *in-situ* after-heating of the grown crystal (see also figure 4).
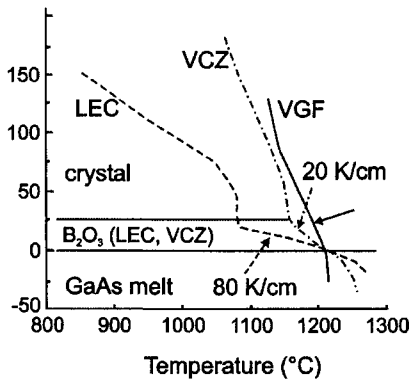

Figure 4. Temperature profiles in the crystal for different methods of GaAs growth.

All variants of the Cz method and also other growth techniques can in principle be combined with static or time-dependent magnetic field devices, which are used for the damping of temperature fluctuations (unsteady convective flow) or for controlling convective species transport.

Another group of important melt growth techniques is based on the principle of directional solidification. The growth of single crystals by melting a charge in a crucible and freezing the melt from one end (seed) to the other (tail) is used for many materials since a long time. However, the strong interaction of the growing crystal with the crucible material limits the yield, and hence the industrial application is limited so far. This method can be carried out by moving the growth interface in a horizontal or a vertical direction. The crystal growth configuration consists typically of a tube furnace, which provides a temperature profile with negative gradients parallel to the growth direction (at least in the vicinity of the growth interface but mainly in the major portion of the melt and crystal). The single-crystal seed is positioned at one end of the horizontal boat or the lower end of the vertical crucible. Crystal growth is carried out by a controlled shifting of the temperature profile relative to the boat or crucible as shown in figure 5. This can be achieved in three ways: by mechanical movement of the crucible relative to the

fixed furnace (introduced by Bridgman in 1925 and Stockbarger in 1936); by mechanical movement of the furnace relative to the fixed crucible; and without any mechanical movement, the shifting of the temperature profile being achieved only by a controlled changing of the powers of the furnace heaters (Gradient Freeze method). The vertical configurations (*Vertical Gradient Freeze (VGF)*, *Vertical Bridgman (VB)*) are preferred for the industrial production of crystals because of the higher yield of round wafers compared to the horizontal methods. The increasing interest in the use of the VGF technique results from the fact that it uses the simplest principle of melt growth and has several advantages compared to the Cz methods. It operates under stable hydrodynamic conditions, it is suitable for computer modeling and for automatic process control, the thermal stress in the growing crystal is lower (see figure 4), round crystals without any diameter control are produced, and the equipment costs are lower.
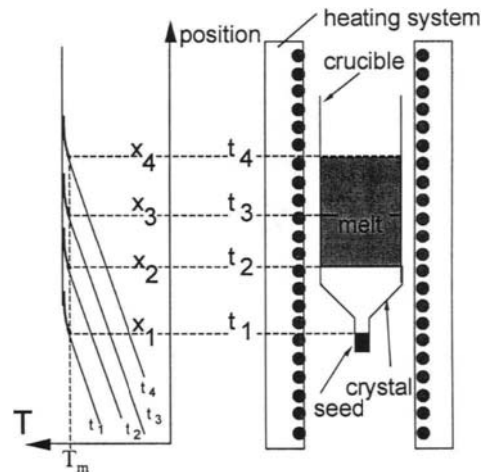
Figure 5. Sketch of furnace temperature profiles (left) for different time steps in a VGF configuration (right).

For more details and descriptions of other melt growth processes, like Kyropulos, continuous feed Czochralski, growth of shaped crystals from dies, Zone Melting, Verneuil, the reader is referred to the literature [3,4,5,6].

3.3. Strategy for optimization

Even though the crude description of the melt growth processes in the previous section looked very simple, the reader should be well aware that optimi-

zation of such processes is a very difficult task. It is necessary in a first step to extract from the crystal growth system the relevant, governing physical and chemical processes (*physico-chemical model*). In a next step, the physical model is transferred into a set of coupled differential equations with certain boundary conditions (*mathematical model*). As the mathematical model cannot be solved directly, methods of numerical treatment have to be used (*numerical model*). Finally, this in principle allows for a *computer simulation* of the crystal growth process. The key question – whether the simulation results are coming close enough to reality – can only be answered by *experimental validation*. This means that the whole strategy consists of sophisticated computer modeling and (!) experimental analysis of a given growth process, which will be shown in the following chapters.

4. PHYSICO-CHEMICAL MODEL

Physico-chemical modeling of melt growth involves many different and interacting processes from thermodynamics, hydrodynamics, chemistry and even electrodynamics. The properties of the grown crystal are determined by the thermal and concentration fields in the close vicinity of the crystal, especially at the phase boundary. However, this vicinity is strongly influenced by the temperature and species transport in the whole growth set-up. So finding a model of crystal growth means finding an appropriate model of the whole crystal growth system.

The set-up can roughly be subdivided into three (sub)systems. This is illustrated in figure 6 for a schematic LEC set-up. The innermost system is the material itself which is to be grown, i.e. the crystal and the melt, including the solid-liquid phase boundary (1). As a second step, the materials directly adherent to crystal and melt can be included to form some kind of "growth environment" (2). The final step is the consideration of all surrounding parts of the system (3). I.e., in different parts of the system different physical mechanisms may be important. The corresponding transport processes are normally represented by a set of coupled differential equations. Now it is important to correctly describe what occurs in the subsystems of figure 6, and what at the interfaces between the different regions and materials.
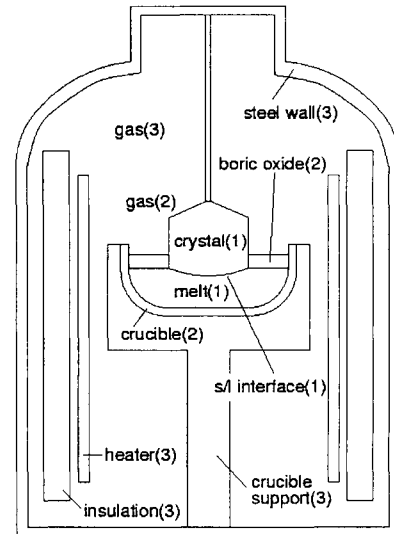
Figure 6. Subsystems of a typical melt growth system.

4.1. Heat transfer

Heat transfer, which determines the temperature distribution in the furnace, is governed by three mechanisms: Heat conduction, radiation, and convection.

Typical sources of heat, i.e. thermal energy, in a crystal growth setup are ohmic heaters and Joule heat from inductive heating with alternating electromagnetic fields.

4.1.1 Heat transfer by conduction

Heat conduction is the most general and maybe the most important mechanism of heat transport in crystal growth set-ups, because it occurs in every material. It means energy transport due to atomic and molecular interaction under the influence of a non-uniform temperature distribution.

The basic law of heat conduction is Fourier's law:

$$\vec{q} = -\lambda \nabla T \qquad (1)$$

It relates the heat flux density $\vec{q}$ to the gradient ∇T of the temperature. The thermal conductivity λ in general depends on temperature and pressure, and is a tensor for anisotropically conducting materials. Typical values for λ vary from 0.01-0.2 W/mK for gases over 0.5-10 for (insulating) solids up to 5-400 for graphite and metallic components of a crystal growth system. The partial differential equation

determining the temperature field is equivalent to the conservation of energy, and for a constant thermal conductivity λ looks like:

$$\rho c_p \frac{\partial T}{\partial t} = -\rho c_p (\vec{u} \cdot \nabla) T + \lambda \nabla^2 T + s \qquad (2)$$

This equation says that the temporal change of the temperature field is influenced by convective (see section 4.1.3) and diffusive transport mechanisms as well as by heat sources s. Without heat sources s and convection ($-\rho c_p (\vec{u} \cdot \nabla) T$) this leaves Fick's law of diffusion.

The following interfaces between different materials are relevant to be treated by a global physical and hence numerical model:

- The *outer boundary of the system* often consists of a water-cooled steel wall. Therefore, the temperature can be assumed to be equal to a fixed value, e.g. room temperature.

- The *boundary condition between two materials*, is described by a heat flux equilibrium:

$$\lambda_1 \frac{\partial T}{\partial \vec{n}}\bigg|_1 = \lambda_2 \frac{\partial T}{\partial \vec{n}}\bigg|_2 \qquad (3)$$

- The *moving solid-liquid interface* between melt and crystal includes some special features, which are discussed in section 4.3.

If only a part of the facility is to be modeled, the heat flux density $-\lambda \nabla T \cdot \vec{n}$ normal to the boundary may be set equal to a given value, to a heat transfer $h(T - T_{ref})$ or a radiative expression $\varepsilon \sigma (T^4 - T_{ref}^4)$.

4.1.2 Heat transfer by radiation

Heat transfer by radiation happens in every fully or partly transparent material. Typical examples in melt growth set-ups are oxide crystals, quartz crucibles, boric oxide melt covers, and the gas space (e.g. argon, nitrogen). Radiation happens very fast (at the speed of light), so any time dependence can be neglected. Moreover, it needs no medium - the energy is transported by photons. In crystal growth for example, this is significant for low pressure gas environments. Radiative heat transport is extremely nonlinear, the radiative heat flux from a surface with temperature T is proportional to T^4. Working at high temperatures of much more than 1000 K, this emphasizes the importance of radiative heat transport in melt growth environments. Radiation is strongly influenced by the geometry of the set-up,

because one point on a surface emits radiation in all directions and receives radiation from all directions, and also by the surface properties of the materials, which are often not very well known. In crystal growth facilities, opaque materials (e.g. graphite, steel) can normally be assumed to have an ideal diffuse grey surface, which means the total emitted radiation power per area S can be calculated as $S = \varepsilon \sigma T^4$, with $0 \le \varepsilon \le 1$ being the emissivity and $\sigma = 5.67 \cdot 10^{-8}$ W/m^{2}K^4 the Stefan-Boltzmann constant. The special case of $\varepsilon = 1$ implies a black body. The most common way of treating radiation in crystal growth set-ups is to calculate the so-called view factors [7]. They describe which fraction of the total radiation power emitted by one radiation tile reaches another radiation tile, determining the net energy exchange between each pair of tiles.

Sometimes there is a spectral dependence of absorption, transmission and reflection, i.e. a material is transparent only in certain wavelength intervals and absorbing or opaque in others, as illustrated in figure 7.
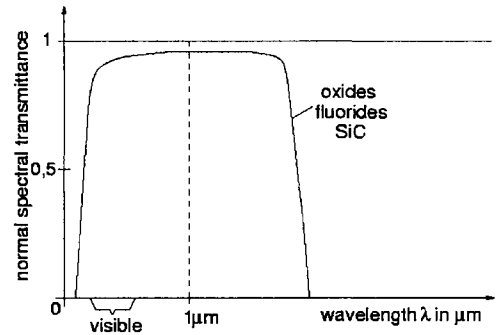
Figure 7. Spectral transmittance of examples for semitransparent materials (schematic).

Semitransparency means that absorption does not only take place at the surface of a material, but energy is partly transmitted through the body, and also absorbed and emitted within the body. Common approaches to model semitransparency are the n-band model, which assumes certain wavelength bands either to be fully transparent or fully opaque [8], or so-called P_N approximations, which describe the direction-dependent radiation intensity in absorbing and scattering media with the help of spherical harmonics [7].

4.1.3 Heat transfer by convection

In a typical melt growth setup, convection happens in the melt of the grown material and in the gas space. With the thermal conductivity of gases being very small, convection besides radiation has an important influence on the global heat transport in the gas space of the furnace. Gas convection is driven by in- and outflows, and especially by buoyancy due to temperature differences in the facility. Especially at high gas pressures, gas convection is turbulent and reaches values of meters per second.

Much more multifarious are the influences driving melt convection, especially in Czochralski set-ups. They can be subdivided into two kinds of forces. Firstly, *contact forces* applying at the boundaries of the melt, like thermocapillary (Marangoni) convection or crystal/crucible rotations. And secondly, *bulk forces* affecting the melt volume, like buoyancy driven by thermal or solutal gradients. Also magnetic fields act as bulk forces in electrically conducting melts. They have become an essential tool to control semiconductor melt convection, either by damping (stationary fields) or by controlled stirring (time-dependent fields) [9,10]. These forces are illustrated in figure 8 for a Czochralski melt.

Mechanism	Configuration	Dimensionless number	Description
Buoyant convection buoyancy due to thermal expansion		Rayleigh number $Ra = \dfrac{g\beta r^3 \Delta T}{\nu \kappa}$	$\dfrac{\text{buoyant forces}}{\text{viscous forces}}$
Marangoni convection force due to temperature dependence of surface tension		Marangoni number $Ma = -\dfrac{\partial \gamma}{\partial T}\dfrac{\Delta T r}{\nu \kappa}$	$\dfrac{\text{resulting surface forces}}{\text{viscous forces}}$
Forced convection crucible rotation centrifugal force, Coriolis-force		Taylor number $Ta = \dfrac{4\omega^2 r^2}{\nu^2}$	$\dfrac{\text{rotational forces}}{\text{viscous forces}}$
crystal rotation centrifugal force			
rotating magnetic field driving of azimuthal flow		magnetic Taylor number $Ta_m = \dfrac{B^2 r^4 \sigma \omega}{2\rho \nu^2}$	$\dfrac{\text{electromagnetic forces}}{\text{viscous forces}}$
alternating magnetic field driving of radial flow			

Damping of convection

axial magnetic field

damping in the entire melt volume

CUSP magnetic field

damping in the periphery of the melt

Figure 8. Convection mechanisms in a Czochralski set-up and dimensionless numbers.

Assuming the melt to be a Newtonian fluid [11] and the Boussinesq approximation [11] to be valid, the velocity field is described by the continuity equation $\nabla \cdot \vec{u} = 0$ (mass conservation) and the Navier-Stokes equations (momentum conservation):

$$\rho \frac{\partial \vec{u}}{\partial t} = -\nabla p - \rho(\vec{u} \cdot \nabla)\vec{u} + \rho \nu \nabla^2 \vec{u} - \rho \vec{g} \beta \Delta T + \vec{f} \quad (4)$$

In addition, turbulence models may be applied for melt convection [11].

In hydrodynamics it is common to make use of dimensionless similarity parameters, that allow a

comparison of systems with different size or materials. When all important dimensionless numbers have the same value, results can thus be transferred from one system to another. Important dimensionless numbers for convection can be found in figure 8. Other common parameters [12] include the Reynolds number $Re = UL/v$, which compares inertia forces and viscous forces. The Prandtl number $Pr = v/\kappa$ is a scaled material property characterizing heat transport. The Grashof number $Gr = Ra/Pr$ besides Ra is another measure for buoyancy, and the Hartmann number $Ha = BL\sqrt{\sigma/\rho v}$ is a measure for stationary magnetic fields.

4.2. Species transport

The partial differential equation describing species transport is completely analogous to heat transport, including temporal change, convective and diffusive transport, and sources, e.g. from chemical reactions:

$$\frac{\partial c}{\partial t} = -(\vec{u} \cdot \nabla)c + D\nabla^2 c + s \qquad (5)$$

Without sources and convection this leaves Fick's 2nd law. The Schmidt number $Sc = v/D$ compares kinematic viscosity v and diffusion coefficient D, and as a dimensionless material property is a measure for the relative weight of convection in species transport. It is completely analogous to the Prandtl number for heat transport. However, while typical Prandtl numbers for semiconductor melts have values around 10^{-2}, typical Schmidt numbers have much higher values of more than 10^1. This implies that convection can not be neglected for species transport at all, in contrast to many cases of heat transport.

Looking for typical examples of species transport in melt crystal growth one can of course take any kind of dopant transport, determining e.g. the electrical properties of the grown crystalline material. In addition, two more complex systems of species transport should be mentioned.

One is oxygen in the silicon Czochralski process, which plays an essential role in the IC fabrication by the so-called intrinsic gettering. The source of oxygen is a temperature-dependent chemical reaction of the Si melt with the quartz crucible. The oxygen is transported through the melt, and a part of it is in-

corporated into the growing crystal, but a high percentage also evaporates as SiO at the free melt surface [4].

The second example is species transport in GaAs LEC set-ups [13]. The electronic properties of gallium arsenide strongly depend on the concentrations of native defects and impurities, with carbon playing a crucial role. The carbon content in the melt is controlled via carbon monoxide in the gas space. The carbon species and other constituents like boron, oxygen or silicon are transported in the gas, in the melt, and through the boric oxide layer, undergoing a complex system of chemical reactions.

4.3. Phase boundary

The phase boundary between crystal and melt is surely the most important interface in the crystal growth system. Many of the crystal properties, like defect concentrations, are directly related to the processes at the interface. With a growing crystal (growth velocity V) heat transport includes the generation of latent heat L. The resulting heat flux boundary condition describes the so-called *Stefan problem* ("free boundary problem"):

$$\lambda_s \frac{\partial T}{\partial \vec{n}}\bigg|_s = \lambda_l \frac{\partial T}{\partial \vec{n}}\bigg|_l + L\rho_s V \qquad (6)$$

Concerning species transport, the phase boundary shows the phenomenon of segregation. According to the phase diagram, the equilibrium concentrations c_s (solid) and c_l (liquid) of dilluted species are different, i.e. the segregation coefficient $k_0 = c_s/c_l$ is not unity. During an actual freezing process the solute is rejected ($k_0 < 1$) or preferentially absorbed ($k_0 > 1$) by the propagating solid-liquid interface, forming an enriched or depleted solute boundary layer in front of it.

4.4. Defect formation

The formation of crystal defects like point defects, line defects (dislocations), and more-dimensional defects (twins, low angle grain boundaries, precipitates, voids) is governed by the temperature fields during the growth process, species transport, and reaction of species. Defect formation, and hence defect engineering is, therefore, strongly coupled to the mechanisms of heat and species transport, which were described in the previous

sections. However, in addition to that it is necessary to use specific physico-chemical models for the formation of a certain kind of defect. Crystal growth often drives research activities on such kind of defect models and makes use of the results. But a crystal grower is sometimes also confronted with the situation that the relation of crystal properties, i.e. defects, and heat/species transport is not well understood.

It is beyond the scope of this article to introduce models of defect formation. But the reader can find in the literature already accepted models for defect formation, for example of dislocations [14], intrinsic point defects in silicon [15,16], twin formation [17].

In the examples of process optimization in chapter 7, we will make use of these models.

4.5. Material parameters

The reader may already have noticed that the physical laws describing heat and species transport (sections 4.1-4.3) contain a set of material parameters. The same holds if one looks carefully on physico-chemical models which are describing the formation of defects (section 4.4). As a consequence, modeling of crystal growth is only possible if the necessary material parameters are available. This fact may sometimes be a hindrance in the use of modeling tools. On the other hand, there exists a tremendous amount of information about material data, it is often just a question of information and transfer.

In any case, it is necessary for a process optimization to have these data either from material handbooks, suppliers, commercial data bases etc. Sometimes it may also be necessary that crystal growers measure important data or initiate their measurement. A famous example for such an activity was the meta-melt project in Japan [18] on Si data.

It should also be mentioned that the work on a process optimization itself helps to confirm or evaluate material data by comparison of numerical and experimental results (see chapter 6).

5. NUMERICAL TREATMENT

It was shown in chapter 4 that the global model of a melt growth process in general implies a complex set of coupled partial differential equations, which can only be solved numerically. Thus, with numerical methods and computational power undergoing a tremendous development in the last decades, nu-

merical simulation has become an essential and powerful tool for developing and optimizing crystal growth processes [19].

5.1. The idea of numerical process modeling

The basic idea of numerical process modeling is illustrated in figure 9. Starting point are the differential equations related to the physico-chemical model, that describe the process or facility to be simulated (see chapter 4). The comparison to experimental results plays an important role in verifying the numerical code. After a successful validation, the numerical model can be applied to improve crystal growth processes as long as the physical model stays correct and the numerical discretization is well suited for the considered problem
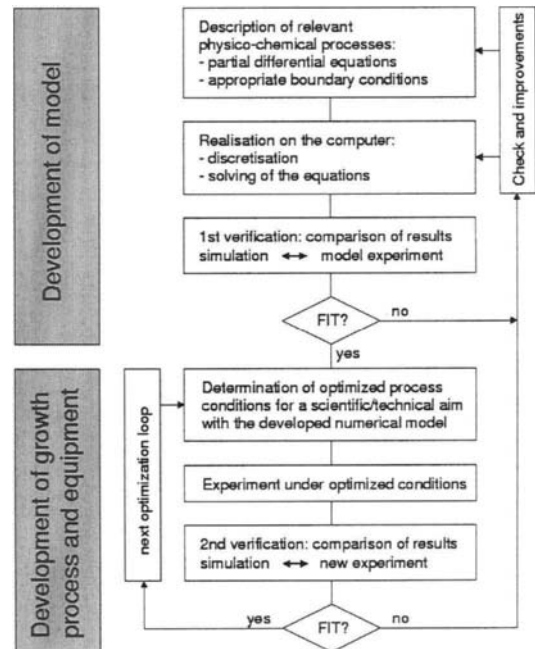
Figure 9. The idea of numerical process modeling.

5.2. Numerical methods and grids

The common basis of all numerical methods is the principle of *discretization*, i.e. both the geometry of the set-up, on which the model equations are to be solved, and the differential equations themselves are transformed from a continuous representation into a discrete one. With respect to geometry this means the values of the physical variables like temperature are known and calculated only in certain discrete points $\bar{x}_i$, which are distributed over the geometry

in such a way that the real temperature field is described as correct as possible by this approximate representation. These points, which constitute the numerical grid, can be arranged either in a structured or in an unstructured manner. This is illustrated in figure 10.

The substantial characteristics of structured grids is that every grid point has a fixed and defined number of neighbor points. Very often, this number is four, which delivers a crossword-like grid. In unstructured grids, which are e.g. made of triangles, the number of neighbors is variable and not known a priori. In that respect, unstructured grids are more flexible for describing complex geometries but are also more complicated concerning the numerical treatment.

The next step is to discretize the differential equations, which means transforming the partial derivatives into discrete expressions of the field values at the grid points. The most common numerical techniques used for this purpose are the methods of Finite Differences, Finite Volumes and Finite Elements. For lucidity, the following illustrations leave out Finite Elements and concentrate on the basic features of Finite Differences and Finite Volumes. To explain it in a very simple way, the use of Finite Differences implies the replacement of derivatives by difference quotients. For a one-dimensional equidistant grid with grid spacing h and a linear numbering of grid points, the first and second derivatives of the temperature at the grid point i are e.g. determined as:

$$\left.\frac{\partial T}{\partial x}\right|_i = \frac{T_{i+1} - T_{i-1}}{2h}, \quad \left.\frac{\partial^2 T}{\partial x^2}\right|_i = \frac{T_{i+1} - 2T_i + T_{i-1}}{h^2} \quad (7)$$

The Finite Volume method assumes the grid points to be the centers of control volumes, which completely cover the geometry without overlapping. Figure 11 shows 2D examples for such control volumes in a structured and in an unstructured grid. For structured grids the neighbors of the grid point P are usually denoted as W(est), E(ast), S(outh) and N(orth). Structured grids do not necessarily have to be orthogonal and equidistant like in figure 10. In the general case (figure 11) this includes that for discretization the differential equations given e.g. in the global coordinates r and z first have to be transformed to the local coordinates η and ξ, which depend on the local direction of the grid lines.
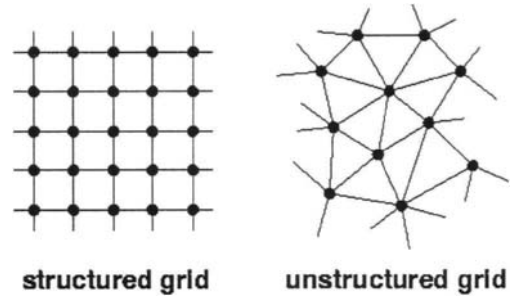
Figure 10. Structured and unstructured numerical grids.
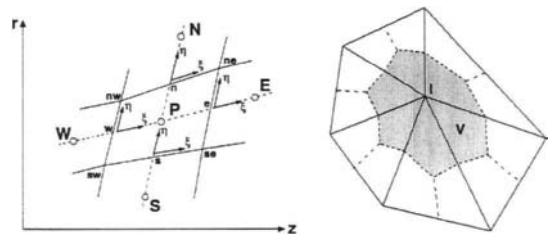
Figure 11. 2D control volumes in a structured (left) and an unstructured grid (right).

The key feature of the Finite Volume method, for which the differential equations are integrated over each control volume, is that it makes use of the Gauss theorem, i.e. volume integrals over the divergence of a vector field are replaced by surface integrals over the vector field. As a typical example, for the Laplace operator this means:

$$\int_V \nabla^2 T \, dV = \int_{\partial V} \nabla T \cdot d\vec{S} \quad (8)$$

Terms transformed in such a way have then only to be evaluated at the boundaries of the control volume.

The discretization of the differential equations and the boundary conditions, and subsequent linearization of the discretized equations delivers a linear equation system with the number of equations equal to the number of grid points. As the differential equations in general are non-linear, the corresponding matrix systems are usually solved iteratively, e.g. [20]. For further information on numerical techniques the reader is referred to literature [11,21].

5.3. Basic approaches of modeling

We have learned that numerical modeling means to transform the physico-chemical model equations by appropriate techniques into a discrete formulation, which is implemented into a computer code to numerically calculate the desired variables and parameters, e.g. the temperature field. However, the philosophy of numerical modeling includes some more basic concepts, which have to be considered before starting the numerical simulation of a crystal growth process.

One aspect to consider is whether model simplifications are justified with respect to the expected numerical effort. The construction of many melt growth facilities (Czochralski, VGF) allows to assume them to be axisymmetric. Then it is possible to restrict the numerical calculation to the meridional r-z-plane, which simplifies the numerical model and reduces the number of necessary grid points and thus computation time and capacity significantly. On the other hand, some problems, like complex buoyant melt convection, still require a fully three-dimensional treatment. Remaining in the field of Finite Volumes, figure 12 illustrates the difference between a quadrilateral 2D axisymmetric control volume and a really 3D structured-grid control volume, where the W, E, S, N neighbors of figure 11 are joined by B(ottom) and T(op).

Another question in process optimization is the use of "direct" and so-called "inverse" simulation. A typical example for this is the optimization of heating powers. In the classical direct approach the temperature field is calculated from given heater powers, which have to be known or optimized empirically. For facility design, this is sometimes not very helpful. In the more advanced inverse case, however, mathematical algorithms deliver the heating powers necessary to achieve a desired temperature field, e.g. prescribed by a certain position and shape of the solid-liquid interface. An example for inverse simulation is treated in section 7.3 (for details see [22]).

5.4. Steps of numerical simulation

In this section, the three practical steps of numerical simulation are presented for a 2D axisymmetric numerical case. The example for a VCz equipment for the growth of GaAs is selected for tutorial purposes, as it has a complex geometry and contains all important features.
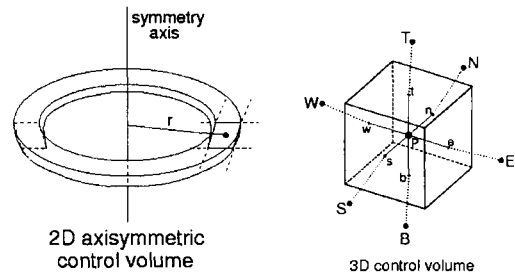
Figure 12. 2D axisymmetric and 3D control volumes.

A. Preprocessing

Preprocessing includes many different aspects and should not be underestimated. A profound preparing of the calculation is absolutely crucial for getting high quality numerical results. First, a detailed CAD drawing of the crystal growth facility is read into the preprocessor tool. It is further edited to include every component which is important for the simulation and to delete details which are negligible. Next, materials are assigned to every region of the geometry, and radiation cavities are defined. This is shown in figure 13a. It is very important to know the material data like thermal conductivity as exactly as possible, including e.g. temperature dependence, because only then simulation can produce realistic results.

Another critical step of preprocessing is the grid generation, illustrated in figure 13b for structured and unstructured grids. The grid must be fine enough and well designed in order to resolve all important physical phenomena. On the other hand, too many points worsen the numerical performance by increasing the computing time unnecessarily.

B. Calculation

After configuration with the preprocessor the numerical calculation is run on the computer making use of the implemented discretization and optimization algorithms. Depending on hardware and the physical and numerical model (2D/3D, direct/inverse) this may take minutes or weeks.

C. Postprocessing

Finally, the calculated numerical results are visualized, analyzed and evaluated for improving the process or for comparison to experimental results. Figure 13c shows the numerically determined velocity and temperature field in the VCz set-up.
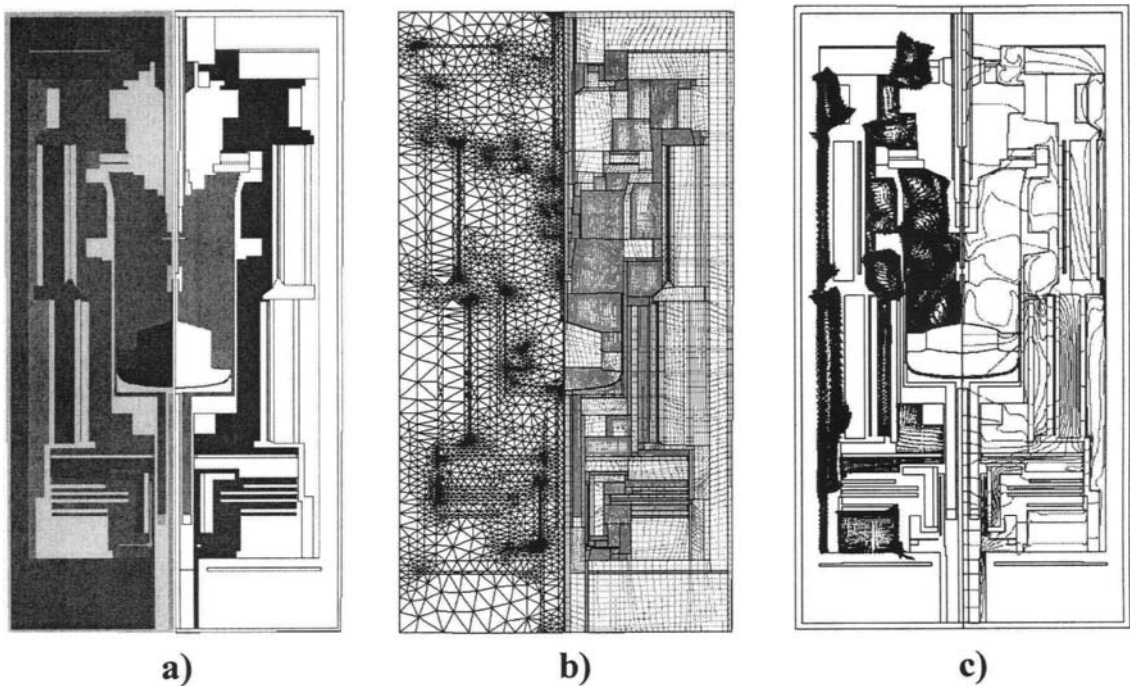
Figure 13. Important steps of numerical simulation in the example of a VCz facility: a) Materials (left) and radiation cavities (shaded areas on the right). b) Numerical grids, unstructured (left) and structured (right). c) Post-processing of the numerical results: gas flow (left) and isotherms (right).

5.5. Required hardware (computers)

As can be seen on the logarithmic scale in figure 14, computer power increases by a factor of ten approximately every five years. In addition, numerical techniques are getting more and more sophisticated. Therefore, numerical simulation today offers vast possibilities, also for industrial-scale process development. However, there are still limits due to a lack of appropriate physical/numerical models or computational power. 2D direct global simulation of a crystal growth facility like in figure 13 works well on a fast PC. This makes simulation interesting for industry, offering a cheap and fast way of process design, compared to expensive and time-consuming empirical and purely experimental developing. The numerical treatment of complex inverse simulations of time-dependent processes or 3D direct numerical simulation of buoyant melt flow in Czochralski set-ups, however, is just beginning to pay off for industry, because up to now the necessary but expensive high performance computing hardware is mostly available only for research. Coming back to figure 14, time will surely change this.

An important step in the development of a powerful numerical process model is its experimental validation. Only by quantitative experimental checks of simulation results computer models can be developed to a level which make them useful. This will be treated in the next chapter.

6. VERIFICATION OF NUMERICAL MODELS BY MODEL EXPERIMENTS

6.1. Definition of model experiments

Typical melt growth processes do not allow for a direct and in-situ measurement of processing data like temperature distribution in the crystal, shape of the solid-liquid interface, concentration of a certain chemical species in the melt etc., which would be necessary for a direct validation of the numerical model. An indirect evaluation of such data by characterizing grown crystals is very time-consuming and often does not provide quantitative data of the growth process precisely enough. On the other hand, the use of sensors and in-situ monitoring devices mostly is disturbing the real growth process.
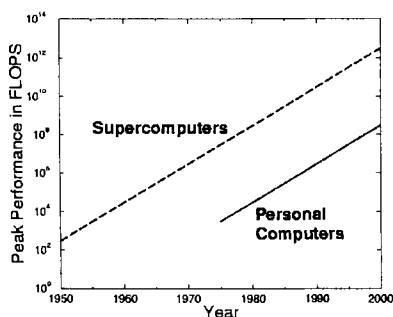
Figure 14. Development of computational power.

In this situation one has to develop appropriate *model experiments* in parallel to the *real* crystal growth configuration. The primary goal of a *model experiment* is to provide extensive quantitative data, which are characterizing important phases of the corresponding bulk growth process. It is *not* important that a high quality crystal is grown during a model experiment; but it is important that sophisticated sensors and in-situ monitoring devices can be used. On the other hand, it is also important that the model experiment comes as close as possible to the real growth set-up and growth process.

In the following sections, different kinds of model experiments will be shown, which have been used in the literature to obtain quantitative data on

- temperature distributions in the growth set-up, including crystal and melt
- the shape of the solid-liquid interface during the whole growth process
- a time-resolved growth rate
- spatial and temporal concentrations of specific dopants in the melt and ambient gas atmosphere.

These quantitative experimental data will then be compared to the numerical results of the simulation of the model experiments (1st verification in figure 9). As a result, a good correlation leads to confidence in the numerical model – called *validation*. In the case of larger discrepancies, the numerical model cannot be used for a process optimization and has to be improved ("no"-loops in figure 9). In that case the experimental data from the model experiments can be used to find out the reason of discrepancies. Typical examples of errors are errors in the geometry or the CAD drawing (e.g. neglecting of gaps between solid parts), an insufficient physical model (e.g. neglecting of convection or 3D effects), an insufficient numerical treatment (e.g. too coarse

grids, too big time steps), and wrong material parameters (e.g. emissivity, semitransparency, temperature dependence of heat conductivity).

6.2. Analysis of temperature distributions by model experiments

6.2.1 Dummy crystals

Precise temperature measurements in melt growth configurations up to about 1800 °C are normally carried out by thermocouples, which are commercially available for the various temperature ranges. For reasons of protection, the thermocouples are shielded by a tube of non-corrosive material. The diameters of these sensors should be as small as possible (typically a few millimeters) to avoid too much disturbance of the thermal field by the sensor itself.

A very easy and useful test of the temperature distribution in a crystal growth set-up can be carried out by using a so-called *dummy crystal*. A dummy crystal is shaped out of a solid material with a heat conductivity which comes as close as possible to that of the crystal. For example, graphite is very useful in the case of semiconductors. The advantages of using a dummy crystal compared to a real crystal are as follows: Any problems with the seeding, the growth process itself, reactive chemistry, hazardous conditions etc. can be avoided. Any shape of the crystal cone, cylinder length, interface etc. can be machined very easily. Holes can be drilled into the dummy to be used for moveable thermocouples (see examples in figures 15,16). The melt can also be replaced by a *dummy melt*, either by a liquid with a low melting point (e.g. gallium in the case of semiconductors) or even by a second solid material. And finally, the dummy can be used for a wide range of thermal heating conditions, also temporal variations, without any danger of "melting" the crystal or even freezing the melt (if it is also replaced by a dummy).

Czochralski-type configurations

Figure 15 shows an example for the use of a dummy crystal (graphite) and a dummy melt (Ga) in an industrial LEC production facility for GaAs.

Five thermocouples are fixed at different radial positions at the pulling rod. Various dummy crystals (e.g. 3" and 4" diameter) with interface shapes according to the experimentally observed ones were used, with holes for the thermocouples as shown in

figure 15. The comparison of the thermo-physical data of GaAs and the graphite dummy crystal, which are given in table 1 [23], shows a good agreement in the relevant thermal diffusivity κ.
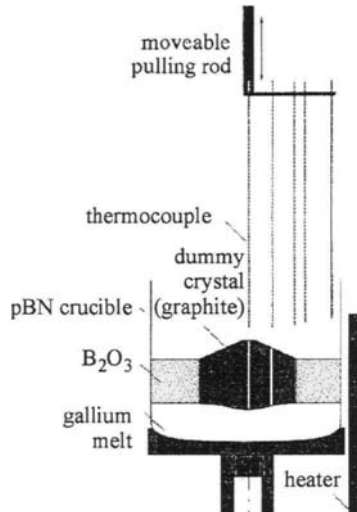
Figure 15. Central part of an industrial GaAs LEC set-up with a graphite dummy crystal and a gallium model melt.

Table 1. Material data for GaAs crystal and dummy.

T=1500 K	ρ [kg/m³]	λ [W/mK]	c_p [J/kgK]	κ [m²/s]
GaAs	5170	7.1	424	$3.2 \cdot 10^{-6}$
Graphite EK40	1720	10	1800	$3.2 \cdot 10^{-6}$

Temperature profiles in the dummy crystal and melt are obtained by a stepwise axial movement of the pulling rod. The heights of the melt and the boric oxide, the crucible dimensions and crucible rotation rate, as well as heating conditions can be varied systematically. The crystal rotation can not be used – due to the thermocouple fixation.

The results of the model experiments were used successfully for process optimization, in a first step for a selection of parameters [23], and in a second step for the validation of a numerical process model [24].

Bridgman-type configurations

Figure 16 shows the case of a VGF or Bridgman growth configuration with a graphite dummy replacing the crystal *and* the melt. With this device the furnace concept can be tested with respect to the range of controllable axial temperature gradients. The good agreement of the measured and calculated

axial temperature profiles (rhs. in figure 16) for typical "growth conditions" of 4" GaAs provides also a good example for a validation of the simulation software CrysVUN++ [25].
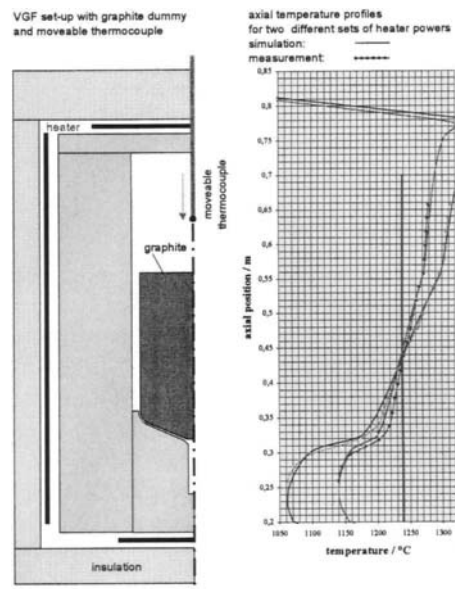
Figure 16. VGF model experiment with 4" dummy crystal and dummy melt (graphite). Various axial profiles were measured and simulated (right).

Deviations from axisymmetric temperature distributions and radial temperature gradients can be analyzed by using a certain number of thermocouples in different radial and azimuthal positions of the dummy (compare figure 15).

6.2.2. Temperature measurements with real crystals

Temperature fields inside crystals

In-situ measurements of temperatures during real crystal growth experiments are much more difficult and expensive than using a dummy crystal (section 6.2.1). On the other hand, they are necessary if one needs more quantitative experimental information about the growth process as can be obtained by dummy experiments.

It has been demonstrated in the literature (e.g. by the authors' group [26]) that thermocouples can be used to "grow in" a crystal. One disadvantage of this procedure is the fact that it provides temperatures of only one fixed position inside the crystal. This

drawback is overcome by a set-up which uses an already grown crystal equipped with a series of thermocouples in various axial positions. The prepared crystal is then mounted in a Cz puller and brought into contact with the melt under conditions of a real growth process as depicted in figure 17.
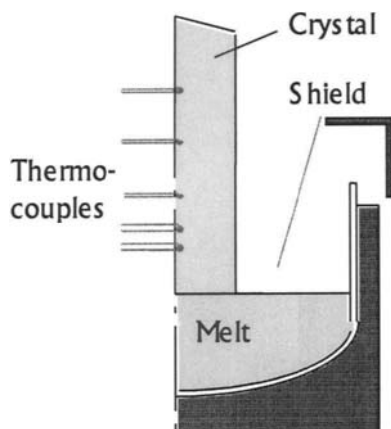
Figure 17. Czochralski model experiment with real crystal and additional thermocouples (see text).

Of course, the data correspond only to one certain situation in the growth run. The results, however, could be used very successfully to validate different software programs for the global modeling of the silicon Cz process [27]. Figure 18 is a nice example for demonstrating the strong influence on the heat transfer of a certain device called "heat shield" and demonstrates the excellent quantitative agreement between simulation and measurement.
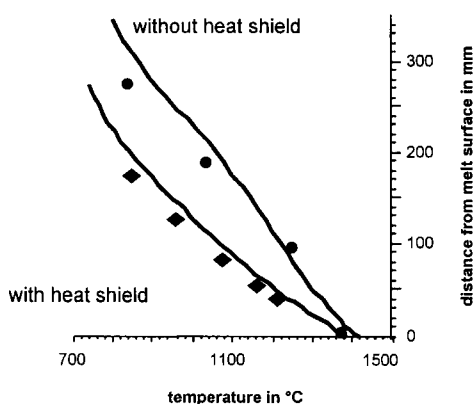
Figure 18. Experimental (symbols) and numerical (lines) axial temperature profiles with and without heat shield for the example of Si Cz.

Temperature fields in the melt

Thermocouples can also be used to detect temperature distributions inside the melt. This is of special relevance in Cz growth. It gives information about the global temperature field and heat transfer inside the melt but also about the status of time-dependent convective flow (temperature fluctuations).

It has been demonstrated by the authors' group [28] that a mapping of temperature distributions inside the melt during growth runs in Czochralski production facilities can be carried out by using specially designed thermocouples. Of special interest is here a detection of temperatures in the vicinity of the growth interface, which can be achieved by a U-like bending of the thermocouple as shown in figure 19. The silica protection tubes of the thermocouples are stabilized by rods of glassy carbon to avoid a deformation due to the rotational and convective movements. With a similar device it is possible to measure temperatures along the crucible wall by pressing the tip of the thermocouple against it – even during crucible rotation. The results are of special interest as boundary conditions for computer modeling (crucible wall temperatures in figure 20) and for validation of numerical models (See e.g. figure 21).
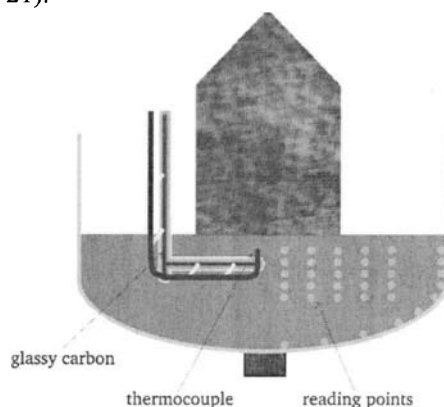
Figure 19. Set-up for melt temperature measurements below the crystal [28].

A further important field of interest, which is analyzed by model experiments using thermal sensors, is the study of melt flow and the state of turbulent convection in Cz crucibles. Unsteady bulk melt flow can be detected by pairs of thermocouples at different positions. The type of flow pattern is evaluated from the phase shift of temperature fluctuation signals between the two fixed monitoring

points [29]. It has to be pointed out, however, that analysis of turbulent flows in large size melts, e.g. in a 250 kg Si melt in a Cz crucible (Grashof number 10^{10}), cannot be carried out by thermocouples due to the limitation of detection frequencies < 0.5 Hz. In such cases, the use of an optical temperature measurement with a sampling rate up to 100 Hz is more useful [30]. A silicon glass rod is dipped into the melt and used as an optical waveguide to transfer the signal outside of the pulling chamber. The upper end of the rod is connected to an optical detector via an optical fiber. With such a device, it was found that *large* Cz melts are in a state corresponding to so-called *hard turbulence* [30].
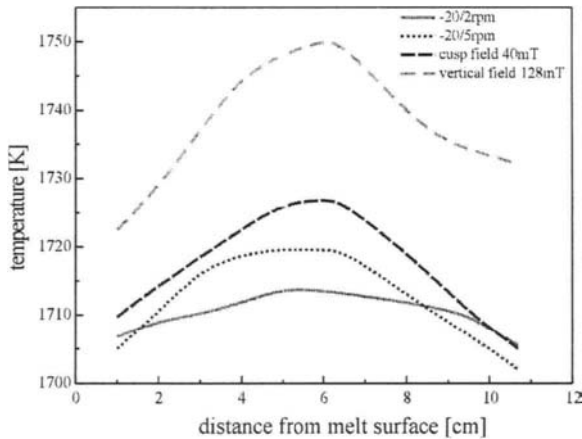
Figure 20. Temperature distributions measured along the crucible wall for different parameter sets in Si Cz [28].

6.3. Monitoring of the shape of the solid-liquid interface and the growth rate

The optimization of a bulk crystal growth process is not possible without controlling the shape of the solid-liquid interface during the whole growth process. Of course, an in-situ monitoring would be optimum but is very difficult and expensive in bulk melt growth because it has to use X-ray monitoring techniques. The most famous model experiments with X-ray imaging of s-l interface shapes of semiconductor growth by Cz were carried out by Kakimoto at NEC laboratories [31].

Much more convenient are methods of interface demarcation, which are using doping striations caused at certain time steps, and analyzing the striation patterns in longitudinal sections of the as-grown crystal. Credit has to be given to Gatos and Witt,

who developed this technique at MIT to a high perfection [32].

Doping striations are caused by stepwise changes of the growth rate, which causes a nearly stepwise change of the segregation of a dopant and hence its striation-like incorporation. A stepwise change of the heating power is the easiest way of changing the growth rate. But also the rotation of the crucible in Czochralski and Vertical Bridgman configurations is efficient. It just uses small deviations from an axisymmetric temperature field.
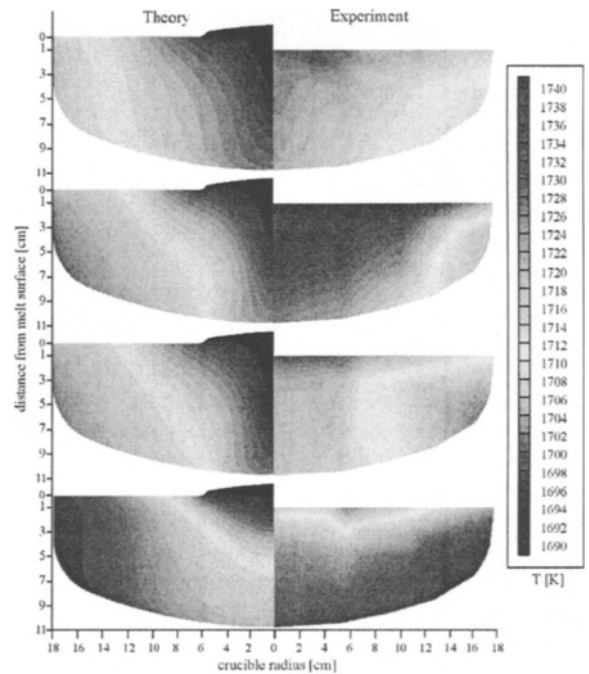
Figure 21. Numerical (left) and experimental (right) temperature fields in a Si Czochralski crucible with 36 cm diameter for (from top to bottom) 2 rpm crucible rotation, 5 rpm crucible rotation, 5 rpm crucible rotation with 40 mT cusp magnetic field, 5 rpm crucible rotation with 128 mT vertical magnetic field.

By correlating the time steps of the interface demarcation to the time scale of the growth process it is possible to evaluate quantitatively the growth rate during the whole growth run. This is demonstrated for example by the results of the growth of 3" diameter GaAs crystals by the VGF method [33,34], shown in figure 22. It shows the striation pattern (dashed lines) in the longitudinal section of the crystal. This *model experiment* is also used to *validate a numerical model* of the VGF process. For that

purpose, the calculated interface shapes (solid lines) are plotted for the same time steps (1-9) where the experimental ones are created (by one crucible rotation). The axial positions of the s-l interface are plotted in figure 23 versus the time scale of the growth process.

The comparison of measured and calculated results shows a good correlation of the shapes of the solid-liquid interface throughout the whole growth run. The maximum deviation of about 20 mm of the positions of the interface in the cone region corresponds to a temperature difference of 14 K if the axial gradient of 7 K/cm is considered. This corresponds to a discrepancy of about 1% if one considers the absolute temperatures of about 1500 K. The observed growth rate deviates from the predicted one by 35%.

The unsatisfactory relation of the growth rates gave reasons to improve the numerical model. Very recently, an agreement of the growth rate within a range of less than 5% was achieved in the authors' laboratory.

6.4. In-situ monitoring of chemical processes and species transport during melt growth

Chemical processes and species transport cannot be treated in a general way like heat transfer and temperature fields (section 6.2), because it depends specifically on the kind of chemical species and crystal material. On the other hand, the quantitative control of a certain chemical reaction, the removal of reaction products or the controlled introduction of a chemical species into a growth system may be decisive for the optimization of a crystal growth process. Important fields are:

- control of composition of a compound (sometimes incorrectly called *control of stoichiometry*)
- control of incorporation of a dopant via the gas phase or by chemical reactions
- removal of reaction products to avoid defect formation or avoid exceeding of concentration limits.

Special sensors were developed for example in the authors' laboratory to detect oxygen in-situ during the silicon Cz process [35]. The results of in-situ measurements were used to validate models of oxygen transport during Si-Cz growth [36].
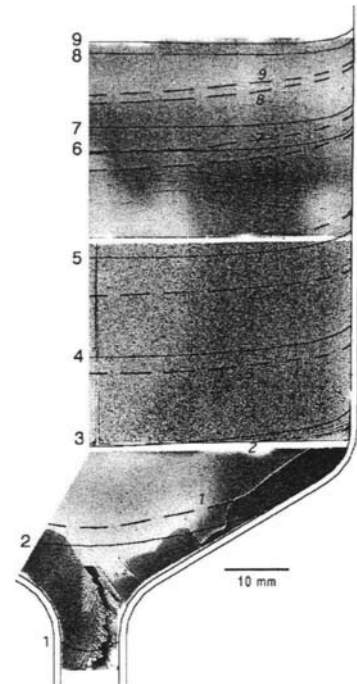
Figure 22. Infrared transmission topographs of longitudinally cut, (110)-oriented wafers of one half of a 3" GaAs crystal with induced rotational striations (dashed lines). For comparison, the corresponding calculated solid-liquid interface shapes (solid lines) are given.
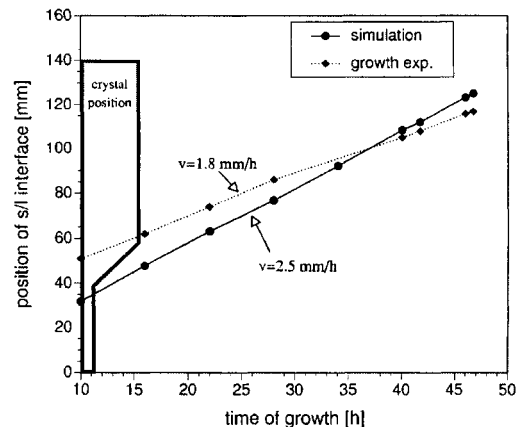
Figure 23. Simulated and measured axial position of the solid-liquid interface during growth of a 3" Si-doped GaAs crystal by the VGF technique in an 8-zone furnace.

7. PROCESS OPTIMIZATION

7.1. General

In the previous chapters all necessary steps were carried out to set up a numerical process model which is experimentally validated. Now the crystal growth processes can be optimized according to the strategy illustrated in figures 1 and 9 by applying the respective models.

It cannot be the goal of such an article to provide the reader with the exact numbers of parameter sets for each of his individual crystal growth problem. On the other hand, the authors' goal is to encourage the reader to use the described tools of process modeling by demonstrating various examples of successful process optimization in important fields of industrial bulk melt growth like Cz, LEC, VCz, VGF and Bridgman. The examples of process optimization, which are shown in the following sections, are taken from actual collaborative projects between the authors' laboratory and internationally leading industrial crystal producers as well as academic institutions. They relate to Czochralski growth of Si (section 7.2) and VGF growth of GaAs (section 7.3).

Other fields of successful applications, like the LEC and VCz processes, the growth of oxides and CaF$_2$, are treated according to the same principles but cannot be shown here due to a lack of space.

7.2. Czochralski growth of Si

7.2.1 V/G optimization

An important criterion to be fulfilled for the optimized growth of Si in the Cz process is the V/G criterion (see [15,16]). The axial temperature gradient G in the crystal along the growth interface is a parameter which cannot be detected experimentally but only deduced from a computer simulation. The growth rate V, i.e. the pulling speed in the Cz process, decides more or less about the economic yield. An overall optimization can, therefore, be deduced from a plot like figure 24 very easily. The optimization looks easy in principle but is difficult in practice as an increase in the growth rate from V_s to V_i leads to an increased creation of heat of crystallization in the growth interface which has to be removed. On the other hand, the axial (!) temperature gradient G has to be increased from G_s to G_i and kept constant across the whole crystal diameter. The latter requirement is very critical in order to achieve the

optimized crystal properties throughout the total diameter of the crystal.

It is clear that the optimization strategy has to consider the heat transfer in and from the growing crystal. According to chapter 5, the tool of inverse simulation seems to be very promising for this case. An efficient optimization can be carried out by setting only three temperature conditions, as sketched in figure 25: T1=T_m (melting point), and T2=T1-G·δ ($G \approx$(T1-T2)/δ) with T3=T2, considers radially uniform heat transfer and (V/G)=C* (constant).
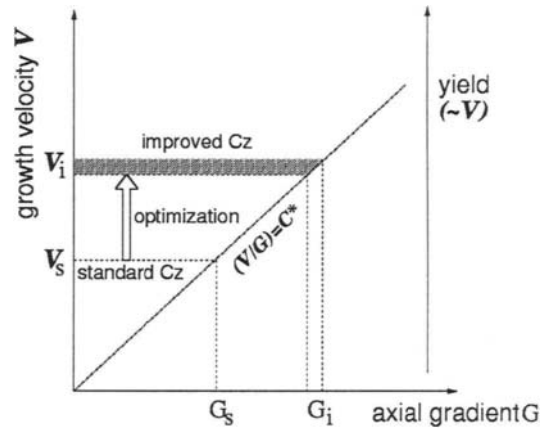
Figure 24. Optimization of the Si-Cz process: Growth rate V and yield versus axial temperature gradient G in the crystal (at the growth interface) by considering the (V/G)=C* criterion.

In order to apply the inverse simulation strategy it is necessary to give the system in figure 25 additional degrees of freedom with respect to the heating and cooling. The control algorithm can optimize heating powers versus time even under consideration of cooling conditions.

The positions and geometry of heating and cooling devices, which are used in industry are of course a matter of secrecy and may also be patented. Nevertheless, there is a huge variety af possible solutions and a creative crystal grower can find its "own" design.

But even for that purpose a geometry optimization can be provided by numerical modeling as demonstrated in the following example. Figure 26 shows the sketch of a Cz set-up containing three heat shields (L1, L2, L3) to be used for an optimized thermal post growth treatment of the crystal. The optimum temperature conditions for the crystal are defined by three temperatures (T1, T2, T3) at certain

positions on the crystal axis (see figure 26). Now the task is to optimize the geometry of the heat shields in order to achieve the set temperatures T1, T2, T3 in the crystal for a given Cz configuration (including heater). The three lengths L1, L2, L3 of the heat shields are used as variables to be optimized by a control algorithm. Figure 27 shows the result of the simulation, which was started with an equal length for all shields. The lower plot shows that after only 6 iterations the control algorithm finds a solution where the differences ΔT of the temperatures from the set values T1, T2, T3 (compare figure 26) are very small. The upper plot in figure 27 illustrates how the lengths L1, L2 and L3 are varied during the iterations until the optimized lengths are found.

Figure 26. Czochralski set-up with three heat shields of different lengths L1, L2, L3, which have to be optimized to achieve certain temperature conditions T1, T2, T3 on the crystal axis.

Figure 25. Control points T1, T2, T3 for the (*V/G*) optimization.

7.2.2 Optimization of the stability of melt temperatures and crucible overheating

It is well known that in larger melt crucibles convective flows are causing temperature fluctuations which may exceed amplitudes of 10 K in the vicinity of the growth interface [37]. Not even that this effect causes non-uniform crystals (striations) [37], but it can also disturb the growth process considerably. Different measures were developed in the past to reduce the fluctuation amplitudes. Increased crucible rotation rates and magnetic fields (for electrically conducting melts) are the most efficient ones [12]. However, both measures are reducing the convective heat transport, especially in radial direction. As a consequence, the temperature of the crucible wall has to be increased in order to provide the same heat flux as under conditions of strong convective heat transfer (e.g. without magnetic damping).

Figure 27. Relative shield lengths L1, L2, L3 (top) and deviations $\Delta T1$, $\Delta T2$, $\Delta T3$ from the desired temperatures T1, T2, T3 (bottom) versus iteration number (compare figure 26).

This phenomenon is illustrated by results of experimental studies and numerical process modeling (see figures 20 and 21). It is obvious that both, increased crucible rotation and (increased) applied vertical magnetic field are increasing the temperature of the crucible wall, especially in the hot range around 5 mm below the melt surface in figure 21. However, increased crucible temperatures are con-

siderably reducing the mechanical stability of the silicon crucible and are injecting too much oxygen into the melt due to increased crucible corrosion.

An optimization of the Czochralski process must, therefore, find a compromise between the benefits of reducing convection and its disadvantages. In figures 28 and 29 it is demonstrated how such an optimum can be found for a certain Si Cz puller by the aid of careful experimental and numerical process analysis.
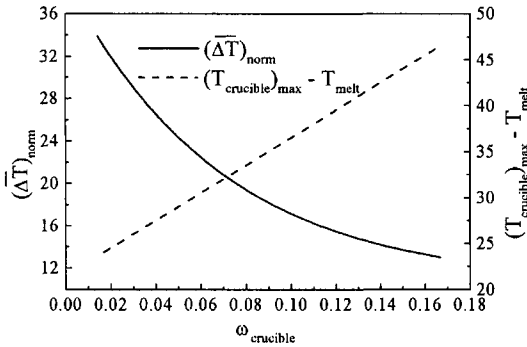
Figure 28. Si Cz: Normalized average temperature fluctuations (left) and maximum crucible temperature (relative to melting temperature) versus crucible rotation rate $\omega_{crucible}$.
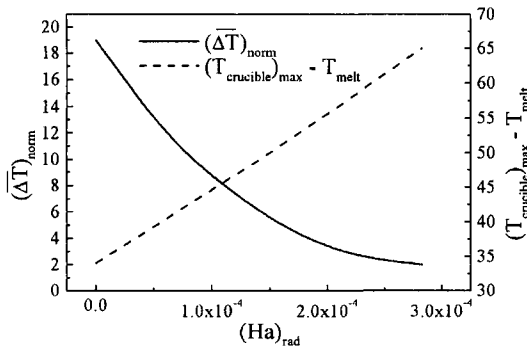
Figure 29. Si Cz: Normalized average temperature fluctuations (left) and maximum crucible temperature (relative to melting temperature) versus Hartmann number Ha_{rad}.

7.3. Optimization of VGF growth of GaAs

In this section it will be demonstrated how the strategy of inverse simulation (see chapter 5) can be used to find an optimized VGF process, e.g. for the growth of GaAs [22].

First one has to define the controlling conditions for the VGF run, which may be as follows according

to figure 30. The position of the solid-liquid interface is fixed in the center of the crucible at a certain height by the condition $\vartheta_1 = 1511$ K $= T_m$. The condition of a fixed temperature gradient e.g. 2 K/cm in the melt in order to avoid instabilities of the interface is achieved by $\vartheta_2 = \vartheta_1 + 2$K with an axial distance of 1 cm above the interface. The condition of a planar crystal-melt interface for a low thermal stress is simulated by setting the temperature $\vartheta_3 = \vartheta_1 = 1511$ K at the same height, but on the periphery. Furthermore, we choose a condition to prevent the overheating of the GaAs melt. For that purpose, we use a temperature condition ϑ_4 in the axial position at the top of the melt. We want to avoid that the temperature at the top exceeds 1526 K, that is 15 K above the melting temperature of GaAs, in order to avoid a great loss of arsenic. This condition is defined by $\vartheta_4 = \vartheta_1 + \min\{2\text{K/cm}\cdot d_{14}, 15 \text{ K}\}$, where d_{14} denotes the distance between the two controlling nodes for ϑ_1 and ϑ_4. This condition means that for small distances d_{14} the melt overheating fits the thermal gradient, while for larger distances the melt overheating does not exceed 15 K.

The strategy of "shifting" the control points through the crucible gives the heating powers in the heaters for each simulated position of the interface. An example for the distribution of heating power in eight heaters is shown in figure 31. The symbols are the calculated results. In a corresponding experiment the interpolated lines are used as time-dependent heater powers. In the simulation of the seeding and the final state, the growth velocity is set to zero. In all other states a fixed value, e.g. 2.5 mm/h, is applied. The numerical calculation was performed at the authors' laboratory with the software program CrysVUN++ on a mesh with about 5000 nodes covering the whole VGF set-up [22].

The temperature distributions in the GaAs (crystal and melt), which are obtained as a part of the inverse-modeling procedure, are extracted for a discussion of the results. These temperature distributions are in fact identical to the results of a direct modeling if one uses the heater powers of figure 31 as input data for a simulation of the VGF process. This direct simulation corresponds to a real crystal growth experiment, where the heater powers are given as input data. The calculated axial temperature profile along the center of the crucible is plotted in figure 32 for 15 time steps of the growth run. These time steps correspond to certain equidistant positions of the growth interface on the axis of the crucible as

the growth rate is constant. The different graphs represent the growth states, which belong to the corresponding set of the heater powers represented in figure 31 by one set of the eight symbols for one fixed position of the solid-liquid interface at the furnace axis.
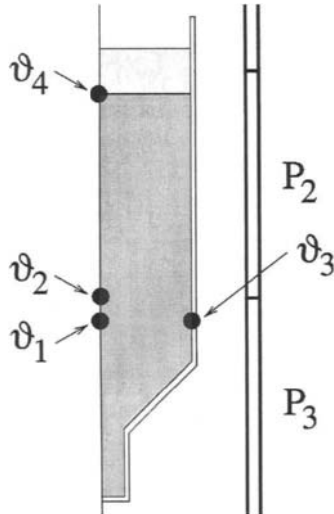
Figure 30. Controlling conditions (fixed temperatures ϑ_1-ϑ_4) for an optimized VGF growth process of GaAs (see text).
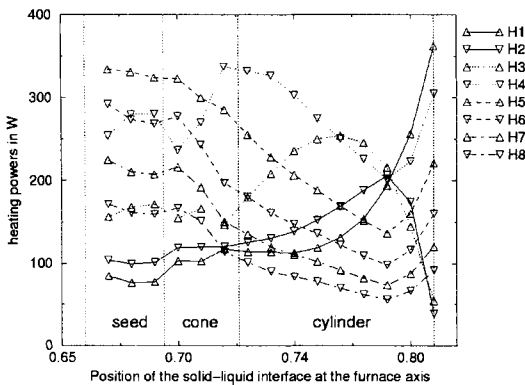
Figure 31. Calculated heating powers (versus the position of the solid-liquid interface) of the eight heaters for 15 different positions of the growth interface under the conditions defined in figure 30.

One can clearly recognize the kink of the temperature profile at the position of the crystal-melt interface $T_m = 1511$ K. The break of the slopes in the axial profiles on the interface corresponds to the Stefan condition (see chapter 4). The curve for the

crystal tail end ($z = 0.795$ m) shows the end of the growth mode, where the temperature gradient in the as-grown crystal is reduced in order to reduce the thermal stress during a following cooling procedure. The results show clearly that important requirements for a good VGF process are fulfilled: The axial temperature gradients $\delta T/\delta z$ are nearly constant in the crystal in the region of the growth interface, which should provide conditions of low thermal stress (see later, figure 35), and the temperature in the GaAs melt does not exceed the given limit of $T \le 1526$ K.

Next the shapes of the crystal-melt interfaces are analyzed. They are plotted in figure 33 again for the different time steps considered corresponding to the set of heating powers depicted in figure 31. Most of them have a planar shape, but deviations in the vicinity of the periphery of the crystal become apparent. This deviation can be explained by the strongly anisotropic conductivity λ of the pBN crucible of 62.7 W/mK (parallel) and 2.0 W/mK (perpendicular) to the axis.
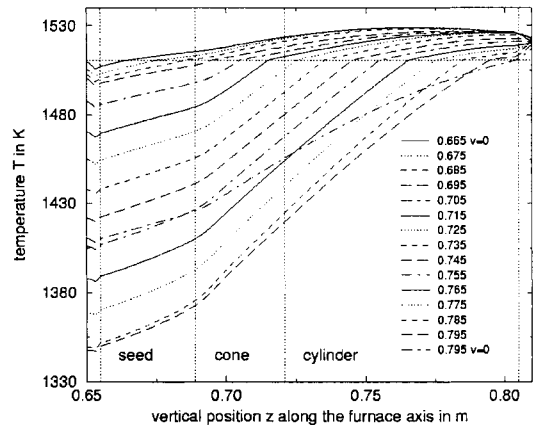
Figure 32. Calculated axial temperature profiles in a GaAs VGF growth run for different growth states, i.e. positions of the solid-liquid interface. The legend shows the positions of the interface at the symmetry axis [22].

This bad fitting of heat conductivities of pBN and GaAs (7.1 W/mK (solid) and 17.8 W/mK (liquid)) always causes a strong bending of the heat flux in the vicinity of the crucible wall and hence a locally concave bending of the interface. This can clearly be seen from the enlarged temperature profile of this region depicted in figure 34. The strong convex bowing of the solid-liquid interface at the tail end of the crystal (top of figure 33) can be explained by the

fact that the top portion of the melt is heated from the side by the top heaters. These deviations from the controlling conditions prescribed by a planar interface show that the calculation cannot fulfill physically unreasonable conditions. But it gives a physically possible solution that comes closest to it – in the meaning of a minimum criterion.

Nevertheless one can demonstrate with a calculation of the thermal stress in the GaAs crystal during the VGF growth run that the inverse modeling provides optimized process conditions. In figure 35 the distribution of the thermal stress in terms of the von-Mises criterion σ_{vm} (see [14]) is shown for three different growth stages. The values of σ_{vm} are the best ones that have been achieved after lots of calculations and empirical optimizations. Especially the value of $\sigma_{vm} = 0.8$ MPa for the solid-liquid interface in the conical region ($z = 0.705$ m) is quite remarkable.

Next it will be demonstrated how such kinds of process optimizations can improve the quality of GaAs crystals considerably and thus improve VGF efficiency. The mean Etch Pit Density (EPD) of investigated wafers cut at two positions of three VGF-grown crystals was found to be 2400 (seed end) and 4200 cm^{-2} (tail) before the process optimization and 570 (seed end) and 240 (tail) after it [34].

Figure 34. 3" GaAs VGF: Enlarged temperature profile around the solid-liquid interface and the crucible, for the position $z=0.755$ m of the interface. The isotherms differ by $\Delta T=2$ K.
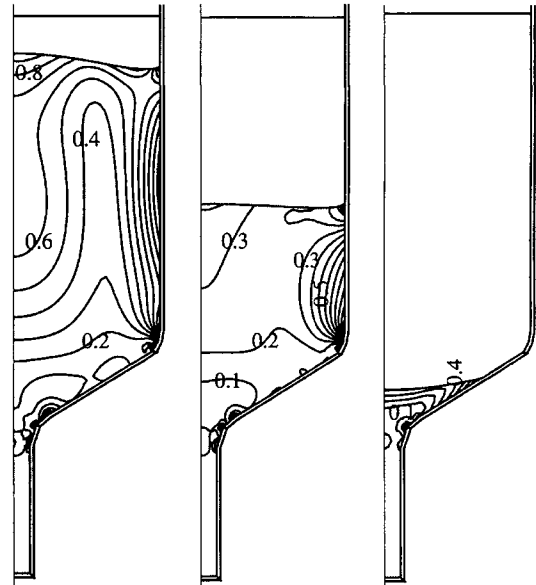
Figure 35. 3" GaAs VGF: Lines of equal von-Mises stress in MPa for three representative growth stages with the interface at the axial positions of 0.705 m (left), 0.775 m (center), and 0.795 m (right), respectively.

To decrease the EPD below 100 cm^{-2} by a further optimization of the VGF process, makes it necessary to consider additional parameters. A detailed study of the initial growth conditions in the seed channel in the authors' laboratory [38] made it obvious that facets at the crystal rim are influencing the crystal

Figure 33. 3" GaAs VGF: Shapes of the solid-liquid interface for various positions during growth calculated with the heating powers shown in figure 31.

quality. It was shown that large facets in the seed channel are increasing the EPD. But the length of facets can only be reduced by increasing the axial temperature gradient, which is in contradiction to minimizing thermal stress. An optimization of the VGF process means in this case to find a compromise with regard to both phenomena. Using this strategy, GaAs crystals of 3" diameter were grown with an EPD of $30 cm^{-2}$ [39].

8. Conclusions

This article provides guidelines for the optimization of melt growth processes both for industrial production and R+D purposes. It is based on the experience in research on melt growth of semiconductors and collaborative projects with leading industrial enterprises, gained during a period of more than 25 years at the Erlangen Crystal Growth Laboratory.

The reader should be convinced that the combined use of experimental analysis and computer modeling is the best strategy to optimize his crystal growth process – even if it is differing from the examples given in chapter 7. The authors are offering to anybody who might be interested their support in developing the appropriate strategy for "his" crystal growth process.

LIST OF SYMBOLS

B	Magnetic induction
c	Concentration
c_p	Specific heat capacity
D	Diffusion coefficient
$\vec{f}$	External force density
$\vec{g}$	Gravity vector
g	Gravitational acceleration
G	Temperature gradient
h	Heat Transfer Coefficient
Ha	Hartmann number
k_0	Segregation coefficient
L	Characteristic Length
L	Latent Heat
$\vec{n}$	Normal vector
p	Pressure
Pr	Prandtl number
$\vec{q}$	Heat flux density
r	Characteristic Length
r	Radial coordinate
Ra	Rayleigh number
Re	Reynolds number
S	Radiation power per area
$\vec{S}$	Surface vector
Sc	Schmidt number
t	Time
T	Temperature
T_m	Melting Temperature
$\vec{u}$	Velocity
U	Characteristic velocity
V	Growth velocity
V	Volume
x	Growth parameter
$\vec{x}$	Coordinate vector
z	Axial coordinate
β	Thermal expansion coefficient
γ	Surface tension
ε	Emissivity
η	Local coordinate direction
κ	Thermal diffusivity
λ	Thermal conductivity
ν	Kinematic viscosity
ζ	Local coordinate direction
ρ	Density
σ	Electrical Conductivity
σ_{vm}	Von-Mises stress
ω	Angular velocity
Ω	Parameter window

REFERENCES

1. F. Rosenberger, Fundamentals of Crystal Growth, Springer Series in Solid State Sciences, vol. 5, Berlin, 1979.
2. R. Fornari and C. Paorici (eds.), Theoretical and Technological Aspects of Crystal Growth, Materials Science Forum Vls. 276-277, Trans Tech Publications, Switzerland, 1998.

190

3. D. T. J. Hurle (edt.), Handbook of Crystal Growth, Elsevier Science, Amsterdam, 1996.
4. Georg Müller, Melt Growth of Semiconductors, in [2], pp. 87-108.
5. K.-T. Wilke and J. Bohm, Kristallzüchtung, VEB Deutscher Verlag der Wissenschaften, Berlin, 1988.
6. G. Müller and P. Rudolph, in K. H. J. Bushow, R. W. Cahn, M. C. Flemings, B. Ilschner, E. J. Kramer, S. Mahajan (eds.), Encyclopedia of Materials, Pergamon Press (to be published).
7. M. F. Modest, Radiative Heat Transfer, McGraw-Hill, New York, 1993.
8. R. Siegel and J. R. Howell, Thermal Radiation Heat Transfer, Taylor&Francis, Philadelphia, 1992.
9. R. Moreau, Magnetohydrodynamics, Kluwer Academic Publishers, Dordrecht, 1990.
10. J. Friedrich, Y.-S. Lee, B. Fischer, C. Kupfer, D. Vizman, and G. Müller, Phys. Fluids, 11 (1999) 853.
11. J. H. Ferziger and M. Perić, Computational Methods for Fluid Dynamics, Springer, Berlin, 1996.
12. G. Müller and A. Ostrogorsky, Convection in Melt Growth, in [3], pp. 709-819.
13. W. A. Oates and H. Wenzl, J. Crystal Growth, 191 (1998) 303.
14. J. Völkl, Stress in the Cooling Crystal, in [3], p. 821.
15. W. von Ammon, E. Dornberger, H. Oelkrug, and H. Weidner, J. Crystal Growth, 151 (1995) 273.
16. V. V. Voronkov and R. Falster, J. Crystal Growth, 204 (1999) 462.
17. D. T. J. Hurle, J. Crystal Growth, 147 (1995) 239.
18. S. Togawa, S. Chung, S. Kawanishi, K. Izunome, K. Terashima, and S. Kimura, J. Crystal Growth, 160 (1996) 41.
19. G. Müller, Numerical Simulation of Crystal Growth Processes, in M. Griebel, C. Zenger (eds.), Numerical Simulation in Science and Engineering, Notes Fluid Mech. 48, Vieweg, Braunschweig, 1994, pp. 130-141.
20. H.-J. Leister and M. Perić, Int. J. Num. Meth. Heat Fluid Flow, 4 (1994) 159.
21. O. C. Zienkiewicz and R. L. Taylor, The Finite Element Method, McGraw-Hill, London, 1997.
22. M. Kurz and G. Müller, J. Crystal Growth, 208 (2000) 341.
23. A. Seidl, PhD thesis, University Erlangen-Nürnberg, Germany, 1995.
24. J. Fainberg, PhD thesis, University Erlangen-Nürnberg, Germany, 2000.
25. J. Stenzenberger, PhD thesis, University Erlangen-Nürnberg, Germany, 2000.
26. J. Völkl and G. Müller, in Y. I. Nissim and P. A. Glasow (eds.), Proc. E-MRS Meeting, Straßbourg, 1987, p. 141
27. E. Dornberger, E. Tomzig, A. Seidl, S. Schmitt, H.-J. Leister, C. Schmitt, and G. Müller, J. Crystal Growth, 180 (1997) 461.
28. O. Gräbner, G. Müller, J. Virbulis, E. Tomzig, and W. von Ammon, Microelectronic Eng., 1 (2002), in print.
29. A. Seidl, G. McCord, G. Müller, and H.-J. Leister, J. Crystal Growth, 137 (1994) 326.
30. O. Gräbner, A. Mühe, G. Müller, E. Tomzig, J. Virbulis, and W. von Ammon, J. Mat. Sci. and Eng. B, 73 (2000) 130.
31. K. Kakimoto, M. Eguchi, H. Watanabe, and T. Hibiya, J. Crystal Growth, 102 (1990) 16.
32. A. F. Witt and H. C. Gatos, J. Electrochem Soc., 115 (1968) 70.
33. J. Amon, P. Berwian, and G. Müller, J. Crystal Growth, 198/199 (1999) 361.
34. G. Müller, P. Berwian, E. Buhrig, and B. Weinert, in R. Diehl (edt.), High Power Diode Lasers, Topics Appl. Physics, 78, Springer, Berlin, 2000, pp.121-171.
35. A. Seidl, R. Marten, and G. Müller, J. Crystal Growth, 166 (1996) 680.
36. A. Mühe, R. Backofen, J. Fainberg, G. Müller, E. Dornberger, E. Tomzig, and W. von Ammon, J. Crystal Growth, 198/199 (1999) 409.
37. G. Müller, Crystal Growth from the Melt, Crystals 12, Springer, Berlin, 1988.
38. B. Birkmann, M. Rasp, J. Stenzenberger, and G. Müller, J. Crystal Growth, 211 (2000) 157.
39. B. Birkmann and G. Müller, paper accepted for ICCG13, Kyoto.

Epitaxial Lateral Overgrowth of GaN

A. Usui[a] and A. Sakai[b]

[a]Photonics and Wireless Devices Research Labs.,NEC Corporation
34 Miyukigaoka, Tsukuba, Ibaraki 305-8501, Japan

[b]Department of Crystalline Materials Science, Graduate School of Engineering,
Nagoya University
Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan

Here we describe an epitaxial lateral overgrowth (ELO) method we developed for reducing the dislocation density of GaN epitaxial layers and producing layers with excellent optical and electrical properties. TEM observations show that dislocation bending plays an important role in reducing the threading dislocation density and suggest that the bending is closely related to the appearance of the facets early in ELO, so we call our method facet-initiated ELO (FIELO). Several phenomena accompanying dislocation bending, including c-axis tilting, are discussed here in detail and the characteristics of FIELO-GaN are shown from various points of view. X-ray measurements show that the decreased dislocation density is accompanied by a reduction of the twist angle of the c-axis, and we show that the dislocation density can be estimated from the twist angle. We also uses scanning reflection electron microscopy (SREM) to investigate the c-axis tilting of the FIELO-GaN surface, and AFM measurements show that very smooth step-flow growth mode occurs during FIELO even at growth rates greater than 100 μm/h.

1. INTRODUCTION

The use of short-wavelength coherent-light sources will enable us to double or triple the capacity of today's optical storage systems. The pioneering work of Nakamura *et al.* [1,2] has shown that they are likely to be made of InGaN, and Nichia's group has recently estimated the lifetime of a 30mW InGaN MQW laser to be 15,000 hours [3]. Although this material system is a potential source of commercial blue/violet LDs, there are still problems to be solved before the mass-production of these devices will be practical. One is the lack of a high-quality substrate that is appropriate for growing a laser structure and is also suitable for mass-production processes. Sapphire and SiC substrates are used presently, but a large number of dislocations then go through the epitaxial layers from the interface between these substrates and the epitaxial layer. Most of dislocations are generated at the interface because the chemical and physical properties of these substrates are very different from those of the epitaxial layer. To make LDs with a long lifetime, we need to reduce the number of these threading dislocations. The ideal way to do this would be to use a lattice-matched substrate, such as GaN bulk crystal, but sufficiently large bulk GaN wafers of high crystal quality cannot be made yet [4,5]. We therefore studied the feasibility of reducing the dislocation density by using epitaxial lateral overgrowth (ELO) because this growth process reported reduces the dislocation density in epitaxial layers of other III-V compound semiconductors, such

as GaP and GaAs [6,7].

Our 1997 paper reporting the use of hydride vapor phase epitaxy (HVPE) in the ELO growth of thick GaN layers with a low-dislocation-density was the first report of GaN-ELO [8]. The dislocation density of 10^8-10^9 cm^{-2} in the original GaN/sapphire hetero-interface was reduced to 10^6-10^7 cm^{-2}, and crack-free GaN layers that had mirror-like surfaces and were 100 μm thick were grown on a 2-inch-diameter sapphire substrate. And because we found that the formation of a facet structure in the early stage is important in suppressing threading dislocations [9], we called this growth method facet-initiated epitaxial lateral overgrowth (FIELO).

An advantage of using the ELO technique to make InGaN MQW LDs was first demonstrated by Nakamura et al., who in 1977 reported a long-lifetime (longer than 1000hr) LD grown on the ELO-GaN substrate [10]. This was remarkable improvement over the InGaN MQW LD made using the conventional growth method, which had worked for only 300h [11]. We also demonstrated an MQW-InGaN LD on a high-quality FIELO-GaN substrate with a back contact and cleaved facet mirror structure [12]. Thus, the ELO technology is now among the most interesting subjects in GaN research and development.

The present paper describes the FIELO process in HVPE. We have used transmission electron microscopy (TEM) to investigate the mechanism of dislocation reduction and have confirmed the superior crystal quality of FIELO-GaN through x-ray rocking curve measurements and optical and electrical measurements. These results are shown here along with the electrical and optical properties of this material.

2. GROWTH PROCEDURE

The growth system used in this study was hydride vapor phase epitaxy (HVPE) with GaCl and NH$_3$, and the GaN was grown in a horizontal quartz reactor. The GaCl partial pressure was set at levels between 5.2 x10^{-3} and 13x10^{-3} atm, and the NH$_3$ partial pressure was set at levels between 0.13 and 0.53 atm. The total flow rate, including the H$_2$ carrier gas, was 3,800 sccm. The GaN deposition reaction in HVPE is described as follows:

$$GaCl + NH_3 \rightarrow GaN(s) + HCl + H_2. \qquad (1)$$
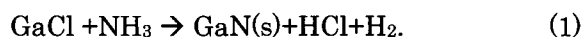
The growth rate measured when samples were grown under GaCl and NH$_3$ at partial pressures of 5.2 x10^{-3} atm and 0.26 atm is plotted against growth temperature in Fig. 1. Below 900°C the growth rate increased with increasing growth temperature. This dependence is a characteristic of kinetically controlled growth. The growth rate reached a maximum value of about 70 μm/h near 950°C and decreased with further increases of growth temperature. The dependence of the growth rate on GaCl partial pressure at a growth temperature of 1000°C is shown in Fig. 2, where it is seen to increase linearly as a function of the GaCl partial pressure and reach 97 μm/h at 1.0x10^{-2} atm. Thus, at the higher growth temperatures the growth rate was limited by the quantity of the supplied source gases. This indicates that growth at the higher temperatures was thermodynamically limited. A growth rate greater that 100 μm/h was easily obtained by increasing the NH$_3$ partial pressure, but the surface became very rough at NH$_3$ partial pressures above 0.6 atm. In most of the experiments described in the rest of this paper, growth was therefore carried out at 1000°C under a GaCl partial pressure of 5.2x10^{-3} atm. and a NH$_3$ partial pressure below 0.6 atm. This resulted in growth rates between 50 and 100 μm/h.

3. ELO PROCESS

Fig. 3 shows the patterned substrate structure used for ELO. The substrates we used were 2-inch-diameter sapphire wafers with a GaN layer 1–1.5 μm thick on top. Window stripes aligned along the $<11\overline{2}0>$

direction of the GaN layer were fabricated by using SiO₂ masks. The mask width was 1–4 μm and one period consisting of a window region and a masked region was 7 μm.

Scanning-electron-microscope (SEM) images of surfaces are shown in Fig. 4. At the beginning of the growth, facet structures were evident in the window regions (Fig. 4(a)). The sidewalls consisted of $\{1\bar{1}01\}$ planes, and there was no extraneous deposition on the SiO₂ mask. Fig. 4(b) to (c) show that the facets in each groove grew laterally and coalesced as growth proceeded. The gaps between the facets were eventually filled, and a flat (0001) surface appeared (Fig. 4(d)).
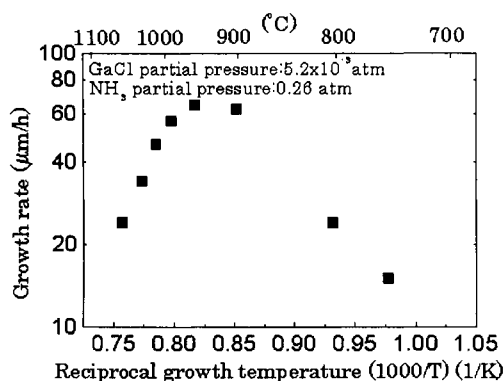
Fig. 1. Growth temperature dependence of the GaN growth rate.
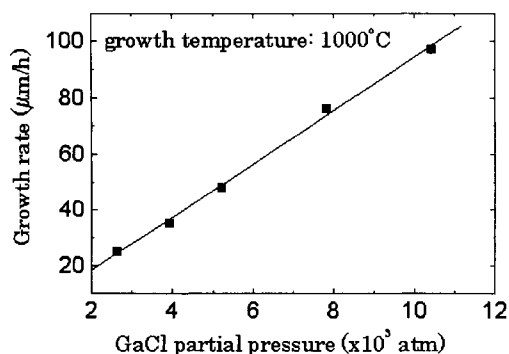
Fig. 2. GaCl partial pressure dependence of the growth rate.

The dislocation density of these samples was found from the TEM observation to be between 10^6 and 10^7 cm^{-2}, which is a range three orders of magnitude smaller that that of the dislocation density in conventionally grown GaN. The detailed mechanism of this reduction in dislocation density is described in Section 4. The thickness of 100-μm-thick FIELO-GaN layers could be kept uniform to within ±3% over a 2-inch-diameter wafer by optimizing the length of the mixing zone where GaCl and NH₃ combine.
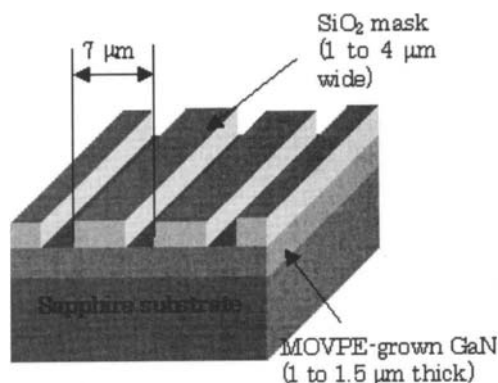
Fig. 3. Substrate structure used for the ELO growth.
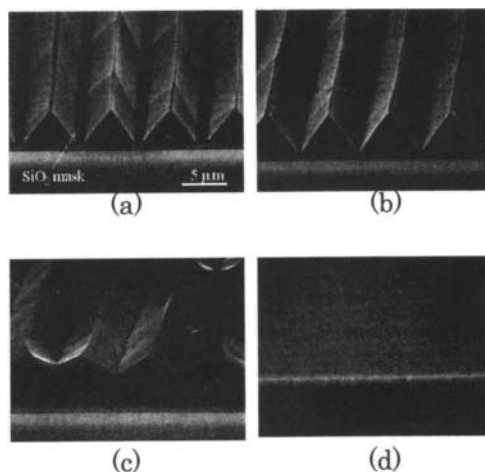
Fig. 4. SEM images of surfaces after growth for (a) 2.5 min, (b) 5 min, (c) 10 min, and (d) 30 min.

4. DISLOCATION BEHAVIOR IN FIELO-GAN

4.1. Dislocation structures

Fig. 5 is a cross-sectional TEM image showing typical defect structures near the film-substrate interface in a sample with the $<11\bar{2}0>$ mask and a 140-μm-thick ELO-GaN layer. The image is a view along the mask stripe direction and was taken under two-beam diffraction conditions with $g=\bar{1}100$ (g is the reflection vector used for imaging). A portion of the ELO layer, a roughly 1-μm-thick MOVPE-grown substrate GaN layer (the substrate layer), SiO$_2$ masks about 1 μm wide, and a sapphire substrate can be seen. Diffraction analyses revealed that the GaN layer had a wurzite structure and an epitaxial orientation relationship with respect to the sapphire substrate given by [0001] GaN || [0001] Al$_2$O$_3$ and [11$\bar{2}$0] GaN || [1$\bar{1}$00] Al$_2$O$_3$. Furthermore, many dislocations in the substrate layer can be seen to be aligned along the [0001] direction of the GaN and distributed uniformly throughout the layer. The density of these threading dislocations was estimated to be greater than 10^9 cm^{-2}.

The Burgers vector notation for dislocations in GaN is shown in Fig. 6, where a GaN crystal has a hexagonal unit structure and a threading dislocation running along the c-axis. When the Burgers vector is one of the six vectors that are equivalent to the *a* vector (1/3 $<11\bar{2}0>$) on three prismatic-planes (for example, the one denoted by A in Fig. 6), the threading dislocation has a pure-edge character. When the Burgers vector is one of the six vectors on one of the six pyramidal planes that are equivalent to the *a+c* vector (1/3 $<11\bar{2}3>$) (denoted by B in Fig. 6), the threading dislocation has a mixed character. In the case of a screw character, the Burgers vector is parallel to the *c* vector ($<0001>$). TEM analyses of the threading dislocations shown in Fig. 5 revealed that the substrate layer has mostly (more than 70%) pure-edge dislocations, relatively few mixed dislocations, and very few screw dislocations. These results are consistent with those of other studies [13,14,15]. The origins of these edge and mixed dislocations in the substrate layer have

been examined by Ning *et al.* [15] and Wu *et al.*[16]. In the initial stage of MOVPE growth, GaN islands are formed on the sapphire substrate. These islands are grown to rotate each other due to the lattice mismatch between GaN and sapphire. As the islands grow, the pure-edge dislocations are formed at the coalesced site, each of which has a Burgers vector that is parallel to the c-plane and runs perpendicular to the surface. The origin of the mixed dislocations is not fully understood but it too may be related mode of the initial growth of the GaN islands.
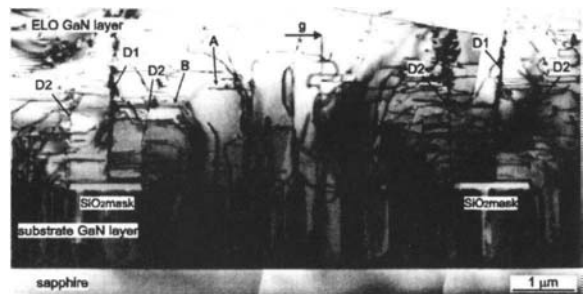
Fig. 5. Cross-sectional TEM [11$\bar{2}$0] image (taken with $g=\bar{1}100$) of a sample grown by HVPE with a $<11\bar{2}0>$ mask.

It can be clearly seen in Fig. 5 that almost all the dislocations in the substrate layer propagate into the ELO layer and that no defects are newly generated at the interface. It should be noted that many dislocations parallel to the interface can be seen in the ELO layer; these dislocations originate mostly from the threading dislocations in the substrate layer, and form a bending configuration. Such bending dislocations piled up along the [0001] direction are especially evident near the SiO$_2$ mask. These results allow us to deduce that threading dislocations vertically aligned in the substrate layer start to propagate laterally during the growth of the ELO layer and that this change of the propagation direction prevents the dislocations from crossing the film and reaching the surface region.

The lateral propagation of dislocations with

b= 1/3 <11$\bar{2}$0> depends less on their proximity to the SiO₂ mask than does that of those with b=1/3 <11$\bar{2}$3 >. For example, we first focus on a bending dislocations with b=1/3 <11$\bar{2}$0 >, indicated by arrow A in Fig. 5 (hereafter we call this type of bending dislocations Type-A dislocations). The structure of a Type-A dislocation is shown in detail in Fig. 7(a). The pure-edge character of the dislocation originating from the threading dislocation in the substrate layer changes into the screw character when the distortion propagates laterally in the ELO layer. Since the morphology of most of the pure-edge threading dislocations in the substrate layer changes in this way and most (more than 70%) of the dislocations in our samples were dislocations of this kind, it is expected that most threading dislocations do not reach the surface of the ELO layer.
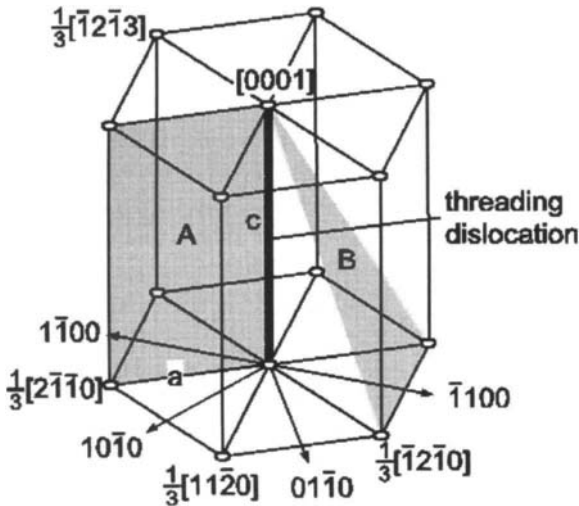
Fig. 6. Notation of directions and Burgers vectors for threading dislocations in the hexagonal unit cell of GaN.

Another type of dislocations with b=1/3 <11$\bar{2}$3 > can also be seen in Fig. 5. The dislocation indicated by arrow B, for instance, (hereafter we call this type of dislocations Type-B) has a Burgers vector of

either b = ±1/3 [2$\bar{1}\bar{1}$3] or ±1/3 [1$\bar{2}$13]. The structure of a Type-B dislocation is shown in detail in Fig. 7(b). Note that the lateral segments of Type-B dislocations no longer lie on their {1$\bar{1}$01} slip planes. This configuration thus may be not brought about solely by the glide process of the dislocation.
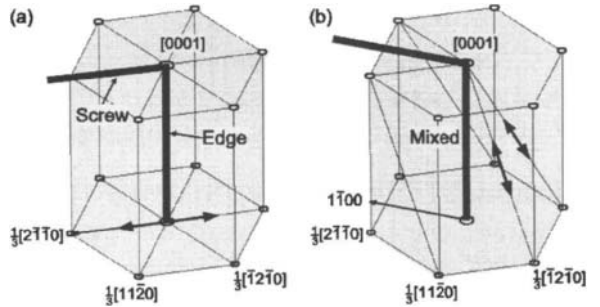
Fig. 7. Detailed structure of (a) Type-A and (b) Type-B dislocations.

Fig. 8 is a cross-sectional TEM image showing the overgrown region near the SiO₂ mask in the same type of GaN film shown in Fig. 5. It is clear that many bending dislocations originating from threading dislocations in the substrate layer are piled up on the mask. Furthermore, there are two types of defects, denoted by D1 and D2, running parallel to the [0001] direction: the D1 defect originates near the center of the SiO₂ mask and the D2 defect comes from both edges of the mask. Fig. 9(a) shows a high-resolution (HR) TEM image of the defect D1 and a periodic array of the end-on dislocations forming a boundary between two neighboring GaN crystals. Closer inspection of the lattice fringes in the adjacent GaN crystals reveals that the c-axis directions in both crystals are slightly tilted toward the boundary. This tilting is schematically illustrated in the inset. The average tilt-angle is estimated from transmission electron diffraction (TED) patterns taken from this region to be 2°. Therefore, it is concluded that the boundary consisting of the arrayed dislocations has the nature of a tilt boundary.
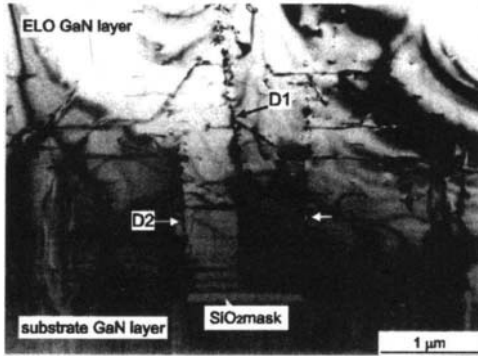
Fig. 8. Cross-sectional TEM [11$\bar{2}$0] image of part of the GaN region overgrown on the SiO$_2$ mask.

Burgers vectors of the dislocations constituting this boundary can be directly determined from their HRTEM images, as in the example shown in Fig. 9(b). When the positive sense of the dislocation line vector is taken to be into the page and a clockwise Burgers circuit is drawn according to the start-finish/right-handed screw convention, a Burger vector is determined by the vector pointing from S to F. As a result, the projection through the [11$\bar{2}$0] direction of the Burgers vector of each dislocation in the D1 defect is determined to be $\sqrt{3}$ a/2, where a is the a vector in a GaN hexagonal unit cell. Furthermore, most dislocations in the D1 defect have the same Burgers vector, and thus the boundary is composed of one set of dislocations. Note that the dislocation line corresponds to the upper edge of the extra-half-plane in the crystal (hereafter we refer to these types of dislocation as negative dislocations). This structure well explains why the c-axis tilting described above occurs when these dislocations pile up along the [0001] direction to form a boundary.

A similar HRTEM characterization for the D2 defect shown in Fig. 9(c) clarifies that it also consists of arrayed dislocations that are parallel to the [11$\bar{2}$0] direction and that form a tilt boundary. But in this case, the mean separation between the dislocations constituting the boundary is 2 to 4 times longer than that for the D1 defect and the tilt-angle for the neighboring crystals caused by the D2 defect was estimated from TED patterns to be 1°. Furthermore, the Burgers vector of each dislocation, measured from the HRTEM images of Fig. 9(d), has the same absolute value as that in the D1 defect ($\sqrt{3}$ a/2) but the opposite sign. This means that the dislocation corresponds to the lower edge of the extra-half-plane (hereafter we refer to this type of dislocation as positive dislocations). The array of the positive dislocations causes the c-axis directions in the adjacent crystals to tilt away from the boundary.
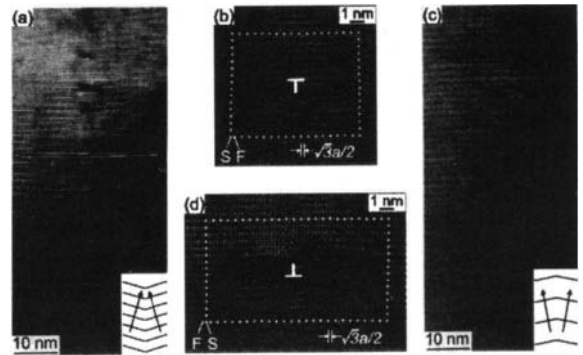
Fig. 9. (a) Cross-sectional HRTEM [11$\bar{2}$0] image of the D1 defect. The inset shows the tilting of the c-axis directions in two adjacent GaN crystals. (b) Burgers circuit around one of the dislocations forming the D1 defect. (c) and (d) are a HRTEM image and a Burgers circuit for the D2 defect.

Fig. 10 shows a simplified model of the D1 and D2 defect structures in the overgrown region. Both types of defects form tilt-boundary, but the signs of the Burgers vectors of the dislocations responsible for them are opposite. This causes crystallographic tilting of the two overgrown-region domains separated by the D1 defect. The mean tilt-angle in each domain, up to about 2 μm from the SiO$_2$ mask, is

about 1°. As seen in Fig. 8, the separation between the end-on dislocation-segments in each defect increases as the film thickness increases. More than about 5 μm away from the mask there are hardly any end-on dislocations in the D1 and D2 defects. Thus, the tilting diminishes and eventually vanishes with increasing film thickness because the tilt-angle depends on the distance between the dislocations in the boundary [17].
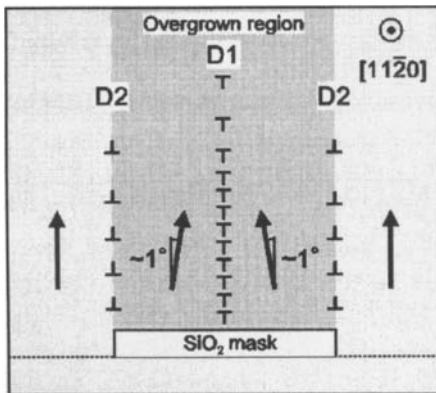
Fig. 10. Simplified D1 and D2 defect structures in the overgrown region.

As mentioned earlier, there are essentially two types of dislocations in the ELO layer: those originating from pure-edge dislocations (Type-A) and those originating from mixed dislocations (Type-B) in the substrate layer. Type-A dislocations have Burgers vectors with $b = 1/3 < 11\bar{2}0 >$, which are equivalent to the a-axis vectors in the hexagonal unit cell. The value obtained for the Burgers vectors of both the D1 and D2 defects, $\sqrt{3}\ a/2$, should be a projection through the $[11\bar{2}0]$ direction of an absolute value of the a vector. Therefore, it is deduced that the origin of the dislocations forming the D1 and D2 defects is Type-A dislocations that were propagated laterally in the ELO layer. The mechanism by which these defects form is described in detail in Section 4.3.

4.2. Bending of pure-edge threading dislocations

Section 4.1 showed that dislocation bending plays an important role in reducing the threading dislocation density in ELO films. The dislocation morphology suggests that this bending is closely related to the appearance of the facets during the earlier stage of ELO. Fig. 11 shows cross-sectional TEM micrographs taken with different **g** from the same area in a sample with the $<11\bar{2}0>$ mask, and these micrographs show dislocation structures at an early stage of ELO. A portion of the ELO-GaN layer with $\{1\bar{1}01\}$ facets, a substrate GaN layer, and SiO₂ masks are evident and it can be clearly seen in Fig. 11(a) that there are bending dislocations and laterally aligned segments of dislocations in the ELO region. Thus the bending of dislocations has already occurred in this growth stage. Especially, vertically aligned dislocations are hardly observed in the upper region. Some of the lateral dislocation segments and the bending dislocations are out of contrast with **g**= $000\bar{2}$ (Fig. 11(b)). From the invisibility criterion **g·b**=0 [18], most of the dislocations that are in contrast in Fig. 11(b) have Burgers vectors of **b**=$1/3 < 11\bar{2}3 >$ and are identified as Type-B dislocations. The dislocations that are out of contrast in Fig. 11(b), however, have Burgers vectors perpendicular to the [0001] direction. That is, **b**=$1/3 < 11\bar{2}0 >$. Thus they are Type-A dislocations. The broken lines in Fig. 11(b) correspond to the initial position of the $\{1\bar{1}01\}$ facets appearing during ELO. For Type-B dislocations, the lateral propagation is observed to occur predominantly around these positions. For Type-A dislocations, on the other hand, such as the bending dislocation indicated by the arrow in Fig. 11(a), bending morphology is always observed regardless of these facet lines. This indicates that the lateral propagation occurs randomly during the growth stage before the formation of the $\{1\bar{1}01\}$ facet is complete.
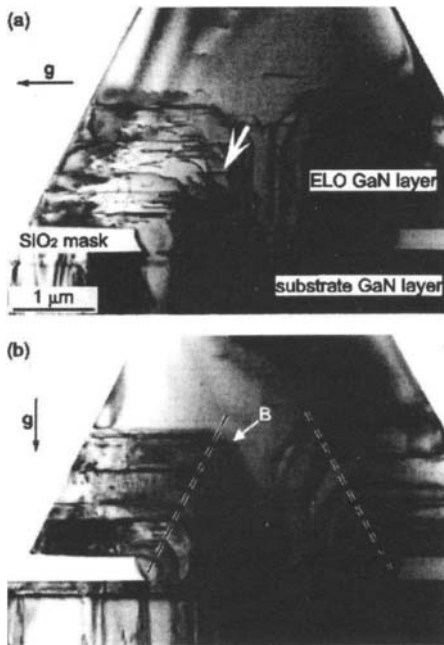
Fig. 11. Cross-sectional TEM images taken of part of the ELO layer of the HVPE $<11\bar{2}0>$ sample grown for 2.5 min: (a) $\mathbf{g}=1\bar{1}00$ and (b) $\mathbf{g}=000\bar{2}$.

In general, the dislocation morphology in crystals is very sensitive to the variation in the shape of growing surfaces because the dislocation lines tend to follow the direction of minimum energy per unit growth length [19]. Since the self-energy of a dislocation segment is determined principally by the Burgers vector and the dislocation line vector, the dislocations tend to penetrate the growing surface nearly perpendicularly because this minimizes their length. This type of dislocation morphology can be seen at a site in the Type-B dislocation indicated by the arrow B in Fig. 11(b). The abrupt bending morphology of Type-A dislocations, in contrast, does not indicate such energy-minimized structures. Their growth must therefore be governed by kinetics rather than the energetics.

To investigate the mechanism of this bending, we examined the surface morphology of ELO films before the formation of the $\{1\bar{1}01\}$ facet was completed. Fig. 12 is a high-resolution SEM image showing the surface microstructure of an ELO film at this early stage of growth [20]. Steps with bright or dark line contrast are evident on the (0001) surface, as are pits with an areal density estimated to be 1.1×10^9 cm^{-2}. This pit density suggests that the threading mixed dislocation density at this growth stage is much higher than that when the overgrowth is completed. It is clearly observed that the surface is substantially composed of the short length steps that terminate at a large number of pits. Thus we deduce that step-flow growth occurs in this early stage of growth.
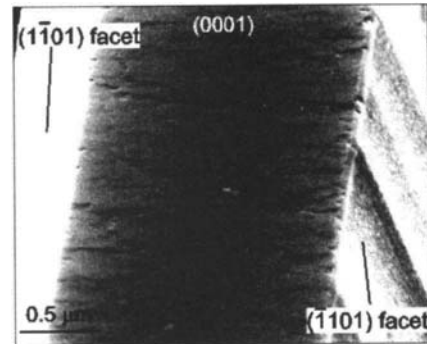
Fig. 12. SEM image of the surface step and terrace structures of a sample grown for 1 min.

A kinetic model based on the observed surface morphology has been proposed to explain the dislocation bending in the early stage of ELO [20]. Before describing it here, we recall that the bending dislocations (Type-A) are randomly distributed within the region surrounded by the $\{1\bar{1}01\}$ facets and that step-flow growth occurs on the (0001) surface of the window regions before the facet formation is completed. Fig. 13 shows one mechanism by which a pure-edge threading dislocation would bend laterally in the direction of its Burgers vector. Fig. 13(a) is a macroscopic perspective view of a hexagonal GaN cell in which the character of a vertically aligned dislocation changes its character from edge to screw when it propagates laterally on

a (0001) surface. Fig. 13(b) is an atomistic illustration of a growing (0001) surface of GaN that includes a pure-edge dislocation perpendicular to the surface. The structure of the dislocation core could be drawn from a recent result obtained by STEM observation [21]. We consider the film to grow as a result of adatoms stacking on hexagonal lattice sites on the (0001) surface. If this stacking occurs simultaneously at each site, the edge dislocation grows perpendicularly as the film does and neither lateral propagation nor character change occur. On the other hand, if the A-A step moves across the dislocation from right to left by step-flow growth, atom No. 2 atom can bond with atom No. 1 atom before bonding with atom No. 3. If the 1-2 bond is more kinetically favorable than the 2-3 bond, the 3-4 and 5-6 bonds form as the step flows, resulting in a screw dislocation parallel to the film surface. This would account for the abrupt bending of the pure-edge dislocation onto the (0001) plane, which forms the lateral segment running along the direction of the Burgers vector and having a screw character.
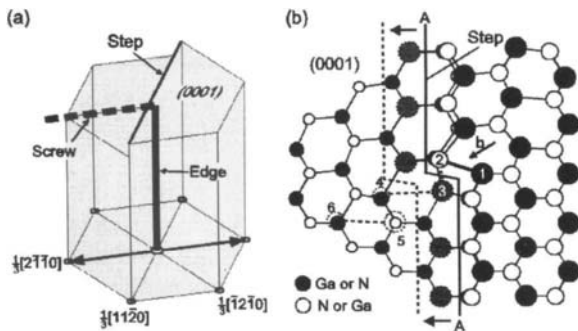
Fig. 13. Schematic diagrams showing a mechanism of the lateral propagation of a Type-A dislocation in the GaN film: (a) perspective structure of the bending dislocation; (b) atomistic representation of the lateral propagation of a dislocation aligned perpendicular to the growing (0001) surface.

The above model cannot, however, be applied to the bending of mixed threading

dislocations since, as mentioned earlier, the lateral segments of Type-B dislocations no longer lie on the slip planes of the $\{1\bar{1}01\}$. The only phenomenon that we confirmed is that the bending occurs when the growth starts on the $\{1\bar{1}01\}$ facet on which the dislocations intersect.

4.3. Self-organized propagation of dislocations during ELO

In this section, ELO films at the intermediate stage in the lateral overgrowth on the mask are examined in order to elucidate the relationship between the bending dislocation and the tilt boundaries in the overgrown regions. Fig. 14(a) is a typical cross-sectional TEM image which shows many end-on dislocations running parallel to the $[11\bar{2}0]$ direction. Some of them (indicated by arrowheads) were found to form a line at nearly regular intervals along the [0001] direction. Individual structures are directly determined from the high-resolution TEM observation of the $[11\bar{2}0]$ projection of their end-on dislocations, as in the example shown in Fig. 14(b). As a result, every end-on dislocation observed in the sample is found to have the form of a positive dislocation with its extra-half-plane pointing towards the film surface, as denoted by ⊥ in Fig. 14(b). Such structures and their regular arrangement are quite similar to those of the D2 defect shown in Fig. 9(c).

Fig. 15(a) shows a plan-view TEM image of the same sample whose cross-sectional TEM images are shown in Fig. 14. Note two types of distinguishable dislocation morphology as well as Type-B dislocations that propagate approximately normal to the mask stripe. One is shown by a bundle of dislocations (denoted BD) corresponding to the end-on dislocations seen in Fig. 14 can be seen along the mask stripe in the overgrown region. The other type, an example of which is indicated by arrow A, is shown by dislocations propagated from the window regions to the bundle. By using systematic contrast analysis, we identified these dislocations as Type-A

dislocations with Burgers vectors of the $b=<11\bar{2}0>$ type. The TEM images in Figs. 14 and 15 unambiguously show that Type-A dislocation segments parallel to the c-plane of GaN, which come from the window regions, propagate along the mask stripe direction and form into a bundle in the overgrown region.
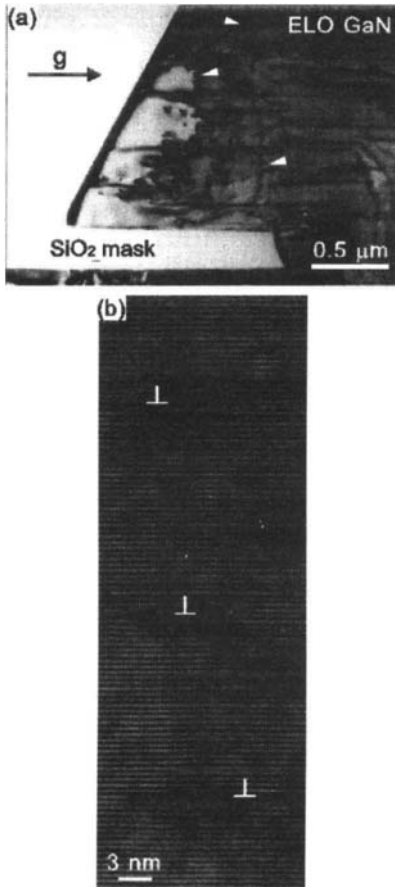

Fig. 14. (a) Cross-sectional TEM $[11\bar{2}0]$ image taken with $g=\bar{1}100$ and showing dislocations in the HVPE $<11\bar{2}0>$ sample at an intermediate stage of lateral overgrowth. (b) Cross-sectional high-resolution image of the end-on dislocations in the same sample.
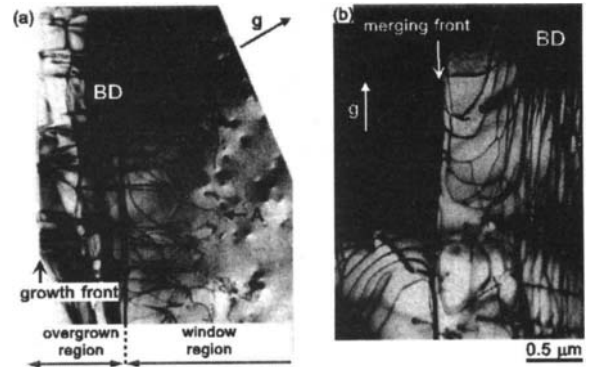
Fig. 15. Plan-view TEM image showing dislocation morphology in the HVPE $<11\bar{2}0>$ sample (a) before and (b) after just after the coalescence of the overgrown region.

We infer from the above results that self-organized propagation of Type-A dislocations results in a bundle consisting of dislocations whose Burgers vectors have the same (positive) sign. Since the Burgers vectors of Type-A dislocations on the prismatic-planes of the hexagonal crystal also lie on the basal c-plane and their lateral segments initially propagate parallel to their Burgers vectors, it is plausible that the lateral screw segments readily cross-slip onto the c-plane. As mentioned earlier, however, dislocation lines tend to follow the direction of minimum energy per unit growth length. Therefore, to explain the existence of prolonged dislocation line running along the mask, we have to take an external force into account, which drives the propagation of Type-A dislocations along the mask stripe direction during the ridge growth in ELO. This has been discussed in detail by Sakai *et al.* [20].

4.4. Mechanism of threading dislocation reduction and c-axis tilting

A plan-view TEM image of the overgrown regions in a HVPE $<11\bar{2}0>$ sample at the growth stage just after coalescence was shown in Fig. 15(b), and several dislocations that run from a dislocation bundle to a

merging front and then propagate in the opposite direction can be seen there. This clearly shows that at the coalesced site are formed negative dislocations along the mask stripe, one of whose origins is the positive Type-A dislocations in the bundle. It is clear that the bundle of positive dislocations causes the overgrown region to bend such that the c-axis tilts toward the center of the mask. Since most of the dislocations in the bundle terminate at the growth front during overgrowth, the termination site nucleates a new dislocation when adjacent facets coalesce and it is buried. In this process, in order to relieve excess strain energy accumulation, there is a driving force on the newly formed dislocation that makes them propagate along the reversed direction to that of the positive dislocations. This makes it difficult for Type-A dislocations to propagate vertically after the coalescence and thus reduces the threading dislocation density in ELO films.

We infer the following scenario for Type-A dislocation propagation and the defect evolution accompanied by c-axis tilting (see Fig. 16). After dislocation bending occurs because of step-flow growth on the (0001) surface of the seeding region grown from the window area (Fig. 16(a)), the film that contains the bending Type-A dislocations starts to overgrow on the mask while being stressed (Fig. 16(b)). The stress induces a force exerted on the Type-A dislocations lying on the c-plane so that "sweep-out" of these dislocations occurs by the self-organized propagation along the mask stripe. The resultant positive dislocations in the overgrown region are the origin of the D2 type of defects. Then the negative dislocations are formed at the coalesced site partly by the reversed propagation of the positive dislocations such that the c-axis tilting caused by their arrangement can be accommodated (Fig. 16(c)). This corresponds to the formation of a tilt boundary in the D1 defect near the center of the mask.
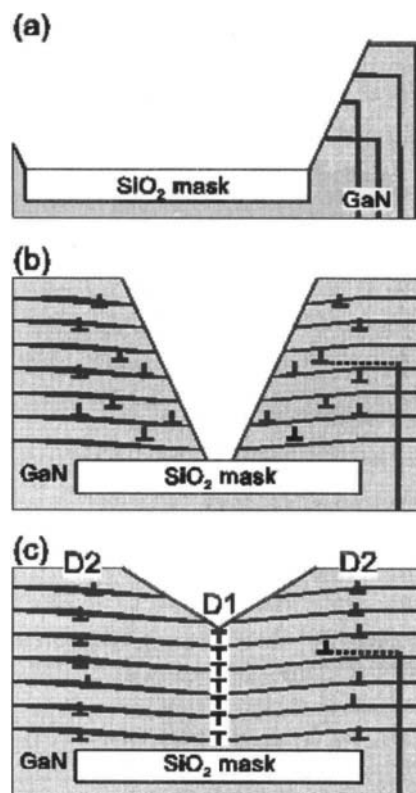
Fig. 16. Defect evolution during ELO.

At the same time, the final arrangement of the D2 defect is determined by the interaction of the positive dislocations in the stress field of the coalesced film (Fig. 16(c)). Typically, in the case of the $<11\bar{2}0>$ mask, all positive dislocations form an energetically stable arrangement—piling up side by side along the c-axis—as a result of the interactive forces due to their edge component as well as either attractive or repulsive forces due to their screw component. This implies that a kind of "polygonization" [17] might occur in the overgrown region during ELO. Hence this polygonization may leads to the formation of the D2 defect having a complete form of a tilt boundary.

5. CHARACTERISTICS OF FIELO-GAN

In this section we show that FIELO-GaN

layers more than 100 μm thick have low dislocation densities and good structural quality. We also describe the excellent optical and electrical properties of FIELO-GaN layers.
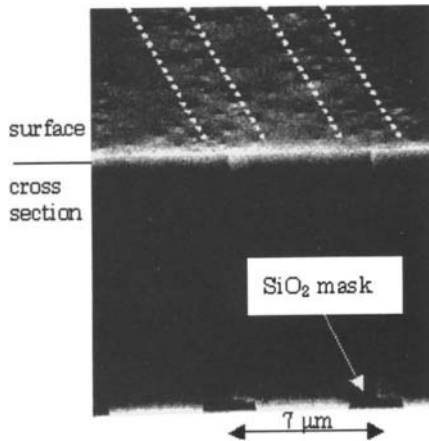
Fig. 17. SEM image showing the relationship between the mask pattern and the etch-pits on the top surface when the FIELO layer is thin (10 μm).

5.1. Thickness dependence of dislocation density

We found that the residual dislocation density of GaN layers grown with the FIELO technique could be reduced by growing thicker layers. Fig. 17 shows SEM images of etch-pits produced by dipping the sample into a hot (230°C) H_2SO_4: H_3PO_3 solution. When the GaN layer was thin (10 μm), the etch-pit-density (EPD) increased just above the mask region, as shown in dotted lines of Fig. 17(a), while the EPD on the window region decreased to 10^7 cm^{-2}. When the GaN layer was thick, on the other hand, the distribution of etch-pits was low and uniform on the entire wafer. The relation between EDP and GaN thickness is shown in Fig. 18 for GaN layers ranging from 10 to 600 μm thick. Dislocation densities determined from plan-view TEM observation of some of the samples are also shown there. When the EPD is high, the counting of etch-pits sometimes results in underestimations because more of the pits overlap. However, good agreement

between the EPD and the dislocation density was obtained in lower EPD region. This Fig. clearly shows that thicker films have lower EPDs. The lowest EPD obtained in this experiment was 8×10^6 cm^{-2}. The lower dislocation densities in thicker films might be due to some of the threading dislocations forming closed loops during the growth. Some dislocations are also thought to be swept out the crystal.
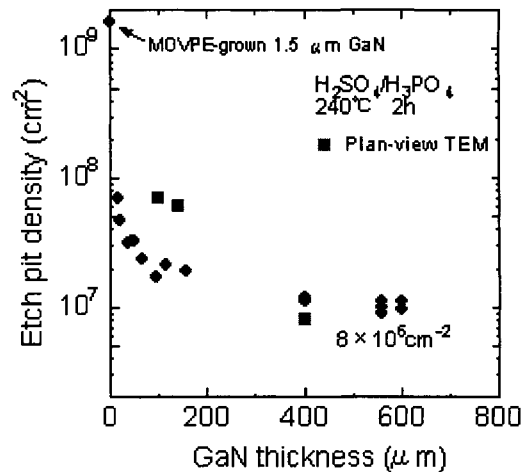
Fig. 18. Thickness dependence of etch-pit density. Also shown are dislocation densities determined from plan-view TEM observation of some samples.

5.2. Scanning reflection electron microscopic evaluation of *c*-axis tilting on the top surface

TEM observation of the mask region of FIELO-GaN revealed that there are two types of defects aligned along the [0001] direction. Both types are tilt boundaries, but the dislocations of which they are composed have Burgers vectors with opposite signs (Fig. 10). This leads *c*-axis tilting of the overgrown region by about 1°. It is interesting to know what extent the tilting remains on the surface when GaN layers are thick. We recently used scanning reflection electron microscopy (SREM) to investigate the dependence of the tilt angle on the thickness of FIELO-GaN [22]. The experimental system was equipped with a field emission electron gun, a secondary

electron detector for SEM, and a fluorescent screen for SREM. A 30-keV electron beam with a diameter of less than 3 nm was used for the SREM and SEM observations. The SREM images were obtained by recording the change in specular reflection spot intensity on the fluorescent screen through an optical lens. SREM shows atomic-steps on solid surfaces as dark contrast lines since the Bragg condition is disturbed at the steps. And because the image contrast is sensitive to the electron diffraction angle, we could see the local surface distortion that accompanies threading dislocations and crystallographic tilting. In this study, we used FIELO-GaN samples with thicknesses of 50 and 100 μm.

Fig. 19 shows SEM and SREM images of a 50-μm-thick FIELO-GaN film with $<1\bar{1}00>$ strips of SiO_2 mask. The upper portions of (a) and (b) are SREM images of the GaN surface, and the lower portion of (a) is an SEM image of the cross-section in which we can recognize the SiO_2 masks grown on the MOVPE-grown GaN layer. Note that the stripe contrast on the FIELO-GaN surface, which runs straight along the $<1\bar{1}00>$ direction, can be seen clearly in these SREM images. Comparing the SREM and cross-sectional SEM images, we observed that the period of the dark stripes was equal to that of the SiO_2 masks at the interface and that the overgrown regions on the masks showed dark contrasts as indicated by the dotted lines in Fig. 19(a). By changing the position of the optical lens from this optimal position, we found that changing the position of the optical lens caused the stripe contrast to disappear or reverse. From the results of previous TEM studies and this SREM study, we can conclude that the stripe contrast in SREM images corresponds to the crystallographic *c*-axis tilting of the overgrown areas on the surface, as illustrated in Fig. 20. The tilt angle estimated from the movement of the optical lens was less than 0.3°. This result is consistent with those of a previous x-ray diffraction study [23]. On the mask, two domains with opposed tilt angles coalesced as shown in Fig. 20 (this behavior was described in Section 4.4.), but the

uniform SREM contrast within most of the overgrown areas indicates that one of the two became dominant and that the GaN surface changed to a single domain.

(a)

(b)
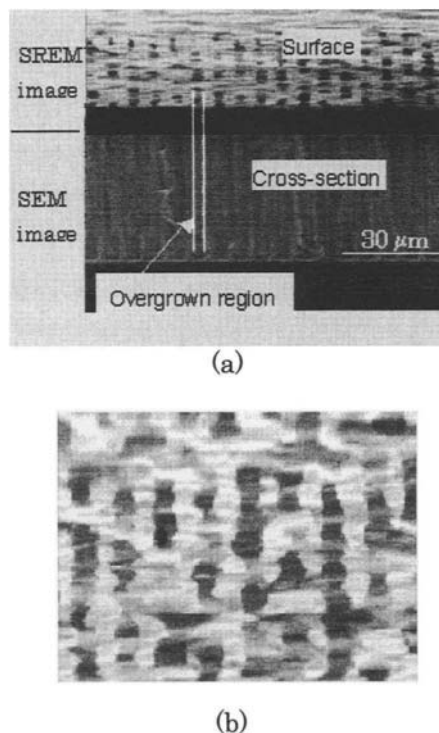
Fig. 19. (a) SEM and (b) SREM images of a 50-μm-thick FIELO-GaN film. The upper portion of (a) and all of (b) are SREM images of the GaN surface, and the lower portion of (a) is an SEM cross-sectional image.

As described in Section 5.1, the dislocation density and crystallographic structure depend on the thickness of the HVPE-GaN layer. We therefore studied GaN layers more than 100 μm thick, which have a dislocation density of roughly 10^7 cm^{-2} on the top surface. Fig. 21 shows SREM images of a 100-μm-thick FIELO-GaN film. Although a few areas showing faint contrast are evident, the stripes due to the crystallographic tilting and twisting that we could see on a thinner GaN film are not. Given the detection limit of the SREM method, we can conclude that the tilt angle of the overgrown areas in 100-μm-thick

GaN was less than 0.05°. These SREM results indicate that a thicker FIELO-GaN film has better crystallographic structure in the overgrown areas and also has a smaller threading dislocation density.
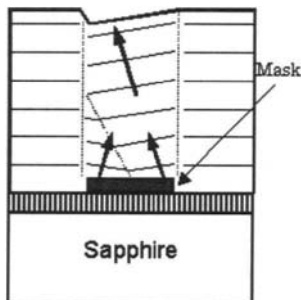

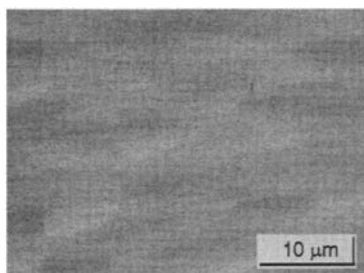Fig. 20. Merging two tilting domains into a single domain.

Fig. 21. SREM image of a 100-μm-thick FIELO-GaN film.

5.3. Surface morphology

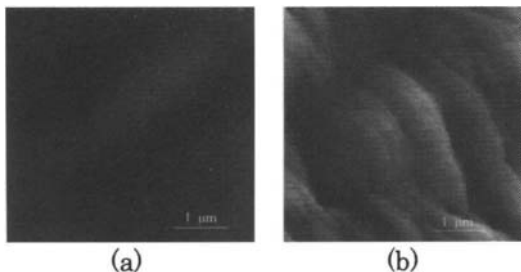We used atomic force microscopy (AFM) to examine the surface morphology of FIELO-GaN layers in order to clarify the growth mode.

Fig. 22. AFM Images of (a) FIELO-GaN and (b) MOVPE-grown GaN surfaces.

Fig. 22(a) shows an AFM image of a sample

grown at 100 μm/h. A clear step-flow mode was observed on this sample, and it was confirmed that the step-flow growth occurred at growth rates up to at least 120 μm/h. We found this growth mode on the entire surface of 2-inch FIELO-GaN wafer, in mask regions as well as window regions. The surface of a GaN layer grown by MOVPE, in contrast, was very rough ((b)). There were many traces of end-on dislocations, and steps were disturbed or terminated by these dislocations.
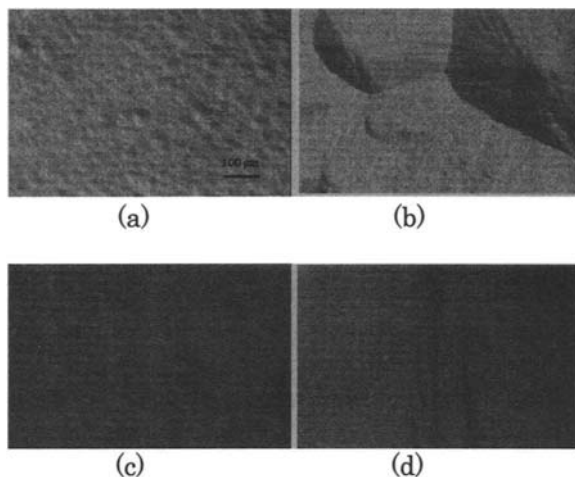
Fig. 23. Photomicrographs showing the relation between the surface morphology of FIELO-GaN ((b) and (d)) and that of the underlying MOVPE-grown GaN ((a) and (c)).

Fig. 23 shows photomicrographs of two MOVPE-GaN layers of the GaN grown on those surfaces by FIELO. Fig. 23(b) shows that the surface of the FIELO-GaN had many hillocks when the underlying MOVPE-grown GaN had the rough morphology shown in Fig. 23(a). And as shown in Fig. 23(c) and (d), the surface of GaN layer FIELO-grown on a MOVPE-grown layer with a smooth surface was markedly better. We found that, even in the worst case, the skirts of hillock were then very long (100 μm to 1 mm) compared with the heights of the hillocks (10 nm to 1 μm). When several layers for LD structure were deposited on such a substrate by MOVPE, the morphology was basically inherited from the

substrate surface. The scale of the undulation, however, was small enough that the undulation did not markedly impair the LD functions. The FIELO-GaN wafer can therefore be used as an "epi-ready" substrate. No polishing or chemical etching is necessary to flatten the surface before it can be used as a substrate

5.4. Crystal quality evaluated by x-ray diffraction

We evaluated the crystal quality of GaN wafers by measuring the full-width-at-half-maximum (FWHM) of the x-ray rocking curve (XRC) [23]. The c-axis tilt angle was measured by the conventional out-of plane (e.g. (0002)) XRC, while the c-axis twist angle was directly determined using grazing incidence angle (GID) x-ray diffraction by measuring the FWHM of $(10\bar{1}0)$ diffraction peak. The typical glancing angle was about 0.1° and the penetration depth of x-ray was about 0.1 nm. The Cu $K\alpha_1$ line (λ=1.5405 Å) was used.

The twist and tilt angles obtained for two FIELO-GaN wafers of different thicknesses and a 10-μm-thick GaN wafer grown by the conventional HVPE are summarized in Table 1, where the improvements due to the FIELO method are clearly shown by the values listed for both angles. The twist angle in the 550-μm-thick FIELO-GaN sample was only 15% of that in the conventional HVPE-grown sample. This value is among the best obtained for a GaN epitaxial layer. It should also be noted that improvement was observed even in the 11-μm-thick FIELO sample. The large twist indicates that there were many columnar grains misoriented to neighboring

grains exist and there were many threading dislocations at grain boundaries. Because the TEM observations had shown that most of the dislocations propagate laterally before the thickness of the FIELO-grown layer reaches 5 μm, we think that the reduction of dislocation density resulted in larger, less misoriented grains.

TEM plan-view observation can be used to evaluate the dislocation density accurately, but is hard to use when the dislocation density regions is less than about 10^7 cm^{-2}. Etch-pit measurement is an alternative method, but the correspondence between etch pits and dislocations is not clearly understood. Therefore, we tried to evaluate the propriety of dislocation density estimated from EPD and TEM measurements by measuring twist and tilt angles obtained from FWHM of XRC as mentioned above. The samples used for the measurements were MOVPE-grown GaN with thicknesses of 1 to 2 μm and FIELO-GaN wafers with thicknesses of 110 to 550 μm. The dislocation densities of MOVPE-grown GaN and FIELO-GaN samples were evaluated from plan-view TEM images and EPD measurements, respectively. Etch pits were obtained by dipping wafers into a hot (230°C) $H_2SO_4+H_3PO_3$ solution for about 2 hours. The relation between dislocation density and the FWHM is shown in Fig. 24, where values for bulk GaN crystal obtained by the solution growth [4] are also shown. In the MOVPE-grown GaN samples, the twist angle was much greater than the tilt angle. This agrees well with the TEM study, which showed that the pure-edge dislocation responsible for the c-axis twisting accounts for the majority of threading dislocations [9].

Table 1. Twist and tilt angles of the c-axis in FIELO-grown and HVPE-GaN sample.

	FIELO-GaN#1	FIELO-GaN#2	HVPE-GaN
Thickness (μm)	11	550	10
Twist angle (arcsec)	288	184	1188
Tilt angle (arcsec)	194	90	317

In the FIELO-GaN samples, on the other hand, the twist angle decreased to the same level as that of the tilt angle. As previously reported [9], about 70% of pure-edge dislocations from the underlying MOVPE-grown GaN layer are bent parallel to the interface by the formation of the facet structure early in the FIELO process of GaN. As a result, the ratio of the pure-edge-dislocation density to the total dislocation density was about 20 percent in the top surface of the FIELO-GaN layer. The result of XRC measurements reflects the improvement of crystal quality by the reduction of pure-edge dislocation density. It is interesting to see the relationship of the dislocation density and FWHM in the case of extremely low dislocation density. Porowski reported that the dislocation density of Mg-doped SG-GaN was less than 100 cm^{-2} [4], and FWHMs corresponding to the tilt and twist angel were around 25 arcsec in both angles. These results suggest that the FWHM should be less than 50 arcsec when the dislocation density is 10^5 's cm^{-2}.
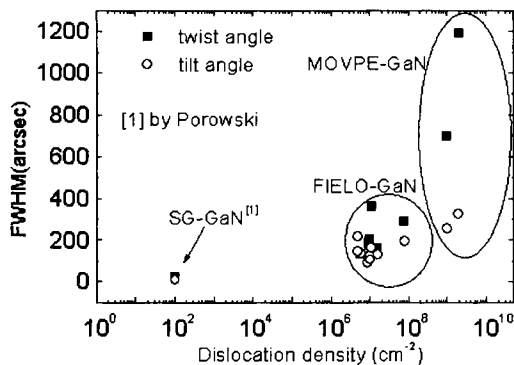
Fig. 24. Relationship between dislocation density and the FWHM of the x-ray rocking curve for MOVPE-grown and FIELO-GaN samples. Values for bulk GaN crystal, reported by Porowski [4], are also shown.

Gay et al have calculated the dislocation density from FWHM by using the equation $D = F^2 /(9b^2)$ [D : dislocation density, F : FWHM (radians), b : Burger's vector] [24].

We used this equation to estimate the dislocation density from XRC results, and Fig. 25 shows the relationship between the dislocation density obtained experimentally and that obtained by calculation. Only the twist angle whose the Burger's vector is $1/3 < 11\overline{2}0 >$ was taken into account in this calculation. Therefore, the value of a-axis lattice-constant (0.3189 nm) was employed as $|b|$. It was found that the calculated result generally explained the experimental data well, although the discrepancy increased slightly at low dislocation densities. Thus, we can roughly estimate the dislocation density from the twist angle obtained from XRC measurement.
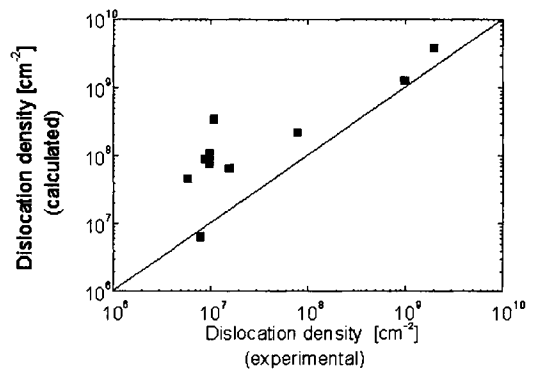
Fig. 25. Relationship between the dislocation density obtained experimentally and that calculated from the twist angle.

5.5. Optical properties

The reflectance spectrum for a 130-μm-thick FIELO-GaN sample is shown in Fig. 26, where two dispersive structures due to A and B excitons are clear. Three other structures can also be seen. These structures are assigned to the first excited-state (n=2) transition of A and B free excitons (A$_2$, B$_2$) and the ground-state transition (n=1) of the C free exciton. To the best of our knowledge, such a clear observation of A$_2$, B$_2$, and C excitons has not been reported in reflectance measurements made without using the modulation technique. Fig. 27 is a low temperature (1.3 K) PL spectrum for a free-standing FIELO-GaN wafer. Strong I_2

emission, which is attributed to excitons bound to neutral donors, was seen around 3.472 eV, and the FWHM of this is only 1.5 meV. The free-exciton emissions from A and B excitons is clear at 3.478 eV and 3.484 eV, and higher-energy emission denoted by A_2 can be resolved into three peaks. In most previous reports, these peaks have not been resolved and this emission has been treated as a single peak.

Yamaguchi *et al.* recently reported a very long decay time for a FIELO-GaN layer (1.3 ns) from the results of their time-resolved and temperature-dependent PL measurements [25]. When that GaN was used as a substrate for InGaN quantum well structure grown by MOVPE, the defect density and the fluctuation of In composition in the QW layers were much lower than those on the sapphire substrate.
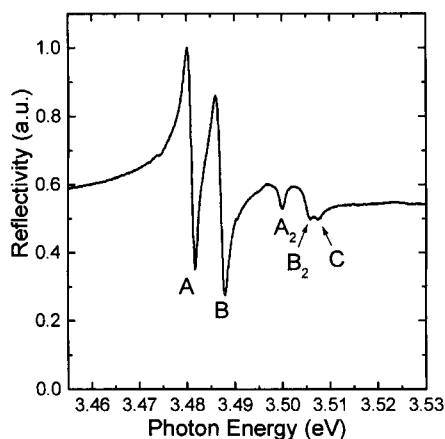
Fig. 26. Reflectance spectra at room temperature for a 130-μm-thick FIELO-GaN sample.

5.6. Electrical properties

The silane SiH_4 is generally used as a precursor for Si-doping (n-type) in MOVPE but is easily decomposed upstream in the hot-wall HVPE reactor, so we used thermally stable SiH_2Cl_2 instead of SiH_4 for the Si-doping into FIELO-GaN layers. The relation between the carrier concentration and mobility at room temperature is shown in

Fig. 28, which also shows the data of Nakamura *et al.* [26]. All of those samples were free-standing GaN wafers 100–130 μm thick from which the mask stripe pattern region had been removed. It is clearly shown that the present data are superior to those of MOVPE-grown samples. Furthermore, the mobilities measured for an undoped GaN layer 270 μm thick were very high: 860 $cm^{-2}/V/s$ (n=4.7x10^{15} cm^{-3}) at 300 K and 2780 $cm^{-2}/V/s$ (n=8.7x10^{14} cm^{-3}) at 77 K.
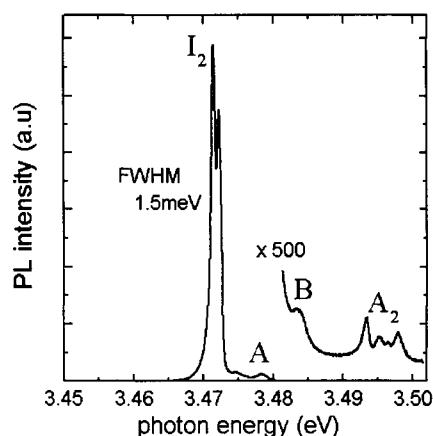
Fig. 27. Low-temperature (1.3 K) PL spectrum for free-standing FIELO-GaN.
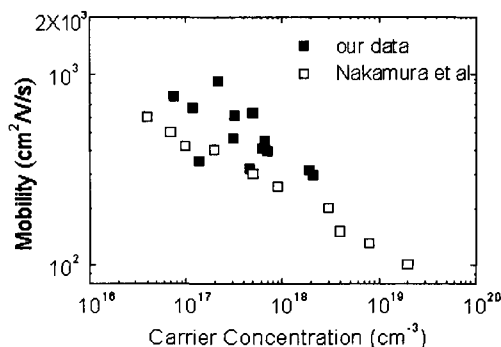
Fig. 28. Relationship between the carrier concentration and mobility at room temperature. Open squares show data by Nakamura *et al.* [26].

6. CONCLUSION

We developed an epitaxial lateral

overgrowth (ELO) method for reducing the dislocation density of GaN epitaxial layers. The results of TEM observations showed that it resulted in threading dislocation densities as low as 6×10^7 cm^{-2}, which is about two orders of magnitude smaller than that of GaN epitaxial layers on a sapphire substrate.

We confirmed that the dislocation bending plays an important role in reducing the threading dislocation density in the FIELO-GaN layers. The observed dislocation morphology suggests that the occurrence of the bending is closely related to the appearance of the facets early in ELO. This is why we call this method facet-initiated ELO (FIELO). We determined the nature of bending dislocations, and here we discussed in detail the formation of a tilt boundary and the resultant c-axis tilting.

Since FIELO-GaN layers more than 100 µm thick were found to have dislocation densities lower than those of thinner layers, thick FIELO-GaN samples were the focus of the studies reported here. Using scanning reflection electron microscopy (SREM), we found the tilt angle on the top GaN surface of these samples to be less than 0.05°. X-ray measurements revealed that the c-axis twist angle decreased along with the dislocation density, and we showed that the dislocation density could be estimated from the twist angle. AFM measurements showed very smooth step-flow growth even when the growth rate was greater than 100 µm/h. Well-resolved dispersive structures for free excitons were clearly be observed in reflectance measurements of thick FIELO-GAN layers, and low-temperature (5 K) PL measurements showed a very strong band-edge emission with a FWHM of only 1.6 meV. An undoped sample of free-standing GaN showed very high mobilities: 860 cm^{-2}/V/s (n=4.7×10^{15} cm^{-3}) at 300 K and 2,780 cm^{-2}/V/s (n=8.7×10^{14} cm^{-3}) at 77 K. These results indicate that the electrical quality and optical quality FIELO-GaN are very high.

ACKNOWLEDGEMENTS

REFERENCES

1. S. Nakamura and G. Fasol, The Blue Laser Diode (Springer, Berlin, 1997).
2. S. Nakamura, M. Senoh, S. Nagahama, N. Iwasa, T. Yamada, T. Matsushita, H. Kiyoku and Y.Sugimoto, Jpn. J. Appl. Phys., 35 (1996) L74.
3. S. Nagahama, N. Iwasa, M. Senoh, T. Matsushita, Y. Sugimoto, H. Kiyoku, T. Kozaki, M. Sano, H. Matsumura, H. Umemoto, K. Chocho and T. Mukai, Jpn. J. Appl. Phys., 39 (2000) L647.
4. S. Porowski, Extended Abstracts of the 1999 Int. Conf. on Solid State devices and Materials, Tokyo, 1999, pp. 58-59.
5. T. Inoue, Y. Seki, O. Oda, S. Kurai, Y. Yamada and T. Taguchi, Extended Abstracts of the 1999 Int. Conf. on Solid State Devices and Materials, Tokyo, 1999, pp. 60-61.
6. S. Zhang and T. Nishinaga, Jpn. J. Appl. Phys., 29 (1990) 545.
7. Y. Ujiie and T. Nishinaga, Jpn. J. Appl. Phys., 28 (1989) L337.
8. A. Usui, H. Sunakawa, A. Sakai and A.A.Yamaguchi, Jpn. J. Appl. Phys., 36 (1997) L899.
9. A. Sakai, H. Sunakawa, and A. Usui, Appl. Phys. Lett., 71 (1997) 2259.
10. S. Nakamura, M. Senoh, S. Nagahama, N. Iwasa, T. Yamada, T. Matsushita, H. Kiyoku, Y. Sugimoto, T. Kozaki, H. Umemoto, M. Sano and K. Chocho, Jpn. J. Appl. Phys., 36 (1997) L1568.
11. S. Nakamura, M. Senoh, S. Nagahama, N. Iwasa, T. Yamada, T. Matsushita, Y. Sugimoto and H. Kiyoku, Jpn. J. Appl.

Phys. Lett., 36 (1997) L1059

12. M. Kuramoto, C. Sasaoka, Y. Hisanaga, A. Kimura, A.A.Yamaguchi, H. Sunakawa, N. Kuroda, M. Nido, A. Usui, and M. Mizuta, Jpn. J. Appl. Phys., 38 (1999) L184.

13. X.H. Wu, L.M. Brown, D. Kapolnek, S. Keller, B. Keller, S.P. DenBaars and J.S. Speck, J. Appl. Phys., 80 (1996) 3228.

14. W. Qian, M. Skowronski, M. De Graef, K. Doverspike, L.B. Rowland and D.K. Gaskill, Appl. Phys. Lett., 66 (1995) 1252.

15. X.J. Ning, F.R. Chien, P. Pirouz, J.W. Yang and M. Asif Khan, J. Mater. Res., 11 (1996) 580.

16. X.H. Wu, P. Fini, E.J. Tarsa, B. Heying, S. Keller, U.K. Mishra, S.P. DenBaars and J.S. Speck, J. Cryst. Growth, 189/190 (1998) 232.

17. J.P. Hirth and J. Lothe, *Theory of Dislocations* (Krieger, Florida, 1992).

18. P.B. Hirsch, A. Howie, R.B. Nicholson, D.W. Pashley and M.J. Whelan, *Electron Microscopy of Thin Crystals* (Butterworths, London, 1965).

19. H. Klapper, Y.M. Fishman and Y.G. Lutsau, Phys., Stat. Sol. (a), 21 (1974) 115.

20. A. Sakai, H. Sunakawa, A. Kimura and A. Usui, J. Electron Microscopy, 49 (2000) 323.

21. Y. Xin, S.J. Pennycook, N.D. Browning, P.D. Nellist, S. Sivananthan, F. Omnes, B. Beaumont, J.P. Faurie and P. Gibart, Appl. Phys. Lett., 72 (1998) 2680.

22. H. Watanabe, N. Kuroda, H. Sunakawa and A. Usui, Appl. Phys. Lett., 77 (2000) 1786.

23. K. Kobayashi, A.A. Yamaguchi, S. Kimura, H. Sunakawa, A. Kimura and A. Usui, Jpn. J. Appl. Phys., 38 (1999) L611.

24. P. Gay, P.B. Hirsch and A. Kelly, Acta Metall., 1 (1953) 315.

25. A.A. Yamaguchi, Y. Mochizuki and M. Mizuta, Jpn. J. Appl. Phys., 39 (1999) 2402.

26. S. Nakamura, T. Mukai and M. Senoh, Jpn. J. Appl. Phys., 31 (1992) 2883.

Effects of buffer layers and advanced technologies on heteroepitaxy of GaN

Kazumasa Hiramatsu

Department of Electrical and Electronic Engineering, Mie University, 1515 Kamihama, Tsu, Mie 514-8507, Japan
e-mail: hiramatu@elec.mie-u.ac.jp

The effects of buffer layers on the heteroepitaxial growth of GaN are reviewed. Heteroepitaxial growth of GaN by metalorganic vapor phase epitaxy (MOVPE), hydride vapor phase epitaxy (HVPE) and molecular beam epitaxy (MBE) have been carried out using various buffer layers. Buffer layers are very important for obtaining high-quality GaN single crystals. The following buffer layers are mainly employed in MOVPE: (1) a low temperature deposited AlN or GaN buffer layer in MOVPE on sapphire (0001), (2) a high temperature deposited AlN in MOVPE on 6H-SiC, (3) a high temperature deposited AlN in MOVPE on Si (111). Other substrates, buffer layers and growth methods (HVPE and MBE) have also been studied. These buffer layers were found to play roles in (a) high-density nucleation of GaN on the buffer layers, (b) arrangement of crystalline directions of GaN islands, (c) quasi two-dimensional growth (lateral growth) of GaN, (d) coalescence of GaN and (e) step flow growth. Advanced technologies related to buffer techniques are also reviewed in this chapter.

1. INTRODUCTION

GaN is one of the most promising materials for optical devices in the region from blue to ultraviolet light and for electronic devices used in high power, high frequency and high temperature. It is extremely difficult to grow a large-scale bulk single crystal of GaN because of the high equilibrium pressure of nitrogen at a growth temperature of more than 1000°C. Therefore, vapor phase epitaxial methods, such as metalorganic vapor phase epitaxy (MOVPE) and hydride vapor phase epitaxy (HVPE) using various substrates such as sapphire (α-Al$_2$O$_3$) or 6H-SiC have been tested for growing GaN single crystal films. However, it was fairly difficult to grow high quality epitaxial films, particularly with smooth surfaces free from cracks, because of the large lattice mismatch and the large difference in the thermal expansion coefficient between GaN and those substrates, as shown in Table 1.

To solve this problem, the deposition of a thin buffer layer before GaN growth via MOVPE, HVPE and molecular beam epitaxy (MBE) was proposed. In the case of MOVPE, sapphire, 6H-SiC and Si substrates are mainly used. Surface morphology as well as electrical and optical properties of GaN films has been improved remarkably by the preceding deposition of an AlN [1] or a GaN [2] buffer layer before MOVPE growth of GaN films. Recently, in order to improve the quality of a GaN layer with low temperature buffer layer, in situ monitoring [3, 4] and low temperature interlayer technique [5, 6] are developed. In the

Table 1 Lattice constants and thermal expansion coefficients for various substrates

Substrate	Lattice Constant a (nm)	Mismatch (%)	Thermal Expansion Coefficient ($\times 10^{-6}$ K^{-1})	Mismatch (%)
α-Al_2O_3	0.4758	-13.9	7.5	34
6H-SiC	0.308	-3.4	4.2	-25
Si	0.5431	20.4	3.6	-36
GaAs	0.56533	25.3	6.0	7.0
ZnO	0.325	1.9	2.9	-48
$MgAl_2O_4$	0.8083	-10.3	7.5	33
$LiGaO_2$	0.3186	-0.09	7.0	25

case of HVPE, sapphire substrates and a ZnO buffer layer are mainly used to grow thick GaN layer [7]. The fabrication of freestanding GaN substrates has also been reported [7]. Besides a sapphire substrate, 6H-SiC [8], GaAs [9, 10] and NdGaO₃ (NGO) [11] substrates are used. Recently, large freestanding GaN substrates were fabricated by HVPE using GaAs substrates [10]. In the case of MBE, buffer layers or nitridation of substrate [12, 13] have been used as the initial growth technique. The first report on the use of a buffer layer technique in MBE was a report on the growth of GaN on a sapphire substrate with an AlN buffer layer in 1983 [14]. In 1993, it was reported that a GaN buffer layer could be used as well as an AlN buffer layer [15]. The use of 6H-SiC [16] and Si [17-19] substrates has also been reported.

In this chapter, I will show how the crystalline properties of GaN films grown by MOVPE and HVPE are improved by using various buffer layers and advanced technologies for buffer layer techniques.

2. MOVPE (metalorganic vapor phase epitaxy)

2.1. Sapphire Substrate

2.1.1. Low Temperature Buffer Technique

A horizontal type MOVPE reactor operated at an atmospheric pressure was used for the growth. An optical grade polished (0001) sapphire was used as the substrate. The off-angle of the sapphire substrate was less than 0.5 degrees. The substrate was placed on a graphite susceptor, which was heated by an r.f. Figure 1 shows an illustration of the growth process with the AlN buffer layer. Prior to the growth, the substrate was heat-treated at 1150°C for 10 min in a stream of H₂ to remove surface damage and impurity. Trimethylgallium (TMG), trimethylaluminum (TMA) and ammonia (NH₃) were used as source materials for GaN or AlN growth. The carrier gas was hydrogen (H₂). Before the growth of the GaN film, a thin AlN buffer layer of about 50 nm in thickness was deposited at about 600°C. Then, the substrate temperature was raised to a growth temperature of about 1050°C, and a GaN film was grown. Figure 2 shows the surface morphology of GaN with and without the AlN buffer. The surface morphology as well as the electrical and optical properties of the GaN film were remarkably improved by the preceding deposition of a thin AlN layer as a buffer layer before the GaN growth [1].

In order to clarify the role of the AlN layer, every stage of GaN growth was studied by SEM. Figure 3 shows the changes in surface morphology during the early growth stage of GaN. After 5 min of growth, as shown in Fig. 3(a), many truncated hexagonal pyramidal mesas are formed. Then,
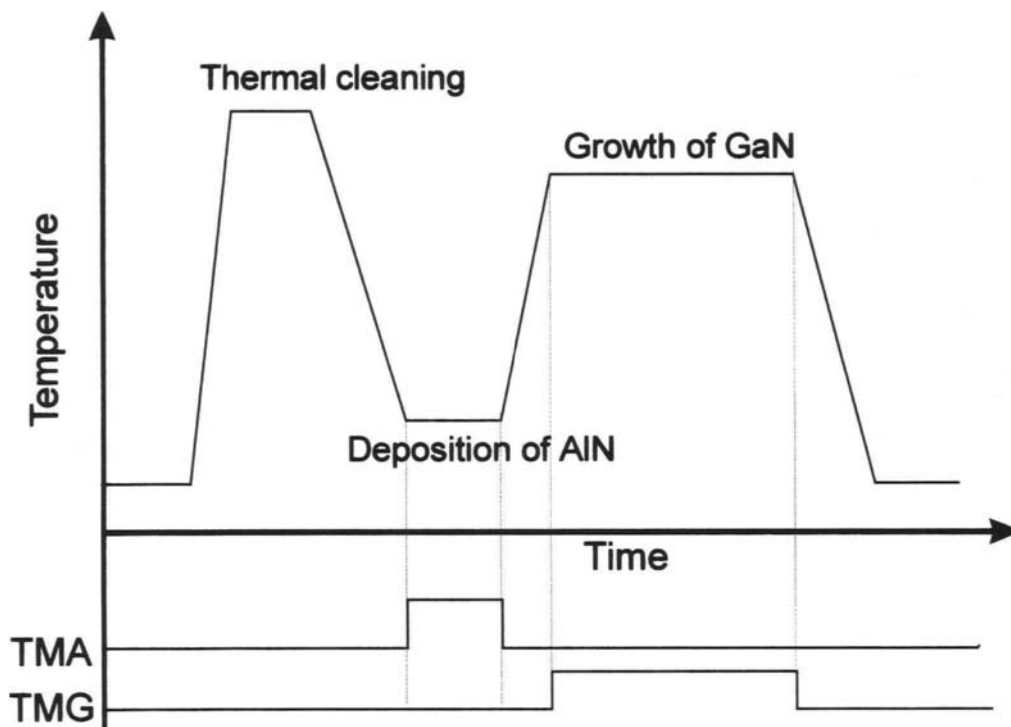

Figure 1. Illustration of the growth process with an AlN buffer layer.
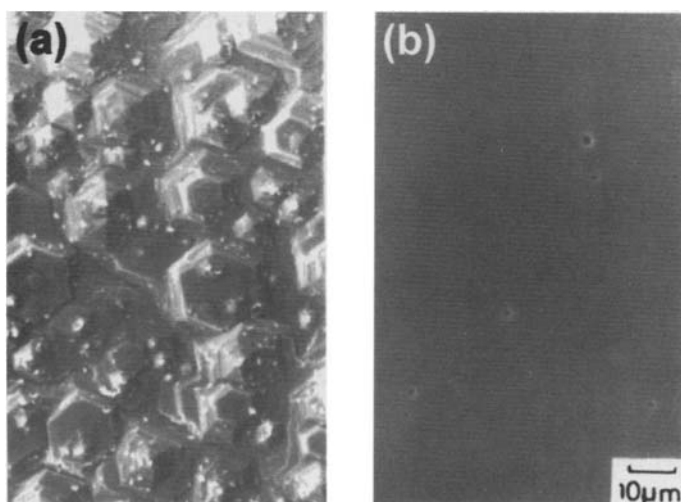

Figure 2. Surface morphology of GaN without (a) and with (b) an AlN buffer layer.

two-dimensional lateral growth proceeds favorably for a certain period of the growth time (Figs. 3(b) and 3(c)). This clearly indicates that quasi-lateral growth dominates at a certain thickness of the GaN film. Finally, the whole area of the substrate is covered by a GaN film with a flat surface, as shown in Fig. 3(d).

From electron diffraction spots, the crystallographic relations between GaN, AlN and α-Al$_2$O$_3$ were found to be [0001] GaN ‖ [0001] AlN ‖ [0001] Al$_2$O$_3$ and [1-100] GaN ‖ [1-100] AlN ‖ [11-20] Al$_2$O$_3$. The sharp spots of AlN indicate that the AlN is crystallized epitaxially on the sapphire substrate during the raising of the temperature and/or the growth of GaN. The above relations agree with those in the case of GaN grown directly
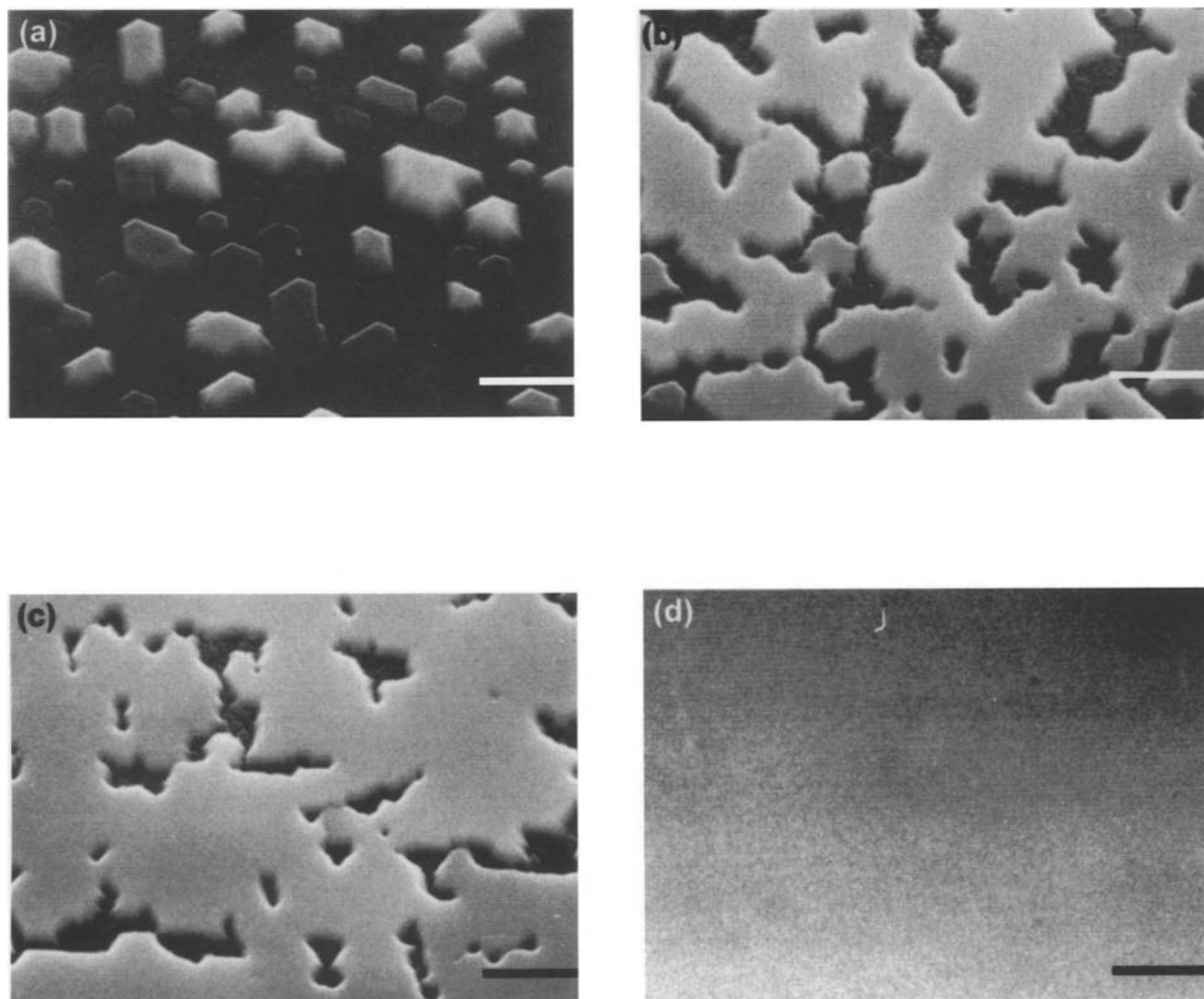
Figure 3. Changes in surface morphology during the growth of GaN (a) 5 min, (b) 10 min, (c) 20 min, (d) 60 min after the deposition of AlN buffer layer.

Bar indicates 2 μm.

214

(a) α-Al₂O₃

[0001]

GaN island

(b) GaN (5 min) C-axis

{10$\bar{1}$0}

3D-growth (0001)

(c) GaN (60 min)

C-axis

Irregular surface

(d) GaN (90 min)

C-axis

pit
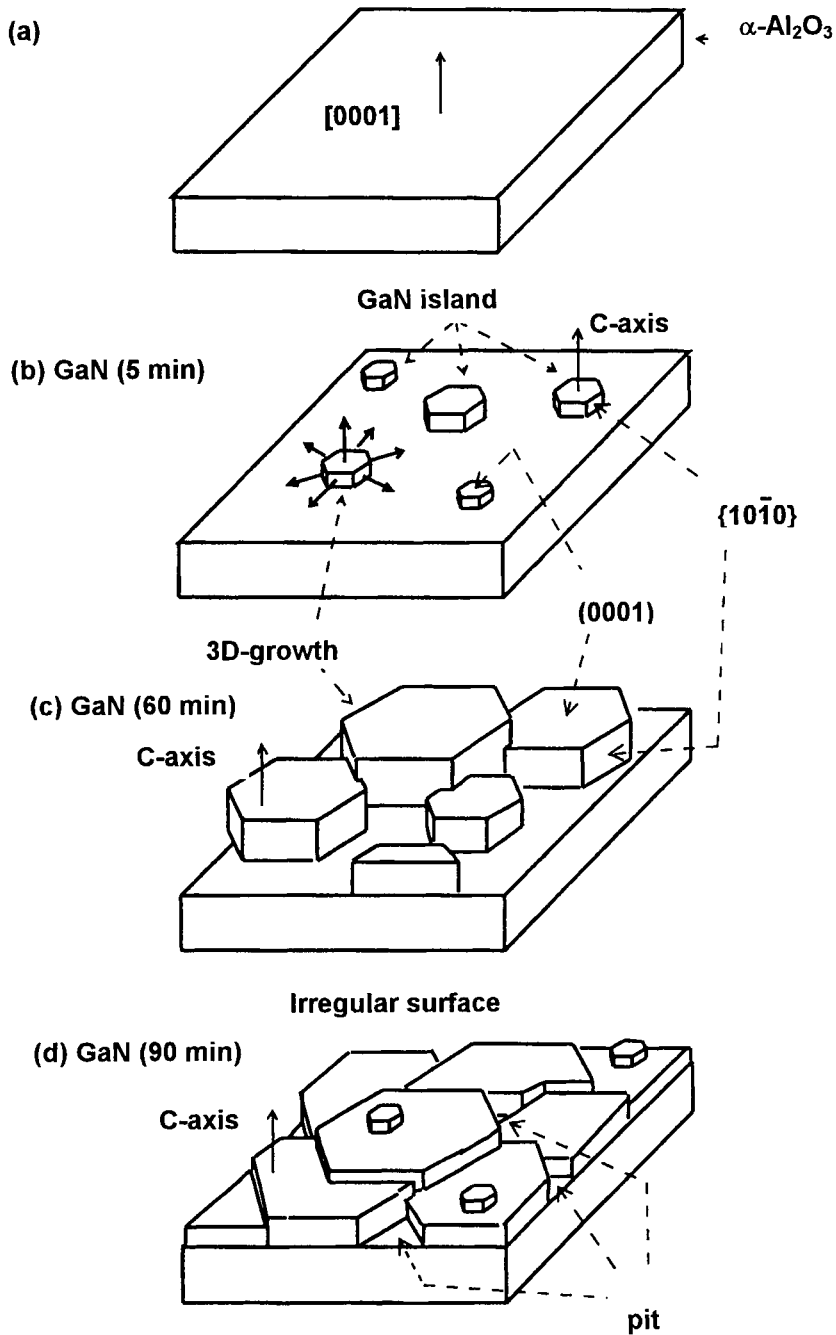
Figure 4. Model of changes in surface morphology during the growth of GaN without an AlN buffer layer.

(a) AlN buffer (5min)

AlN buffer

α-Al₂O₃

[0001]

Lateral growth

GaN island "truncated hexagonal pyramidal

AlN

mesa"

(b) AlN buffer (5min)

+ GaN (5min)

[0001]

{1̄011}

(0001)

Coalescence

(c) AlN buffer (5min)

+ GaN (10min)   [0001]

[0001]   AlN

Step flow growth

(d) AlN buffer (5min)

+ GaN (60min)
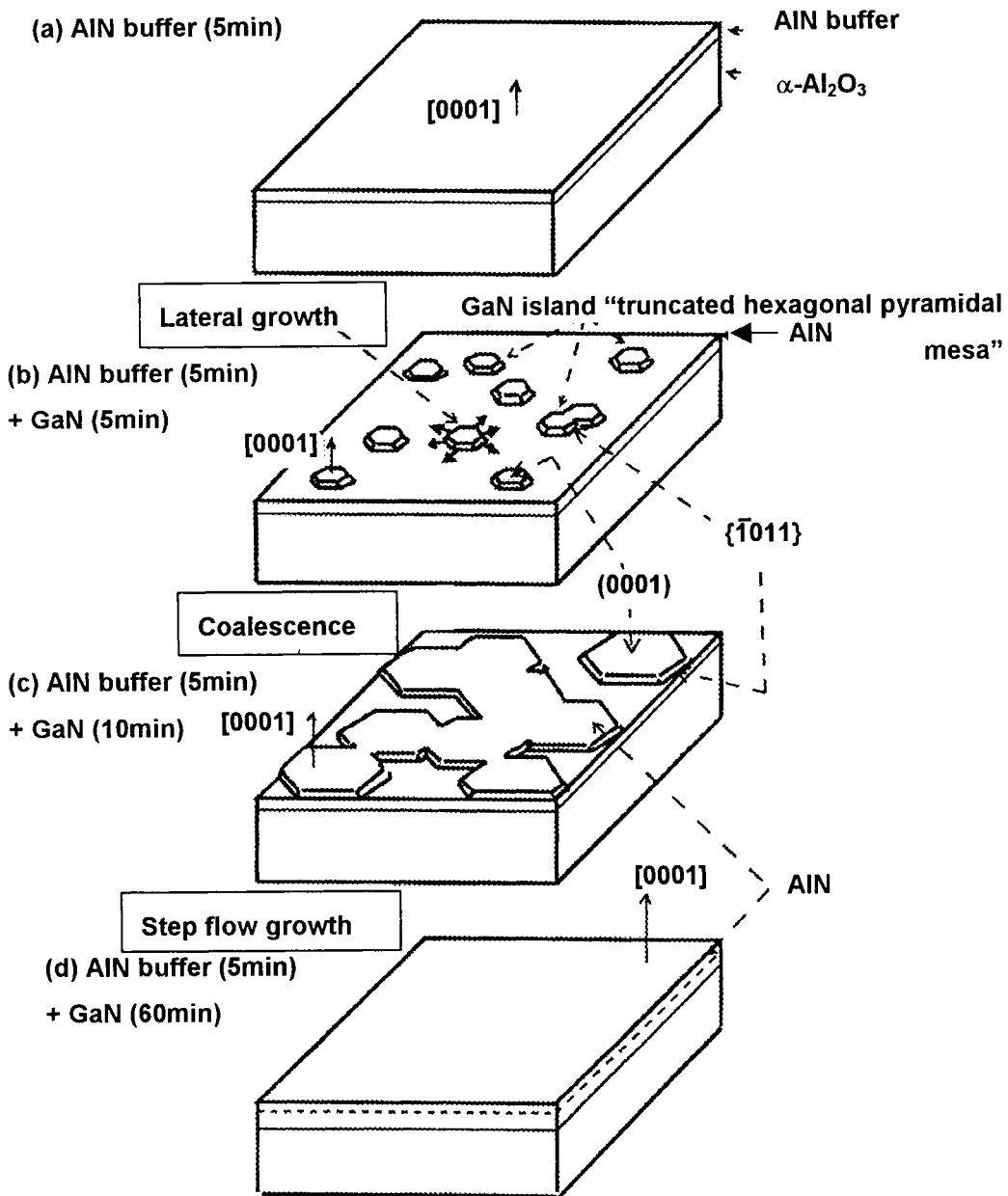
Figure 5. Model of changes in surface morphology during the growth of GaN after the deposition of an AlN buffer layer.

on (0001) sapphire without AlN.

On the basis of these results, growth models can explain the difference between the modes of growth of the two cases were proposed (see Figs. 4 and 5).

In the case of a conventional GaN film without an AlN layer (Fig. 2(b)), many hexagonal GaN columns with different sizes and heights are formed, as depicted in Fig. 4(b). They grow three-dimensionally (figs. 4(b) and (c)), resulting in a rough surface and many pits at their boundaries, as shown in Fig. 4(d).

On the other hand, in the case of a GaN film grown with an AlN layer under optimum conditions, many AlN crystallites may be formed together with an amorphous-like structure. These fine crystallites were found to have the [0001] axis parallel to that of the sapphire and may act as a nucleation center for the following growth of a GaN island, which is not a column, but a truncated pyramidal mesa as shown in Figs. 3(a) and 5(b). In further growth, lateral growth and coalescence of GaN islands are seen (Figs.3(c) and 5(c)). These may be caused by a decrease in interfacial free energy between the GaN islands and the amorphous-like AlN layer.

2.1.2. Growth Mechanism

Observation of the initial growth stage and cross-sectional TEM images of GaN films revealed the growth processes of GaN without and with an AlN buffer layer. Figure 6 shows a bright field image with a low magnification to observe a wide region of the specimen. It can be seen that the GaN layer has a three-zone texture. The zone nearest to the AlN buffer layer has fine image contrasts due to a number of faults. Consequently, this zone is called here a "*faulted zone* " (Z_F), the thickness of which is about 50 nm. Above the

faulted zone, there is another zone that has a number of trapezoid crystals. This trapezoid structure was revealed by SEM observations of the GaN islands to be a truncated hexagonal pyramidal mesa, as shown in Fig. 3. As the density of defects in this zone is much lower than in the faulted zone, it is called here a "*semi-sound zone* " (Z_{SS}), the thickness of which is about 150 nm thick. The remaining area is called a "*sound zone* " (Z_S), as it contains only a small amount of defects. In the sound zone, the defect density decreases abruptly in the case of a layer of GaN thicker than about 300 nm, and a high quality and uniform GaN is obtained.

Figure 7 shows a dark field image with slightly higher magnification to clarify the microstructure inside the AlN layer and the zones of the GaN layer. It can be seen that the AlN layer gives an image contrast of stripes perpendicular to the interface. Figure 8 shows a lattice image of AlN on α-Al$_2$O$_3$. The difference in image contrast with location is attributed to the small difference in orientations of crystallines. The moiré images are due to crystal boundaries. The image contrasts in Figs. 6 and 7 reveal that the AlN layer is composed of columnar fine crystals like frost columns, the diameters of which are in the order of 10 nm.

The initial growth models of GaN are considered on the basis of these results. Figs. 9 and 10 show schematic diagrams of the growth processes of GaN on sapphire without or with an AlN buffer. In the case of a GaN film without an AlN buffer layer, the nucleation density of GaN on the sapphire substrate is low, as shown in Figure 9 (1). Many hexagonal GaN columns with different sizes and heights are formed and they grow three-dimensionally (Figure 9 (2)), resulting in a rough surface and many pits at their boundaries as shown in Figure 9 (3). Furthermore, many crystalline defects are generated near the boundaries between

Figure 6. Bright field image with a low magnification to observe a wide region of a GaN epitaxial layer with an AlN buffer layer.
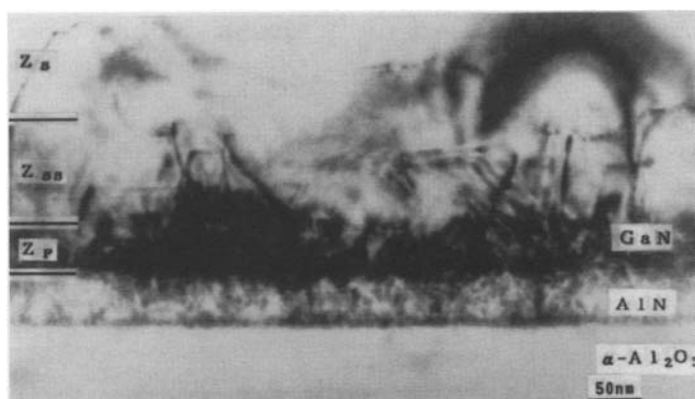

Figure 7. Dark field image with slightly higher magnification to clarify the microstructure inside the AlN layer and the zones of the GaN layer.
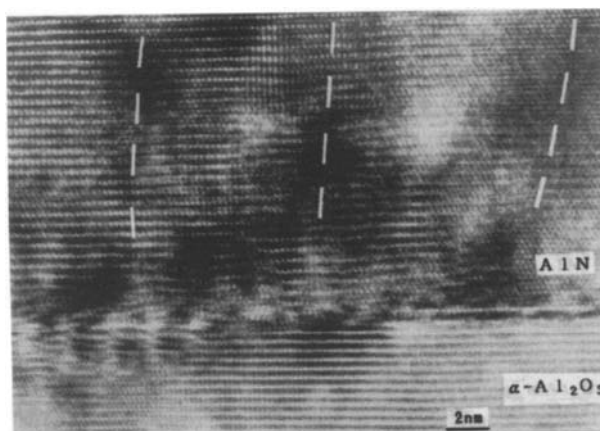

Figure 8. Lattice image of an AlN buffer layer on an α-Al$_2$O$_3$. substrate.

GaN grains due to misorientation of each island.

The growth process of GaN film with AlN buffer layer is shown in Figure 10. The AlN buffer layer has an amorphous-like structure at the deposition temperature, but when the temperature is raised to the growth temperature of GaN, AlN is crystallized by solid phase epitaxy and then it exhibits a columnar structure. Since the AlN films were single-crystal-like from electron diffraction spots, orientations of AlN columnar crystals were

(1) Nucleation

(2) Island growth

(3) Coalescence
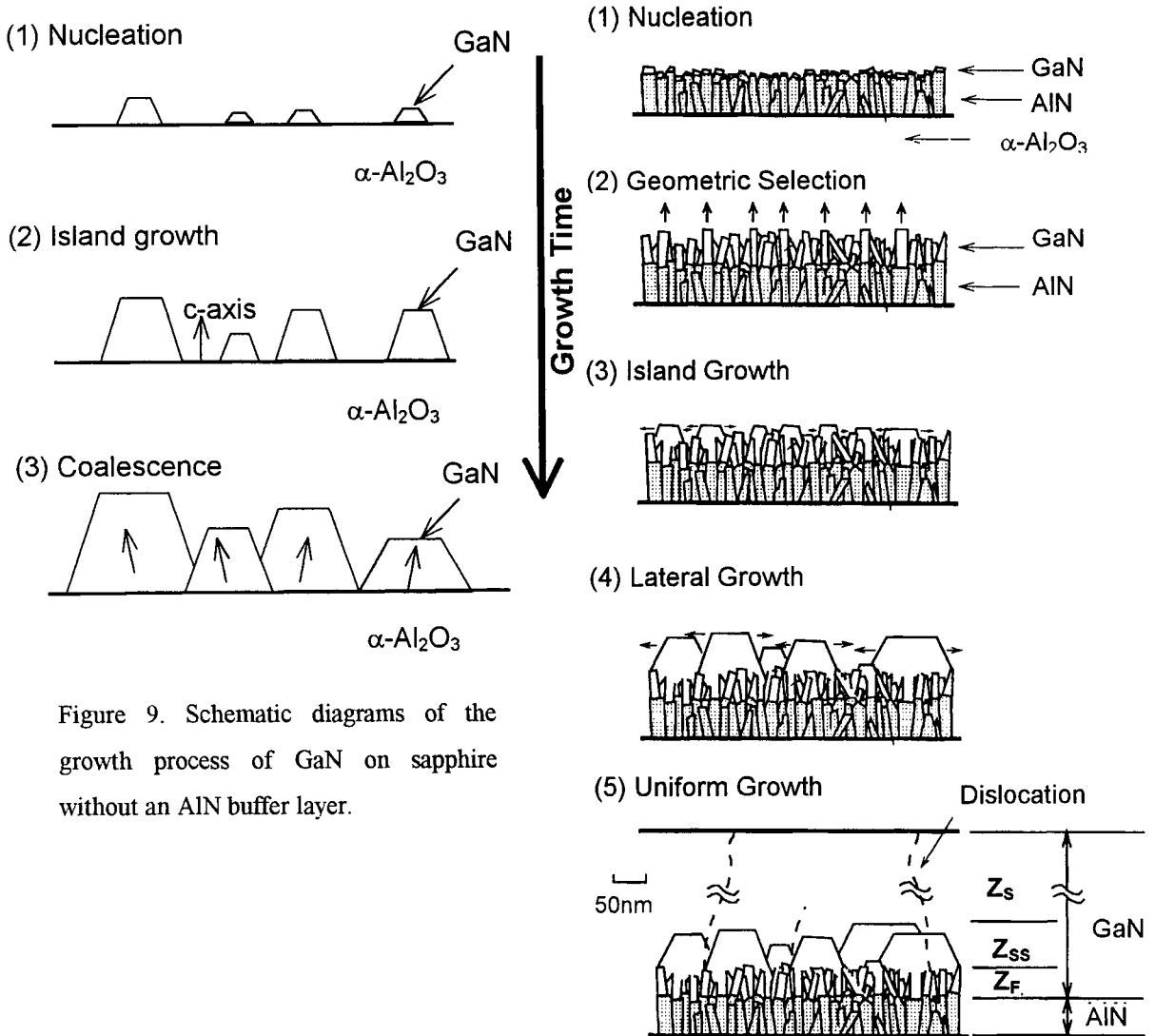
Figure 9. Schematic diagrams of the growth process of GaN on sapphire without an AlN buffer layer.

(1) Nucleation

(2) Geometric Selection

(3) Island Growth

(4) Lateral Growth

(5) Uniform Growth

Figure 10. Schematic diagrams of the growth process of GaN on sapphire with an AlN buffer layer.

found to be arranged each other.

Each GaN column is grown from GaN nuclei that have been generated on the top of each columnar fine AlN crystallite. Therefore, it is thought that high-density nucleation of GaN occurs owing to the high density of the AlN columns, as shown in Figure 10 (1), compared with the nucleation density of GaN grown directly on the sapphire substrate.

The columnar fine GaN crystals increase accordingly in size during the growth process and the crystalline quality of GaN is improved in this stage. It is thought that geometric selection of GaN fine crystals occurs, as shown in Figure 10 (2). Each first fine crystal of GaN begins to grow along the c-axis, forming a columnar structure. The columns have various random orientations and do not continue to grow uniformly. The number of columns emerging at the front gradually decreases as the front area of each column increases. Since the only columns that survive are those that grow along the fastest growth directions (i.e., those columns for which c-axis are normal to the substrate), then all columns become arranged in the direction normal to the substrate, as indicated by the arrows in Figure 10 (2).

In the next stage, trapezoid crystals are formed on the columnar crystals. As the front area of the column increases due to geometric selection, a c-face appears in the front of each column, and trapezoid islands with c-faces are formed, as shown in Figure 10 (3). These islands preferentially grow to become larger trapezoid crystals, which cover the small islands nearby. The stage in Figure 3 (a) corresponds to the generation of the trapezoid island crystals after geometric selection of columnar crystals.

Subsequently, lateral growth and coalescence of the islands occur in the stages shown in Figure 3

(b) to Figure 3(c). The pyramidal trapezoid crystals grow at a higher rate in a transverse direction, as shown in Figure 10 (4), because the growth velocity of the c-face is much slower. After the lateral growth, the islands repeat coalescence each other very smoothly.

Finally, since the crystallographic directions of all islands agree well with each other, a smooth GaN layer with only a small amount of defects can be obtained as a result of the uniform coalescence, as shown in Figures 10 (5) and 3 (d). Thus, uniform layer-by-layer growth occurs, thus creating high-quality GaN film with a low defect density and a smooth surface.

Based on the above results, the roles of an AlN buffer layer in GaN MOVPE growth are summarized as follows:

(1) High-density nucleation of GaN occurs on the AlN columnar crystals.

(2) Geometric selection occurs among the GaN fine crystals, which are able to arrange the crystallographic directions of GaN columnar crystals.

(3) Crystalline defects near the interface between the grains are greatly reduced because of lateral growth and coalescence among GaN crystals that have been arranged in crystallographic directions.

(4) The surface is covered at an early growth stage, and a smooth surface is easily obtained because of the higher lateral growth velocity of the pyramidal trapezoid islands with c-faces on the top of the columnar crystals.

Thus, the non-uniform growth of GaN on AlN mentioned above plays an important role in the realization of uniform growth and in enabling a high-quality GaN film with few defects to be obtained even on a highly mismatched (13.8%) substrate.

220

2.1.3. Advanced Technology for buffer layers

The use of in situ monitoring for optimization of the growth process has recently been studied [3, 4]. This method involves measurement of reflectivity of the epitaxial layer with crystal growth. The monitoring system is put together with crystal growth system. Figure 11 shows a schematic diagram of the reflectivity spectrum for
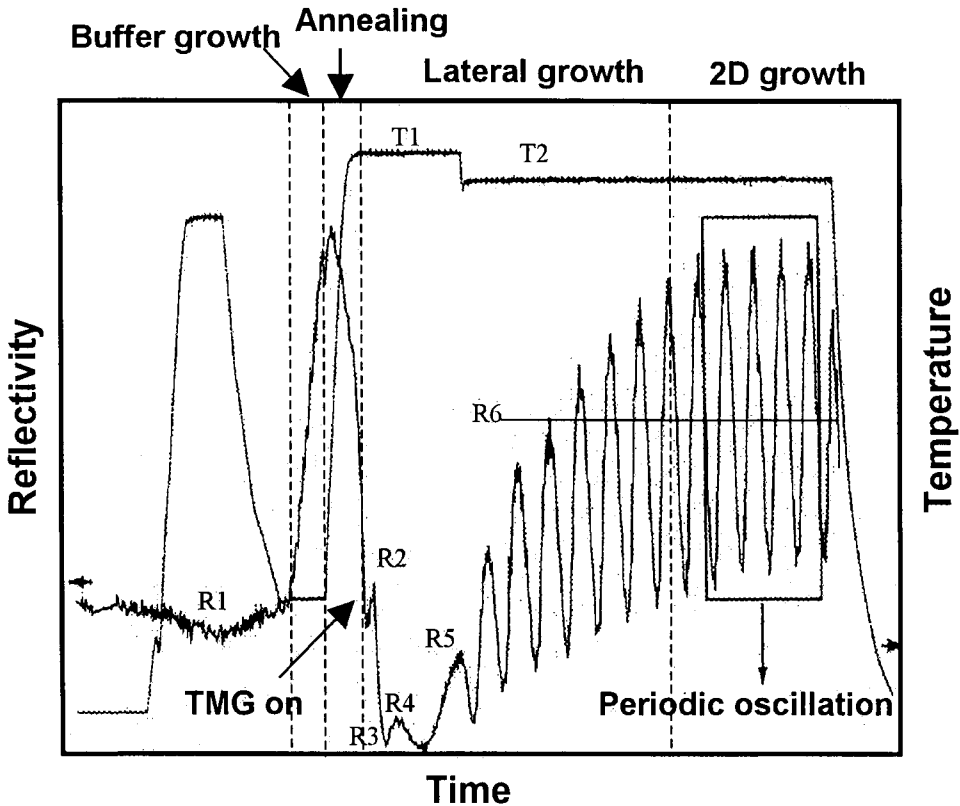

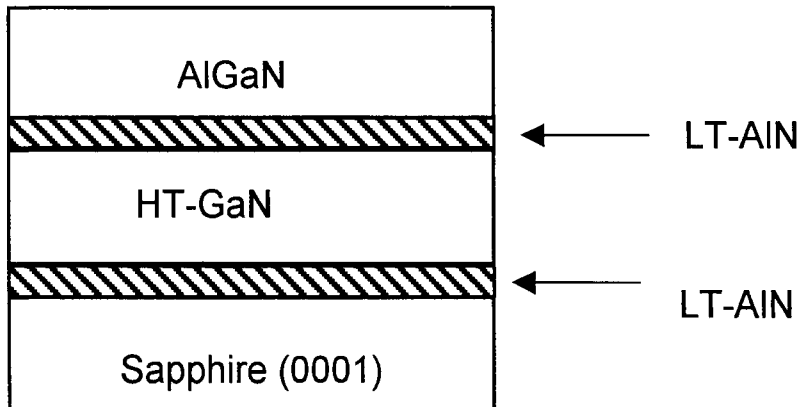Figure 11. Schematic diagram of the reflectivity spectrum for typical GaN growth.

Figure 12. Structure of samples using a second low-temperature-deposited buffer layer

typical GaN growth. During desorption under hydrogen, no significant change in reflectivity occurs; the reflectivity stays on the value of bare Al_2O_3 substrate at a constant value (R1). During buffer layer growth, a steep increase in reflectivity is observed. This increase depends on the thickness of the deposited buffer layer. During annealing, reflectivity drastically decreases and then, stabilizes at a constant value (R2). Switching on the TMG induces a further decrease in reflectivity; this "Ga nose" is typical for all measured spectra. The following shape depends on the temperature profile during annealing and determines both growth rate and the beginning of subsequent GaN growth. The Ga nose lies at R2 at T1, the following decrease to R3 reflectivity is steep, the first oscillation maximum (R4) follows soon. The loss of reflectivity can be explained by the re-crystallization of the buffer layer. After the second maximum (R5), the reactor temperature was reduced to T2, which is convenient for a smooth surface. The lateral growth of GaN is characterized by a steep ascending envelope of the reflectivity to its maximum value, corresponding to the average reflectivity of R6. The periodic oscillation means 2-dimensional growth of GaN.

The crystalline quality of a GaN epitaxial layer has been improved by introducing an AlN buffer layer; however, GaN films grown on sapphire with a low-temperature buffer layer contain large amounts of dislocations of the order of 10^8-10^{10} cm^{-2}, which might affect the performance of nitride-based devices. To solve this problem, the insertion of a second low-temperature-deposited buffer layer between high-temperature-grown GaN was proposed [5, 6]. Figure 12 shows the structure of samples in which a second low-temperature-deposited buffer layer was used. Creation of high-quality and crack-free thick AlGaN samples was possible by using this technique. This is because tensile stress in the AlGaN layer is reduced by the AlN interlayer.

2.2. 6H-SiC substrate

As mentioned above, sapphire substrates are conventionally used for nitride growth; however, the cleavage is not so strong and thermal and electrical conductivity are low. Thus it is difficult to fabricate laser diodes, FETs (Field Effect Transistors) and so on. Unlike sapphire, GaN and 6H-SiC have the similarity of the lattice structure, including closer lattice constant and thermal expansion characteristics along the basal planes. In addition, 6H-SiC has high thermal and electrical conductivity and strong cleavage. Thus, the use of a 6H-SiC substrate to grow GaN is expected to lead to improvements in structural and optelectronic properties, it is expected that 6H-SiC will become the material used for nitride laser substrates in the future. AlN has the closest structural characteristics to those of SiC within the $Al_xGa_{1-x}N$ system and has been used as a buffer layer for GaN epitaxy on SiC.

Recently, the number of reports on nitrides grown on SiC substrates has increased. Uniform growth of GaN on SiC substrates has been reported with high temperature (about 1100°C) AlN or AlGaN buffer layers [20-22]. The growth of AlGaN on SiC without a buffer layer has also reported. The $Al_xGa_{1-x}N$ alloys containing even low (x>0.05) concentrations of AlN showed smooth surfaces and high crystal quality [23].

2.3. Si substrate

Silicon is a very promising substrate material for the growth of GaN layers, allowing future integration of well-established Si electronics with

GaN-based photonic devices. To date, several groups have succeeded in the fabrication of GaN-based devices on Si substrates. However, the growth of GaN on Si layers of crystalline quality comparable to that of GaN layers grown on sapphire or SiC substrates is still a challenging issue.

It is difficult to grow GaN on a Si substrate because of the poor nucleation of GaN on Si in addition to the large lattice mismatch and thermal expansion coefficient. To overcome this difficulty, a buffer layer or an intermediate layer, such as AlN [24], SiC [25], AlAs [26] and γ-Al_2O_3 [27], grown prior to GaN growth, is usually used. High temperature AlN and AlGaN buffer have also been used in selective area growth of GaN on a Si substrate [28, 29].

2.3.1 Growth using a 3C-SiC buffer layer

3C-SiC was grown in a vertical-type, low-pressure chemical vapor deposition (CVD) reactor. The substrate was a Si (111) wafer misoriented 4 degrees off toward the <110> direction. The Si and carbon (C) source gases were dichlorosilane (SiH_2Cl_2) and isobutene (i-C_4H_{10}), respectively. Hydrogen chloride (HCl) was added to these source gases during the 3C-SiC growth in order to control the stoichiometry of the SiC layer. Typical flow rates for SiH_2Cl_2, i-C_4H_{10}, HCl and hydrogen (H_2) were 6, 10, 90 and 6000 sccm, respectively. The 3C-SiC layer was grown at 925°C. Under these conditions, a highly (111)-oriented 3C-SiC layer of about 200 nm in thickness was grown on the Si (111) substrate. Growth of GaN films on these highly (111)-oriented 3C-SiC layers, which cover the whole surface of the Si substrates, were carried out in a horizontal-type MOVPE reactor operated at atmospheric pressure.

Figure 13 shows SEM photographs of the surfaces and cross sections of GaN films grown on Si (111) covered with a 3C-SiC layer (Figure 13a) and grown directly on Si (111) (Figure 13b). As shown in Figure 4a, a GaN film with a smooth surface is obtained on Si covered with a 3C-SiC layer, indicating that layer growth of GaN can be achieved on a Si substrate using 3C-SiC as an intermediate layer. On the other hand, a GaN film grown directly on Si (111) has very poor surface morphology. An island-like structure similar to that shown in Figure 1b is observed. It is thought that GaN is grown three-dimensionally on the Si substrate without the 3C-SiC intermediate layer. The cracks that can be seen in Figure 12a were caused by the large difference between the thermal expansion coefficient of GaN and Si, as shown in Table 1.

The RHEED patterns for GaN using a 3C-SiC buffer layer were spotty, indicating that the GaN film is single crystalline. On the other hand, the RHEED patterns for GaN without a 3C-SiC intermediate layer were very weak and hallow-like. Evidence showing that the GaN film grown directly on Si is a single crystal was not obtained. These results indicate that the quality of a GaN film grown on Si is improved by using 3C-SiC as an intermediate layer.

The epitaxial relationship between a GaN film and a Si substrate using a 3C-SiC intermediate layer as revealed by X-ray Laue patterns is shown in Figure 14. It can be seen in the figure that the (0001) plane of GaN is parallel to the (111) plane of Si and that the [11-20] direction of GaN is parallel to the [110] direction of Si.

The large lattice mismatch between GaN and Si is reduced by using a 3C-SiC buffer layer and, hence, growth of GaN is changed from a three-dimensional mode to a two-dimensional mode. The 3C-SiC buffer layer was not a single crystal but highly (111)-oriented crystals. There have been no
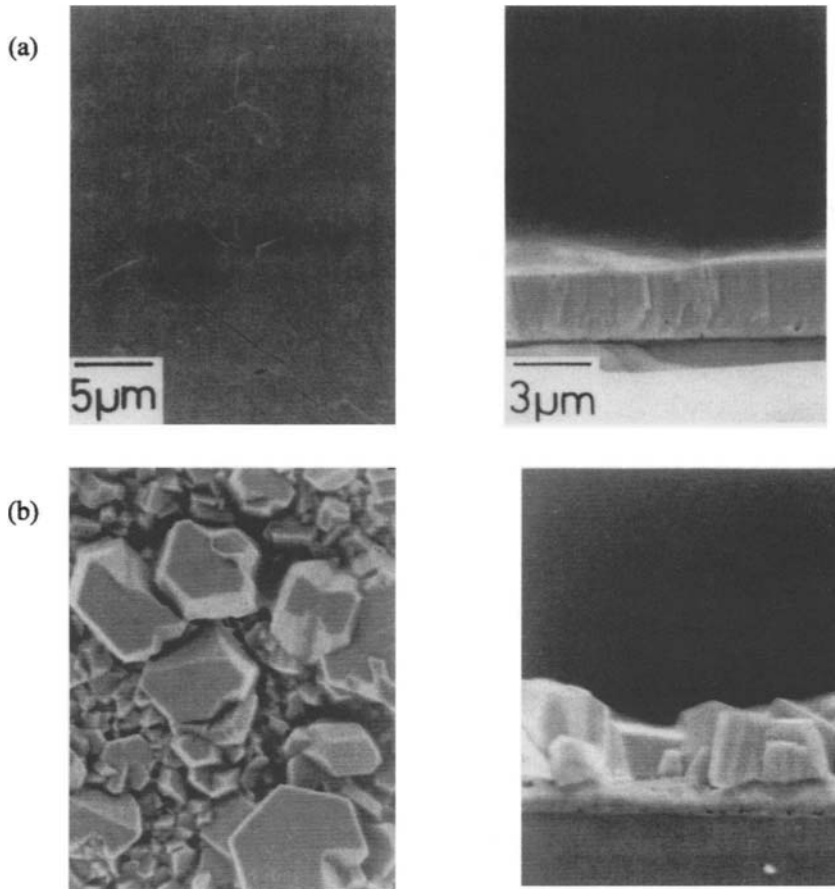
Figure 13. SEM photographs of the surface and cross sections of GaN films grown on Si (111) covered with a 3C-SiC buffer layer (a) and grown directly on Si (111) (b).
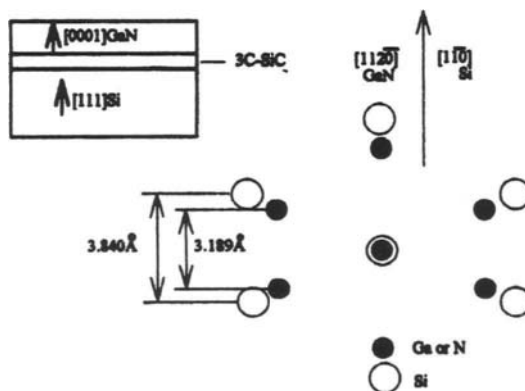

Figure 14. Schematic illustration of crystallographic orientations of a (0001) GaN layer and a (111) Si substrate using a 3C-SiC buffer layer.

experiments on GaN growth on a single crystalline 3C-SiC buffer layer. Further studies on the growth process of GaN on a SiC buffer layer and on the effects of a 3C-SiC buffer layer are needed.

2.3.2. Growth using AlN buffer layer

AlN buffer layers were deposited on Si (111) substrates at layer thickness of 100 - 200 nm. At the deposition temperature is between 800 and 1250°C, the surfaces of the AlN layers were rather smooth with good coverage on the Si surfaces. At a deposition temperature above 1100°C, the RHEED patterns had spots, indicating that the AlN layers are single crystalline. On the other hand, when the temperature was lower than 800°C, the RHEED patterns had rings, indicating that the layer is polycrystalline. At 900°C, polycrystalline AlN crystals were included in the single crystal of the AlN layer, and the AlN film became polycrystalline with increasing temperature. Formation of Si_3N_4 due to reaction between NH_3 and Si substrates might be possible; however, there is no indication from the RHEED patterns.

Figure 15 shows SEM photographs of the surfaces and cross sections of GaN films grown on Si (111) substrates without and with AlN intermediate layers deposited at 1150°C. As shown in Figure 15a and 15b, a GaN film with a smooth surface is obtained on Si (111) covered with an AlN layer, although cracks are observed on the surface. On the other hand, an island-like structure is observed on the GaN film grown directly on Si (111) (Figure 15c and 15d). The film thickness of GaN was about 2 μm. When the deposition temperature of the AlN layer was higher than 900°C, strong streaks and Kikuchi lines are clearly observed in the RHEED pattern, indicating thiat indicates that high-quality single crystalline GaN films had been

grown. On the other hand, when the temperature was 800°C, a spotty pattern was observed, indicating that the surface of GaN had become rough.

Figure 16 shows the variation in the FWHM of double-crystal X-ray rocking curves (XRCs) for GaN films grown on Si with AlN intermediate layers which predeposited at different temperatures. When the AlN deposition temperatures were higher than 1050°C, the FWHM of XRC is about 10 - 20 min, also indicating good crystallinity of the GaN films. On the other hand, the FWHM of XRC increased abruptly when the deposition temperatures were lower than 900°C. These results are consistent with those obtained RHEED analyses.

The epitaxial relationship between GaN and a Si substrate was revealed by X-ray diffraction measurements. It was found that the (0001) plane of GaN is almost parallel to the (111) plane of Si and that the [11-20] direction of GaN is parallel to the [110] direction of Si. The cracking plane of GaN is mainly the {1010} plane. This result is the same as that obtained in the case of GaN on Si (111) using a 3C-SiC buffer layer.

As mentioned above, we can obtain an extremely smooth surface of GaN by using an AlN buffer layer. This is mainly because coverage of AlN is increased more than that of GaN. It is thought that the Si surface is easily covered with AlN because chemical reaction between an Al species and a Si substrate occurs more easily than that between a Ga species and a Si substrate. Another reason is that the lattice mismatch of 16.9% between GaN and Si is greatly reduced to 2.5% in the case of GaN and AlN, resulting in realization of two-dimensional growth of GaN instead of three-dimensional growth.

The crystalline quality as well as the surface morphology of GaN films are greatly improved with AlN buffer layers deposited at a temperature over
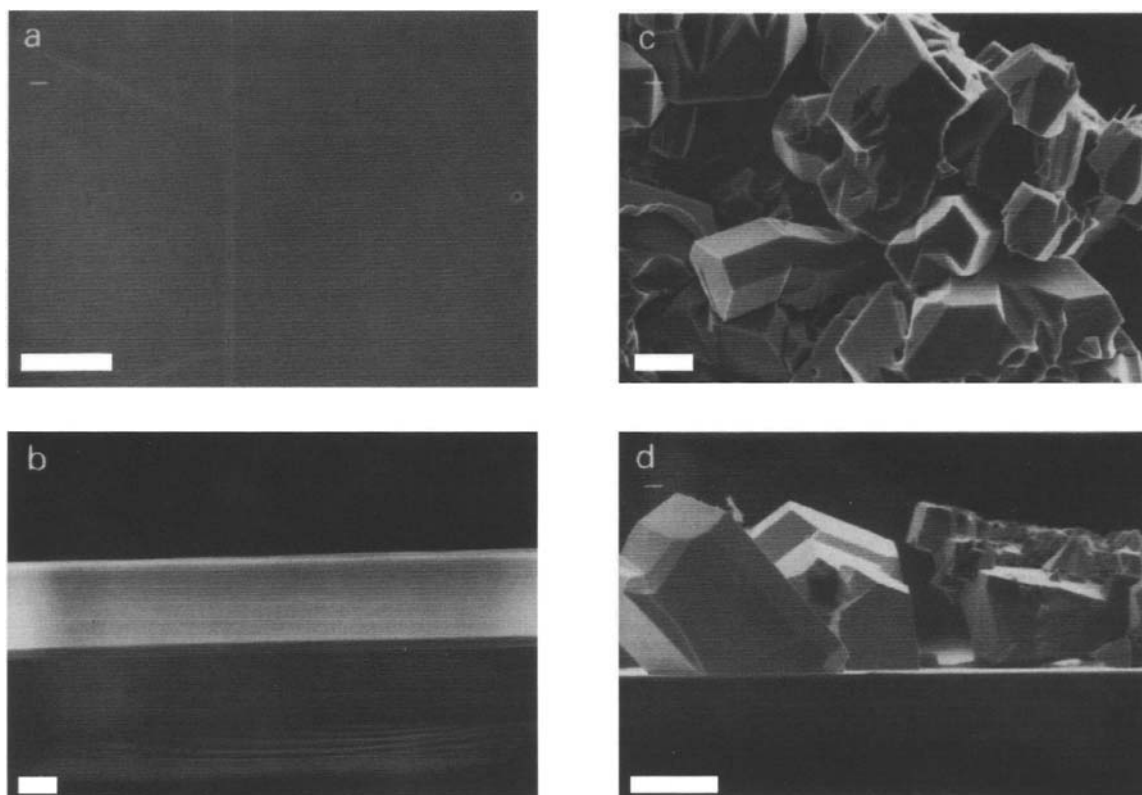
Figure 15. SEM photographs of the surfaces and cross sections of GaN films grown on Si (111) covered with an AlN buffer layer (a) (b) and grown directly on Si (111) (c) (d). Bar indicates 5 μm in (a) (b) (c) and 1 μm in (d).
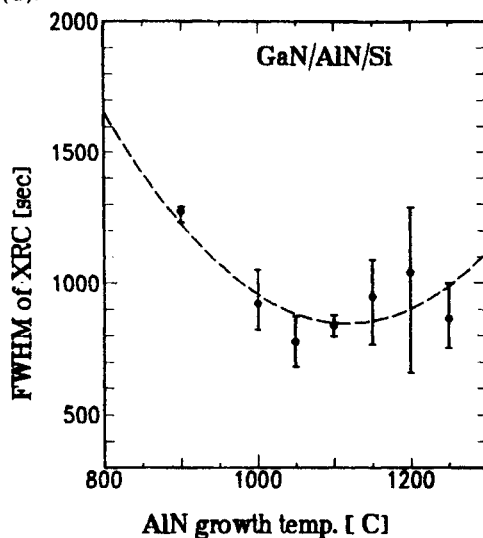

Figure 16. Dependence of the FWHM of double-crystal X-ray rocking curves of GaN layers on AlN buffer layers deposited at different temperatures.

900 °C. Since AlN buffer layers fabricated at a temperature over 900°C are single crystals, a single crystal of AlN is more important as a buffer layer on Si substrate to obtain a high-quality GaN film. Recently, high-quality GaN films have been grown by MOVPE on 6H-SiC using high temperature buffer layer of AlN. In this case, it was confirmed that the AlN buffer layer is a single crystal and coalescence among GaN islands occurs and that the surface becomes smooth in less than only one minute. Thus, GaN can be grown on a buffer layer epitaxially from the early growth stage. Therefore, this AlN buffer layer is thought to be a kind of single crystalline AlN substrate.

In contrast, an AlN buffer layer on a sapphire substrate, as mentioned in the previous section, is not single crystalline. The initial growth stage of GaN on an AlN buffer layer has several nonuniform growth stages as shown in Figure 15.

Thus, the crystalline structures of an AlN buffer layer are different in the case of a sapphire substrate and in the case of a Si substrate. Further studies on the initial growth process and observations of fine crystalline structures near the interfaces between the GaN films and the buffer layers are needed to determine the reason for the crystalline structures of the buffer layers.

3. HVPE (hydride vapor phase epitaxy)

3.1. Thermal strain in a GaN grown layer

Heteroepitaxial growth on a sapphire substrate is usually employed for the preparation of single crystalline GaN. Because of the large difference between the thermal expansion coefficients of GaN and sapphire, the sample bends if a thick film is grown and cracking often occurs near the heterointerface due to the thermal stress during cooling after the growth. Cracks occurred in a GaN (0001) / α-Al$_2$O$_3$ (0001) heterostructure fabricated by HVPE, and it was found that the cracking was caused by tensile stress applied to the sapphire substrate due to the large difference between the thermal expansion coefficients (α_{GaN} < $\alpha_{Sapphire}$). Lattice constants and strains of GaN have been measured for film thickness in the range of 0.6 to 1200 μm in order to clarify the process of strain relaxation [30, 31]. GaN (0001) films were grown on sapphire (0001) substrates by the MOVPE and HVPE methods. The thickness of the sapphire substrate was 250 μm. The detailed growth conditions of MOVPE were reported in the previous section. The growth rate was about 3 μm/h. A conventional HVPE system was also used for the growth of GaN. HCl was reacted with Ga at 850°C and then Ga + HCl gases were introduced into the growth zone with NH$_3$ and a purified N$_2$ carrier gas at 1030°C. The growth rate was high (30 – 70 μm/h) and the film thickness varied from 11 to 1200 μm.

Figure 17 shows the lattice constant as a function of the film thickness of GaN in the range of 0.6 to 1200 μm. The lattice constant for thin GaN films of a few microns in thickness is relatively large. This originates from lattice deformation due to thermal stress after cooling from the growth temperature because of the difference between the thermal expansion coefficients of GaN and sapphire. Then the value of c decreases gradually until the film thickness reaches about 100 μm. Beyond 100 μm, the value of c becomes constant. Thus the strain in the GaN film is almost completely relaxed at thickness greater than 100 μm. From this value, the intrinsic lattice constant c_0 is determined to be 5.1850Å.

The relaxation mechanism has been studied by comparison of experimental strain with calculated strain, considering the relaxation due to
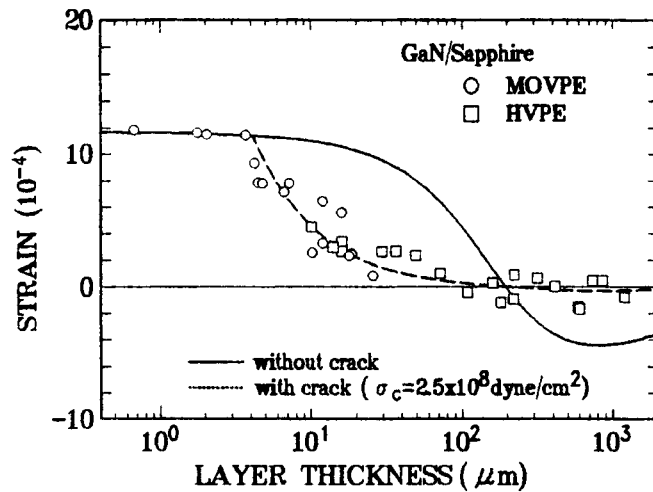
Figure 17. Comparison of the calculated and experimental strains along the c-axis. The solid and dotted lines represent the stresses calculated without and with consideration of the relaxation due to cracking, respectively. The circles and squares correspond to MOVPE-grown and HVPE-grown GaN, respectively.
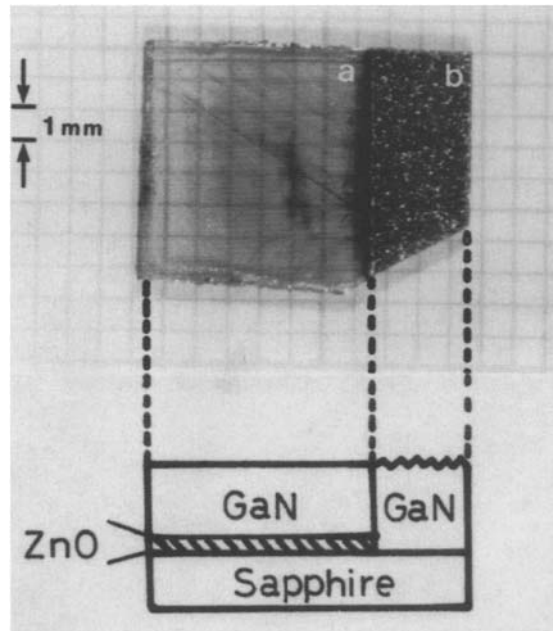

Figure 18. Photographs taken from the top of a GaN layers grown on a sapphire substrate with (a) and without (b) ZnO buffer layers, with corresponding side views of the layers illustrated below.

cracking of the sapphire substrate. A comparison of calculated and experimental strains at GaN surface along the c-axis is shown in Figure 8. The strain measured experimentally is constant until the thickness reaches 4 μm. The strain decreases abruptly and then is almost completely relaxed at a thickness greater than 100 μm. These experimental data do not agree with the theoretical curve (solid line) calculated under the condition of no cracking. However, the data are in good agreement with the curve (dotted line) calculated with consideration of the relaxation due to cracking at the critical stress of 2.5×10^8 dyne/cm^2. Thus, the sudden relaxation of the strain is attributed to the cracking, not the bending of the heterostructure.

When the film thickness of GaN is 4 to 20 μm, "macrocracks" are not observed, suggesting that interface defects such as "microcracks" and/or dislocations result in sudden relaxation of the strain. When the film thickness of GaN is greater than 20 μm, "macrocracks" occur near the interface, especially on the sapphire side, and play an important role in the relaxation of the thermal strain.

3.2. Growth using a ZnO buffer layer

Growth of thick GaN substrates is not only desirable for making high performance optelectronic devices but is also required for measuring the intrinsic properties of GaN, which are still not clear. Because of the high equilibrium pressure of nitrogen at the growth temperature, it is extremely difficult to grow a large-scale "bulk" single crystal of GaN. Preparation of thick GaN crystalline films (bulk) by the MOVPE method is very difficult owing to the low growth rate. HVPE has therefore been employed to prepare thick GaN single crystalline films. However, the reproducibility of growing GaN single crystals by the HVPE method is poor. A ZnO buffer layer sputtered on a basal plane sapphire substrate has been used for the preparation of a GaN film by the HVPE method. ZnO is thought to be an excellent buffer because (1) its physical properties, shown in Table. 1, are similar to those of GaN and (2) it can be etched by any acid, e.g., aqua regia, and it is therefore possible to separate a GaN film from the sapphire substrate by etching away the ZnO buffer layer.

Films grown with a ZnO buffer layer were
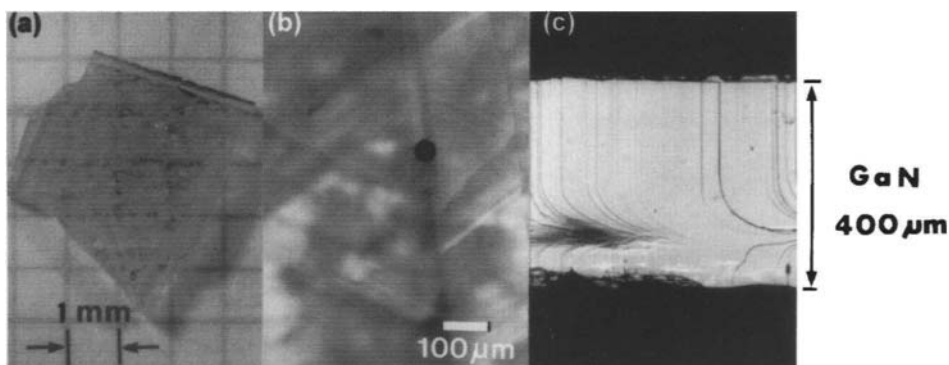
Figure 19. Photograph of a top view (a) and differential interface micrographs of the surface (b) and cross section (c) of the single crystalline GaN prepared by using a ZnO buffer layer.

229
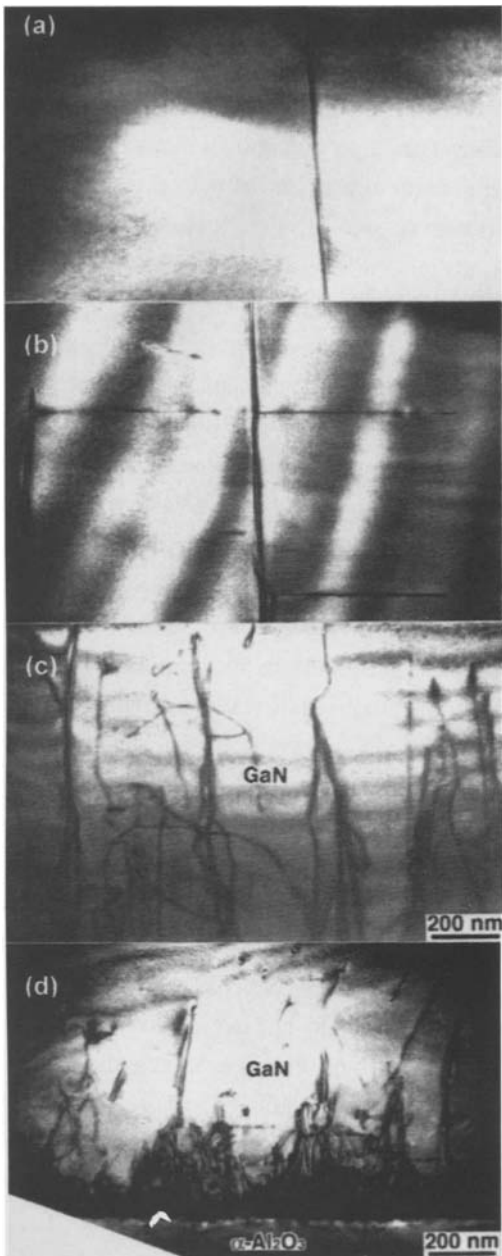
Figure 20. TEM images of HVPE-GaN on sapphire with a ZnO buffer layer taken at 85 μm from interface (a), 55 μm from interface (b), near the interface (c), and at the interface (d).
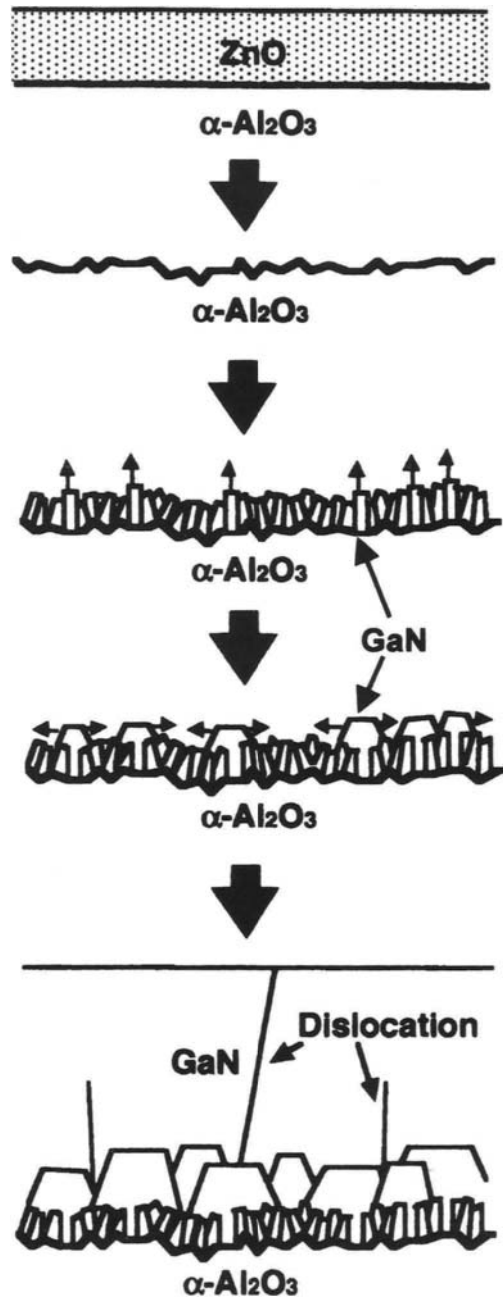

Figure 21. Growth process of GaN on sapphire using a ZnO buffer.

found to be single crystalline, while most of the films grown directly on a sapphire substrate by HVPE were polycrystalline (see Figure 18). These results imply that a sputtered ZnO layer forms an excellent buffer layer and also greatly improves the reproducibility of the growth of GaN by HVPE. In addition, it has been observed that the growth using substrates with ZnO buffer layers can be reproduced.

Single crystalline films of GaN alone were obtained by etching the sputtered ZnO layer away using aqua regia with an ultrasonic cleaner. After etching, the GaN films did not peel off easily, and the sapphire substrates had to be pulled away using forceps. Figure 19 shows an example of a thick single crystalline film of GaN after etching away the ZnO buffer layer. It confirms that the single-crystal films of GaN can be prepared by this method. Recently, using a thick GaN film as a substrate, homoepitaxial growth of GaN by MOVPE and application to a blue light emitting device have been realized [32, 33].

Next, the crystalline structure and the distribution of the threading dislocations of a GaN layer with a ZnO buffer were characterized by TEM. Figure 20 shows TEM images of HVPE-GaN on sapphire with ZnO buffer layer. There are many threading dislocations in the vicinity of the interface, as shown in Fig. 20(d), while the number of dislocations decreased with the layer thickness of GaN, as shown in Figs. 20(a), (b), (c). In Fig. 20(a), the density of threading dislocations is in the order of 10^8 cm^{-2}.

The growth process of a GaN film with an AlN buffer layer is shown in Figure 21. In Fig. 20 (a), the interface between GaN and sapphire is rough. It is thought that the sapphire substrate is attacked by a reactive compound caused by reaction between the ZnO buffer and a reactive gas (HCl, GaCl or NH$_3$). Thus, the ZnO buffer disappeared during high temperature growth of GaN, and high-density nucleation of GaN occurred due to the roughness of the surface of the sapphire substrate. Once columnar fine GaN crystals had been formed, a high quality GaN layer was grown in a similar manner to the growth of GaN with a low temperature AlN buffer layer.

4. SUMMARY

Studies on heteroepitaxial growth of GaN films by MOVPE and HVPE using buffer layers and related advanced technology were reviewed.

When a buffer layers was not used in the heteroepitaxy of GaN, the crystalline quality of the GaN film weas very poor and the surface morphology was very rough. The poor crystalline quality is attributed to the growth process, which can be summarized as (1) low-density nucleation of GaN on the substrate, (2) formation of many crystalline defects due to inhomogeneous coalescence among large GaN islands having the same crystalline directions, and (3) three-dimensional growth mode after the completion of inhomogeneous coalescence.

The buffer layer plays role in (1) high-density nucleation of GaN on the buffer layer, (2) homogeneous coalescence among the GaN islands of which crystalline direction is much arranged, (3) formation of low crystalline defects near the interface between the GaN grains due to the homogeneous coalescence, (4) the two-dimensional growth mode after the homogeneous coalescence. Thus, the change in the growth process due to the use of a buffer layer plays an important role in the growth of high-quality GaN epitaxial films with a low density of crystalline defects.

Advanced technology related with buffer

layers, such as in situ monitoring and a low temperature interlayer technique, has also been described. These techniques are very effective for improving crystal quality of GaN or AlGaN with buffer layers

ACKNOWLEDGEMENTS

REFERENCES

1. H. Amano, N. Sawaki, I. Akasaki and Y. Toyoda: Appl. Phys. Lett. 48 (1986) 35.
2. S. Nakamura: Jpn. J. Appl. Phys., 30 (1991) L1705.
3. M. Lünenbürger, H. Protzmann, M. Heuken and H. Jürgensen.: Phys. Stat. Sol. (a), 176 (1999) 727.
4. S. Figge, T. Bottcher, S. Einfeldt, and D. Hommel; J. Cryst. Growth, 221 (2000) 262.
5. M. Iwaya, T. Takeuchi, S. Yamaguchi, C. Wetzel, H. Amano and I. Akasaki: Jpn. J. Appl. Phys., 37 (1998) L316.
6. T. Kashima, R. Nakamura, M. Iwaya, H. Katoh, S. Yamaguchi, H. Amano and I. Akasaki: Jpn. J. Appl. Phys., 38 (1999) L1515.
7. T. Detchprohm, K. Hiramatsu, H. Amano and I. Akasaki: Appl. Phys. Lett., 61 (1992) 2688.
8. P. Ruterna, B. Beaumont, P. Gibart and Yu. Melnik: Phys. Stat. Sol. (b) 216 (1999) 697.
9. F. Hasegawa, M. Minami, K. Sunaba, and T. Suemasu: Jpn. J. Appl. Phys., 38 (1999) L700.
10. K. Motoki, T. Okahisa, N. Matsumoto, M. Matsushima, H. Kimura, H. Kasai, K. Takemoto, K. Uematsu, T. Hirano, M. Nakayama, A. Koukitsu and H. Seki: Jpn. J. Appl. Phys., 40 (2001) L140.
11. A. Wakahara, T. Yamamoto, K. Kawano, A. Yoshida, Y. Seki, and O. Oda: Proc. of 2nd Int. Symp. On Blue Laser and Light Emitting Diodes, (1998) 190.
12. C. Heinlein, J. Grepstad, T. Berge and H. Riechert: Appl. Phys. Lett., 71 (1997) 341.
13. K. Balakrishnan, H. Okumura and S. Yoshida: J. Cryst. Growth, 189/190 (1997) 244.
14. S. Yoshida, S. Misawa and S. Gonda: Appl. Phys. Lett., 42 (1983) 427.
15. T. D. Moustakas, T. Lei and R. J. Molnar: Physicaca B, 185 (1993) 36.
16. J. S. Hwang, S. Tanaka, S. Iwai, Y. Aoyagi and S. Seong: J. Cryst. Growth, 200 (2000) 63.
17. N. P. Kobayashi, J. T. Kobayashi, D. P. Dapkus, W. J. Choi, A. E. Bond, X. Zhang and D. H. Rich: Appl. Phys. Lett., 71 (2000) 3569.
18. Y. Hiroyama and M. Tamura: Jpn. J. Appl. Phys., 37 (1998) L630.
19. S. A. Nikishin, N. N. Faleev, V. G. Antipov, S. Francoeur, L. G. Peralta, G. A. Seryogin, H. Temkin, T. I. Prokofyeva, M. Holtz and S. N. G. Chu: Appl. Phys. Lett., 75 (1999) 2073.
20. F. A. Ponce, B. S. Krusor, J. S. Major, Jr., W. E. Plano and D. F. Welch: Appl. Phys. Lett., 67 (1995) 410.
21. B. Moran, M. Hansen, M. D. Craven, J. S. Speck and S. P. DenBaars: J. Cryst. Growth, 221 (2000) 301.
22. T. W. Weeks, M. D. Bremser, K. S. Ailey, E. Carlson, W. G. Perry and R. F. davis: Appl. Phys. Lett., 67 (2000) 401.
23. K. Horino, A. Kuramata and T. Tanahashi: Mat. Res. Soc. Symp. Proc., 449 (1997) 73.
24. A. Watanabe, T. Takeuchi, K. Hirosawa, H.

Amano, K. Hiramatsu and I. Akasaki: J.Crystal Growth, 128 (1993) 391.

25. T. Takeuchi, H. Amano, K. Hiramatsu, N. Sawaki and I. Akasaki: J. Cryst. Growth , 115 (1991) 634.

26. A. Strittmatter, A. Krost, J. Bläsing, and D. Bimberg: Phys. Stat. Sol., 176 (1999) 611.

27. L. Wang, X. Liu, Y. Zan, J. Wang, D. Wang, D. C. Lu and Z. Wang: Appl. Phys. Lett., 72 (1998) 109.

28. Y. Kawaguchi, Y. Honda, H. Matsushima, M. Yamaguchi, K. Hiramatsu and N. Sawaki: Jpn. J. Appl. Phys., 37 (1998) L966.

29. H. Marchand, N. Zhang, L. Zhao, Y. Golan, S. J. Rosner, G. Girolami, P. T. Fini, J. P. Ibbetson, S. Keller, S. DenBaars, J. S. Speck and U. K. Mishra: MRS Int. J. Nitride Semicond., Res.4, 2 (1998) .

30. T. Detchprohm, K. Hiramatsu, K. Itoh and I. Akasaki: Jpn. J. Appl. Phys., 31 (1992) L1454.

31. K. Hiramatsu, T. Detchprohm and I. Akasaki: Jpn. J. Appl. Phys., 32 (1993) 1528.

32. T. Detchprohm, K. Hiramatsu N. Sawaki and I. Akasaki: J. Crystal Growth, 137 (1994) 170.

33. T. Detchprohm, K. Hiramatsu, N. Sawaki and I. Akasaki: J. Crystal Growth, 145 (1994) 192.

Self-assembled quantum dots systems : the case of GaN

B. Daudin

CEA/Grenoble, Département de Recherche Fondamentale sur la Matière Condensée, SP2M/Physique des Semi-Conducteurs, 17 rue des martyrs, 38041 Grenoble cedex, France

The role of the various parameters governing the controlled growth of self-assembled quantum dots systems is illustrated through the example of GaN. Both polytypes of GaN, i.e. hexagonal and cubic phase, deposited under specific growth conditions on AlN obeys the Stranski-Krastanov mechanism. The interplay between thermodynamical and kinetical parameters is investigated. In particular, it is shown that kinetical effects may be dominant and completely inhibit the 2D/3D transition. Under conditions allowing one for its observation, it is shown that the growth of self-organized three-dimensional islands can be achieved. These islands behave as quantum dots, exhibiting optical properties dominated by localization effects.

1. INTRODUCTION

Self-assembled quantum dots are commonly observed in a wide variety of semiconductor systems ranging from IV-IV (Ge/Si) to II-VI (CdSe/ZnSe) and III-V (InAs/GaAs, InAs/InP, GaN/AlN, GaInN/GaN) materials. For all these various combinations, the deposition of material A (possibly a ternary alloy) in compressive strain on material B leads to the formation of nanometric islands, provided that the elastic mismatch between A and B is larger than a limit value, depending on the material class, which exceeds several percent. If the dot-forming material exhibits a bandgap energy value lower than the bandgap energy value of the substrate / capping material, then carrier confinement is expected. Depending on the semiconductor family, as shown in figure 1, the resulting self-organized nanostructures are then expected to exhibit peculiar optical properties for wavelenghts ranging from infrared to ultraviolet, making them objects of prominent interest in view of applications.

The driving force responsible for 3D islands formation when depositing A over B is still a subject of controversy and has given rise to innumerable studies following the historical work of Stransky and Krastanov [1]. Although this work refers to the case of two materials with the same lattice parameter, the Stransky-Krastanov (SK) growth mode, as usually presented, consists of a two stage-mechanism, i.e. deposition of a strained bidimensional layer matched to the substrate (the wetting layer), followed by 3D islanding and elastic strain relaxation through free surface formation. In a very crude approximation, the 2D/3D transition results from a competition between elastic energy, interface energy and surface energy leading to the formation of a 3D island when the elastic strain accumulated during coherent deposition of A over B exceeds a critical value. However, this view is very approximate. First, it neglects the kinetical aspects which may be determinant in the strain relaxation process. Second it does not take into account the detailed mechanisms of 3D island nucleation which are often

parasitized by interdiffusion or by the presence of structural defects such as (but not limited to) steps or dislocations.

The aim of this article is not to give an extensive overview on 2D/3D transition. It will be rather focused on the self-assembling mechanisms of GaN quantum dots grown by plasma-assisted MBE. Although being a relatively newcomer in the family of dot-forming materials [2-5], the GaN/AlN system is an attractive one for illustrating the SK growth mode experienced by numerous combinations of semiconductor materials. Then, we believe that analyzing the various aspects of 2D/3D transition in nitrides allows one, more generally, for identifying the relevant parameters governing this important growth mechanism, whatever the class of materials under consideration. For instance, among the advantages of GaN/AlN one should mention 1) the very weak interdiffusion between GaN and AlN which greatly simplifies the interpretation of data and 2) the experimental possibility to emphasize kinetics or thermodynamics aspects by changing the growth conditions, as it will be dicussed to some extent in section 3. In this sense, nitrides can almost be considered as a model system for describing the self-formation of semiconductor quantum dots.
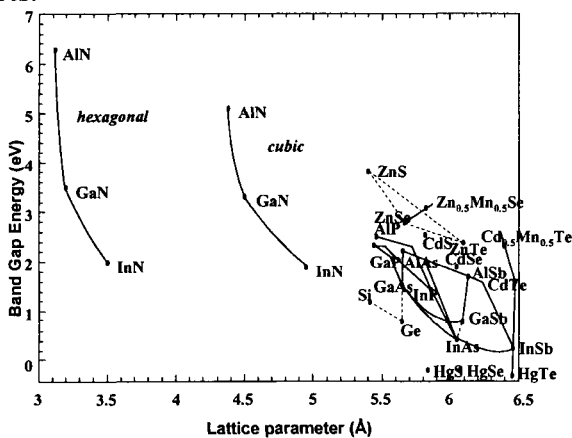
Figure 1 : bandgap energy versus lattice parameter for the most important semiconductors

2. GROWTH MODE OF GaN

The three growth modes most usually found when depositing a material A over a material B are illustrated in figure 2. Note first that the Volmer-Weber mechanism is rather observed for metals or for systems with very high mismatch whereas the Frank-van der Merwe and the Stranski-Krastanow modes are commonly observed for semiconductor combinations.

In both the Frank-van der Merwe and the Stranski-Krastanow mode, the growth is layer-by-layer up to a critical thickness. Above the critical thickness, two different elastic strain relaxation mechanisms are possible : either plastic relaxation through the formation of misfit dislocations or elastic relaxation through the formation of 3D islands. Actually, the ability to form islands or not depends on the balance between interfacial energy, surface energy and elastic energy. In the particular case of nitride heteroepitaxy, it has been found that the deposition of GaN on AlN obeys the SK growth mode whereas the deposition of AlN on GaN occurs according to the Frank-van der Merwe mode [6].
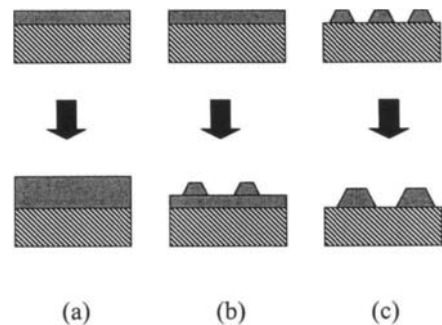
Figure 2 : Schematics of the 3 possible growth modes of A over B : a) Frank-Van der Merwe, b) Stranski- Krastanow, c) Volmer-Weber.

The change in surface morphology and strain state experienced by GaN deposited on AlN have been in-situ analyzed using reflection high energy electron diffraction (RHEED). The result is shown in figure 3. In

this experiment, both in-plane lattice parameter (indicative of the strain state) and Bragg spot intensity (indicative of the surface morphology) have been measured during the deposition of the first layers of GaN. The intensity of the RHEED specular spot was also measured, allowing for a determination of the growth rate during deposition of the wetting layer which was found to be about 2 monolayers (MLs) thick.
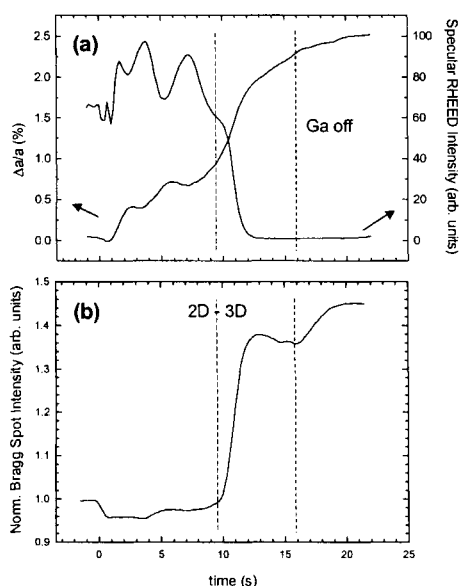

Figure 4 : RHEED diagram of GaN grown on AlN. a) during deposition of the wetting layer ; b) after 3D island formation. Note the presence of facets.

Figure 3 : RHEED analysis of the growth of GaN on AlN : a) in-plane lattice parameter and specular spot intensity variation as a function of deposition time, b) Bragg spot intensity variation as a function of deposition time.

After completion of the wetting layer, a sharp increase in Bragg spot intensity is associated with 3D islanding (see figure 4). Simultaneously, an increase of in-plane lattice parameter up to about 2.5 percent was observed, indicating an almost complete elastic relaxation of the GaN island with respect to AlN substrate.

Interestingly, figure 3 also reveals an oscillatory behaviour of the in-plane lattice parameter during the deposition of the wetting layer, dephased of 180° with respect to RHEED specular spot intensity variation. Such a behaviour in the sub-critical thickness regime has been previously observed for InAs on GaAs [7] and for CdTe on ZnTe [8]. It is assigned to the elastic strain relaxation associated with the formation of quasi-bidimensional platelets. Such an interpretation also holds for GaN on AlN. The presence of these platelets is the signature of kinetical effects related to adsorbate nucleation mechanisms [9]. Their role in the relaxation of the built-in elastic energy will be discussed in more details in section 3.1.

Despite the fact that elastic energy does not depend on the sign of the deformation, i.e. does not change for compressive or tensile strength, the deposition of B over A does not obey in the general case the same mechanism as for the deposition of A over B. This furthermore emphasizes the role of surface and interface energy in governing the elastic strain energy relaxation. As an extreme illustration of these considerations, it has to be recalled that InAs grown on (001) GaAs obeys the SK mode but obeys the Frank-Van

der Merwe mode when deposited on A (111) GaAs [10].

In the specific case of nitrides, as previously mentioned, the growth of AlN on GaN obeys the Frank-Van der Merwe mode. This is illustrated in figure 5 which shows both the in-plane lattice parameter and the RHEED specular spot intensity variations. In this case, the persistent RHEED intensity oscillations are a signature of a layer-by-layer growth mode. Consistent with this view, the oscillations of in-plane lattice parameter reveal that elastic strain relaxation first occurs through the formation of flat platelets, likely followed by the formation of misfit dislocations.
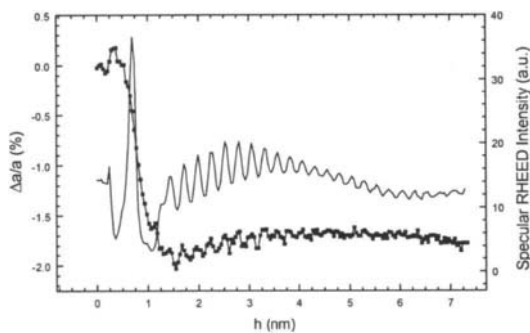
Figure 5: Relative variation of the in-plane lattice parameter $\Delta a/a$ and specular RHEED intensity as a function of layer thickness h during the deposition of a single AlN layer on a GaN substrate at T_S = 640 °C. Oscillations, characteristic of a layer-by-layer growth, are visible on both signals.

At this stage, it is worth recalling that nitrides can be grown either in the wurtzite or in the zinc-blende phase, depending on the substrate. As a matter of fact, the use of sapphire and of (0001) SiC allow for the growth of material in the wurtzite phase whereas the use of (001) GaAs or (001) SiC results in the growth of cubic material. However, the above considerations on the SK growth mode of GaN on AlN and on the Frank Van der Merwe growth mode of AlN on

GaN hold for both cristallographic phases. This is a rather unique experimental situation as semiconductors listed in figure 1 experiencing a SK growth mode are generally grown in the cubic phase.
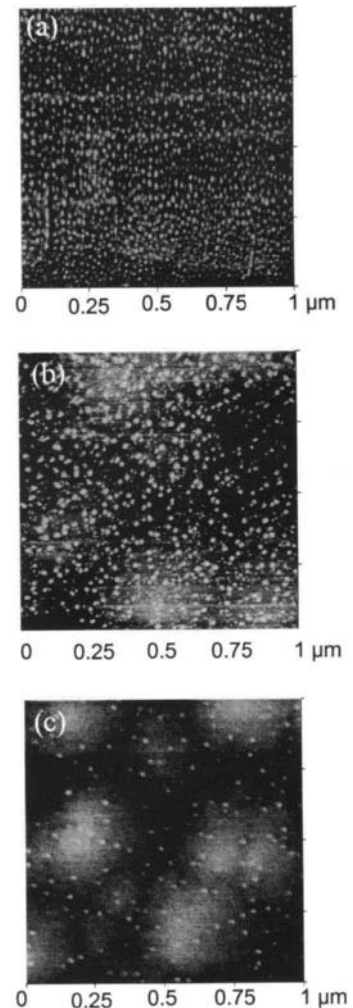
Figure 6 : AFM images of GaN QDs grown on AlN buffer layer for cubic (a) and hexagonal (b and c) crystallographic phases. (a) cubic case, QD density of 3×10^{11} cm^{-2} (b) hexagonal case, QD density of 2×10^{11} cm^{-2}, GaN coverage of 4.8 monolayers (c) hexagonal case, QD density of 2×10^{10} cm^{-2}, GaN coverage of 3.3 monolayers.

In order to illustrate this statement, figure 6 shows, for both zinc-blende (a) and wurtzite (b,c) phases, atomic force microscopy (AFM) images of a single uncovered GaN island layer grown on a 10 nm thick AlN buffer layer. The extracted average island height above the wetting layer for the cubic GaN (Fig.6a) is 1.6 ± 0.5 nm, the average dot diameter is 15 nm. From these dimensions, it is calculated that the aspect ratio (height/diameter) of cubic GaN dots is about 1/10. By comparison, for hexagonal GaN QDs, an aspect ratio of about 1/5 is found, as an evidence that the detailed shape of dots depends on the crystallographic phase.
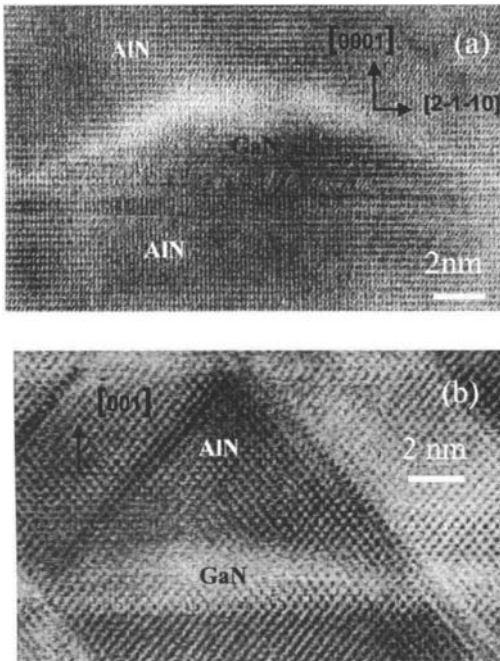
Figure 7: High-resolution TEM images of GaN islands. (a) hexagonal phase (b) cubic phase.

The difference in aspect ratio between cubic and hexagonal GaN quantum dots is further illustrated in figure 7, which shows high-resolution transmission electron microscopy (TEM) images of GaN dots in both

crystallographic structures. Consistent with AFM experiments, it is confirmed that the aspect ratio (height/diameter) is about 1/5 in the case of hexagonal dots and about 1/10 in the case of cubic dots [11].

In view of applications, the control of the dot density is almost as important as the control of their size. The relevant parameters to achieve this purpose are GaN coverage value, growth temperature and growth rate. However, these three parameters are not independent, as their role is closely related to nucleation process and adatom kinetics which eventually govern the amount of material present in 3D islands under specific growth conditions.

In the case of hexagonal dots, the density can be changed by a factor of ten between 2×10^{11} cm^{-2} and 2×10^{10} cm^{-2} (Fig. 6b and 6c), by varying the GaN coverage at a fixed growth temperature. In contrast, when changing the GaN coverage in the case of cubic GaN, the QD density, typically 3×10^{11} cm^{-2}, does not change significantly, and only the QD size varies, as revealed by the energy shift of the wavelength emission. This different behavior is possibly related to the high density of nucleation sites, correlated to the high density of stacking faults always present within cubic materials.

As far as the growth temperature is concerned, figure 8 shows the variation in wurtzite GaN dot density, from $5 \ 10^{11}$ cm^{-2} to $2 \ 10^{11}$ cm^{-2}, when increasing growth temperature from 660 to 760 °C, the decrease in dot density being associated with an increase in their size.

It has been checked, in the case of hexagonal phase, that the aspect ratio does not depend on dot size. This is shown in figure 9 which summarizes AFM results obtained for dots of various size. In addition, results in figure 9 reveal that the dot aspect ratio is the same when measured either by AFM or TEM. This is an indication that capping dots with AlN does not affect their shape, as a clue that no interdiffusion occurs during capping. However, it should be mentionned that this situation, specific to

nitrides, does not correspond to the general case. In particular, in the case of arsenides, it has been found that capping InAs with GaAs [12, 13] or GaInAs with GaAs [14] results in a change in dot shape, which is partly assigned to In segregation phenomena and leads to drastic modification of optical properties.
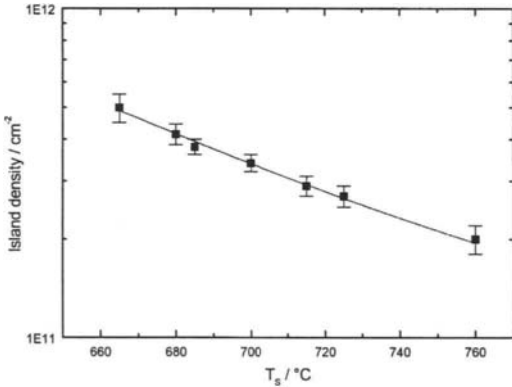

Figure 8 : GaN dot density in the wurtzite phase as a function of the substrate temperature

As a partial conclusion of this section, it appears that GaN dot density can be hardly changed in the case of cubic dots and only by about one order of magnitude in the case of hexagonal phase. This result suggests, at least for nitrides, that the dot nucleation process, far from being intrinsic, is dependent on the presence of extrinsic nucleation centers. This is actually the case for GaN dots and will be discussed in next section. More generally, the lateral positioning of island and the control of their size distribution is an issue of current interest. Among the various attempts to achieve this goal, it is worth mentioning, for the case of InAs dots, the use of vicinal surfaces such as B (311) GaAs [15]. Besides this use of natural steps for dot nucleation, one should also mention the use of patterned substrates, the ultimate goal being to control the surface strain field responsible for dot nucleation and positionning.
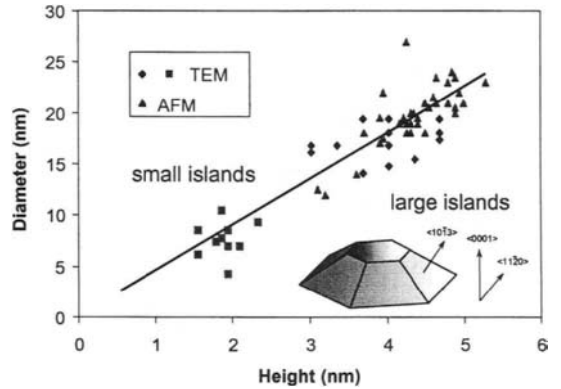
Figure 9 : GaN dot diameter as a function of height as measured by AFM and TEM. The aspect ratio is $1/(5.2 \pm 0.2)$. The inset shows the schematic shape of the GaN dots.

2.1 Nucleation of GaN islands

As mentioned in the previous section, one practical challenge faced when considering QDs in view of applications is the precise control of their size distribution which drastically depends on the nucleation mechanism. As a matter of fact, any structural defects corresponding to a local minimum in the surface elastic energy will likely behave as a preferential nucleation center. Then, the lateral distribution of dots is expected to strongly depend on the presence of structural defects. This is particularly true in the case of nitrides as the lack of homo-substrates results in the presence of high densities of structural defects. In the hexagonal phase, these defects mostly consist of screw and threading edge dislocations. The presence of screw dislocations results in the formation of spiral-like features associated with the presence of steps (see figure 11). In the cubic phase, the dominant structural defects are stacking faults in the (111) planes [16].

In the case of hexagonal material, plane-view TEM experiments have revealed that GaN dots nucleate in the close vicinity of

threading edge dislocations (see figure 10a). As schematized in figure 10b, this is consistent with the view that the distorted AlN lattice in the vicinity of the dislocation core corresponds to a local minimum of elastic potential for nucleating GaN island. Actually, the nucleating GaN adsorbate takes advantage of the local expansion of the AlN lattice on one side of the dislocation which is associated with the presence of one extra-plane [17].

In the case of cubic GaN, AFM experiments have revealed that dots tend to align themselves along (110) directions (see figure 10c). These directions correspond to emerging stacking faults in the (111) planes which likely act as local minima of elastic potential for the nucleation of GaN dots [18].

Besides threading edge dislocations (hexagonal phase) or stacking faults (cubic case), it has also been found that steps might act as efficient nucleation centers. This statement is illustrated in AFM image displayed in figure 11a which shows a spiral-like arrangment of steps which develop themselves around screw dislocations, a characteristic feature of (0001) GaN or AlN grown by MBE. Note in figure 11a the additional presence of threading dislocations acting as pining centers for steps. Consistent with the role of these threading dislocations as nucleation centers for dots, figure 11b shows a spiral-like arrangement of dots.

Then, at this stage, it appears that the high density of defects present in both hexagonal and cubic nitrides strongly affects the nucleation mechanism of dots. Of course, such nucleation centers are randomly distributed on the surface. As a consequence, the nucleation of GaN dots themselves is mostly random, which accounts to a large extent for the wide size distribution extracted from AFM observations, emphasizing the need for high quality substrates which are the key factor for controlling dot distribution, whatever the material under consideration.

However, despite of the presence of structural defects in their close vicinity, GaN dots themselves are virtually free of defects [19]: this is illustrated in figure 12 which shows high resolution TEM images of a GaN dot. From Fourier filtering achieved by selecting (2-1-10) fringes, it has been concluded that, in the general case, dots are perfectly lattice-matched to the surrounding AlN matrix which is of importance for optical properties, as it will be discussed in section 4.
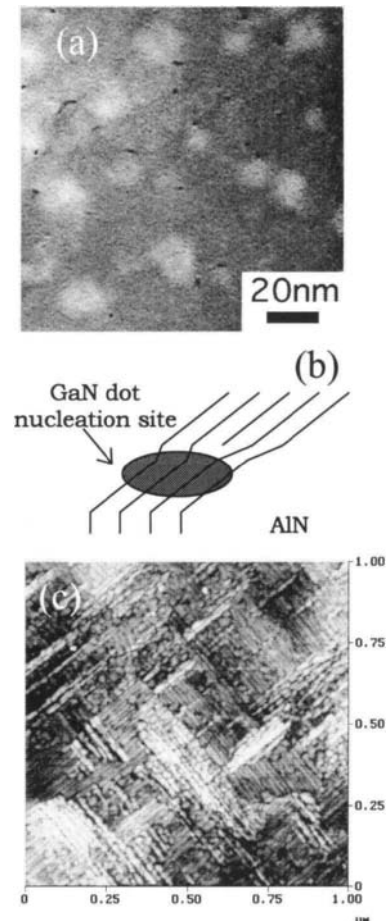

Figure 10 : a) plane-view TEM image of GaN dots nucleated on AlN. Note the black spots besides the dots which are threading edge dislocations ; b) schematics of the GaN dot nucleation mechanism in extended region of the AlN substrate ; c) AFM image of cubic GaN dots nucleated on cubic AlN. Note the alignement of dots along 2 perpendicular directions
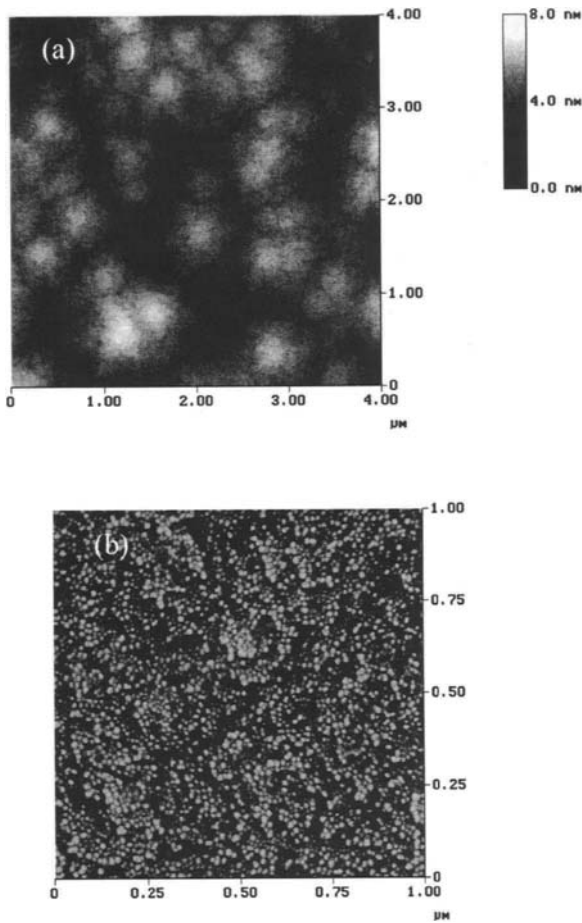
Figure 11: a) spiral-like steps on GaN surface. b) AFM image of the spiral-like arrangement of GaN dots in the wurzite phase.

It might be noted here that the situation is somewhat different in the case of InN quantum dots grown on GaN. In this case, the large elastic strain resulting from the epitaxial growth of inN on GaN (10 percent mismatch) cannot be relaxed by 3D islanding only but requires the additional formation of misfit dislocations. These dislocations are formed underneath the islands leading to the so-called incoherent SK growth mode, as illustrated in figure 13. Fourier filtered image reveals the presence of interface dislocations around the InN island embedded in GaN. Note that the periodicity of the dislocations is

roughly of one over 10 atomic planes, consistent with the mismatch value between InN and GaN.
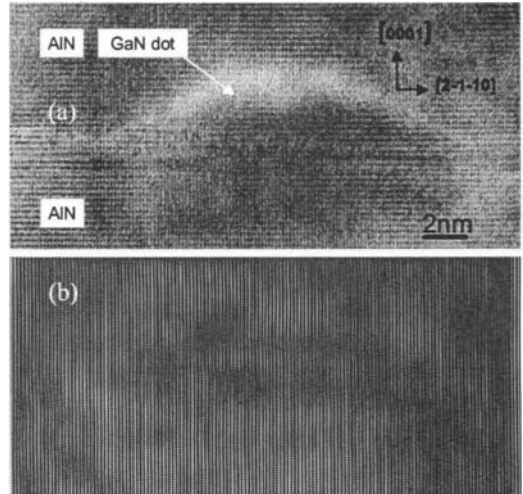
Figure 12: a) High resolution TEM viewgraph of a GaN dot embedded in AlN b) Fourier filtering image of the GaN dot.
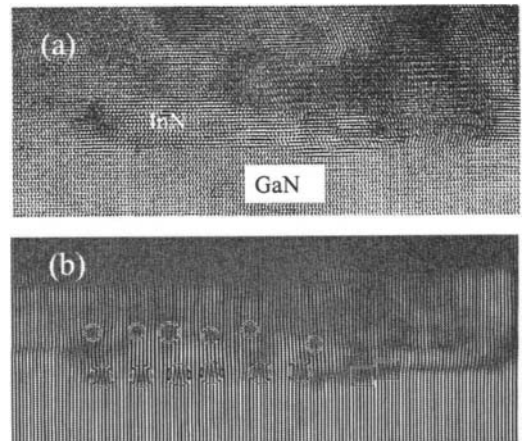
Figure 13: a) High resolution TEM viewgraph of a InN dot embedded in GaN b) Fourier filtering image of the InN dot showing interface dislocations (white circles).

2.2 Correlation of dots

Although the lateral distribution of dots is hard to control, many experimental results on vertical correlation of stacked dot planes are available in literature. In most cases, the vertical correlation is assumed to result from the presence of a surface strain field originating from buried islands. Again the case of nitrides provides an illustration of this mechanism which actually allows for an efficient homogeneization of dot size.

As a matter of fact, taking advantage of the 2D growth mode exhibited by AlN deposited on GaN, it is possible to fabricate periodic, stacked layers of GaN dots separated by an AlN spacer. Depending on the thickness of the AlN spacer separating the island layers, vertical correlation between the GaN dots has been observed by Cross-section Electron Microscopy. While no correlation was found for an AlN spacer thickness of 20 nm [16], such a correlation is present for an AlN spacer thickness typically smaller than 10 nm. This is illustrated in figure 14 which show a stacking of GaN dot layers (with the 2D wetting layer, about two MLs thick,) separated by spacers of AlN which are 10 nm thick. It is worth noting the presence of the vertical dislocation line going through the vertical line of dots on the left hand side of figure 14. This dislocation originates from the AlN buffer layer and, consistent with the results of figure 10, has acted as a nucleation center for dots.

In the case of cubic material, the results of figure 10c have also been confirmed by TEM : as shown in the cross section image of figure 14b, cubic GaN dots stacked one over each other tend to align along inclined directions corresponding to stacking faults in the (111) plane. As far as GaN dots in the hexagonal phase are concerned, the vertical correlation observed for thin enough AlN spacer is consistent with a mechanism proposed by Tersoff et al. [20], i.e. QDs preferentially nucleate on top of each other due to strain field modulations created by the dots in the AlN spacer layer. During this process, a size filtering mechanism is expected to operate, as

the strain field above smaller dots is weaker than above bigger ones, eventually leading to the disparition of smaller dots. This filtering mechanism is shown in figure 15, which puts in evidence merging of two adjacent columns of dots.
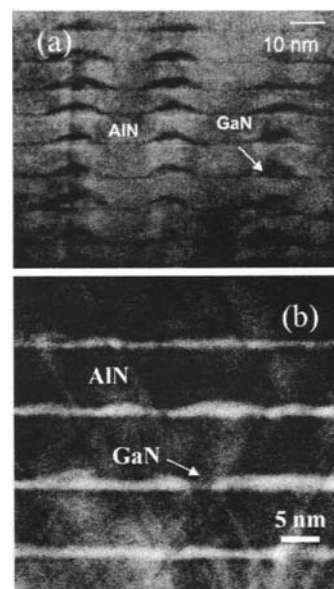

Figure 14: a) TEM cross section image of vertically correlated hexagonal GaN dots . b) TEM cross section image of cubic GaN dots aligned along directions parallel to stacking faults in the (111) plane
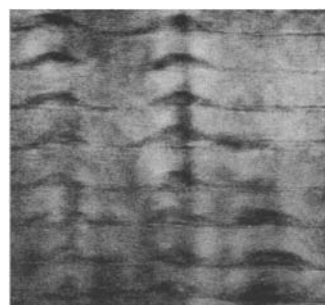

Figure 15: Stacking of GaN dot planes embedded in AlN. Note the merging of two columns of dots, as predicted by the model of Tersoff and coworkers [20]
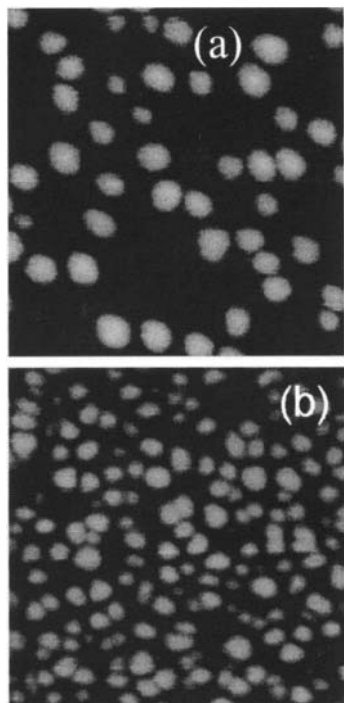
Figure 16: AFM image of GaN dots. a) vertically correlated ; b) uncorrelated. The image dimensions are 200x200 nm.

The effect of vertical correlation and size filtering is further illustrated in figure 16 which shows AFM images of GaN dots grown at the same temperature. Figure 16a corresponds to a stacking of 20 layers of GaN dots, the AlN spacer being thin enough for vertical correlation to be present. By contrast, figure 16b corresponds to one layer of uncorrelated dots. Clearly, vertical correlation results in a smaller density of bigger dots compared to the case of uncorrelated dots. Furthermore, figure 16 also reveals that ' the size distribution of correlated dots is narrower than the size distribution of uncorrelated dots, as an evidence that the vertical correlation mechanism is an efficient way of homogeneizing the dot size distribution.

The results in figure 14b tend to indicate that the mechanism proposed by Tersoff and coworkers is inappropriate to describe the

alignment of GaN dots in the cubic phase. Alternatively, it may suggest that the strain field created by buried layers is less important than the strain field created by the presence of stacking faults which eventually governs the nucleation process. However, it is reasonable to predict that a reduction in the density of stacking faults should minimize their role as dot nucleation center and would allow for an obervation of vertical alignment also in the cubic case.

3. KINETICS OF GaN GROWTH

In first approximation, the theoretical analysis of the SK growth mode described in section 2 relies on a detailed balance of interface, surface and elastic energy. The additional possibility of relaxing the elastic strain through misfit dislocation formation has also been examined, as discussed in the case of InN dots deposited on GaN which illustrate the so-called incoherent SK mode. Besides the parameters mentioned above, the elastic strain relaxation mode of nitride heterostructures and, more generally, the growth mode of nitrides in MBE also depends on kinetical parameters which have a drastic influence on island formation. In particular, when decreasing the growth temperature, the island formation mechanism may be completely inhibited and the growth mode of GaN on AlN may be quasi-2D.

The HRTEM cross section image in figure 17 illustrates this statement. It shows that the SK growth mode of GaN, far from being unique strongly depends on growth conditions. Furthermore, figure 17 also shows that 2D growth can be very well controlled.

This was confirmed by quantitative TEM image analysis which has shown that interfaces are very sharp and that the AlN on GaN interface is larger by one monolayer (ML) than the corresponding GaN on AlN one [21]. It will be shown below that the suppression of the SK mode in favor of another mode of elastic strain relaxation mechanism associated with the formation of misfit dislocations (see figure 17b and 17c)

directly results from the change in the adatoms kinetics due to the change in growth conditions.
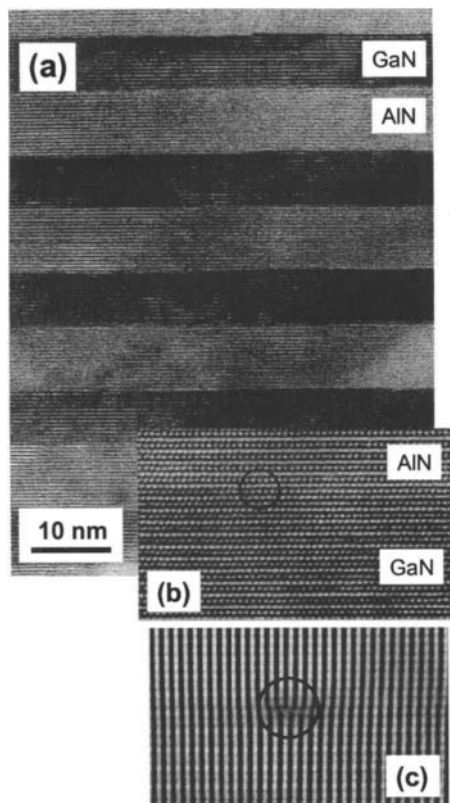

Figure 17: Transmission electron micrograph of a 7.0 nm/9.3 nm GaN/AlN superlattice. (a) general view, (b) atomic imaging showing perfect coherency at the interface except for one dislocation well visible on filtered image (c). Zone axis <2-1-10>.

3.1 Kinetics : the role of the Ga/N ratio

In addition to the effect of growth temperature, the value of the metal/active nitrogen ratio is of primary importance in governing the adtoms kinetics on growing GaN or AlN surface. As a matter of fact, it has been found that slightly Ga-rich conditions are associated with the observation of a streaky RHEED pattern, characteristic of a smooth surface. Intense and persistent RHEED oscillations typical for a layer-by-layer (Frank-Van der Merwe) are observed in these conditions for both AlN and GaN. By contrast, N-rich conditions lead to the roughening of the growing surface associated with the observation of a spotty RHEED pattern. This change in morphology when changing the metal/N ratio value has been assigned to the drastic dependence of the surface diffusion barrier as a function of the metal coverage. For, instance, it has been calculated that the surface diffusion barrier for Ga is far higher for a N-rich surface (1.8 eV) than for a Ga-rich surface (0.4 eV) [22].

In the specific case of GaN, it has also been recently demonstrated that, for extremely Ga-rich conditions, a continuous Ga film can eventually form on the surface[23]. This Ga film, about 2 MLs thick, opens a very efficient diffusion path for N adatoms and is therefore expected to drastically change the surface morphology of the growing layer.

Such an interplay between stoichiometry of the growing surface, adatom kinetics and growth mode are not specific to nitrides. It should be recalled here that the morphology of InAs grown on InP depends on In/As ratio value, leading to the formation of quantum wires or quantum dots with concomitant variations in optical properties [24]. In the extreme case of InGaAs grown on InP in In-rich conditions, it has also been found that 3D islanding could be completely inhibited [25], emphasizing again the determinant role of growth conditions in the QD self-formation process.

In order to relate more precisely the role of the Ga/N ratio value in the growth kinetics of GaN on AlN to the elastic strain relaxation mechanism, figure 18 shows the relative variation of the in-plane lattice parameter, $\Delta a/a$, as a function of time for different Ga fluxes at a substrate temperature of $T_S = 740$ °C.

Consider, first of all, the case corresponding to a Ga effusion cell temperature of $T_{Ga} = 1040$ °C. In this condition, the GaN growth is nearly stoichiometric, i.e. Ga/N ≈ 1. Initially, the in-plane lattice parameter shows almost

no variation before a rapid increase after 8 s of GaN deposition. The apparent relaxation after 8 s, corresponding to the deposition of approximately 2 MLs, is assigned to the elastic relaxation by GaN islanding, characteristic of the Stranski-Krastanov growth mode described in section 2.

For lower Ga fluxes, i.e. for increasingly N-rich growth, the phenomenology remains essentially the same. The 2D-3D transition becomes smoother as the GaN growth rate decreases for lower Ga flux. Note that below T_{Ga} = 1030 °C a small lattice parameter relaxation is observed during wetting layer deposition, that is assigned to a weak surface roughening due to N-rich growth associated with the emergence of very flat 2D platelets likely 1-2 MLs high.

It is worth noting that the existence of such very flat platelets has been confirmed by the oscillatory variation of the in-plane lattice parameter when depositing the wetting layer (see figure 3). Results shown in figure 3 and 18 tend to indicate that these platelets allow for a partial elastic relaxation of the built-in elastic strain. However, after deposition of about 2 MLs, the accumulated elastic energy is so large that the platelet-based mechanism turns to be uneffective and it is followed by 3D island formation.

The situation is quite different when decreasing the growth temperature, as illustrated in figure 19 which shows the relative variation of the in-plane lattice parameter obtained for a substrate temperature of T_S = 660 °C.

Then, in contrast to the high temperature results described above, a significant increase of the lattice parameter is observed from the very beginning of the growth for most Ga flux values. This behavior is distinctly different from that of the SK growth mode because of the lack of a 2D wetting layer. Again, this effect is assigned to the formation of GaN platelets that relax elastically at their borders [9].

Actually, an AFM image of these platlets is shown in figure 13. They appear as flat islands with heights of around 4 ml (1 nm)

and diameters of around 15 nm. The aspect ratio of the platelets is thus about 1/15, which is to be compared to the value of 1/5 for the pyramidal-shaped SK islands obtained at higher temperatures. However, due to their very flat shape, platelet formation alone cannot account for the large elastic strain relaxation occuring soon after strating GaN deposition, as shown in figure 19. Actually, as it will be discussed in section 3.2, such a large strain relaxation is assigned to the additional formation of misfit dislocations. One example of such a dislocation has been shown in figure 17b and 17c.
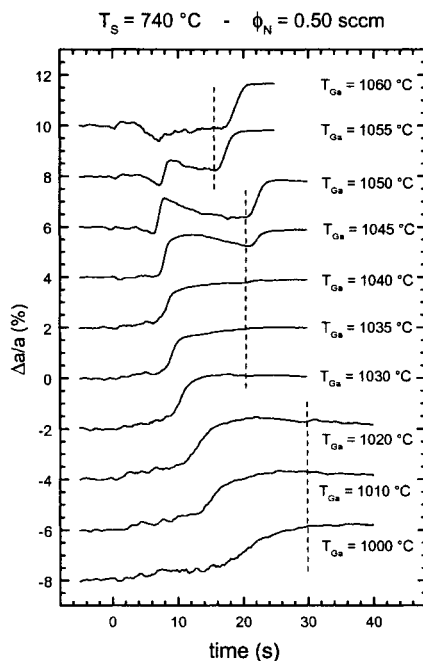

Figure 18: Relative variation of the in-plane lattice parameter a during the growth of GaN on AlN. The substrate temperature was T_S = 740 °C, the N_2 flow 0.50 sccm at 300 W rf power, and the Ga cell temperature as indicated. The dashed lines indicate the growth interruption under N flux.
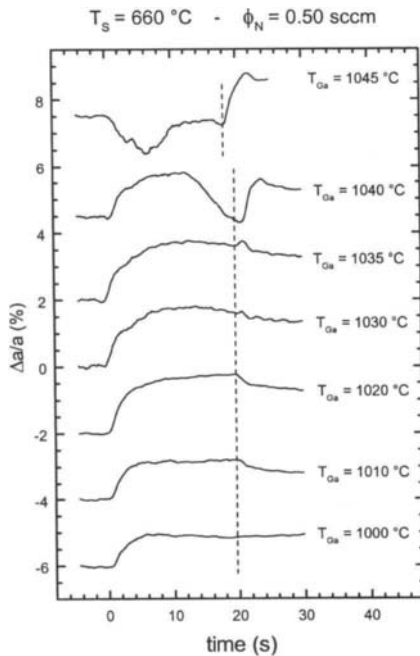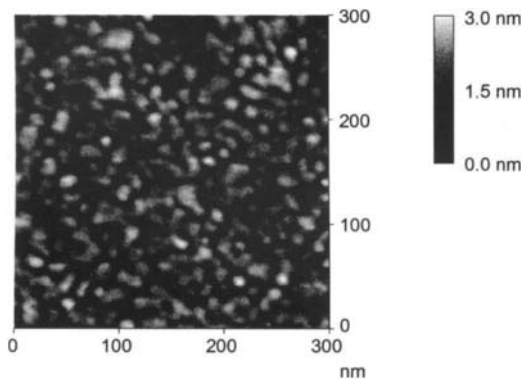
T$_S$ = 660 °C   -   φ$_N$ = 0.50 sccm

Figure 19: Relative variation of the in-plane lattice parameter a during the growth of GaN on AlN. The substrate temperature was T$_S$ = 660 °C, the N$_2$ flow 0.50 sccm at 300 W rf power, and the Ga cell temperature as indicated. The dashed lines indicate the growth interruption under N flux.

Figure 20 : AFM image of a 4 nm thick GaN layer grown on AlN in Ga-rich conditions. T$_S$ = 640°C

The role played by the Ga/N ratio value in the growth kinetics of GaN can be summarized in a phase diagram depicting the different modes of elastic relaxation during the first stages of GaN growth on AlN (0001) as a function of the growth parameters. Figure 21 shows the summary diagram. The lines are intended as guides to the eye indicating the boundaries between the different growth modes. The symbols represent the measured Ga cell temperatures T$_{Ga}$, at which the transition between two growth modes occurs, and were extracted from the series of Δa/a curves obtained at different substrate temperatures T$_S$.

First, it appears that for all substrate temperatures, the growth is always 2D for a given Ga flux. The critical flux increases weakly with temperature, probably due to enhanced Ga reevaporation. For a lower Ga flux, transiently formed islands are found to subsequently coalesce. Then, growth becomes 2D again and remains so.

For even lower Ga fluxes, two regimes are found: at higher temperatures, SK growth is observed, whereas at lower temperatures platelets are formed. At intermediate temperatures (T$_S$ = 680 °C) SK growth is only possible as a transient under slightly Ga-rich conditions. Otherwise, the surface diffusion is sufficiently low so that platelets are formed and relax elastically, then reducing the mismatch and leading to a complete disappearance of the 2D-3D transition.

As expected, the stability region for SK growth increases with increasing substrate temperature. This is consistent with the view that thermally-activated adatom diffusion mobility and reduced platelet nucleation rate lead to a vanishing roughening of the wetting layer by platelets and consequently to a vanishing elastic relaxation before the 2D-3D transition takes place.
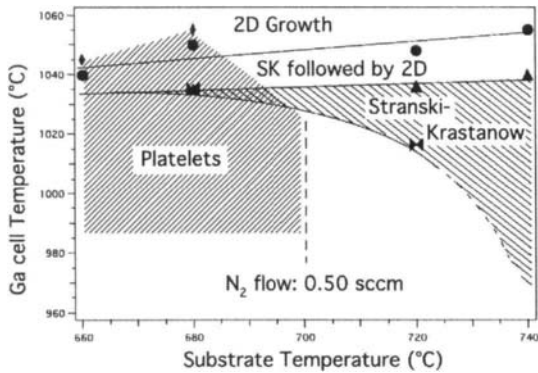
Figure 21: Growth mode phase diagram of GaN on AlN at a N_2 flow of 0.50 sccm at 300 W rf power, i.e. at maximum growth rates of 0.3 ML/s.

So, as a partial conclusion of this section, it appears that the growth mode of GaN deposited on AlN can be manipulated by changing either growth temperature or Ga/N flux ratio, putting in evidence the interplay between thermodynamical and kinetical parameters in the growth mode and/or in the elastic strain relaxation mechanism.

3.2 Platelets and strain relaxation

It should be emphasized that the platelets are to be unambiguously distinguished from the truncated pyramidal islands observed at higher temperature, characteristic of the SK mode. The height of the platelets considered here is smaller, their density much higher and the coverage close to one. They do appear right at the beginning of the deposition and, as they are flat, the growth remains in a pseudo-2D mode, with nuclei formation on top of the platelets. Also, the RHEED pattern maintains its overall streaky character.

It is worth stressing that, due to the very small height of the platelets, the amount of strain being elastically relaxed is small. Then, the formation of dislocations during the growth of GaN is expected as an alternative way of relaxing the built-up strain. As schematically shown in figure 22, dislocations are likely introduced during the platelet coalescence process. Then, the rate of dislocation introduction is proportional to the

local misfit experienced by the deposited material [26]
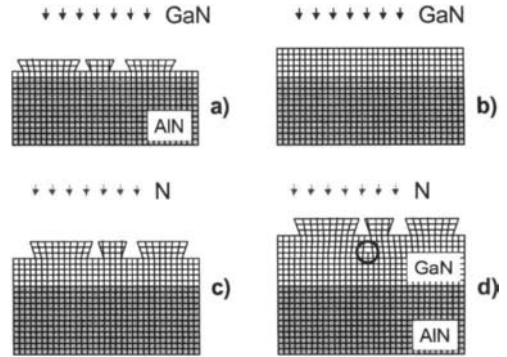
Figure 22: Schematic representation of the growth of GaN on AlN. (a) Early formation of flat GaN platelets in compression and outward elastic relaxation and (b) smoothing during further deposition. (c) Platelets reappear after N only deposition. (d) For larger thickness, dislocations are introduced and irreversible relaxation is observed

A more microscopic model for the nucleation of dislocations at the junction of two growing platelets is illustrated in figure 23. When two strained platelets are close together, the distance between the edge atoms of each platelet is modified compared to that between inner atoms due to edge relaxation. For platelets in tension (compression), this distance is larger (smaller). Thus, it is favorable for adatoms to nucleate an edge dislocation with a positive (negative) Burgers vector at the trough between two islands.

In the limit case, when this mechanism for introduction of misfit dislocations is dominant, the RHEED pattern of the growing GaN film remains smooth all along growth. Simultaneously, the $\Delta a/a$ variation exhibits a shape as shown in figure 24, i.e. a transient corresponding to platelets formation followed by their immediate coalescence. Next, the monotonous increase in $\Delta a/a$ is assigned to the progressive introduction of dislocations.
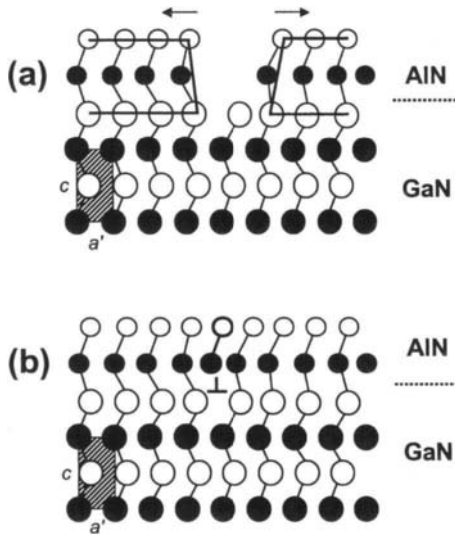
Figure 23: Schematic model of the nucleation of a misfit dislocation between two platelets in tension (AlN on GaN): (a) before coalescence, (b) after coalescence. Note that inward relaxation opens up the gap between the two platelets and favors adatom nucleation at this site, promoting dislocation nucleation.
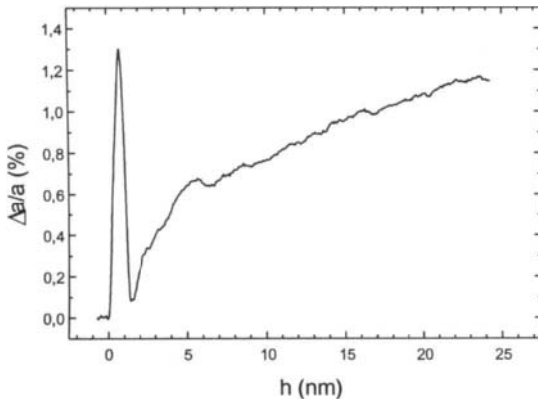
One may ultimately wonder about the reasons for which the shape of GaN islands correspond either to platelets or to truncated pyramids. One clue to adress this issue is provided by Duport and coworkers [27]. It has been found by these authors that the thermodynamically stable shape for an adsorbate cluster depends on the number of atoms it is constituted from. Then, for increasing cluster size, the stable shape evolves from a flat platelet to truncated pyramid and, finally, to a full pyramid, as shown in figure 25. Such a calculation remains to be done for GaN dots in the wurtzite structure. However, the evolution of GaN islands from platelets to truncated pyramids as a function of temperature can be qualitatively interpreted along similar views. In this model, the reduction in adatom mean free path at low temperature results in a decrease in size of nucleated adsorbate, leading it to adopt the platelet shape. At high temperature, by contrast, the reduction in the number of nucleation centers and the concomitant increase in the mean size of nucleated adsorbate tends to favor the truncated pyramid shape associated with the SK growth mode.

Figure 24: Relative variation of the in-plane lattice parameter as a function of layer thickness h during the growth of a single GaN epilayer on AlN. Ga rich conditions, T_S = 640 °C.
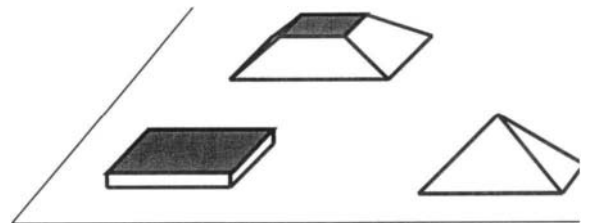

Figure 25: Island stable shapes as a function of their size. The stable shape evolves from flat slab to truncated pyramid and, next, to full pyramid (after ref. [27])

As a summary of this section, it appears that, during the first stages of the deposition of GaN on AlN, the accumulated elastic

energy may be accomodated either plastically or elastically. It has also been shown that the change from elastic to plastic relaxation mostly depends on kinetics which accounts for the formation of flat gaN platelets.

Of course, in the case of elastic relaxation through 3D islanding, further deposition will also lead to the formation of dislocations but this scope is far beyond the subject of the present article

4. EMISSION PROPERTIES OF GaN QDs

One major reason for the current interest in quantum dots relies on the possibility to realize optical devices such as lasers with low threshold current and with a weak temperature dependence of the threshold current [28]. More generally, the discrete density of states exhibited by quantum dots leads to no significant broadening as a function of temperature. In addition, the discrete density of states is expected to result in very sharp gain curves. Based on these principles, several QD lasers have been realized and reported during the last few years [29-34]. Compared to these InGaAs/GaAs devices, future lasers based on InGaN and GaN QDs are expected to operate in wavelengths ranging from visible to UV.

Two additional features specific to nitrides need to be stressed : first of all, the high energy barrier between AlN and GaN, exceeding 2.7 eV, is expected to result in reduced thermal escape of carriers when increasing temperature. This is illustrated in figure 26 which shows the photoluminescence (PL) spectra of a stack of 10 QD layers embedded in AlN for both cubic and hexagonal cristallographic phase. The samples were excited at a photon energy of 3.7eV and 4.67eV for the hexagonal and cubic GaN QDs, respectively. For these photon energies, the dots are excited by absorption directly in their excited states. Figure 26c shows the variation in photoluminescence intensity as a function of temperature, compared to the case of quantum wells and of bulk layer. Almost constant PL intensity is

observed in the case of dots, consistent with the enormous carrier localization energies involved.
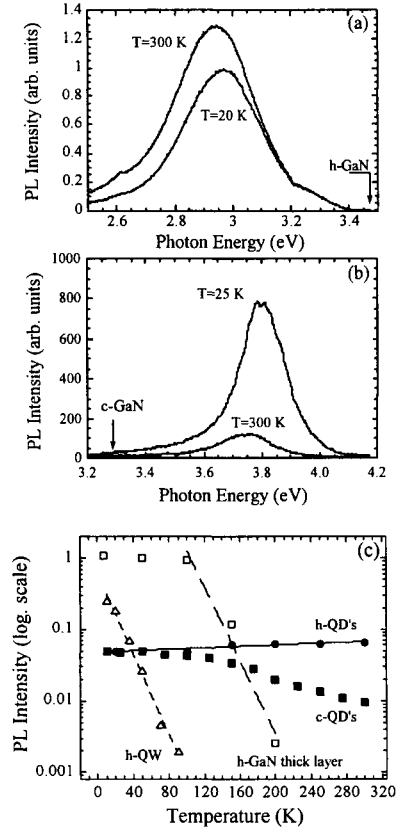

Figure 26: Photoluminescence spectra of GaN quantum dots for both crystallographic phases, hexagonal (a) and cubic (b). (c) shows the different temperature dependence of PL intensity for a thick GaN epilayer (□) and hexagonal quantum well (Δ), as compared to cubic QDs (■) and hexagonal QDs (●).

The linewidths of the broad luminescence spectra shown in figure 26a and 26b are related to the large dispersion of the dot size, likely due to the badly controlled nucleation process, as discussed in section 2.1.

Another specificity of nitrides with respect to other semiconductors is the presence, in the hexagonal phase, of a very large internal

electric field which is a combination of both piezoelectric and spontaneous polarization components [35]. As the value of the resulting internal field reaches several MV/cm, it dominates to a large extent the optical properties, leading to a significant red shift of the energy position of the peak assigned to QDs PL. As a matter of fact, figure 26a shows a peak at 2.95eV, that is to say, 500 meV below the bulk hexagonal GaN energy bandgap due to the presence of an internal electric field of about 7MV/cm. By contrast, in Fig. 26b corresponding to cubic QDs, the PL peak is typically centered at 3.8eV, blue-shifted by more than 500meV with respect to the bulk cubic GaN energy band gap (E_g = 3.26eV at 10K). This confinement energy is much larger than that observed for hexagonal GaN QDs, which is consistent with the absence of a strong internal electric field in these cubic GaN nanostructures.

red shift induced by the presence of the built-in electric field. For higher dots, the red shift is dominating and PL is observed for an energy lower than the GaN band gap energy value [37].

As a consequence of the presence of the internal field, the PL decay time is found to strongly depend on the QD size, as the electric field tends to separate the electron-hole pairs, reducing their overlap and increasing the radiative recombination time. This is shown in figure 28 which exhibits the spectrally-resolved decay time as a function of photon energy. As expected, the decay time is longer for GaN dots than for GaN thick layer. The continuous increase of the decay time when decreasing photon energy is consistent with the presence of the electric field since this one is larger in bigger dots luminescing at lower energy [38].

Figure 27: Quantum dot emission energy as a function of height (after ref.[33]). The dots correspond to experiments
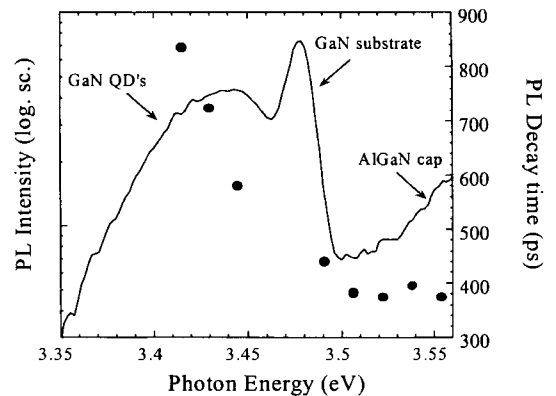

Figure 28: PL spectrum of a QD sample (left-hand side axis) and spectral dependence of the PL decay time as going through the various identified peaks (right-hand side axis)

Further insight in the role of the internal electric field has been provided by theoretical calculations of transition energy as a function of QDs height. Although controversy still remains concerning the value of spontaneous polarization, the calculation of Andreev et al. [36] leads to a reasonable agreement with experimental data (see figure 27). It is then deduced that, for dots about 2 nm high, the confinement is almost compensated by the

Of course, such a dependence of the decay time is not expected in cubic dots, due to the absence of electric field. As a whole the recombination efficiency is expected to be far

higher in cubic dots than in hexagonal ones, making cubic dots even more promising candidates for applications, provided the growth difficulties inherent to the cubic phase are overcome.

5. CONCLUSION

To summarize, we have revealed in this review some of the parameters governing the fabrication of GaN quantum dots in the Stranski-Krastanov mode by plasma-assisted molecular beam epitaxy. We have pointed out the interplay between thermodynamical and kinetical parameters, emphasizing the experimental evidence that kinetics may be dominant and eventually lead to a suppression of the SK mode. Parameters governing the nucleation of dots have been examined. In the hexagonal case, the nucleation sites are mainly determined by the residual dislocation density, giving rise to a vertical alignment. These hexagonal dots present an aspect ratio of 1/5. In the cubic case, the dots grow preferentially within regions bounded by stacking faults which are present in high density in the AlN layers. The aspect ratio of cubic dots is as low as 1/10. The large inhomogeneities in the size distributions are certainly related to the nucleation process likely governed by inhomogeneous strain distribution on the surface. As a consequence, wide PL peaks have been observed. In both cubic and hexagonal phases, the PL intensity of the quantum dots is found to be temperature-independent, as a result of strong localization effects.

As a general conclusion, it appears that nitrides, the self-organizing properties of which have been demonstrated only recently, exhibit features common to other dot-forming semiconductors, despite of the relatively low lattice mismatch, about 2.5 percent, between GaN and AlN. With the further advantage of weak interdiffusion and high confinement barrier they provide a model system for identifying the various parameters involved in self-organization of nanostructures. The ultimate goal, still far from being routinely reached for nitrides as well as for other semiconductors, consist of controlling the dot size distribution which would allow for a full use of the practical potentialities presented by these objects.

REFERENCES

1 I. N. Stranski und L. Krastanow, Sitzungsberichte der Akademie der Wissenschaften Wien, Mathematisch-naturwissenchaftliche Klasse, Abteilung IIb, Chemie, Band 146 (1938) 797

2. B. Daudin, F. Widmann, G. Feuillet, Y. Samson, M. Arlery and J. L. Rouvière, Phys. Rev. B 56 (1997) R7069

3. S. Tanaka, S. Iwai and Y. Aoyagi, Appl. Phys. Lett. 69 (1996) 4096

4. B. Damilano, N. Grandjean, F. Semond, J. Massies and M. Leroux, Appl. Phys. Lett. 75, (1999) 962

5. N. Grandjean and J. Massies, Appl. Phys. Lett. 72 (1998) 1078

6. F. Widmann, B. Daudin, G. Feuillet, Y. Samson, J. L. Rouvière, and N. Pelekanos, J. Appl. Phys. 83 (1998) 7618

7. J. Massies and N. Grandjean, Phys. Rev. Lett. 71 (1993) 1411

8. J. Eymery, B. Daudin, D. Brun-Le Cunff, N. Boudet and S. Tatarenko, Appl. Phys. Lett. 66 (1995) 3456

9. A..Bourret, C. Adelmann, B.Daudin, G.Feuillet, Guido Mula, Phys. Rev. B, in press

10. H. Yamaguchi, M. R. Fahy and B.A. Joyce, Appl. Phys. Lett. 69 (1996) 776

11. E. Martinez-Guerrero, C. Adelmann, F. Chabuel, J. Simon, N.T. Pelekanos, B. Daudin, G. Feuillet and H. Mariette, Appl. Phys. Lett. 77 (2000) 809

12. J. Y Marzin, J. M. Gérard, A. Izraël, D. Barrier and G. Bastard, Phys. Rev. Lett. 73 (1994) 716

13. X. W. Lin, J. Washburn, Z. Lilienthal-Weber, E. R. Weber, A. Sasaki, A. Wakahara and Y. Nabetani, Appl. Phys. Lett. 65 (1994) 1677

14. N. Grandjean, J. Massies and O. Tottereau, Phys. Rev. B55 (1997) R10189

15. Mitsuo Kawabe, Yong Jin Chun, Sigeru Nakajima and Kouichi Akahane, Jpn. J. Appl. Phys. 36 (1997) 4078

16. B. Daudin, G. Feuillet, J. Hübner, Y. Samson, F. Widmann, A. Philippe, C. Bru-Chevallier, G. Guillot, E. Bustarret, G. Bentoumi and A. Deneuville, J. Appl. Phys. 84 (1998) 2295

17. B. Daudin, G. Feuillet, Guido Mula, H. Mariette, J. L. Rouvière, N. Pelekanos, G. Fishman, C. Adelmann and J. Simon, Phys. Stat. Sol (a) 176 (1999) 621

18. B. Daudin, G. Feuillet, H. Mariette, Guido Mula, N. Pelekanos, E. Molva, J. L. Rouvière, C. Adelmann, E. Martinez-Guerrero, J. Barjon, F. Chabuel, B. Bataillou and J. Simon, Jpn. J. Appl. Phys. 40 (2001)

19. M. Arlery, J. L. Rouvière, F. Widmann, B. Daudin, G. Feuillet and H. Mariette, Appl. Phys. Lett. 74 (1999) 3287

20. J. Tersoff, C. Teichert and M. G. Lagally, Phys. Rev. Lett. 76 (1996) 1675

21. M. Arlery, PhD thesis, Université Joseph Fourier Grenoble I, 1998.

22. T. Zywietz, J. Neugebauer and M. Scheffler, Appl. Phys. Lett. 73 (1998) 487

23. J. E. Northrup, J. Neugebauer, R. M. Feenstra and A. R. Smith, Phys. Rev. B 61 (2000) 9932

24. F. Fossard, F.H. Julien, E. Peronne, A. Alexandrou, J. Brault and M. Gendry, Infrared Science and Technology, Proceedings QWIP 2000, Elsevier (2001)

25. Y. Robach, A. Solere, M. Gendry, and L. Porte, J. Vac. Sci. Technol. B 16 (1998) 1786

26. G. Feuillet, B. Daudin, F. Widmann, J. L. Rouvière and M. Arlery, J. Cryst. Growth 190 (1998) 142

27. C. Duport, C. Priester, J. Villain, in Morphological organization in epitaxial growth and removal, Zh. Zhang and M. Lagally (Eds), World Scientific series on directions in condensed matter physics (1997)

28. Y. Arakawa and H. Sakaki, Appl. Phys. Lett. 40 (1982) 939

29. N. N. Ledentsov, V. A. Shchukin, M. Grundmann, N. Kirstaedter, J. Böhrer, O. Schmidt, D. Bimberg, V. M. Ustinov, A. Yu. Egorov, A. E. Zhukov, P. S. Kop'ev, S. V. Zaitsev, N. Yu. Gordeev, Zh. I. Alferov, A. I. Borovkov, A. O. Kosogov, S. S. Ruvimov, P. Werner, U. Gösele, and J. Heydenreich, Phys. Rev. B 54 (1996) 8743

30. K. Kamath, P. Bhattacharya, T. Sosnowski, T. Norris and J. Phillips, electron. Lett. 32 (1996) 8743

31. H. Shoji, Y. Nakata, K. Mukai, Y. Sugiyama, M. Sugawara, N. Yokoyama and H. Ishikawa, Appl. Phys. Lett. 71 (1997) 193

32. L. Harris, D. J. Mowbray, M. S. Skolnick, M. Hopkinson and G. Hill, Appl. Phys. Lett. 73 (1998) 969

33. A. Patane, A. Polimeni, M. Henini, L. Eaves, P. C. Main and G. Hill, Appl. Phys. Lett. 85 (1999) 625

34. A. Patane, A. Polimeni, L. Eaves, M. Henini,. P. C. Main, P. M. Smowton, E. J. Johnston, P. J. Hulyer, E. Herrmann, G. M. Lewis and G. Hill, J. Phys. Lett. 87 (2000) 1943

35. F. Bernardini, V. Fiorentini and D. Vanderbilt Phys. Rev. B 56 (1997) R10024

36. A. D. Andreev, E. P. O'Reilly, Phys. Rev. B 62 (2000) 15851

37. F. Widmann, J. Simon, B. Daudin, G. Feuillet, J. L. Rouvière, N. T. Pelekanos and G. Fishman, Phys. Rev. B 58 (1998) R15989

38. C. Adelmann, M. Arlery, B. Daudin, G. Feuillet, G. Fishman, Le Si Dang, H. Mariette, N. Pelekanos, J. L. Rouvière, J. Simon, F. Widmann, C. R. Acad. Sci. Paris t. 1, Série IV, (2000) 61

Self-organised Growth of Silicon Nanocrystals in Nanocrystalline Si/SiO₂ Superlattices

Leonid Tsybeskov [a] and David J. Lockwood [b]

[a] Department of Electrical and Computer Engineering, University of Rochester
Rochester, New York 14627, USA

[b] Institute for Microstructural Sciences, National Research Council of Canada
Ottawa, Ontario, Canada K1A 0R6

The self-organized growth of silicon nanocrystals in nanocrystalline Si/SiO₂ superlattices results in Si nanocrystals nearly monodispersed in size with a distinct shape and preferred crystallographic orientation. A comprehensive structural, optical, and electrical characterization of this system indicates flat and atomically abrupt Si/SiO₂ interfaces and a low defect density. We present a convincing demonstration of quantization in dispersion of acoustic phonons and of electron-hole pairs. This novel fabrication technique, with angstrom accuracy, is very promising for nanoscale electronic and optoelectronic device manufacturing.

1. INTRODUCTION

The current worldwide interest in nanocrystalline Si (nc-Si) [1] has resulted primarily from Canham's 1990 proposal [2] that efficient visible light emission from highly porous Si structures arises from quantum confinement effects. This is particularly intriguing for three reasons. First, because nc-Si is a Si structure and bulk Si is spectacularly inefficient at emitting light, even at cryogenic temperatures. Second, because the light emitting nanostructure (e.g., porous Si) can be "made in a bucket" within minutes, without resort to either the costly lithographic or epitaxial techniques that were, at the time, the conventional approaches to realizing exceedingly small structures. Third, because Si is the most technologically important material known to mankind, dominating the microelectronics revolution, that influences our everyday lives. The original concept that light-emitting Si devices could eventually result in a new generation of Si chips and extend the functionality of Si technology from microelectronics into optoelectronics was exciting and encouraging. Thousands of papers and much knowledge concerning the nature and properties of light emission in nanoscale Si were generated [1]. The question now of concern is where are we ten years on? Can nanoscale Si meet expectations and be used

in Si nanoelectronics and optoelectronics? Can a reliable device be fabricated in a research laboratory and, eventually, transformed into real commercial manufacturing?

In addition to anodic etching [2], silicon nanocrystals can be produced by selective-size precipitation [3] spark erosion [4], ion implantation [5], and chemical vapor deposition [6]. Although the role of quantum confinement in porous Si has been demonstrated [7] the recombination mechanism in nc-Si is not yet completely understood. The price for the simplicity in porous Si fabrication is high: the structure is fragile, and the Si nanocrystal size distribution is broad. Also, there is no manufacturing technique that is able to control precisely the size distribution, surface passivation, and packing density of the Si nanocrystals. Controlled growth mechanisms at the nanometer length scale are required. Recently, the precise control afforded by molecular beam epitaxy and magnetron sputtering methods have been applied to this problem and definitive observations of quantum confinement induced light emission were obtained from ultrathin-layer disordered Si/SiO_2 superlattices (SLs) [8, 9]. In this system, it proved impossible to crystallize the nanometer-thick Si layers because of the $Si-SiO_2$ interface strain effects.

Among the many different methods of fabricating Si nanocrystals, the thermal crystallization of amorphous Si (a-Si)/SiO_2 layered structures [10] shows the greatest potential. The Si nanocrystals are packed in the form of ordered layers, and their size

and packing density can be controlled precisely. This type of structure has been termed a nc-Si SL, because it consists of periodically alternating layers of Si nanocrystals and SiO_2. In this chapter we show that these novel nc-Si SLs prepared by controlled crystallization exhibit nearly perfect $nc-Si/SiO_2$ interfaces and possess a narrow nanocrystal size distribution. In addition, we discuss in detail the vibrational, light-emitting, and electronic properties of nc-Si SLs. The thorny questions posed above will be answered at a later date, when this novel nc-Si SL technology is more mature.

## 2. CONTROLLED CRYSTALLIZATION OF NANOMETER THICK a-Si LAYERS: A NOVEL GROWTH TECNIQUE

It is well known that atomic-scale growth control of man-made structures can be achieved using techniques such as molecular beam epitaxy (MBE) and different modifications of chemical vapor deposition (CVD) [11]. Historically, a successful growth has been achieved in a family of materials with essentially the same structural parameters (e.g., lattice constant and thermal expansion coefficient). Therefore, the substrate and the epitaxial film usually belong to the same class of materials (e.g., GaAs/AlGaAs). The growth of materials with a significant lattice mismatch (e.g., Ge and Si) is much more challenging, and requires surface preparation and the formation of a so-called wetting layer [11]. The fabrication of a continuous layer always

254

utilizes the Frank-van der Merwe (FM) growth mode to avoid layer cracking and dislocation formation. More recently, enormous attention has been paid to the island-layer or the Stranski-Krastanow (S-K) growth mode, which under proper conditions may provide three-dimensional clusters with a narrow size distribution and vertical self-alignment in superlattice structures. The latter can be achieved because of the specific cluster shape (domes or pyramids) and the strain induced nucleation mechanism.

In addition to the hetero-growth of crystalline materials, a layered structure can be fabricated in a form of an amorphous superlattice (e.g., a-Si:H/a-SiN:H) [12]. This fabrication is less expensive, does not require an ultra-high vacuum, and can be achieved using plasma-enhanced CVD (PECVD) methods. However, in a large majority of reported structures the interface roughness and interface defect density are considerable [12]. Also, the use of PECVD methods results in a high concentration of loosely bound hydrogen atoms. These relatively weak silicon-hydrogen bonds can be broken at temperatures as low as 450°C and post-treatment steps cannot be used to improve the quality of the structure and to reduce the interface defect density.

Here we present a novel fabrication technique that utilizes inexpensive deposition methods such as radio-frequency (RF) magnetron sputtering or low-pressure CVD to produce a-Si/a-SiO$_2$ layered structures. We use several post-treatment steps such as (1) rapid-thermal annealing to nucleate Si crystallites; (2) quasi-equilibrium furnace annealing to complete crystallization; and (3) controlled oxidation to eliminate the residual a-Si and passivate the Si nanocrystals. The completed structures exhibit excellent periodicity, nearly atomically flat SiO$_2$/nc-Si interfaces, and a very low interface defect density.
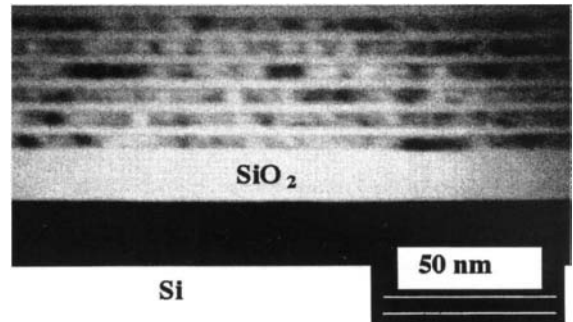
(a)
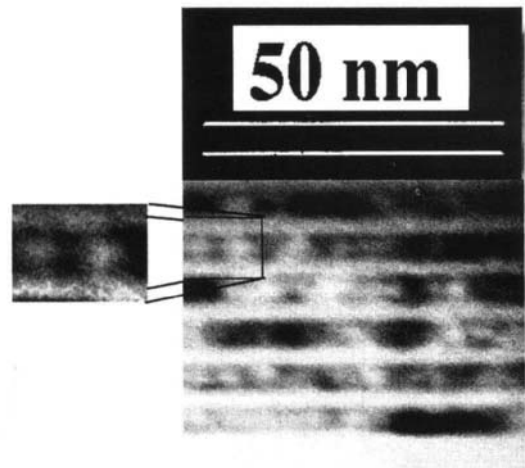
(b)

Figure 1. TEM images showing (a) an example of a nc-Si SL and (b) a magnified picture detailing two individual Si nanocrystals.

3. PREPARATION AND CHARACTERIZATION

The magnetron sputtering of the a-Si/SiO$_2$ multilayers was performed at Rochester in a Perkin-Elmer 2400 sputtering system by RF sputtering and plasma oxidation. In the sputtered samples, the a-Si thickness ranged from ~ 2 to 25 nm and that of the SiO$_2$ from 2.5 to 6 nm. The number of periods varied from 1 to 60. The crystallization was performed by rapid (40–60 s) thermal annealing (RTA) at 800–900°C followed by furnace annealing. In the furnace annealing step, the temperature was increased by ~10°C per minute from 600 to 1050°C. This annealing is defined as quasi-equilibrium annealing (QEA).

The thermal crystallization of a-Si/SiO$_2$ multilayers is controlled by several factors. Silicon has a very low diffusivity in SiO$_2$ and the initial RTA process forms nanocrystals without disturbing the SL periodic structure. The shape of the Si nanoclusters is expected to be close to spherical due to the competition between surface and volume tension. The strain in nc-Si/SiO$_2$ multilayers can be released by QEA. High temperature annealing also results in a decrease of the Si/SiO$_2$ interface defect density (see Ref. 13 for details).

The role of the double-step annealing process is to first create crystalline nuclei by RTA (nucleation stage) and then to complete the crystallization of the Si nanocrystals and improve their surface passivation by QEA (growth stage). After nucleation, the nanocrystalline nuclei are surrounded by an amorphous tissue and occupy not more than ~20–30% by volume of the Si layer (estimated from Raman scattering). After furnace annealing and completed crystallization, nearly 90–95% of the Si layer volume was found to be crystallized in samples with Si nanocrystals greater than 3 nm in diameter [13, 14].

A wide variety of characterization techniques have been applied to study the nc-Si SLs, including transmission electron microscopy (TEM), X-ray specular reflectivity measurements, and Auger spectroscopy [13-14]. The transmission electron microscopy (TEM) analysis was performed with a Philips EM 430 microscope operated at 300 kV. Samples were prepared by a cleavage method [15]. Auger elemental microanalysis (AEM) was performed with a Physical Electronics Industries PHI 650 instrument with a 5 keV electron beam 30° off normal. An argon ion gun operating at 1 keV and 51° off normal was used for depth profiling. The Si LMM and O KLL Auger lines at 96 eV and 510 eV, respectively, were used to obtain the SL composition profile. X-ray specular reflectivity measurements were carried out using a Philips PW1820 vertical goniometer and simulated using the Philips HRS calculation package. Electrical measurements were performed at temperatures ranging from 20 to 300 K in a closed-cycle CTI-Cryogenic system using a Keithly 220 current source and a Keithly 595 quasi-static CV meter. High frequency capacitance-voltage (C-V) measurements were obtained using a Boonton

72BD 1 MHz capacitance meter. Raman experiments were carried out at room temperature in a quasi backscattering geometry using a Jobin-Yvon triple spectrometer or a Spex 14018 double spectrometer. Low-frequency Raman scattering was also measured using a SOPRA DMDP 2000 double monochromator with a cooled Hamamatsu R928P photomultiplier detector. The excitation source was 300 mW of 457.9 nm Ar^+ laser light with an angle of incidence of 77.7°. The incident light was polarized in the scattering plane, while the scattered light was collected without polarization analysis. Time-resolved photoluminescence (PL) data were obtained using a pulsed N_2 laser with a nanosecond pulse duration. Steady state PL spectroscopy was performed using a Ti:sapphire laser and an Ar^+ laser.

Figure 1 shows a TEM picture of a crystallized and annealed a-Si/SiO_2 sample. The nc-Si/SiO_2 layers are very flat with an average roughness less than 0.4 nm. The size of the Si nanocrystals within the layers is close to the thickness of the a-Si layer while the volume packing density is controlled by the thickness of the SiO_2 layer. Figure 1(b) shows the same sample at higher magnification and depicts individual nanocrystals of nearly perfect spherical shape. The nanocrystal diameter is around 4.2 nm with a size distribution of ~ 3–5 %. A similar diameter of ~ 4.6 nm was deduced from the X-ray diffraction analysis.

4. VIBRATIONAL PROPERTIES

Figure 2 shows Raman scattering data in nc-Si SLs with different crystallite sizes and compares it with similar measurements in crystalline Si (c-Si). For example, peaks that correspond to first-order optical zone-center phonon (ZCP) scattering in samples with 8.6 nm or 4.3 nm Si nanocrystals are broader and are slightly shifted toward lower wavenumber compared to the Raman peak in c-Si (Fig. 2(a)). The nanocrystal size of > 4 nm is too large to provide any significant effects in phonon dispersion due to phonon confinement [16], and the observed differences can be explained as being due to strain, an influence of the nc-Si/SiO_2 interface or "bulk" defects (including grain boundary, dislocation, etc.), and non-stoichiometric SiO_{2-x} (the shoulder near 505 cm^{-1}). Figure 2(a) shows progressive changes in the first-order ZCP Raman scattering during high-temperature (1050°C) annealing: the Raman peak narrows, shifts to ~519 cm^{-1}, and the shoulder at ~505 cm^{-1} continuously decreases in intensity. Therefore, we can conclude that annealing decreases the Si defect density and improves the SiO_2 stoichiometry in the nc-Si SLs.

Second-order Raman scattering near 150 and 300 cm^{-1} in c-Si has been explained as phonon scattering from acoustic phonons at wave vectors away from the zone center [17], and the overtone at 302 cm^{-1} is associated with the 2TA (X) phonon. This feature is relatively sharp in c-Si. It increases in intensity, but changes shape in the nc-Si SL with ~ 13 nm Si nanocrystals, and becomes much broader in samples with 4 nm Si nanocrystals (Fig. 2 (c)).
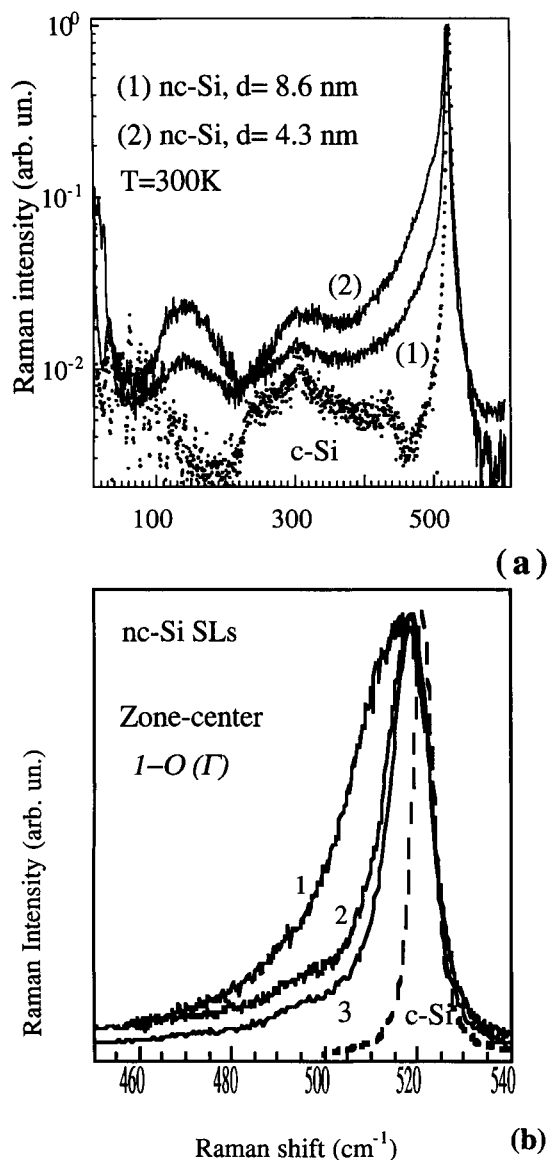
(1) nc-Si, d= 8.6 nm

(2) nc-Si, d= 4.3 nm

T=300K

(a)

(b)

Figure 2. Raman spectra in crystallized nc-Si SLs: (a) The first-order (1-O) zone-center phonon mode and second-order transverse acoustic (TA) phonon scattering in c-Si and SLs with different size Si nanocrystals. (b) The spectra labeled from 1 to 3 show progressive changes due to 20 min annealing steps at 1050°C and are compared with a c-Si spectrum.

This weak scattering in c-Si is also enhanced in intensity when the translational symmetry is disrupted. However, in all our Raman measurements, a pure a-Si phase with peak at 480–490 cm^{-1} was clearly absent. Thus, our Raman data are consistent with selected-area electron diffraction TEM and X-ray diffraction data [13]: the treated material is not amorphous, but crystalline, with a more or less random crystallographic orientation of the Si nanocrystallites. This conclusion is important because a-Si is a highly defective material and these defects can strongly affect the performance of a nanoscale device.

Also, our Raman data clearly shows that the structural properties of nc-Si SLs can be improved by high temperature annealing. This post-treatment step, however, could be dangerous for our layered structure: diffusion at high temperature can be efficient and the abrupt nc-Si/SiO$_2$ profile could be damaged. To address this question we have performed a series of post-treatment experiments such as annealing and oxidation at different temperatures. Samples subjected to annealing have been investigated by TEM (see, e.g., Fig. 1). These measurements have demonstrated that the nc-Si SL structure retains its layered structure with sharp nc-Si/SiO$_2$ interfaces.

A quantization effect of the phonon vibrations could be observed in the case when the phonon coherence length is larger than the nc-Si layer thickness. Low frequency acoustic modes fulfill this

try. Low frequency acoustic modes reflect the average bulk elastic properties of the materials and can be very useful to help estimate the periodicity and the interface roughness in a multilayer structure.

Figure 3(a) shows low-frequency Raman scattering in nc-Si SLs having different nanocrystal layer thicknesses. The acoustic phonon dispersion at small wave vectors in a bulk material is, to first approximation, given by $\omega_m(k) = v_s k$. This is modified in the case of a SL structure to

$$\omega_m(k) = v_s|k + 2\pi m/L|,$$ where v_s is the SL sound velocity, k is the phonon wave vector, $m = 0, \pm 1, \pm 2...$ is the reduced-Brillouin-zone folding index, and L is the SL period. The periodicity of the SL introduces an artificial Brillouin zone boundary at π/L along the growth direction, which results in the folding of the phonon dispersion [16]. The choice of the excitation wavelength in Raman backscattering experiments $\lambda = 4\pi n/k_p$, where n is the SL average refractive index and k_p the photon wave vector, allows excitations to be probed near the center or boundary of the artificial Brillouin zone. In our samples, the first pair of the folded longitudinal acoustic (FLA) phonon modes with $m = \pm 1$ is observed, and the mode frequencies are in a good agreement with theory (our calculations are based on Ref. 18 and are discussed elsewhere [19].) More importantly, the post treatment steps such as steam oxidation at 750°C following annealing at 1050°C do not change the

Figure 3. Low-frequency Raman scattering in nc-Si SLs: (a) SLs with different Si nanocrystal sizes and (b) Raman spectra in a nc-Si SL before and after steam oxidation (750°C) and annealing at ~ 1050°C post-treatment steps.

condition, because they are an in-phase motion of a large number of atoms and are not strongly influenced by the disruption of the translational symme-

folded mode spectra (Fig. 3(b)). The small peak frequency shift of ~0.5 cm^{-1} evident in this figure is not a systematic one, and is due to a small lateral non-uniformity in the SL thickness across the wafer arising from the growth conditions. A simple analysis shows that FLA phonon modes would disappear from Raman spectrum in the case of an average interface roughness larger than ~ 0.4 nm [18]. Thus the observation of FLA modes in these nc-Si SLs indicates sharp interfaces that are retained even after high temperature annealing.

5. PHOTOLUMINESCENCE

The PL spectra in a nc-Si/SiO$_2$ SL with Si nanocrystal sizes larger than the exciton Bohr radius is found to be similar to bulk Si [20, 21]. However, several feature of the PL spectra prove that the PL originates in the nc-Si/SiO$_2$ SLs, and not in the c-Si substrate. First, the quenching of the PL with temperature and electric field are unusually weak [20]. Second, the p+ c-Si substrate on which the nc-Si/SiO$_2$ SLs are deposited for these measurements does not luminesce at room temperature, whereas the PL efficiency of the nc-Si/SiO$_2$ SLs is estimated to be greater than 0.1% at 300 K [20]. A short-time steam oxidation produces a relatively high low-temperature (LT) PL external quantum efficiency of ~0.3%. This makes it possible to examine details of the PL phonon replicas and the no-phonon (NP) line (Fig. 4(b)), that are extremely sensitive to impurities and intrinsic defects. The LT PL fine structure in the nc-Si/SiO$_2$ SLs proves that the Si nanocrystals are extremely pure, and no defect or impurity related PL has been found. Preliminary doping experiments show that nc-Si/SiO$_2$ SLs containing large Si nanocrystals can be doped by phosphorus diffusion. The well-known phosphorus signature in the LT PL spectrum of bulk c-Si (see Ref. 22) was identified in the P-doped nc-Si/SiO$_2$ SLs (Fig. 4(c)). Since the nc-Si/SiO$_2$ SLs are separated from the c-Si substrate by a ~100 nm thermally grown SiO$_2$ layer, P diffusion into the c-Si substrate is improbable. Finally, a shift of the PL peak toward shorter wavelengths has been found in smaller Si nanocrystals (d < 5 nm).

For small-size Si nanocrystals (d < 5 nm), the inhomogeneous broadening of the PL spectrum does not permit a moderate resolution spectroscopic identification of the radiative transitions in Si nanocrystals. The inhomogeneity can be partially lifted by selectively exciting only a subset of the luminescent states, and spectrally sharp PL features can be obtained. Figure 5 shows resonantly excited PL spectra in the nc-Si SL with 4.3 nm Si nanocrys-tals. A moderate resolution PL spectrum shows a strong dependence on the excitation wavelength. The step-like structure reflects phonon-assisted radiative transitions, and the energies of the characteristic c-Si phonons are shown (Figure 5). The decrease of the excitation photon energy makes all features signifi-cantly sharper. The use of the asynchronously modulated excitation technique allowed the detection of a slow, in the range of > 100 μs, PL component.
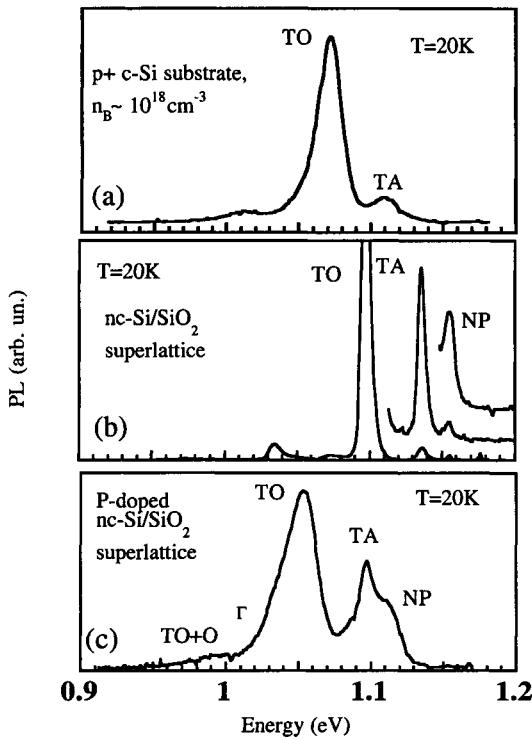
Figure 4. The PL spectra of (a) the B-doped Si substrate, (b) a crystallized nc-Si/SiO$_2$ SL after short-time steam oxidation, and (c) a P-doped nc-Si/SiO$_2$ SL. The PL phonon replicas and the no-phonon (NP) line are indicated and shown in detail (Fig. 4(b)). The LT PL spectrum of the crystallized nc-Si SL(Fig.4(b)) is similar to the PL spectrum of bulk c-Si with an impurity level of 10^{12} cm^{-3} (see Ref. 23 for details). The Si nanocrystal size is ~13 nm.

It is well known that phonon-assisted recombination is a slow process, with a typical lifetime in the 1 ms range [22]. We studied in detail the PL kinetics and found at least three different PL components: (a) a slow PL with > 1 ms lifetime at low temperature, decreasing to 10^{-4}–10^{-5} s at room
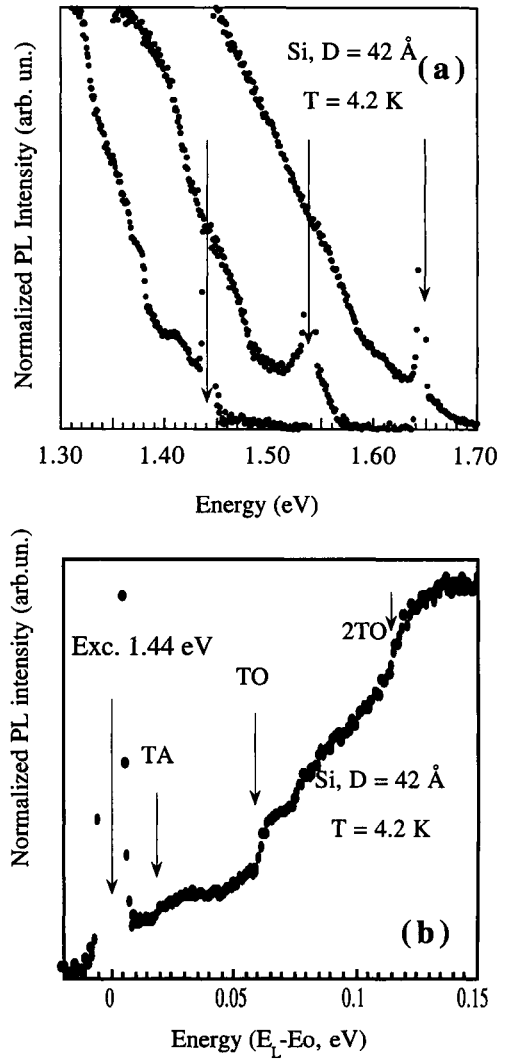
Figure 5. (a) Moderate and (b) high-resolution PL spectra in a nc-Si/SiO$_2$ SL under resonant excitation at 4.2 K. The step-like structure in the PL spectrum shows the signature characteristic of c-Si phonons (indicated by the arrows).

temperature; (b) a faster PL with a 200–500 ns lifetime; and (c) the fastest PL with a lifetime of the order of ~10 ns. The fastest PL component clearly

shows a superlinear intensity dependence, which can be expected in Auger recombination. The PL with 200-500 ns lifetime also shows a significant dependence in the excitation intensity, which could be explained by saturation of the recombination channels associated with specific recombination centers. The measurements show that the origin of the PL could be more complex than just one dominant mechanism.

6. CARRIER TUNNELING

The usefulness of layered semiconductor structures and SLs consisting of alternately deposited materials very often depends on the quality of their interfaces. Ideally, each interface must be chemically abrupt and free of structural defects. Unfortunately, this is not the case in the majority of Si-based multilayer structures: a significant difference in lattice constant and thermal expansion coefficient strongly affects the interfaces of crystalline Si-Ge, Si-CaF$_2$, and other Si-based SLs [23]. Silicon oxide, a wide gap native insulator for c-Si, can not be used in traditional epitaxial growth because of its amorphous nature, i. e. absence of long-range order. In general, only disordered Si can be grown on a substrate with a nanometer thick amorphous SiO$_2$ layer [8]. On the other hand, Si continues to be the major material in the semiconductor industry, and the question of the interface quality in Si based layered structures is critically important.
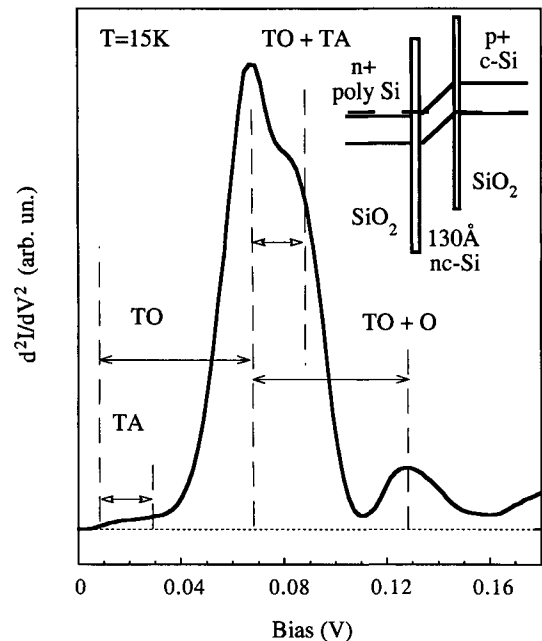
Figure 6. Second derivative of the I-V curve in a structure consisting of Al / 300 nm n+ polycrystalline Si contact / 1.5 nm a-SiO$_2$ / 13 nm nc-Si / 1.5 nm a-SiO$_2$/ p+ c-Si substrate. Different c-Si phonon energies are shown. The tunnel current threshold of ~ 46 mV indicates a small, built-in potential, possibly due to a localized charge.

A series of 8- to 20-period SLs were grown for this study. Details of the sample preparation conditions are described in Refs. 10–12. Several techniques have been used to study the nc-Si/SiO$_2$ interface, but current tunnel spectroscopy (CTS) is one of the most sensitive techniques available to characterize junctions and interfaces [24-27]. In order to apply CTS, we fabricated a double-barrier structure with a single ~13 nm-thick nc-Si layer sandwiched between two, tunnel-transparent SiO$_2$ layers of thickness ~1.5 nm. The structure was deposited on a heavily doped p+ Si substrate (resistivity of ~0.01 Ω

cm) with an Al/n+ doped polycrystalline Si top contact (Fig. 6). The structure is similar to the Esaki diode, because the electron injected from the top contact should tunnel across the 13 nm layer of Si to the heavily B-doped substrate. The experiment was performed at 15 K to minimize thermal broadening and the electron-phonon interaction. In order to clarify the details, we have used the second derivative of the I-V data. Several features corresponding to different phonon and phonon-combination energies (TO, TA, etc.) are observed (Fig. 6). Note that in a Si Esaki diode of perfect quality, up to twelve transitions with different phonons and phonon combinations have been reported [28].

The observation of phonon-assisted transitions in the interband tunneling indicates clean nc-Si/SiO$_2$ interfaces. A silicon nanocrystal greater than 10 nm in size has an indirect bandgap structure, and the interband tunneling requires phonon assistance to conserve momentum. The interband tunneling is always accompanied by the excess current. In the majority of cases the origin of the excess current has been associated with tunneling via interface defect states. The inelastic scattering due to interaction with structural defects changes the electron momentum. Thus, the phonon-assisted tunneling is no longer the dominant mechanism, and the features in the I-V curve should be washed out. The observation of phonon-assisted tunneling indicates a very low density of interface defects in the nc-Si/SiO$_2$ structure.
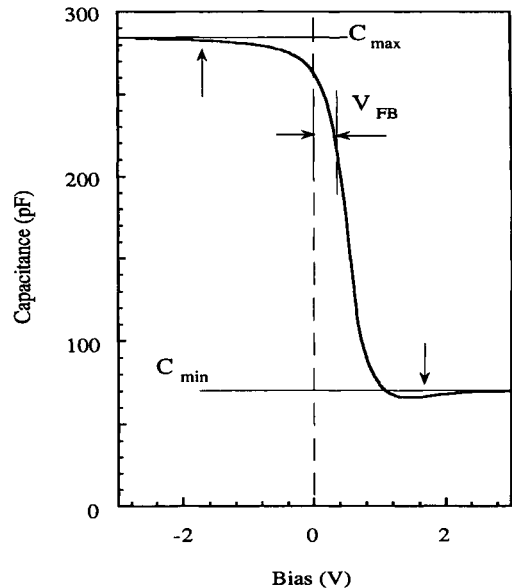
Figure 7. High-frequency (1 MHz) room temperature C-V characteristic measured in an 8-period nc-Si/SiO$_2$ SL with ~ 2 nm thick nc-Si layers separated by ~ 2 nm of SiO$_2$. The arrows show the calculated voltages for C$_{max}$ and C$_{min}$ in the ideal MIS structure. The value of the flat-band voltage V$_{FB}$ ~ 500 mV indicates a small amount of localized charge.

To estimate the interface defect density in a nc-Si/SiO$_2$ SL, we used conventional C-V measurements. The sample used for this work consisted of 20 nm of thermally grown amorphous SiO$_2$ on an n-type (~1-5 Ω cm) c-Si substrate followed by an 8-period SL. The nc-Si layers with a thickness of ~2 nm are separated by amorphous SiO$_2$ layers with a thickness of ~2 nm. The 20 nm of thermally grown SiO$_2$ blocks the direct current, and the 1 MHz frequency C-V characteristic is very similar to the C-V curve in a standard metal-insulator-semiconductor (MIS) structure (Fig. 7). Figure 7 shows the C-V

data with a saturation at C_{max}, which corresponds to the capacitance of the insulating layer (including the thermally grown SiO_2 and nc-Si/SiO_2 SL) with an average dielectric constant of ~1.8–2.0 (depending on the volume ratio between Si and SiO_2). In addition, the C-V plateau at C_{min} is found to be in excellent agreement with the calculated value of the maximum depletion layer thickness (~0.5 μm) for the ~10^{16} cm^{-3} concentration of free carriers that corresponds to the resistivity of ~5 Ω cm for an n-type c-Si substrate. This result justifies the use of a standard high-frequency C-V analysis: comparing the experimental C(V) data with the C(V) characteristic calculated for the ideal MIS structure, the value of the interface defect density is estimated to be ~10^{11} cm^{-2}.

What does this defect density value mean in the case of the 8 nc-Si layers (or 9 interfaces) separated by SiO_2? Considering a Si nanocrystal area density of 10^{12}-10^{13} cm^{-2} and assuming that all the defects are on the nc-Si/SiO_2 interfaces, we estimate that only one per 100-1,000 nanocrystals is defective. This estimate is supported by our recent observation of a relatively high PL quantum efficiency in nc-Si/SiO_2 SLs. Our latest results indicate that the defect density can be significantly reduced using an optimized preparation condition and a cleaner manufacturing environment. In other words, the density of nc-Si/SiO_2 interface defects is approaching the level of defect tolerance in standard CMOS processing.

7. CONCLUSION

Precisely controlled layer-by-layer growth of a-Si/SiO_2 SLs followed by annealing stages to induce Si crystallization and defect reduction has produced a novel system for studying the optical and electronic properties of regular arrays of ~5 nm Si nanocrystals. The sharp nc-Si/SiO_2 interfaces and the low density of defects have allowed tunneling spectroscopy to be applied to these structures. Efficient (> 0.1%) infrared light emission is observed at room temperature, although phonon-assisted recombination is still required. A high thermal stability of these structures has been demonstrated. These results show that nc-Si/SiO_2 structures have considerable potential for nanoelectronic and optoelectronic device applications.

Acknowledgements

264

References

1. D.J. Lockwood, Light Emission in Silicon: From Physics to Devices, Academic Press, San Diego, 1997.

2. L.T. Canham, Appl. Phys. Lett. 57 (1990) 1048.

3. W. L. Wilson, P. J. Szajowski, and L. E. Brus, Science 262 (1993) 1242.

4. R. E. Hummel and S.-S. Chang, Appl. Phys. Lett. 61 (1992) 1965.

5. T. Shimazu-Iwayama, S. Nakao, and K. Saitoh, Appl. Phys. Lett. 65 (1994) 1814.

6. H. Takagi, H. Ogawa, Y. Yamazaki, A. Ishizaki, and T. Nakagiri, Appl. Phys. Lett. 56 (1990) 2379.

7. P. D. J. Calcott, K. J. Nash, L. T. Canham, and M. J. Kane, Mater. Res. Soc. Symp. Proc. 358 (1995) 465.

8. Z. H. Lu, D. J. Lockwood, and J.-M. Baribeau, Nature (London) 378, 258 (1995).

9. B.T. Sullivan, D.J. Lockwood, H.J. Labbé, and Z.-H. Lu, Appl. Phys. Lett. 69, (1996) 3149.

10. L. Tsybeskov, K. D. Hirschman, S. P. Dutta-gupta, M. Zacharias,, P. M. Fauchet, J. P. McCaffrey and D. J. Lockwood, Appl. Phys. Lett. 72 (1998) 43.

11. R. Hull and J. C. Bean, Principles and Concepts of Strained-Layer Epitaxy, in Strained-Layer Superlattices: Materials Science and Technology, Semic. and Semimet., Vol. 23, (ed. T. P. Pearsall), Academic Press, New York, NY, 1991, 1-67.

12. R. A. Street, Hydrogenated Amorphous Silicon, Cambridge University Press, NY (1991) 417.

13. L. Tsybeskov, K. D. Hirschman, S. P. Dutta-gupta, P. M. Fauchet, M. Zacharias, P. Kohlert, J. P. McCaffrey and D. J. Lockwood, in Quantum Confinement IV: Nanoscale Materials, Devices, and Systems, edited by M. Cahay, J.-P. Leburton, D.J. Lockwood, and S. Bandyopadhyay, the Electrochemical Society, Pennington, NJ, 1997, 134-145.

14. L. Tsybeskov, G. F. Grom, P. M. Fauchet, J. P. McCoffrey, J.-M. Baribeau, H. J. Labbe, G. I. Sproule, and D. J. Lockwood, in Quantum Confinement : Nanostructures, edited by M. Cahay, J.-P. Leburton, D.J. Lockwood, and S. Bandyopadhyay, the Electrochemical Society, Pennington, NJ, 1999, 76-94.

15. J. P. McCaffrey, Microsc. Res. Tech. 24 (1993) 180.

16. I. H. Campbell and P. M. Fauchet, Solid State Commun. 58 (1986) 739.

17. P. A. Temple and C. E. Hathaway, Phys. Rev. B. 7 (1973) 3685.

2. D. J. Lockwood, M. W. Dharma-wardana, J.-M. Baribeau, and D. C. Houghton, Phys. Rev. B 35 (1987) 2243.

3. G. F. Grom, D. J. Lockwood, J. P. McCaffrey, H. J. Labbe, P. M. Fauchet, B. White, J. Diener, H. Heckler, D. Kovalev, F. Koch, and L. Tsybeskov, Nature 407, (2000) 358.

4. L. Tsybeskov, K. D. Hirschman, S. P. Dutta-gupta, P. M. Fauchet, K. L. Moore, and D. G. Hall.Appl. Phys. Lett. 69 (1996) 3411.

5. L. Tsybeskov, K. L. Moore, D. G. Hall and P. M. Fauchet,, Phys. Rev. B., Rapid Communications, 54 (1996) R8361.

6. G. Davis, Physics Reports, 176 (1989) 84.

7. See for example, Silicon Molecular Beam Epitaxy, Materials Research Society Symposium Proceedings 220, J. C. Bean, S. S. Iyer, and K. L. Wang (ed.), Pittsburgh, PA, 1991, 651.

8. T. Hiramoto, in Mesoscopic Physics and Electronics, ed. T. Ando, Y. Arakawa, K. Furuya, S. Komiyama, H. Nakashima, Mesoscopic Physics and Electronics, Springer, Tokyo, 1998, 213-219.

9. A. H. Dayem and R. J. Martin, Phys. Rev. Lett. 8 (1962) 246.

10. R. S. Becker, J. A. Golovchenko, D. R. Hamann, and B. S. Swartzentruber, Phys. Rev. Lett. 55 (1985) 2032.

11. H. P. Zeindl, et. al., Appl. Phys. Lett. 50 (1987) 1164.

12. A. G. Chynoweth, R. A. Logan, and D. E. Thomas, Phys. Rev. 125 (1962) 877

Growth and Characterization of Semiconductor Silicon Carbide
for Electronic and Optoelectronic Applications: An Industrial Perspective

H. McD. Hobgood, M. Brady, W. Brixius, G. Fechko, R. C. Glass, D. Henshall, J. Jenny, R. Leonard, D. Malta, St. G. Mueller, V. Tsvetkov, and C. H. Carter, Jr.

CREE , Inc., 4600 Silicon Drive, Durham, NC, 27703, USA

Within the last several years SiC semiconductor device technology for optoelectronic and power electronic applications has made tremendous progress resulting primarily from the commercial availability of SiC substrates of ever increasing diameter and quality. Achieving the full potential of this important semiconductor continues to be critically dependent on the production of large diameter SiC single-crystals of high crystalline quality and controlled impurity content. The objective of this paper is to discuss from an industrial viewpoint the current state of SiC crystal growth technology and to present empirical results that reflect recent advances in SiC crystal growth, including: substrate diameter enlargement to 100-mm; improvements in crystal purity yielding optically transparent 6H and 4H-SiC crystals exhibiting residual impurities in the 10^{15} cm^{-3} range and corresponding improvement in thermal conductivity approaching 5.0 W/cmK; a continuous improvement in micropipe defect density to as low as 0.9 cm^{-2} over a full 50-mm diameter 4H-SiC substrate; and control of electrical behavior of SiC substrates.

1. INTRODUCTION

During the past decade SiC semiconductor device technology for electronic and optoelectronic applications has made tremendous progress resulting primarily from the commercial availability of SiC substrates of ever increasing diameter and quality. Examples of current state of the art devices include [1]: high brightness and ultra-bright blue and green InGaN-based LEDs which take full advantage of the electrical conductivity of the 6H-SiC substrate by employing a conductive AlGaN buffer layer; microwave MESFETs on semi-insulating 4H-SiC substrates with power densities as high as 4.6 W/mm at 3.5 GHz and total CW output power of 80 W at 3.1 GHz from a single chip; 8.6 kV p-i-n diodes fabricated on high quality SiC epitaxial layers; thyristors conducting 12 A at 6.5 V with 2600 V blocking voltage; and GaN/AlGaN HEMTs fabricated on semi-insulating 4H-SiC substrates exhibiting power densities of 6.9 W/mm at 10 GHz. These exciting device results stem primarily from the exploitation of the unique electrical and thermophysical properties offered by SiC compared to Si and GaAs. Among these are (Table 1.): a large band gap for high temperature operation and radiation resistance, high critical breakdown field for high power output, high saturated electron velocity for high frequency operation, and significantly higher thermal conductivity for thermal management of high power devices.

1.1. Considerations for SiC crystal growth

Throughout the technical evolution of semiconductor SiC, the fabrication of SiC crystals exhibiting the desired electrical and crystalline properties has played a central role in the realization of the full potential of this important semiconductor material. However,

Table 1
Comparison of fundamental properties of Si, GaAs, and SiC for device applications

	$E_g(eV)$	$E_b(MV/cm)$	$v_{sat}(10^7 cm/s)$	$\kappa_T(W/cmK)$	ε
Si	1.12	0.6	1.0	1.5	11.8
GaAs	1.42	0.6	2.0	0.5	12.8
6H-SiC	3.02	3.2	2.0	3.0*	10.0
4H-SiC	3.26	3.0	2.0	3.0*	9.7

*For substrate doping density of $\sim 5 \times 10^{18}$ cm^{-3}

the preparation of large diameter monocrystals of SiC in industrial quantities presents significant engineering challenges. In spite of the extensive terrestrial abundance of its elemental components, the natural occurrence of the compound (mineralogically called moissanite) in the earth's crust is scant to almost non-existent and is due mainly to chance collisions with SiC-containing debris such as meteoritic material and interstellar dust particles associated with stellar evolution [2]. Thus, the industrial fabrication of the compound relies exclusively on its chemical synthesis from the elemental components Si and C, a process first developed in the late1800s [3]. Transformation of this process into one capable of producing crystals first in the form of small spontaneously seeded platelets [4] and ultimately in the form of seeded monocrystalline boules [5,6,7,8] with diameters attaining 100-mm [9] has been the focus of intensive efforts over the last two decades.

While most single crystal semiconductor boules are grown by crystal pulling or seeded solidification from melts consisting of elemental (e.g. Si and Ge) or compound semiconductors (such as III-Vs and some II-VIs), the thermodynamics of SiC render these approaches impractical at present for the production of industrial quantities of monocrystalline SiC. The SiC phase diagram [10] exhibits a peritectic at 2830°C and a total pressure of $\sim 10^5$ Pa. Calculations suggest that stoichiometric melting occurs only at pressures exceeding 10^5 atm and temperatures > 3200°C [11]. Although it is feasible to create similar growth conditions, for example those used for growth of small diamond crystals, such a process is currently not feasible for commercial production of large diameter semiconductor grade SiC boules.

Another possibility is to exploit the solubility of carbon in a silicon melt. The solubility of C in liquid Si is 1-19 at.% over the temperature interval from 1412-2830°C. This implies that crystal pulling by top seeded solution growth from a carbon enriched silicon melt might be feasible. Recent studies employing over pressures of Ar (100-120 atm) to control the evaporation of silicon in carbon enriched Si melts have resulted in the growth of thick epitaxial layers at growth rates of 10-15 microns/hr; however, at the high growth rates required for production of large diameter bulk SiC crystals the technique is currently limited by silicon inclusions and polycrystallinity associated with process instabilities [12]. It is possible to increase the solubilty of carbon by adding other metals to the melt (e.g. Sc or various rare earth elements) which in principle would allow the use of pulling techniques. However, there are currently no crucible materials which would be stable with these melts and solvent evaporation remains a problem. In addition, the likelihood of incorporation of the metal additives in the growing crystals precludes their use as a
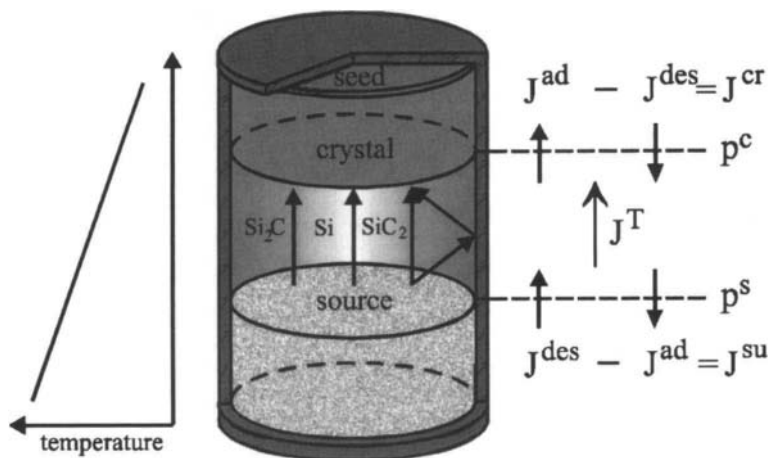
Figure 1. Schematic representation of a crucible configuration for SiC PVT growth and the mass transfer processes (see Section 2.1) inside the crucible. J^{ad} and J^{des} are adsorption and desorption fluxes, respectively. p^s and p^c are local partial pressures at the source and crystal, respectively.

semiconductor substrate. In light of these difficulties, physical vapor deposition via seeded sublimation has been the most successful technique for growth of SiC boules in production quantities.

2. SIC CRYSTAL GROWTH

2.1. The Physical Vapor Transport Process

Large diameter SiC crystal growth is based on a modification of the original SiC sublimation method first developed by Lely [4] and extended later as a seeded sublimation technique by Tairov and Tsvetkov [5]. This latter method, more generically termed physical vapor transport (PVT), was further refined by Carter [6], Stein [7], and Barrett [8] for producing large diameter SiC boules, and various modifications of these techniques are now used at many laboratories worldwide. As shown schematically in Figure 1, SiC crystal growth at temperatures of 2500K to 2700K proceeds by the transport of subliming molecular species across a temperature gradient resulting in deposition upon a monocrystalline SiC seed of the desired orientation held at a lower temperature. Growth is typically performed in a system in which various graphite components constitute the hot-zone of the furnace. The SiC source disposed in a growth crucible is usually sublimed under partial vacuum conditions in the presence of an inert gas ambient (e. g. Ar, He, N_2) which fills the growth chamber. Sublimation occurs non-congruently from the solid to the vapor and yields four principal molecular vapor species consisting of Si, Si_2, Si_2C, and SiC_2 [13]. The total vapor pressure of the subliming species in the temperature range typical for SiC growth (2500-2700K) has been estimated to be approximately 1.5-12 torr corresponding a particle density of $(0.6-4) \times 10^{16}$ cm^{-3} [14]. The sublimation rate is to first order a function of the source temperature, the pressure of the ambient gas, and the surface area of the source material. Figure 2 shows the dependence of the sublimation rate on pressure reported for one particular growth configuration [15].

Because the thermophysical conditions inside the crucible are not directly accessible during SiC crystal growth, an increasing
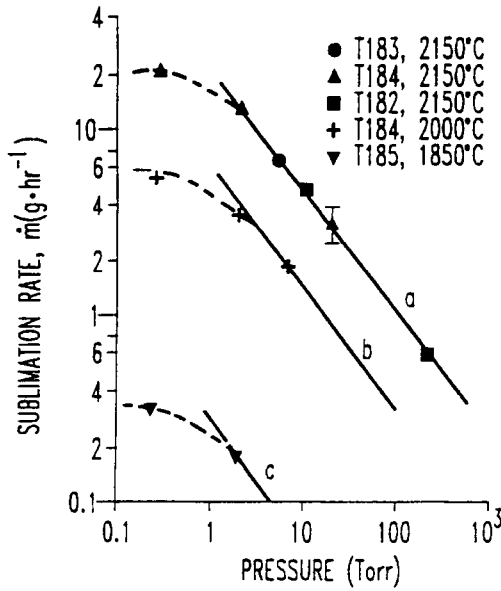
Figure 2. Sublimation rate vs pressure
and temperature (after [15]).

Figure 3. Kinetics of SiC crystal growth
in general coordinates.

number of research groups are beginning to utilize the results of numerical simulations of heat and mass transfer mechanisms for improvement of the growth process [16]. Such approaches, while desirable for the accuracy they can ultimately provide, must be customized to the individual growth system used and are highly dependent upon the availability of reliable thermophysical data of all materials inside the growth chamber. Although analytical models of SiC crystal growth are only approximately valid for describing the PVT growth of a dissociative compound like SiC, when taken with caution, they can provide important clues for the growth rate behavior observed in real SiC PVT systems. Following this approach, Mueller [17] has recently developed an analytical model which predicts with reasonable accuracy the growth rate behavior observed in their PVT system. In this model, the molar fluxes representing source sublimation (J^{su}) and

crystallization at the growth interface (J^{cr}) are linked by the transport of molecular species (J^{T}) through the intervening gas phase consisting of Si-C species and inert gas (Fig. 1). The relevant mass transfer mechanisms inside the crucible are diffusion and advective flow, while convection is negligible. A simplified version of this model assumes diffusion to be the dominant rate- limiting step for transport. The average diffusion coefficient D of the Si-C components in the vapor phase, taking into account the inert ambient gas, is inversely proportional to the total gas pressure in accordance with the classical expression for the bi-molecular diffusion coefficient:

$$D = A/(p^{*} + p^{Ar}) \qquad (1)$$

where p^{*} and p^{Ar} are the partial pressure of the subliming species and the inert gas ambient respectively and A is constant. Based on this simple model, it is predicted
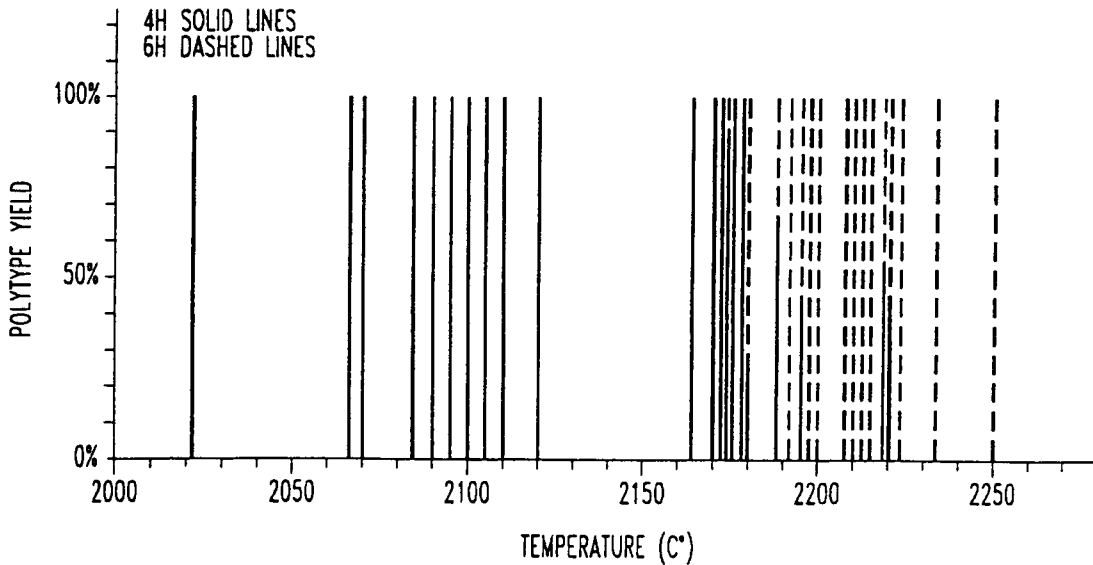
Figure 4. Polytype yield for PVT 4H-SiC growth as a function of growth temperature (after [15] and [23]).

that the total flux of J^T and thus the growth rate, is related to the partial pressure of the subliming species and the inert gas pressure by the reciprocal relation:

$$(J^T)^{-1} = B/p^* + C(1 + p^{Ar}/p^*) \qquad (2)$$

where B and C are constants in temperature. According to Eq. 1), the plot of experimental data in the coordinates $1/J^T = f(p^{Ar}/p^*)$ should be linear with a temperature independent slope. It also predicts a decreasing intercept with increasing temperature, since p* increases with temperature. The data of Figure 3 exhibits a close agreement with the predictions of eqn.2) and demonstrates that even a relatively simple analytical model can provide valuable insight into the SiC PVT process.

Controlling nucleation, composition, and thermal geometry in the PVT process requires automated pressure and temperature control. Precise control over nucleation is critical, and crystal quality is greatly affected by seeding parameters. One of the most important is the surface quality of the seed crystal. The loss of

Si from the seed surface during nucleation can result in a C-rich surface of poor quality, causing defect generation and loss of monocrystallinity.

2.2. Polytype stability

Silicon carbide occurs in over 170 different polytypes, the most common of which are the cubic 3C form, the hexagonal 4H and 6H forms, and the rhombohedral 15R structure. Various models have been put forward to account for the polymorphism of SiC [18,19]; however, to date, no single model offers a complete explanation for the mechanism of polytype formation and the stability regions for their existence during crystal growth. Early in the development of SiC it was noted qualitatively that certain polytypes tended to exist over certain temperature regimes [20]. Birnie and Kingery observed that the cubic polytype tends to be associated with silicon-rich conditions [21]. Recent studies in the technologically important 6H and 4H polytype systems have shown that temperature and pressure [22,23], crystal growth face [24], and
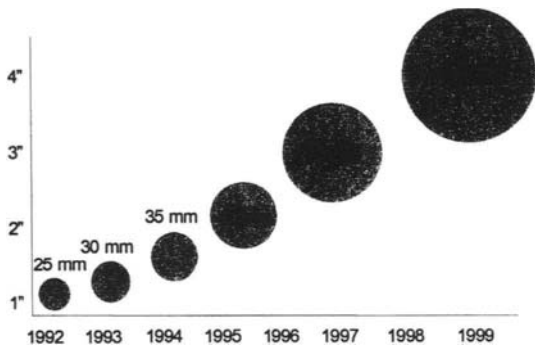
Figure 5. Progression of SiC diameter enlargement with time.

Figure 6. 100-mm diameter (0001)-oriented 6H-SiC wafer grown by physical vapor transport.

most recently the temperature gradient imposed across the growing crystal [25] have a major influence on the resulting polytype. Figure 4, for example, shows the data of Augustine, et al. [15,23] demonstrating that growth temperature exerts a major influence on the relative yield of 4H and 6H polytype. For their growth process, a critical growth temperature was observed below which the 4H polytype was favored over the 6H. This observation implies that to achieve high yields of the 4H polytype, reduced temperatures should be applied. However, reduction in temperature can lead to reduced growth rates. To sustain production growth rates, they suggest a corresponding reduction in process pressure [15]. Similar observations are reported by Straubinger et al. [25]. Additionally they demonstrate that the critical temperature for polytype conversion depends upon the axial gradient imposed during the growth. All of these studies suggest that the process parameters for seeded sublimation growth by PVT must be critically set to achieve optimum yields of the desired polytype.

2.3. Crystal diameter enlargement

Increasing substrate diameter is crucial for reducing the cost of SiC devices through economies of scale and the use of silicon or GaAs device fabrication equipment. However, growth of crack-free large diameter SiC crystals with high crystalline quality requires increased attention to system design and optimization of the thermal distribution in the growth environment to minimize excessive mechanical and thermoelastic stresses. At our laboratories, the diameter and quality of SiC substrates produced using the PVT technique have constantly increased. Figure 5 shows schematically the progression of our SiC diameter enlargement with time. It should be noted that milestones indicated in Figure 5 are those associated with research and development efforts. SiC substrate development has been characterized by the continuous increase in crystal diameter from less than 25-mm in the early 1990s up to current crystals with diameters of 100-mm. Production volumes of 6H and 4H-polytypes are now available at diameters up to 75-mm. Early prototype 75-mm diameter SiC wafers suffered from excessive low-angle grain boundaries near the wafer peripheries. This problem has been circumvented in the
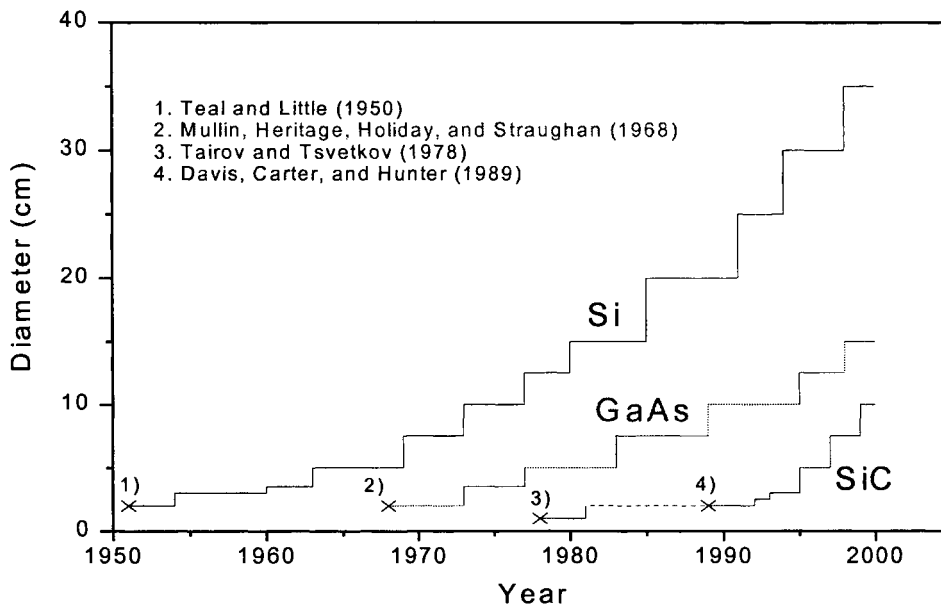
272

Figure 7. Evolution of SiC diameter enlargement compared to silicon and GaAs technologies.

intervening years allowing for the first time the extension of the PVT technique to the fabrication of fully single-crystal substrates with diameters up to 100-mm [9], as shown in Figure 6. This significant development in wafer enlargement represents a major step toward bringing semiconductor SiC to parity with III-V compound semiconductors, such as GaAs, in available wafer area for device fabrication. It is of interest to compare the rate at which SiC diameter enlargement has proceeded technologically compared to the corresponding development for silicon and GaAs. As shown in Figure 7, progress in the growth of SiC crystals of ever increasing diameter has proceeded at a rate well in excess of that experienced in the silicon and GaAs industries, largely as a result of the crystal growth experience gained from these foregoing technologies.

3. CRYSTALLOGRAPHIC DEFECTS

3.1. Growth mechanisms and defect generation

In parallel with development efforts to enlarge SiC crystal diameter, studies have been directed toward improvement in overall crystal perfection. Silicon carbide substrates are currently characterized by a wide range of crystalline defects, the most significant of which are open-core dislocations (called micropipes), low-angle boundaries, and conventional dislocations. Micropipe defects that occur to a varying extent in all SiC wafers produced to date are seen by many as preventing the commercialization of many types of SiC devices, especially high current power devices, and their origin and elimination in semiconductor SiC are subjects of intense interest to crystal growers and device engineers alike.

Table 2
Primary mechanisms of micropipe formation in SiC sublimation growth

fundamental

1. *thermodynamic*	2. *Kinetic*
a) thermal field uniformity	a) nucleation processes
b) dislocation formation	b) inhomogeneous supersaturation
c) solid-state transformation	c) constitutional supercooling
d) vapor phase composition	d) growth face morphology
e) vacancy supersaturation	e) capture of gas phase bubbles

technological

1. *process instabilities*	2. *seed preparation*	3. *Contamination*

Figure 8. Step flow around a platelet due to system instability.

Figure 9. SWBXRT showing coalescence of micropipes along length of boule: a) seed end, b) growth end.

Most discussions about micropipe formation mechanisms revolve around Frank's theory [26] which attributes the micropipe to the hollow core of an associated screw dislocation with a large Burgers vector several times the unit cell dimension. The dislocation content of micropipes is indicated by the existence of growth spirals originating at micropipes, characteristic stress patterns around micropipes visible in stress birefrigence [17], and by the results of synchrotron white beam X-ray topography (SWBXT) studies [27] and their interpretation based on kinematical diffraction theory [28]. Mounting evidence supports the Frank-model of hollow dislocation cores in SiC which allows a quantitative correlation of the micropipe radius with the magnitude of the Burgers vector of the screw dislocation associated with the pipe [26]. Recent studies have indicated that the critical Burgers vector magnitude above which micropipes tend to be generated is 3x the unit cell length in the 6H-SiC crystal lattice [29].

Several mechanisms, or combinations of mechanisms, related to or specifically causing micropipes in SiC boules grown by the seeded
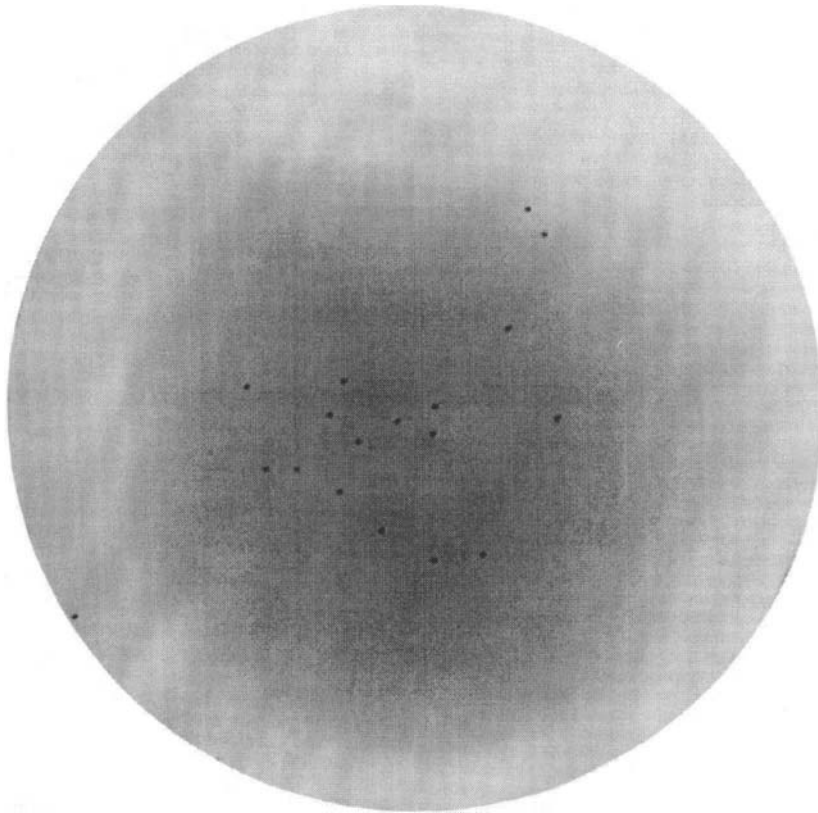
Figure 10. 50-mm KOH-etched 4H-SiC wafer with micropipe density of 0.9 cm^{-2}.

sublimation PVT method have been identified (Table 2) in our work. In all these cases, one must consider the seed surface quality, the growth process stability and cleanliness, as well as the specific parameters controlling nucleation density and growth rate. In our work we have found many examples where instabilities during growth contribute to micropipe formation through the formation of graphite particles and Si droplets. A unique feature observed on PVT grown SiC surfaces is the existence of clearly discernable growth spirals and steps. The density of these can vary, but the PVT growth rates (0.5 to 2 mm/hr) require a relatively high supersaturation of the growth constituents that can result in a high density of these spirals. If local instabilities occur in temperature or pressure and surface diffusion limitations prevail, the potential for secondary nucleation increases. Under this type of growth the opportunity for defect formation is high. As steps move across one another, islands contact each other at slightly different angles and dislocations are easily created. In this process, Figure 8 shows a growth surface exhibiting a high density of steps and the impingement of these steps on a platelet. These formations may initiate at defects arising from the seed or from other sources incurred during growth. In our work, we have seen that while micropipes form during growth, simultaneous processes
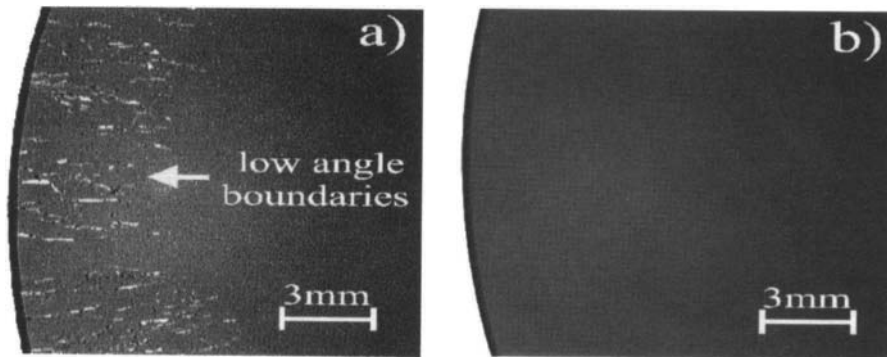
Figure 11. Elimination of low angle boundaries at periphery of 4H-SiC wafers: a) non-optimized process, b) optimized process.
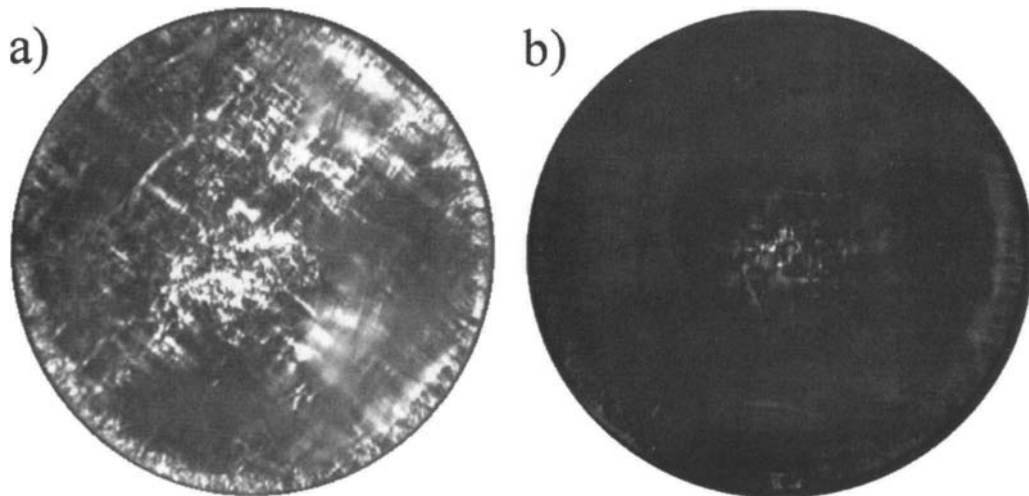
Figure 12. Quality improvement in 75-mm 4H-SiC as indicated by stress birefringence: a) early R&D wafer, b) recent wafer.

such as pipe movement and coalescence occurs which tends to reduce the number of pipes. This is evident in Figure 9 where a series of SWBXT topographs show the agglomeration of hollow cores in wafers analyzed along the length of a boule. In SWBXT the strain fields around the pipe disrupt the diffraction process to the extent that holes appear in the topographs.

Despite the different fundamental and technological reasons for micropipe formation, results at Cree made it possible to steadily
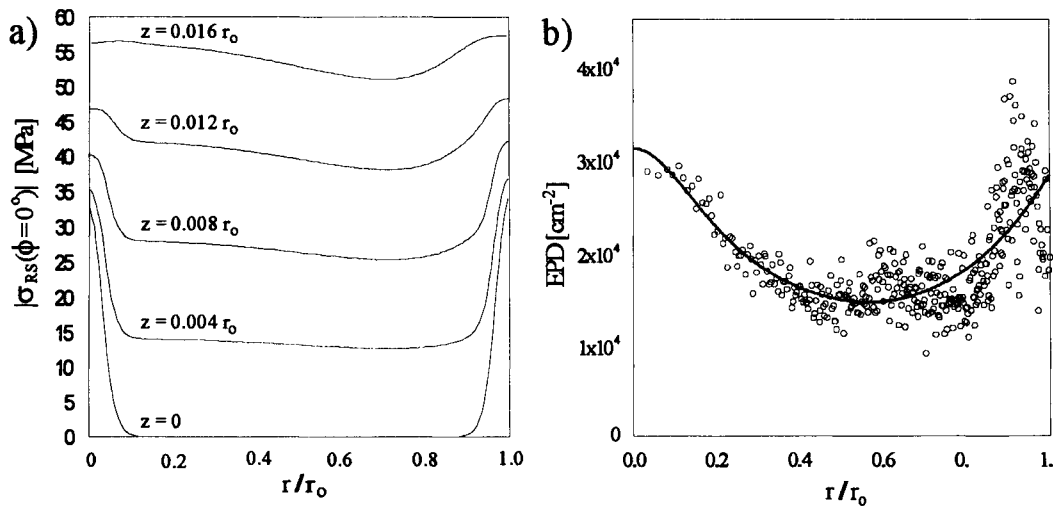
Figure 13. a) radial dependence of the resolved shear stress in the (0001) <11$\underline{2}$0> glide direction as a function of distance from the growth interface (z=0); b) radial variation of the local etch pit density within arears of 949 µm x 714 µm on a 6H-SiC wafer.

decrease the micropipe density over the past several years. The analysis of KOH-etched 4H-SiC wafers from low micropipe density boules has revealed areas up to 25-mm in diameter that are entirely free of micropipes. Figure 10 corresponds to a digitized image of a 50-mm diameter 4H-SiC KOH-etched wafer with a total of 18 micropipes corresponding to an overall density of 0.9 pipes/cm^2.

Concomitant with the growth of large diameter crystals grown under non-optimized process conditions is the tendency of the crystal to form low-angle boundaries near the crystal periphery. In SiC substrates, low angle boundaries are visible as void-like linear crystallographic features extending radially inward from the wafer edge and generally following low-index planes (see Figure 11a). They can sometimes extend through the entire thickness of the wafer. Recent work has resulted in a dramatic reduction in these defects to < 3 with a length of <3 mm each for production 50-mm wafers, while they are totally eliminated in our current R&D substrates up to diameters of 100-mm (see Figure 11b). The significant quality

improvement of our 4H- and 6H-SiC 3-inch (75-mm) wafers is visible in overall reduction of stress variation in recent wafers relative to earlier 75-mm diam. wafers, as demonstrated in Figure 12.

3.2. Thermally induced stress and dislocation generation

Besides the elimination of micropipes the overall reduction of dislocation densities in SiC substrates is a major technological goal. While typical etch pit densities (EPD) in SiC wafers range from 10^3-10^4 cm^{-2}, dislocation densities as low as 10 cm^{-2} have been observed in small Lely platelets [30], clearly indicating the room for technological improvement. Generally the following mechanisms are considered to be dominant for the generation of dislocations during crystal growth [31]: 1) nonuniform heat flow during solidification and ensuing thermal stress causing plastic deformation; 2) dislocation propagation and multiplication from an imperfect seed; 3) condensation of excess point defects near the growth temperature to form prismatic dislocation loops. Based on an axisymmetrical
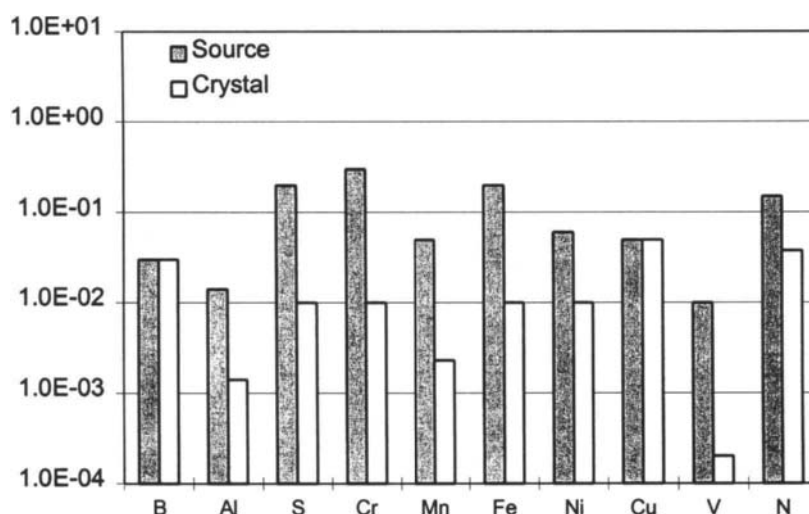
Figure 14. Impurity concentration (ppmwt) for SiC sublimation source material and crystal grown from that source vs principal elements observed in bulk SiC.

global heat transfer model, the thermal profile inside the growth chamber and the thermally induced stress in the growing SiC crystal was calculated (for details see [32]). The calculated resolved shear stress in the (0001) <$11\underline{2}0$> glide direction (Figure 13a), relevant for the primary slip system in SiC, significantly exceeds the reported values for the critical resolved shear stress for plastic deformation in SiC [33]. The radial variation shows a W-shape over the crystal diameter, which is also reflected in the radial variation of local etch pit densities found on SiC wafers (Figure 13b). These results strongly suggest that thermal stress plays a key role in the generation of dislocations during SiC growth.

4. CRYSTAL PURITY

4.1. Electrical and thermal effects

As with any semiconductor material, purity and its impact on electrical and thermo-physical properties are very important for SiC boules and substrates. Nitrogen and aluminum are the main n- and p-type dopants,

respectively, used for doping SiC boules because they create relatively shallow donor and acceptor levels within the SiC bandgap. Our results do not show any distinctions for doping processes between 6H and 4H crystals. However, as reported by a number of workers [23,34], there is a significant difference for N and Al incorporation in sublimation boule growth for the crystal faces (0001)Si and (0001)C. For nitrogen there is an increase in incorporation from (0001) Si to (0001)C while for Al it is the reverse. Nitrogen incorporation is approximately a function of the square root of the N_2 partial pressure in the growth environment [4], and for Al the dependence is roughly linear with Al concentration up to the 10^{20} cm^{-3} range. Nitrogen incorporation decreases with growth temperature while Al increases.

The maximum attainable resistivity in undoped n-type crystals is limited principally by residual nitrogen in the growth environment, while that for undoped p-type crystals is limited by boron. Reduction of boron content in SiC is particularly difficult

Table 3
Thermal Conductivity (T= 300 K)

Crystal	Direction	Carrier (cm-3)	K_{th} (W/cm K)
4H-n-type	a	5.0×10^{15}	4.9
4H n-type	c	5.0×10^{15}	3.9
4H n-type	c	2.0×10^{18}	3.3
4H s. i.	a	--	4.7
4H s. i.	c	--	3.7
6H n-type	a	3.5×10^{17}	3.8
6H n-type	c	3.5×10^{17}	3.2
6H n-type	c	1.5×10^{18}	3.0
6H n-type	a	1.4×10^{16}	4.0

since it exhibits a transfer coefficient of near unity [35]. The extent to which these impurities dominate the electrical behavior of the undoped SiC crystals grown by PVT is largely a function of the purity of the source materials and the cleanliness of the growth system. Figure 14 shows the results of quantitative analysis of source material and as-grown crystals based on secondary-ion mass spectroscopy. These results show that concentrations of nitrogen and boron in bulk crystals can be reduced to mid 10^{15} cm^{-3}. In material of the highest purity, elemental metallic impurities typically exhibit concentrations less than or equal to 0.01 ppmwt and transfer coefficients less than unity.

For heat dissipation in high-power density device structures, the thermal conductivity of the substrate is a critical parameter. Our data (Table 3) indicates that at sufficiently low overall impurity content where isotope scattering can be neglected, the thermal conductivity in bulk SiC shows a significant dependence on the free carrier concentration (electron-phonon scattering [36]). Thus, in bulk crystals exhibiting low impurity content, e.g. showing semi-insulating (s.i.) behavior, the thermal conductivity approaches 5 W/cmK for heat flow directions perpendicular to the c-axis. This value is close to that determined by Slack for Lely-grown samples [37]. However,

an appreciable reduction in heat conduction parallel to the c-direction is also observed, consistent with previous results [38].

The availability of semi-insulating (s.i.) 4H-SiC substrates removes the final impediment for development of SiC microwave devices and MMIC circuits. Cree has developed a process for producing very high resistivity 2-inch 4H-SiC wafers, typically using a combination of deep-level dopants and/or intrinsic point defects. From resistivity measurements at high temperatures, high resistivity SiC substrates fabricated by intentional deep-level incorporation exhibit an activation energy of 1.7 eV [39], which is consistent with deep-level compensation in 4H-SiC with a bandgap of 3.2eV. The extrapolation of this data to room temperature indicates a resistivity of 10^{20} ohm-cm. In an alternative process, high resistivity behavior has been achieved at Cree in 4H-SiC without using intentional doping with deep-level dopants. For these substrates an activation energy of approximately 1.1 eV has been extracted from high temperature Hall-effect measurements [40]. The corresponding thermal conductivity of semi-insulating 4H-SiC was measured as 4.7 W/cmK at 300K (a-direction).

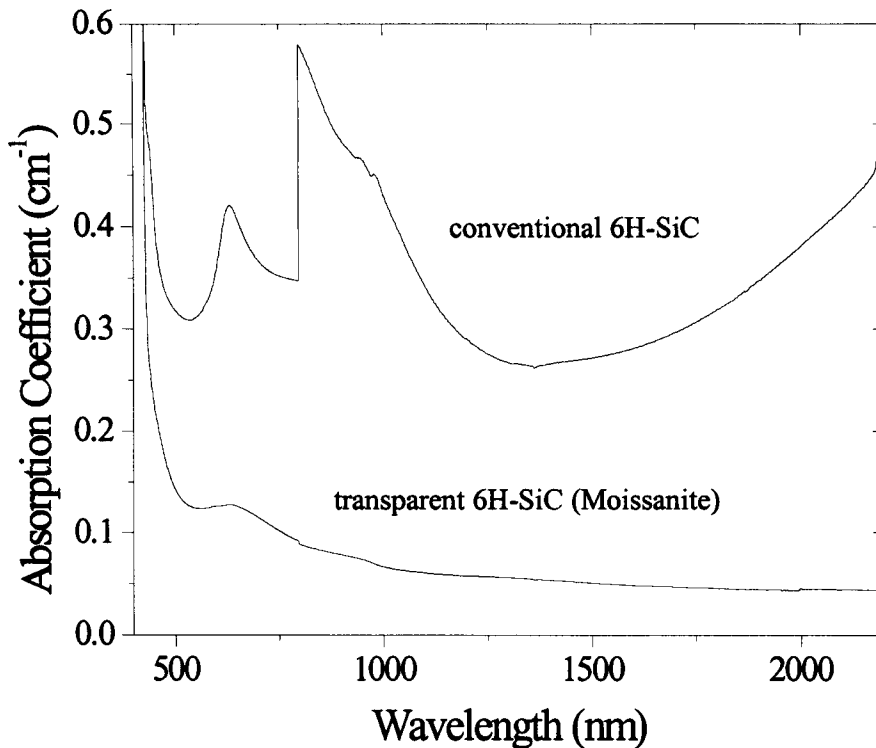## 4.2. Optically transparent SiC

Figure 15 Absorption spectra at 293 K comparing conventional 6H-SiC (upper curve) and near colorless Moissanite (lower curve).

Up to the present, SiC crystal growth has focused almost exclusively on the fabrication of SiC crystals targeted toward electronic devices. However, additional interest in silicon carbide has arisen recently for purely optical applications. This interest stems from the similarities in refractive index, dispersion, mechanical hardness, and thermal conductivity, between silicon carbide and diamond, which suggest the utility of silicon carbide as a unique gemstone material (moissanite). Moissanite, like diamond, is characterized by its broad transparency over the visible spectrum, optical brilliance, and resistance to chemical and abrasive attack. Recent advances at Cree have allowed the fabrication of monocrystalline moissanite in a near colorless form at crystal diameters up to 75-mm. Figure 15 shows absorption spectra for near colorless moissanite compared to conventional 6H-SiC. Conventional 6H-SiC exhibits large absorption bands in the visible region at energies below the band edge. These absorption bands give rise to the characteristic green color observed in n-type 6H-SiC [41]. The origin of the bands is attributed to interband transitions from states near the conduction band edge to states in higher lying conduction bands [42]. [The large offset in the absorption spectrum (upper curve) near 800 nm is an artifact associated with a detector change in the spectrometer at this wavelength.] By contrast, near colorless moissanite exhibits a broad transparency associated with a relatively featureless absorption spectrum.

In view of the potentially large materials volumes involved in the manufacture of bulk SiC crystals for the gemstone industry, semiconductor SiC for the electronics industry is expected to benefit greatly from this parallel

development of high quality SiC crystals.

5. SUMMARY

Recent progress in the development of the PVT technique for SiC bulk growth at our laboratories has led to substrate diameters up to 100-mm, residual impurities in the 10^{15} cm^{-3} range, thermal conductivity approaching 5.0 W/cmK in bulk crystals, transparent 6H and 4H-SiC at crystal diameters up to 75-mm, and micropipe densities as low as 0.9 cm^{-2} over a 50-mm diameter 4H-SiC wafer. These advances help to position SiC for an exciting future and provide a sound foundation for the realization of the full potential of SiC for high power density electronic devices, optoelectronic devices of high brightness, and SiC materials applications requiring low optical absorption.

ACKNOWLEDGEMENTS

REFERENCES

1. C.H. Carter, Jr., V.F. Tsvetkov, R.C. Glass, D. Henshall, M. Brady, St. G. Müller, O. Kordina, K. Irvine, J.A. Edmond, H.-S. Kong, R. Singh, S.T. Allen, and J. W. Palmour, Matls. Sci. and Eng., B61-62 (1999) 1.
2. see for example, V. Orfino, A. Blanco, V. Mannella, E. Bussoletti, and S. Fonti, Astronomy and Astrophysics, 252 (1991) 315.
3. E.G. Acheson, U. S. Patent No. 17911 (1892).
4. J.A. Lely, Ber. Dt. Keram. Ges., 32 (1955) 299.
5. Yu. M. Tairov and V.F. Tsevtkov, J. Crystal Growth, 43 (1978) 209.
6. R.F.Davis, C.H. Carter, Jr., and C.E. Hunter, U. S. Patent No.Re 34,861 (February 14, 1995).
7. R.A. Stein and P. Lanig, Mat. Sci. Eng. B, 11 (1992) 69.
8. D.L.Barrett, J.P. McHugh, H. McD. Hobgood, R.H. Hopkins, P.G. McMullin, and R.C. Clarke, J. Crystal Growth, 128 (1993) 358.
9. H. McD. Hobgood, M. Brady, W. Brixius, G. Fechko, R. Glass, D. Henshall, J. Jenny, R. Leonard, D. Malta, St. G. Mueller, V. Tsvetkov, and C. H. Carter,Jr., Matls. Sci. Forum, 338-342 (2000) 3.
10. R.I. Scace and G.A. Slack, J. Chem. Phys., 30 (1959) 1551.
11. V.F. Tsvetkov, S.T. Allen, H. S. Kong, C.H. Carter,Jr., Inst. Phys. Conf. Ser. No. 142 (1996).
12. B.M. Epelbaum, D. Hofmann, M. Mueller, and A. Winnacker, Matl. Sci. Forum, 338-342 (2000) 107.
13. J. Drowart, G. DeMaria, and M.G. Inghram, J. Appl. Phys., 29 (1058) 1015.
14. D.I. Cherednichenko, Y.I Khlebnikov, I.I. Khlebnikov, S. I. Soloviev, and T.S. Sudershan, Matl. Sci. Forum, 338-342 (2000) 35.
15. R.H. Hopkins, G. Augustine, H. McD. Hobgood, U. S. Patent No. 5,873,937 (Feb. 23, 1999).
16. see for example, M. Selder, L. Kadinski, F. Durst, T. Straubinger, D. Hofmann, and P. Wellmann, Matl. Sci. Forum, 338-342 (2000) 31.
17. St. G. Mueller, Herstellung von Siliziumkarbid im Sublimationsverfahren, Shaker Verlag, Aachen (1998) p. 90.
18. A.R. Verma and P. Krishna, Polymorphism and Polytypism in Crystals, Wiley, New York (1966).
19. F. Bechstedt, P. Kaeckell, A. Zywietz, K. Karch, B. Adolph, K. Tenelsen, and J. Furthmueller, phys. stat. sol. (b), 202 (1997) 35.
20. W.F. Knippenberg, Phillips Research Repts., 18 (1963) 161.

21. D.P. Birnie and W.D. Kingery, J. Matl. Sci., 25 (1990) 2827.

22. M Kanaya, J. Takhashi, Y. Fujiwara, and A. Moritani, Appl. Phys. Lett., 58 (1991) 56.

23. G. Augustine, H. McD. Hobgood, V. Balakrishna, G. Dunne, and R. H. Hopkins, phys. stat. sol. (b), 202 (1997) 137.

24. R.A. Stein and P. Lanig, J. Cryst. Growth, 131 (1993) 71.

25. T. Straubinger, M. Bickermann, D. Hofmann, R. Weingaertner, P.J. Wellmann, and A. Winnacker, Matl. Sci. Forum, 353-356 (2001) 25.

26. F.C. Frank, Acta Cryst., 4 (1950) 497.

27. W. Si, M. Dudley, R. Glass, V. Tsvetkov C.H. Carter, Jr., J. Elect. Matls., 26 (1997) 128.

28. X.R. Huang, M. Dudley, W.M. Vetter, W. Huang, W. Si, and C. H. Carter, Jr., J. Appl. Crystallography, 32 (1999) 516.

29. W.M. Vetter and M. Dudley, J. Matl. Res., 15 (2000) 1649.

30. M. Tuominen, E. Prier, R. Yakimova, R.C. Glass, T. Toumi, and E. Janzen, Inst. Phys. Conf. Ser. 142 (1996) 409.

31. B. Mutaftschiev, Dissociation in Solids, Vol. 5, North-Holland, Amsterdam, 1983.

32. St. G. Mueller, R.C. Glass, H.McD. Hobgood, V.F. Tsvetkov, M. Brady, D. Henshall, J.R. Jenny, D. Malta, and C.H. Carter, Jr., J. Cryst. Growth, 211 (2000) 325.

33. A.V. Samant and P. Pirouz, Int. J. Ref. Metals Hard Mater., 16 (1998) 277.

34. V.I. Levin , G.I. Pozdnyakova, Yu. F. Tsvetkov, and Yu. M. Shashkov, Inorg. Matl., 13 (1977) 212.

35. R.C. Glass, G. Augustine, V. Balakrishna, H. McD. Hobgood, R.H. Hopkins, J. Jenny, M. Skowronski, and W.J. Choyke, Inst. Phys. Conf. Ser. No. 142 (1996) 37.

36. D.T. Morelli, J.P. Heremans, C.P. Beetz, W.S. Yoo, and H. Matsunami, Appl. Phys. Lett., 63 (1993) 3145.

37. G.A. Slack, J. Phys. Chem. Solids 34 (1973) 321.

38. L.A. Burgemeister, W. von Muench, and E. Pettenpaul, J. Appl. Phys., 50 (1979) 5790.

39. V.F. Tsvetkov, S.T. Allen, H.S. Kong, and C.H. Carter, Jr., Inst. Phys. Conf. Ser. No. 142, (1996) 17.

40. W.C. Mitchel, A. Saxler, R. Perrin, J. Goldstein, S.R. Smith, A.O. Evwaraye, J.S. Solomon, M. Brady, V. Tsvetkov, and C.H. Carter, Jr. Matl. Sci. Forum, 338-342 (2000) 21.

41. E. Biedermann, Solid State Comm., 3 (1965) 343.

42. W.R.L. Lambrecht, S. Limpijumnong, S.N. Raskeev, and B. Segall, Matl. Sci. Forum, 264-268 (1998) 271.

Crystal growth of SiC II. Epitaxial growth

H. Matsunami[a]* and T. Kimoto[a]

[a]Department of Electronic Science and Engineering, Kyoto University, Kyoto 606-8501, Japan

Chemical vapor deposition (CVD) of silicon carbide (SiC) on SiC substrates is reviewed. Polytype-controlled epitaxial growth of SiC, which utilizes step-flow growth on off-axis SiC {0001} substrates (*step-controlled epitaxy*), is demonstrated, and the growth mechanism is discussed. In step-controlled epitaxy, SiC growth is controlled by the diffusion of chemical reactants in a stagnant layer. Critical growth conditions where the growth mode changes from step-flow to two-dimensional nucleation are predicted as a function of growth conditions using a model describing SiC growth on vicinal {0001} substrates. Step bunching on the surfaces of SiC epilayers, nucleation, and step motion are investigated. The high quality of SiC epilayers is elucidated through photoluminescence and Hall-effect measurements. Excellent doping controllability in the wide range has been obtained by *in situ* doping of a nitrogen donor and aluminum/boron acceptors. Recent progress in SiC epitaxial growth is also presented.

1. INTRODUCTION

Silicon carbide (SiC) is the most famous material to show polytypes with various stacking sequences, but the same chemical composition. The polytypes are represented by the number of layers in the unit cell and the crystal system (C for cubic, H for hexagonal, and R for rhombohedral). In general, 3C-SiC is known as a low-temperature polytype, and 6H-SiC and 4H-SiC are high-temperature polytypes. It is very crucial to control polytypes during crystal growth of SiC for real applications. The wide band gaps of SiC give a very high breakdown field, about ten times higher than that of Si or GaAs [1]. Strong Si-C bonding yields optical phonons with high energy, as high as $100{\sim}120$ meV [2], which leads to a high electron saturation drift velocity (2×10^7 cm/s in 6H-SiC) [3] and a high thermal conductivity (4.9 W/K-cm) [4].

Controllable n- and p-type doping can be done during crystal growth. Selective doping of both donor- and acceptor-type impurities can be achieved by ion implantation. Besides, SiC is the only compound semiconductor that can be thermally oxidized to grow high-quality SiO_2. These

outstanding properties make SiC a real candidate for high-power, high-frequency, high-temperature and radiation-resistant devices, which cannot be realized with Si or GaAs. In particular, a theoretical simulation has predicted that SiC power switching devices will replace present-day Si power devices on account of extremely low power dissipation and reduced chip sizes [5].

However, the difficulty in growth of high-quality and large SiC crystals without polytype mixing had prevented electronic applications, though the potential of SiC has been recognized for a long time. An innovation in bulk crystal growth that uses a seeded sublimation method (a modified Lely method), by which large 6H-SiC and 4H-SiC single crystals with controlled polytypes can be grown [6], has changed research and development of SiC devices.

As for epitaxial growth of SiC, homoepitaxial growth of 6H-SiC has been performed by liquid phase epitaxy (LPE) and chemical vapor deposition (CVD) for blue light-emitting diodes [7]. Although CVD principally has excellent control in uniformity of thickness and impurity doping, there has been a serious problem of polytype mixing. In case of growth on 6H-SiC {0001} basal planes, a temperature as high as 1800°C was required for homoepitaxy of 6H-SiC, and 3C-SiC twin crystals were usually grown at low

temperatures [8-10]. The authors' group developed a new technique, in which single crystalline 6H-SiC can be homoepitaxially grown reproducibly on off-axis 6H-SiC {0001} at low temperatures of 1400~1500°C [11,12]. This technique was named step-controlled epitaxy, because surface steps formed by bringing off-axis on substrates replicate the epilayer polytype same as the substrate polytype. This technique was epoch-making for two reasons: (i) growth temperature can be reduced by more than 300°C, and (ii) epilayers have a quality high enough for device applications. The growth of high-quality SiC on off-axis SiC {0001} substrates has been followed by other groups [13-18]. Rapid progress of recent SiC device fabrication has been supported by step-controlled epitaxy, key technology for the research and development of high-performance SiC devices.

In this paper, step-controlled epitaxy of SiC is reviewed [19]. The growth mechanism of SiC, such as the growth modes, rate-determining processes, and surface kinetics, is described. Optical and electrical properties of undoped SiC epilayers are discussed. Impurity doping during growth is shown for real semiconductor technology. Recent progress in epitaxial growth of SiC is presented [20].

2. STEP-CONTROLLED EPITAXY OF SiC

2.1. Chemical vapor deposition

Homoepitaxial growth of SiC has been carried out by atmospheric-pressure CVD in a horizontal cold-wall reaction tube. Source gases of SiH_4 (1% in H_2) and C_3H_8 (1% in H_2) were carried by purified H_2 gas. The flow rates of SiH_4 and C_3H_8 were typically 0.30 sccm and 0.20 sccm, respectively. The H_2 flow rate of 3.0 slm provides a linear gas velocity of 6~10 cm/s. N_2 was used for n-type doping, and trimethylaluminum (TMA) or B_2H_6 for p-type doping. Hydrogen chloride (HCl) gas was used for etching of the substrate surfaces before CVD growth.

Two kinds of substrates have been used, crystals grown by the Acheson method and a modified Lely (sublimation) method. Because Acheson crystals have naturally-grown {0001} basal planes (perpendicular to c-axis), off-axis substrates were prepared by angle-lapping. As for crystals grown by a modified Lely method, both commercial and home-made wafers were used. The off-orientation was 0~10° toward $\langle 11\bar{2}0 \rangle$. Both (0001)Si and $(000\bar{1})$C faces were used to investigate substrate polarity effects. The surface polarity was identified using the difference in oxidation rates between both faces. $((000\bar{1})$C faces have faster oxidation [21,22].) The polytype of the substrates was roughly identified by their absorption edges in ultraviolet to visible-light transmission spectra and photoluminescence. Their identification was confirmed by X-ray diffraction and Raman scattering.

Substrates were heated by radio-frequency induction on a SiC-coated graphite susceptor. Before CVD growth, in situ HCl etching was carried out to remove surface damage introduced by polishing processes. The growth temperature was varied in the range of 1100~1600°C (typically 1500°C).

2.2. Step-controlled epitaxy

Figure 1 shows the surface morphology of as-grown epilayers, their reflection high-energy electron diffraction (RHEED) patterns ($\langle 11\bar{2}0 \rangle$-azimuth), and their etched surface morphology with molten KOH [19]. These epilayers were grown at 1500°C on on-axis (just) and off-axis (off) 6H-SiC (0001) substrates under a typical gas flow condition. Here, the off-angle is 6° toward $\langle 11\bar{2}0 \rangle$, and the thickness of the epilayers is about 5 μm. The epilayer on the on-axis (0001)Si face shows a smooth surface with a mosaic pattern having several domains separated by step- or groove-like boundaries. The RHEED pattern identifies the grown layer as 3C-SiC (111) with twinning. Besides, triangular etch pits with 3-fold symmetry indicate the growth of the cubic phase. The etch pits are 180°-rotated relative to each other across the groove-like boundaries, which means neighboring domains separated by the boundaries have the twin relationship that is called double-positioning twin [23]. On the on-axis $(000\bar{1})$C face, many island structures were observed, and the grown surface was
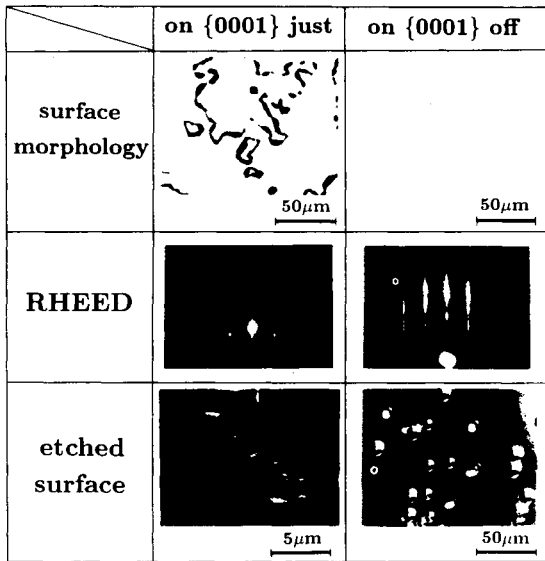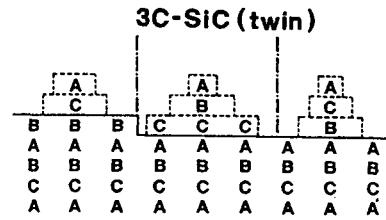
Figure 1. Surface morphology, RHEED pattern, and etched surface of layers grown on on-axis (just) and off-axis (off, 6°) 6H-SiC (0001)Si substrates [19].
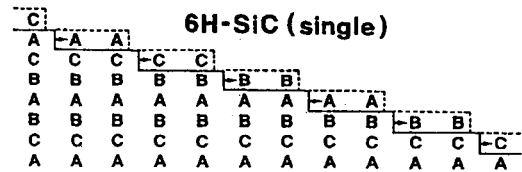

Figure 2. The relationship between growth modes and polytypes of layers grown on 6H-SiC. (a) 3C-SiC by two-dimensional nucleation, and (b) homoepitaxy of 6H-SiC by step-flow growth [19].

rough. The RHEED pattern revealed the grown layer as twinned crystalline 3C-SiC.

In contrast, the epilayers on the off-axis (0001)Si face exhibit a specular smooth surface. The RHEED pattern shows the growth of single crystalline 6H-SiC (0001). The polytype of grown layers was also identified by transmission electron microscope (TEM) observation and photoluminescence. Thus, homoepitaxial layers with excellent surface morphology can be obtained. Also homoepitaxial layers were reproducibly obtained on the off-axis (000$\bar{1}$)C face. This fact is of great interest, because the growth on on-axis (000$\bar{1}$)C faces yields very poor morphology due to island-like growth.

Crystal growth on off-axis 6H-SiC (0001)Si substrates with different off-directions has been studied [13,24]. On a 6H-SiC (0001)Si face with a low off-angle inclined toward $\langle 1\bar{1}00\rangle$, stripe-like morphology appeared, which is caused by pronounced step bunching, and the inclusion of 3C-SiC domains was observed after long-time growth.

Thereafter, epitaxial growth on off-axis (0001)Si faces inclined toward $\langle 11\bar{2}0\rangle$ has been widely studied.

The features of crystal growth are schematically shown in Fig. 2 [19]. On a growing surface in CVD, adsorbed chemical species migrate and are incorporated into a crystal at steps or kinks where the surface energy is low. There exists, however, a competitive nucleation process which takes place on crystal terraces, when the supersaturation is high enough. On on-axis {0001} faces, the step density is very low and wide terraces exist. Therefore, crystal growth occurs mainly on terraces through two-dimensional nucleation under the high supersaturation condition at low temperatures. Thus, 3C-SiC crystals grow on wide terraces at 1500°C, because 3C-SiC is stable at low temperatures. The growth of 3C-SiC on 6H-SiC has been predicted by theoretical studies using a quantum-mechanical energy calculation [25] and an electrostatic model [26]. Because the stacking order of 6H-SiC is ABCACB..., the

growing 3C-SiC takes two possible stacking orders of ABCABC... and ACBACB..., which gives twinning as shown in Fig. 2(a). Here, A,B, and C denote the occupation sites in the hexagonal close packed structure.

On off-axis {0001} faces, the step density becomes high, and the terrace width is narrow enough for adsorbed chemical species to reach steps even at 1500°C. At steps, the incorporation site is uniquely determined by chemical bonds from the bottom and the side of substrate as shown in Fig. 2(b). Therefore, the same polytype as that of substrate can be grown through step-flow growth, inheriting the stacking order of substrate. This technique can be applied to homoepitaxy of any other polytypes such as 4H-, 15R-, and 21R-SiC.

Stable homoepitaxy without 3C-SiC inclusions was performed on off-axis substrates at 1600~2200°C by a sandwich growth method [27]. The improvement of 6H-SiC crystal quality using off-axis substrates has also been reported in LPE [28]. Thus, the use of off-axis substrates is a key technique to grow high-quality homoepitaxial layers at rather low temperatures in various SiC growth methods.

Epitaxial growth on off-axis (vicinal) substrates has been extensively studied for many materials. Though the use of off-axis substrates in step-controlled epitaxy of SiC seems similar to previous studies in other materials, this technique possesses a special meaning that the polytypes of grown SiC layers can be controlled by introducing steps on substrates. The surface steps serve as a template that enforces the replication of substrate polytype in the grown layers. This is why it is called step-controlled epitaxy.

Homoepitaxial growth of 6H-SiC was achieved on off-axis {0001} faces at temperatures as low as 1200°C. The growth temperature for 6H-SiC homoepitaxy can be reduced by more than 600°C. The surface morphology deteriorated by lowering the growth temperature: triangle-shaped pits and hillocks increased in the epilayers. Surface morphology became rough. In addition, twinned crystalline 3C-SiC was grown at 1100°C, even on off-axis substrates, which was attributed to the suppressed migration of adsorbed chemical
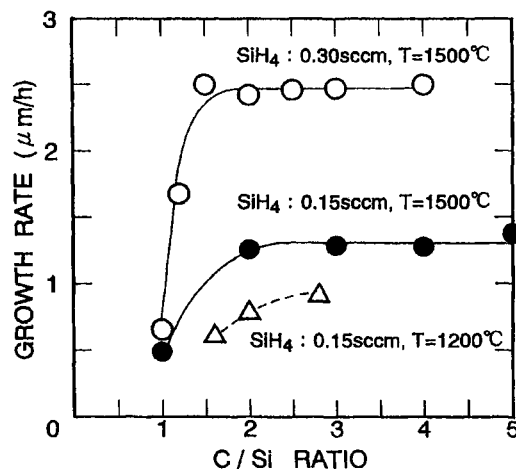
Figure 3. Dependence of the growth rate on the C/Si ratio. Closed and open circles are results for SiH_4 flow rates of 0.15 sccm and 0.30 sccm at 1500°C, and open triangles at 1200°C. Substrates are 6° off-axis 6H-SiC $(000\bar{1})C$ faces [19].

species and the occurrence of two-dimensional nucleation on terraces. Further reduction in the SiH_4 flow rate may lead to 6H-SiC homoepitaxial growth below 1100°C, because two-dimensional nucleation on terraces can be suppressed by lowering supersaturation. In fact, homoepitaxial growth of 6H-SiC was carried out at 1050°C on off-axis 6H-SiC (0001)Si faces by gas source molecular beam epitaxy (MBE) with a very low growth rate of 50~100 Å/h [29].

3. GROWTH MECHANISM OF STEP-CONTROLLED EPITAXY

3.1. Rate-determining process

The growth rate depends on the C/Si ratio as shown in Fig. 3 [19]. Substrates were 6H-SiC {0001} with a 6° off-angle toward ⟨11$\bar{2}$0⟩, and growth was done at 1200°C and 1500°C. The flow rate of C_3H_8 was varied in the range of 0.10~0.40 sccm by keeping the SiH_4 flow rate at 0.15 sccm or 0.30 sccm. In Fig. 3, closed and open circles are the growth rates when the SiH_4 flow rates were 0.15 sccm and 0.30 sccm, respectively at 1500°C [30]. In the region of C/Si > 1.4, the growth rate

has an almost constant value, and it increases proportionally with the flow rate of SiH_4, indicating the rate determining by the supply of SiH_4. In an early study of SiC growth, even on on-axis {0001} faces using SiH_4 and C_3H_8 at 1500~1700°C, the growth rate of SiC was controlled by the supply of SiH_4 [31]. In the region of C/Si < 1.4, the growth rate decreases remarkably, probably due to the lack of C sources. At 1200°C, the growth rate starts to saturate at a higher C/Si ratio, as shown by open triangles in Fig. 3, due to insufficient decomposition of C_3H_8.

Through studies on gas-phase and surface reactions at 1200~1600°C in a SiH_4-C_3H_8-H_2 system [32] and gas-phase kinetics [33], the dominant chemical species contributing to SiC growth are revealed as Si, SiH_2, Si_2H_2 from SiH_4, and CH_4, C_2H_2, C_2H_4 from C_3H_8. In the growth of SiC, Si (or SiH_2) may be preferentially adsorbed and migrate on the surface and make chemical bonds with substrates. In fact, no deposition occurs without SiH_4 supply.

Surface morphology depends on the C/Si ratio. Epilayers grown at 1500°C on (0001)Si faces showed specular smooth surfaces for the C/Si ratios between 2 and 6. On (000$\bar{1}$)C faces, the optimum window of C/Si ratio to obtain specular smooth surfaces is from 2 to 3, which is relatively narrow. Surface morphology is very sensitive to substrate defects and polishing damage. Triangular pits and macrosteps are easily formed at defect sites, and often 3C-SiC nucleation occurs on most triangular pits [34]. Defects in surface morphology can be suppressed by reducing surface damage and improving CVD procedures [35]. It is noteworthy that these defects appear only on (0001)Si faces and epilayers on (000$\bar{1}$)C faces are very flat by using the optimum window of C/Si ratio.

The growth rates of epilayers grown at 1500°C on 6H-SiC substrates with various off-angles from {0001} are shown in Fig. 4 [19]. The flow rates of SiH_4 and C_3H_8 are 0.30 sccm and 0.20 sccm, respectively. Open (on a (0001)Si face) and closed ((000$\bar{1}$)C face) triangles indicate the growth of 3C-SiC, and open and closed circles indicate the growth of 6H-SiC. Homoepitaxial growth of 6H-SiC can be done on off-axis substrates with more
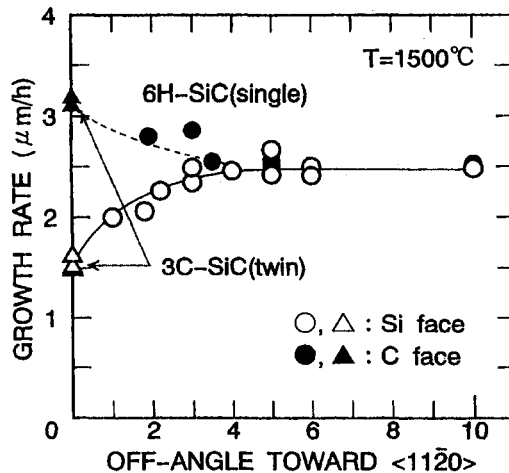
Figure 4. Dependence of the growth rate at 1500°C on the off-angle from {0001}. The flow rates of SiH_4 and C_3H_8 are 0.30 sccm and 0.20 sccm, respectively. Open and closed triangles show the growth of 3C-SiC, and circles denote that of 6H-SiC [19].

than 1°. On on-axis substrates, higher growth rates are obtained on (000$\bar{1}$)C faces, as have been reported [36]. This is ascribed to the higher nucleation rate on (000$\bar{1}$)C terraces, as described later. With increasing off-angle, the growth rates on both faces become almost the same value (in this case, 2.5 μm/h) for off-angles from 4° to 10°.

In Fig. 5, the temperature dependence of the growth rate is shown in the range between 1200°C and 1600°C, in which homoepitaxial growth of 6H-SiC occurs [19]. The flow rates of SiH_4 and C_3H_8 are 0.15 sccm and 0.10~0.14 sccm, respectively, and the off-angle of the substrates is 5~6°. There is little difference between the growth rates on (0001)Si and (000$\bar{1}$)C faces even at low temperatures. The data of growth rate yield an Arrhenius-type curve with a small activation energy of 2.8 kcal/mole. In early CVD growth on on-axis {0001} faces, activation energies of 12 [36], 20 [8] and 22 kcal/mole [30] were reported. Similar insensitivity of growth rate to temperature in CVD was reported on 6H-SiC substrates with 2° off-angle toward ⟨1$\bar{1}$00⟩ [37]. The near in-
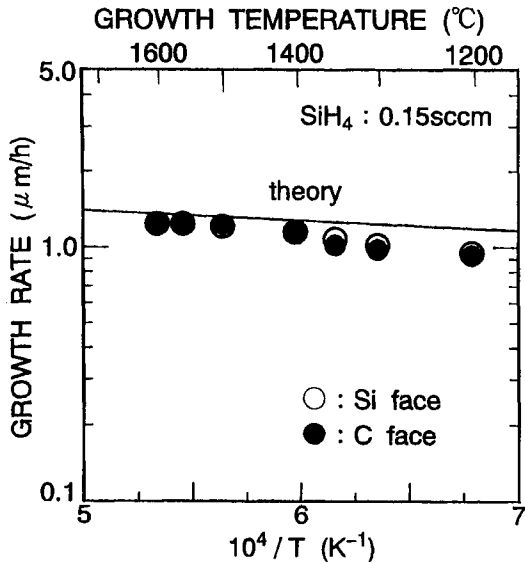
GROWTH TEMPERATURE (°C)

Figure 5. Temperature dependence of growth rate on 6° off-axis 6H-SiC (0001)Si and (0001̄)C faces. The flow rates of SiH_4 and C_3H_8 are 0.15 sccm and 0.10~0.14 sccm, respectively. The solid curve shows the calculated result based on a stagnant-layer model [19].

dependence on polarity and small activation energy are notable features caused by the introduction of off-angle.

Because the growth rate of SiC is determined by the supply of Si species as shown in Fig. 3, a stagnant-layer model [38] can be applied to the present system with some modification. The model has been developed to improve the uniformity of Si epilayer thickness based on experiments of gas flow patterns. In the model, the growth is governed by the diffusion of chemical species from the main gas stream to a crystal surface. The growth rate at sample position x_0 (2 cm from the front of the susceptor) is given by

$$R(x_0) = \frac{h_0 D_0 T_s P_0}{n_0 k T_0^2 \delta(x_0)} \exp(-\frac{D_0 T_s}{T_0 v_0 b} \int_0^{x_0} \sqrt{\frac{v_0 \rho}{\mu x}} dx), (1)$$

where D_0 is the diffusion coefficient of SiH_4 in H_2 at 300 K (0.20 cm^2/s) [38], P_0 the partial pressure of SiH_4 at the inlet (4.95×10^{-5} atm), and v_0 the mean gas velocity under a non-heated condition

(10 cm/s). T_s is the substrate temperature, T_0 = 300 K, k the Boltzmann constant, b the free height of the reactor (2.0 cm), h_0 the spacing of the 6H-SiC {0001} face (0.252 nm), and n_0 the density of adsorption sites on the surface (1.21 × 10^{15} cm^{-2}). ρ (8.13×10^{-5} g/cm^3) and μ (8.7 × 10^{-5} P) [39] are the gas density and gas viscosity, respectively. $\delta(x_0)$ is the stagnant layer thickness (0.51 cm). Here, the condensation coefficient, the ratio of the number of atoms incorporated into a growing layer to that arriving on a surface, is assumed as unity.

The temperature dependence of growth rate calculated using the model is shown by a solid curve in Fig. 5. The absolute value of growth rate calculated from the simple model shows surprisingly good agreement with experimental data. The growth rate increases gradually with temperature increase owing to, mainly, the enhancement of diffusion in a stagnant layer. Although the calculated curve is not activation-type, the curve in the range of 1200~1600°C yields an apparent activation energy of 2.4 kcal/mole, which is very good agreement with the experimental result. Therefore, the SiC growth in step-controlled epitaxy would be limited by mass transport. This fact explains why there is little difference in the growth rates between (0001)Si and (0001̄)C faces because no polarity effect appears in the growth controlled by mass transport [30].

3.2. Prediction of step-flow growth condition

In order to predict the condition for step-flow growth, a simple surface diffusion model based on the BCF (Burton, Cabrera, and Frank) theory [40] is considered, where steps with height h are separated by equal distance λ_0 as shown in Fig. 6 [19]. Here, J denotes the flux of chemical reactants arriving on a surface. Adsorbed chemical species diffuse (migrate) on terraces, and some of them are incorporated into a crystal at steps, and others re-evaporate. As an assumption, nucleation does not occur on terraces. Steps are assumed as uniform and perfect sinks for incoming chemical species indicating that the capture probability of those species at steps is unity regardless the direction from which they approach
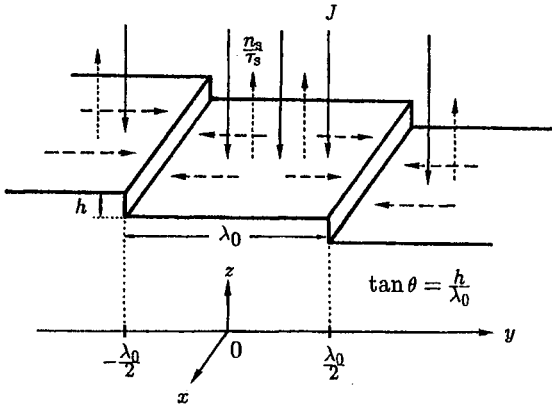
Figure 6. The surface diffusion model in which steps with height h are equally-spaced with separation distance λ_0 [19].

Figure 7. Temperature dependence of n_{s0}/τ_s for several reactions between Si and hydrocarbon species.

the steps. The continuity equation for chemical species is solved under the boundary condition that the supersaturation ratio α equals unity at steps. The α is defined as the ratio of $n_s(y)/n_{s0}$, where $n_s(y)$ is the number of chemical species per unit area on the surface and n_{s0} is that in equilibrium. The number of chemical species on the terraces is calculated based on an equation involving the flux of reactants arriving on the surface J, the mean residence time τ_s of chemical species, and surface diffusion length λ_s. The surface diffusion length is the average distance that chemical species migrate on a "step-free" surface before desorbing.

The spatial distribution of chemical species and the supersaturation ratio α on a surface can be obtained based on the above calculation. Since α takes maximum value α_{max} at the center of a terrace, nucleation most easily occurs at this position. The α_{max} strongly depends on experimental conditions such as the growth temperature, growth rate, and terrace width. This is an essential parameter which determines whether the growth mode is step-flow or two-dimensional nucleation. Thus, the growth modes are determined by the following relationship between α_{max} and α_{crit}:

$\alpha_{max} > \alpha_{crit}$: two-dimensional nucleation,
$\alpha_{max} < \alpha_{crit}$: step-flow.

Under the critical condition, α_{max} should be equal to α_{crit}, which brings the following relation:

$$\frac{\lambda_0}{4\lambda_s} \tanh(\frac{\lambda_0}{4\lambda_s}) = \frac{(\alpha_{crit} - 1)h}{2n_0 R} \frac{n_{s0}}{\tau_s}, \qquad (2)$$

where n_0 is the density of adatom sites on the surface and R is the growth rate. This is the basic equation describing the critical condition. Here, the growth rate R in step-flow growth was thought to be given by the product of the step velocity and $\tan\theta$ $(=h/\lambda_0)$, where θ is the substrate off-angle. Thus,

$$R = v_{step} \tan\theta = \frac{2h\lambda_s}{n_0\lambda_0}(J - \frac{n_{s0}}{\tau_s}) \tanh(\frac{\lambda_0}{2\lambda_s}) \qquad (3)$$

is given. In eq.(2), λ_0 and R are determined by growth conditions, and n_0 and h are inherent parameters of the material, SiC. If the values of n_{s0}/τ_s and α_{crit} are known, and several critical conditions are found experimentally, the surface diffusion length can be estimated from eq.(2). The value of n_{s0}/τ_s can be calculated from the equilibrium vapor pressure P_0 using Knudsen's
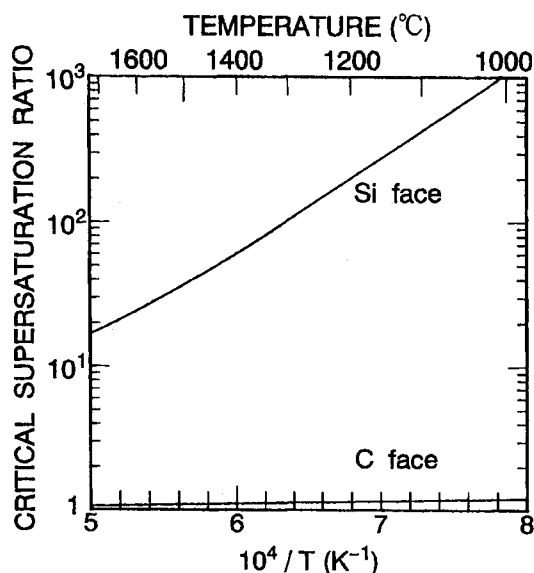
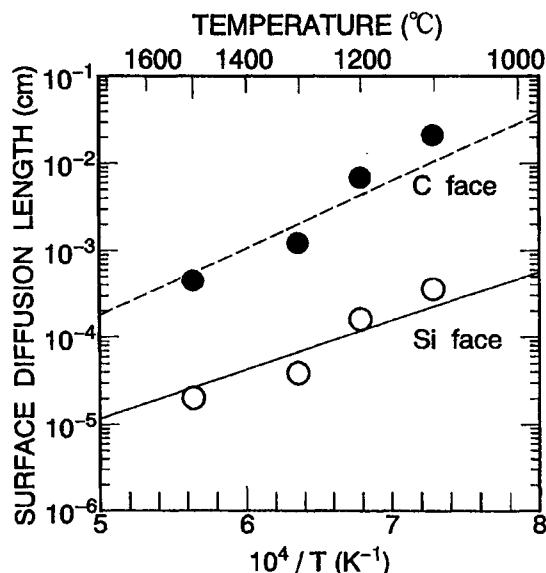Figure 8. Temperature dependence of critical supersaturation ratio for a disk-shaped nucleus [19].

Figure 9. Temperature dependence of surface diffusion length λ_s of adsorbed chemical species on (0001)Si and (000$\bar{1}$)C faces [19].

equation in the kinetic theory of gases [41], and P_0 from the chemical equilibrium constants of the reaction system. Because the supply of Si-related chemical species controls the growth of SiC, the equilibrium vapor pressure of Si-related species is assumed P_0 in SiC growth. The equilibrium vapor pressure of Si P_{Si} is obtained as a function of temperature from the equilibrium equations for several reactions between Si and hydrocarbon species [42]. The temperature dependence of n_{s0}/τ_s is shown in Fig. 7 [42].

Nucleation on terraces becomes dominant when α_{max} exceeds a critical supersaturation ratio α_{crit}. When the critical nucleation rate is assumed as 10^{12} cm^{-2}s^{-1}, which corresponds to one nucleation per unit time on a 10 nm$\times$10 nm area (the average terrace widths in this experiment are 3$\sim$30 nm), α_{crit} for a disk-shaped nucleus [43] is calculated as in Fig. 8 [19]. Here, the volume of a Si-C pair (2.07×10^{-23} cm^3), the bond length between Si and C atoms (0.252 nm), and the surface energy were necessary in the calculation.

Although no reports have been given for the surface energy of 6H-SiC {0001}, values of 2220

and 300 erg/cm^2 for 3C-SiC (111)Si and ($\bar{1}\bar{1}\bar{1}$)C faces, are given, respectively [44]. Because 6H-SiC (0001)Si and (000$\bar{1}$)C faces have similar atomic configurations as 3C-SiC (111)Si and ($\bar{1}\bar{1}\bar{1}$)C faces, those values were used in the calculation. The α_{crit} on (000$\bar{1}$)C faces takes quite low values (nearly unity), indicating that nucleation much frequently occurs on (000$\bar{1}$)C faces under a fixed supersaturation condition.

In the growth of SiC, polytypes of grown layers roughly reflect the growth modes, either step-flow or two-dimensional nucleation: Homoepitaxy of 6H-SiC on 6H-SiC is achieved owing to step-flow growth, whereas 3C-SiC is heteroepitaxially grown on 6H-SiC through two-dimensional nucleation. Then, the growth modes are distinguished by identifying the polytypes of grown layers. The critical conditions have been examined through CVD growth under various growth conditions by changing the growth temperature and off-angle in the range of 1100$\sim$1500°C and 0$\sim$10° (toward $\langle 11\bar{2}0 \rangle$), respectively. Some of the critical growth conditions at various growth temperatures are given in Table 1 [19]. Higher temper-

Table 1
Typical critical growth conditions obtained experimentally [19].

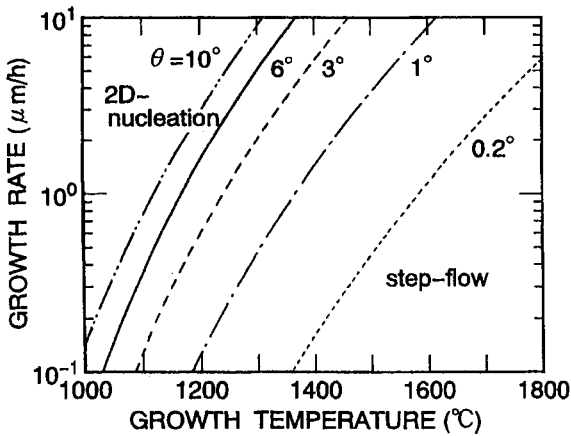| Growth temperature (°C)	Off-angle (deg)	Terrace width (nm)	Growth rate (μm/h)
1100	6	7.2	0.80
1200	3	14.4	0.95
1300	6	7.2	2.50
1500	1	43.3	2.00

Figure 10. Critical growth conditions as a function of growth temperature, growth rate, and substrate's off-angles (θ=0.2°, 1°, 3°, 6°, and 10°). The lower-right and upper-left regions of individual curves correspond to step-flow and two-dimensional nucleation conditions, respectively [19].

atures, larger off-angles, and lower growth rates are preferable to homoepitaxy of 6H-SiC (step-flow growth). Very little dependence of the critical conditions on the substrate polarity was observed.

In Fig. 9, the temperature dependence of surface diffusion length λ_s is given on both (0001)Si and (000$\bar{1}$)C faces calculated from eq.(2) using several data described above [19]. The terrace width λ_0 was calculated from $h/\tan\theta$, and h was assumed as 3 Si-C bilayer height (0.252 nm ×3 = 0.756 nm), based on the result of step

bunching described later. In the figure, solid and dotted lines denote fitting curves using a derived equation. The surface diffusion lengths decrease at high temperatures due to high desorption rates. The values on (000$\bar{1}$)C faces are longer than those on (0001)Si faces by more than one order of magnitude. Although nucleation easily occurs on (000$\bar{1}$)C faces, no significant difference appears in the surface morphology and critical growth conditions in the growth on off-axis substrates. The longer surface diffusion lengths on (000$\bar{1}$)C faces may compensate frequent nucleation on terraces.

The critical growth conditions given by eq.(2) can be predicted under various conditions using the temperature dependencies of n_{s0}/τ_s, α_{crit}, and λ_s. For example, if the growth temperature and off-angle of substrates (terrace width) are fixed, the critical growth rate (maximum growth rate for step-flow growth) can be calculated. For off-angles of 0.2°, 1°, 3°, 6°, and 10°, the critical growth conditions are given by curves shown in Fig. 10 [19]. In the figure, the lower-right and upper-left regions separated by individual curves correspond to step-flow growth and two-dimensional nucleation, respectively. Almost no difference exists in the critical growth conditions for (0001)Si and (000$\bar{1}$)C faces. At higher temperatures, even the higher growth rate and lower off-angle produce step-flow growth. At 1800°C, a very small off-angle of 0.2°, which yields almost the on-axis growth condition, is enough to achieve step-flow growth with a moderate growth rate of 6 μm/h. This may be one of the reasons why 6H-SiC can be homoepitaxially grown on on-axis substrates in early studies [8-10] when the temperature is raised up to 1700~1800°C. As a pre-

diction, large off-angles more than 5° are needed to realize homoepitaxy of 6H-SiC at a low temperature of 1200°C.

4. STEP KINETICS IN SiC EPITAXY

4.1. Step bunching

Step bunching (the formation of multiple-height steps) has been an attractive subject in crystal growth and surface science. Quite a few studies have been reported on the step structure of SiC surface [27, 45-49]. The step structure has been studied using an AFM (atomic force microscope) and a TEM (transmission electron microscope) for as-grown epilayers on SiC {0001} substrates with off-angles of 3~10° toward ⟨11$\bar{2}$0⟩.

In AFM observations, a distinctive difference in surface structures was observed on (0001)Si and (000$\bar{1}$)C faces. Epitaxial growth on a (0001)Si face of 6H-SiC yielded apparent macrosteps with a terrace width of 220~280 nm and a step height of 3~6 nm. Each macrostep was revealed, by higher resolution observations, not to be a single multiple-height step but composed of a number of microsteps [50]. On a (000$\bar{1}$)C face, however, the surface was rather flat and no macrosteps were observed. Epilayers on 4H-SiC (0001)Si faces exhibited real macrosteps with 110~160 nm width and 10~15 nm height in some regions.

The mechanism of apparent macrostep formation has not been elucidated up to now. The surface is quite similar to the so-called hill-and-valley (or faceted) structure [51,52]. The off-axis surfaces may spontaneously rearrange to minimize their total surface energy by increasing surface area, which brings the formation of hill-and-valley structure on off-axis (0001)Si faces.

In order to clarify the step structure, cross-sectional TEM images were examined for 4H-SiC epilayers on (0001)Si and (000$\bar{1}$)C faces. On a (0001)Si face, the number of Si-C bilayers at bunched steps was four, which corresponds to exactly the unit cell of 4H-SiC. On a (000$\bar{1}$)C face, however, single bilayer-height steps dominated with few bunched steps. Through examination of more than 200 steps for each sample, histograms of step height have been made as shown in Figs. 11 and 12 [19]. In 6H-SiC, as shown in

Fig. 11(a), 88% of steps are composed of 3 Si-C bilayers (half of unit cell), and 7% of 6 Si-C bilayers (unit cell). On the other hand, 4-bilayer-height (unit cell) steps are the most dominant (66%) and 2-bilayer-height steps have the second highest probability (19%) on 4H-SiC as shown in Fig. 11(b). Single Si-C bilayer-height steps are relatively few (5% or less) on both 6H- and 4H-SiC. In contrast to the bunched steps on (0001)Si faces, single Si-C bilayer-height steps are dominant (70~80%) on (000$\bar{1}$)C faces for both polytypes, as shown in Fig. 12. Even on (000$\bar{1}$)C faces, some bunched steps have 3- or 6-bilayer-heights in 6H-SiC, and 2- or 4-bilayer-heights in 4H-SiC. Investigation on migrating species, surface coverage, and exact bond configuration at step edges may elucidate the origin of this striking polarity dependence.

On a (0001)Si face with a 3.5° off-angle, the average terrace widths were experimentally revealed as 12.4 nm for 6H-SiC and 16.8 nm for 4H-SiC. The different terrace width between 6H-SiC and 4H-SiC, in spite of the same off-angle, may originate from the different height of multiple steps (3 Si-C bilayer height for 6H-SiC, and 4 Si-C bilayer height for 4H-SiC). To achieve step-flow growth, narrow terraces are preferable, which is crucial in SiC growth, because supersaturation increases on wider terraces leads to 3C-SiC nucleation. In this sense, when the off-angle is fixed, 4H-SiC may have a disadvantage of relatively higher probability for nucleation on terraces owing to wider terrace widths. To overcome this disadvantage, a slightly higher growth temperature would be useful because the longer surface diffusion length of adsorbed chemical species and lower supersaturation on terraces are expected. In a different way, larger off-angles of substrates might be effective in 4H-SiC growth [53]. On the other hand, (000$\bar{1}$)C faces showed much narrower average terrace widths owing to fewer bunched steps for both polytypes.

A similar observation has been reported for 6H-SiC surfaces grown by the Lely method [46] and MBE [29]. Thus, the formation of 3 or 6 bilayer-height steps seems to be inherent in 6H-SiC growth, and 2 or 4 bilayer-height steps in 4H-SiC growth. The origin of step bunching in
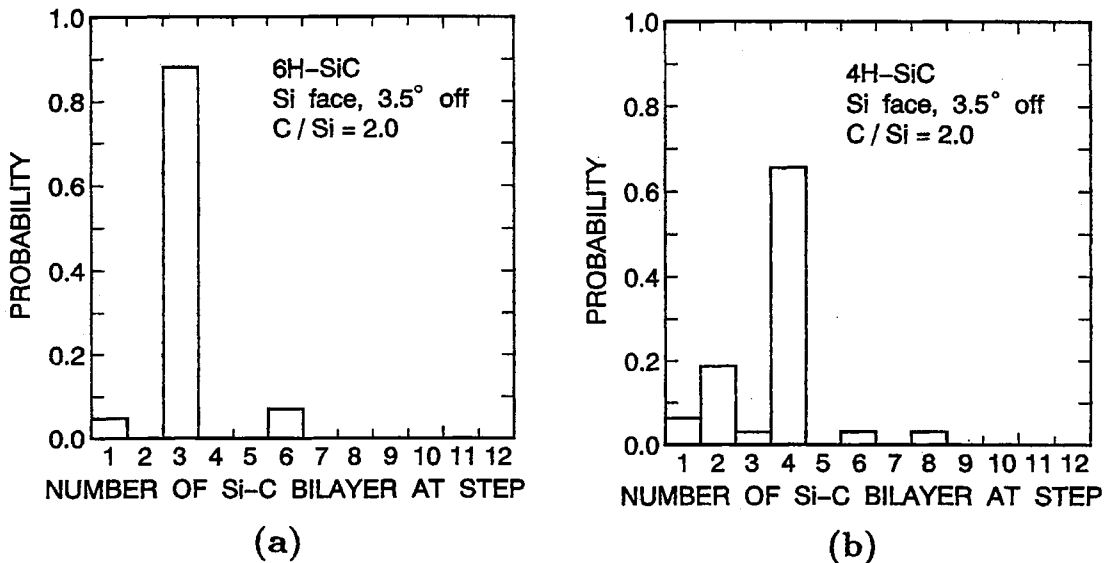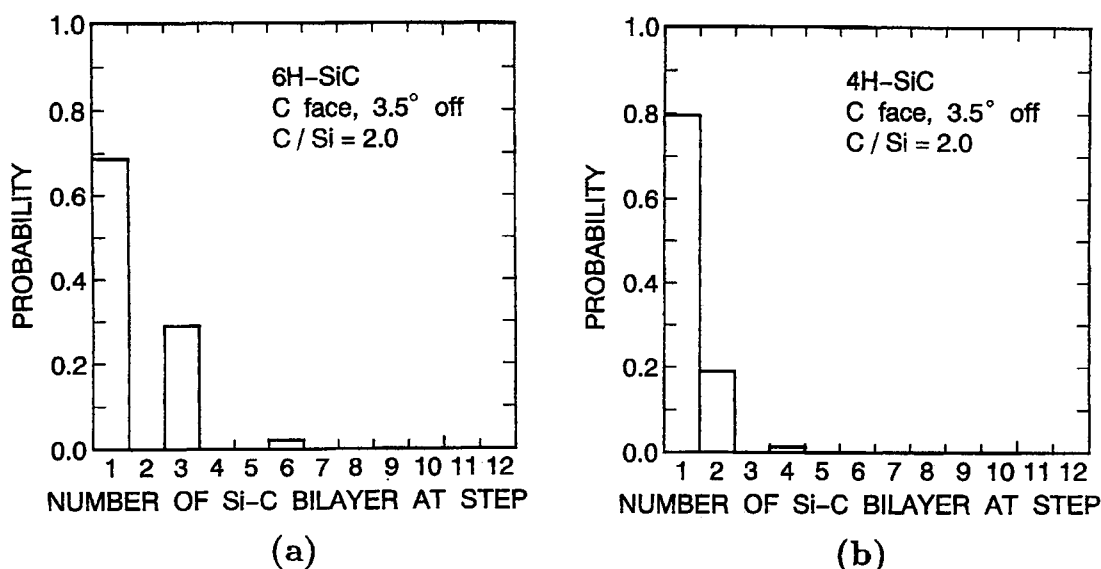
Figure 11. Histograms of step height for surfaces of (a) 6H-SiC and (b) 4H-SiC epilayers. Substrates are (0001)Si 3.5° off-axis toward $\langle 11\bar{2}0 \rangle$ [19].

SiC may be correlated with the surface equilibrium process. Surface energies are different for each SiC bilayer plane owing to the peculiar stacking sequence [28]. Different surface energies may lead to different step velocities for each Si-C bilayer, and thereby causes structurally-induced macrostep formation [54].

4.2. Nucleation and step motion

To understand the mechanism of epitaxial growth of SiC, the formation of nuclei and lateral growth rates of steps have been studied through short-time growth experiments. Single crystals of 6H-SiC {0001} by the Acheson method were used as substrates. Because nucleation is sensitive to surface conditions such as defects, impurities, and damage produced during polishing processes [43], naturally grown surfaces having basal planes (on-axis {0001} faces) were used. Temperatures for SiC growth were varied in the range of 1200~1600°C, and the growth time was very short, 10~600 sec. Nucleus formation has been analyzed after short-time growth for other materials [55,56].

The surfaces after 30-sec growth at 1600°C and

300-sec growth at 1200°C were analyzed. Growth at high temperatures of 1500~1600°C gave nuclei with distorted hexagonal shapes on on-axis {0001} substrates. Nuclei with various sizes from 20 to 50 μm appeared with facets arranging parallel to {1$\bar{1}$00} faces. Small critical nuclei were formed at the initial stage of crystal growth, and they laterally grew by capturing chemical species. The facet formation may be ascribed to the fact that the hexagonal shape inherits from critical nuclei and/or that anisotropy in the lateral growth rate distorts the shape. Although the shapes were hexagonal, Raman scattering analysis showed that the nuclei contain a 3C-SiC component [57].

In low-temperature growth at 1200~1300°C, triangular nuclei were formed, the size of which was less than 5 μm even though their growth time was longer than that for the high-temperature growth. The facets of nuclei were also arranged parallel to {1$\bar{1}$00} faces, with two different orientations, each rotated by 180°. The nuclei may be the origin of twin crystals of 3C-SiC with double-positioning domains, which have been observed in the growth on 6H-SiC {0001} basal (on-axis)

Figure 12. Histograms of step height for surfaces of (a) 6H-SiC and (b) 4H-SiC epilayers. Substrates are (000$\bar{1}$)C 3.5° off-axis toward $\langle 11\bar{2}0 \rangle$ [19].

planes at temperatures lower than 1700~1800°C in early studies [8-10].

Figure 13 shows the temperature dependence of nucleus density (the number of nuclei per unit area) on on-axis (0001)Si and (000$\bar{1}$)C faces [19]. At low temperatures, the nucleus density is significantly high, which can be attributed to higher supersaturation caused by the reduced equilibrium vapor pressure and also by the suppressed surface diffusion of adsorbed chemical species. The nucleus density is higher on (000$\bar{1}$)C faces than on (0001)Si faces by more than one order of magnitude because the former has much lower surface energy than the latter [44].

To measure the lateral growth rates as in a reference [56], circular mesa-shaped features with 120 nm height and 20~300 μm diameter were formed on SiC substrates by conventional photolithography and reactive ion etching (RIE). To remove damaged layers formed during RIE, samples were oxidized at 1100°C for 5 h, and the thermal oxide layers were etched off with HF. The substrates are on-axis 6H-SiC (0001)Si faces. The lateral advance of mesa edges had a nearly linear dependence on time and was independent

of the mesa diameters. The temperature dependence of lateral growth rate on on-axis (0001)Si faces is shown in Fig. 14 [19]. The values are very large compared with the vertical growth rate of 1.2 μm/h at 1600°C. The lateral growth rate decreases significantly at low temperatures due to the decrease in the surface diffusion length of the chemical species. It increases with temperature giving an activation energy of 85 kcal/mole in the range of 1200~1500°C. The slight decrease at 1600°C may be attributed to enhanced reevaporation of chemical species and/or etching by H_2. The anisotropy in lateral growth, which is fastest along $\langle 11\bar{2}0 \rangle$ and slowest along $\langle 1\bar{1}00 \rangle$ [57], reflects the surface diffusion length of adsorbed chemical species on 6H-SiC {0001} faces [58].

5. CHARACTERZATION OF EPITAXIAL LAYERS

5.1. Structural characterization

Very smooth surfaces can be obtained on both (0001)Si and (000$\bar{1}$)C faces when crystal growth is carried out under optimum con-
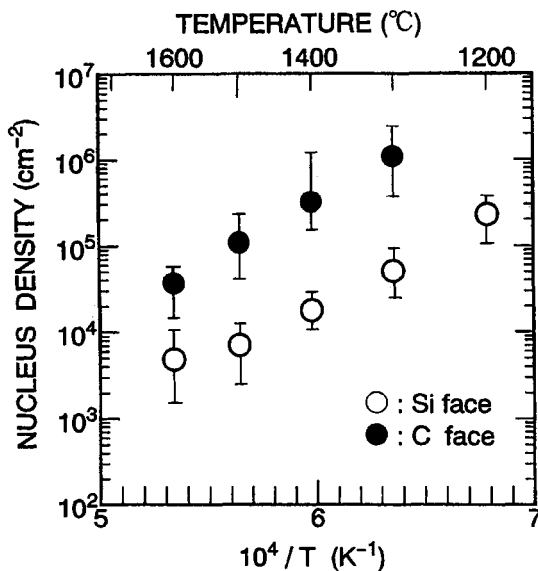
TEMPERATURE (°C)

Figure 13. Temperature dependence of nucleus density on on-axis (0001)Si and (000$\bar{1}$)C faces. The flow rates of SiH_4 and C_3H_8 are 0.15 sccm and 0.10 sccm, respectively [19].
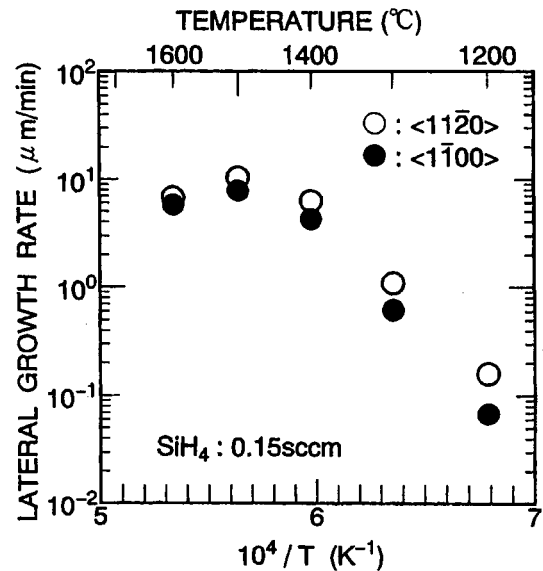
TEMPERATURE (°C)

Figure 14. Temperature dependence of lateral growth rate on on-axis (0001)Si faces. The flow rates of SiH_4 and C_3H_8 are 0.15 sccm and 0.10 sccm, respectively [19].

ditions. Micropipes, which are pin-holes with 0.1~5 μm diameter align along the c-axis [59-62]. Though small micropipes are difficult to detect on as-polished surfaces (before growth), triangular pits accompanied with "shadows" due to the impedance of step-flow are formed at the micropipe positions after epitaxial growth. Defects on the surface are examined after etching by molten KOH at 420°C for 5 min. Large hexagonal pits correspond to micropipes, and small hexagonal pits originate from screw dislocations with Burgers vectors along the c-axis [59]. "Shell" pits are revealed to originate from slip dislocations in the basal plane [62]. The typical etch pit density is 10^4 cm^{-2} for epilayers grown on commercial wafers, and $10^5 \sim 10^6$ cm^{-2} for epilayers on Acheson crystals. Slip dislocations can be eliminated by epitaxial growth [63], which has been elucidated by X-ray topography.

5.2. Photoluminescence

To characterize the quality of epilayers, photoluminescence measurements were carried out at low temperatures using a suitably filtered He-Cd laser (325 nm). The photoluminescence spectrum at 2 K from a 10 μm-thick 6H-SiC epilayer is shown in Fig. 15 for a (a) wide range, and (b) band edge region [19]. The sample is unintentionally doped n-type, and the estimated donor concentration is 1×10^{16} cm^{-3}. Peaks of P_0 (412.3 nm), R_0 (414.3 nm), and S_0 (414.5 nm) are zero-phonon lines due to the recombination of an exciton bound to neutral nitrogen substituting at one hexagonal and two cubic sites. Subscripts on the labels of individual peaks denote phonon energies in meV related to the recombination. Free exciton peaks are marked as I with related phonon energies. As in Fig. 15(a), the band edge luminescence is dominant; also, the donor (N: nitrogen) - acceptor (Al: aluminum) pair luminescence band (440~550 nm) is negligibly small for epilayers, though the pair luminescence is dominant for

(a)

(b)
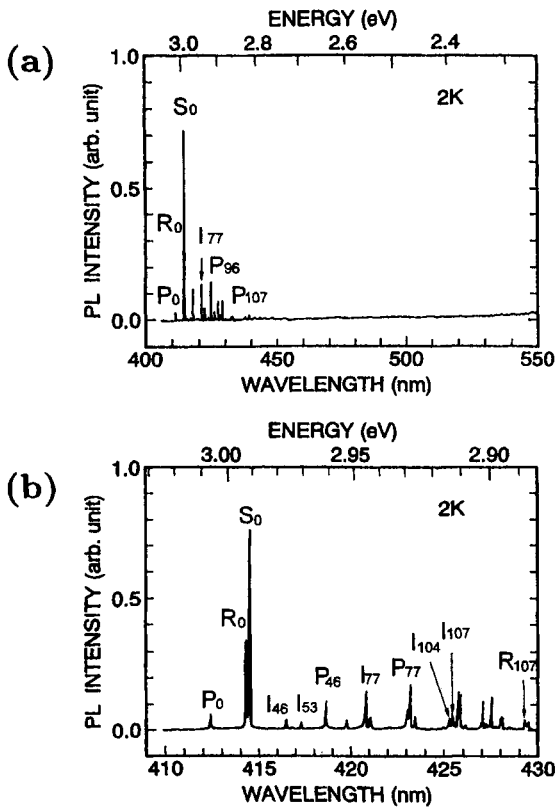
Figure 15. Photoluminescence spectra from a 10μm-thick 6H-SiC epilayer at 2K. (a) wide range spectrum and (b) band edge spectrum [19].

Figure 16. Temperature dependence of electron mobility of 6H-SiC and 4H-SiC epilayers [19].

substrates. No Ti-related peaks, of which zero-phonon peaks would appear at 433.5, 439.8, and 445.1 nm, were observed, though Ti acts as an effective luminescent center in SiC [64,65]. Besides, no "L-lines" attributed to intrinsic defects (probably divacancies) [66] were seen, which usually appear at 472.3, 478.9, and 482.3 nm.

Definite free exciton peaks were observed at 420.6 nm (I_{77}), 423.5 nm (I_{96}), and 425.2 nm (I_{107}) in Fig. 15(b), indicating that the quality and purity of the epilayer was high. The lack of zero-phonon lines for free exciton peaks reflects the indirect band structure of SiC. From the photoluminescence spectrum, very little contamination of Al acceptor is expected owing to no peaks

at 412.7~413.5 nm related to Al bound excitons [67].

5.3. Hall-effect measurements

Electrical properties were characterized by Hall-effect measurements for undoped n-type epilayers which are electrically isolated from substrates using epitaxial p-n junction. Ohmic contacts were prepared with Ni annealed at 1100°C in Ar. The carrier concentration of the epilayer at room temperature was 2.5×10^{16} cm^{-3}. Curve fitting for the temperature dependencies of carrier concentration in 10-μm-thick 6H-SiC and 4H-SiC epilayers was done using an electrical neutrality equation [68]. A two-donor model was adopted, in which each dopant substitutes at two inequivalent (cubic and hexagonal) sites with different energy levels [69,70]. The energy levels for hexagonal and cubic N donors are 97 meV and 141 meV for 6H-SiC, and 71 meV and 124 meV for 4H-SiC. The acceptor concentration is as low as $8 \sim 9 \times 10^{13}$ cm^{-3}, indicating a very low compensation ratio of 2×10^{-3}.

The electron mobility is 351 cm^2/V-s for 6H-SiC and 724 cm^2/V-s for 4H-SiC at room temperature. The mobility increases with decreasing
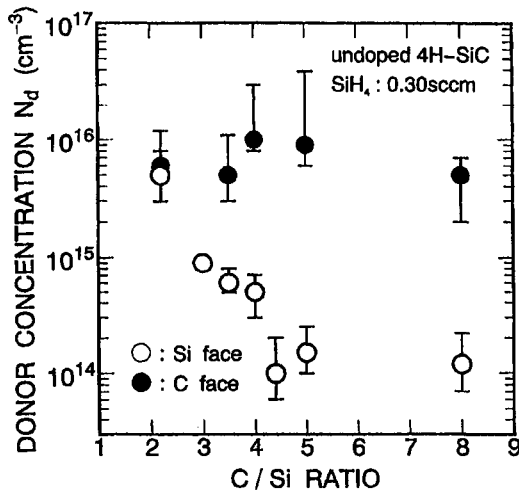
Figure 17. C/Si ratio dependence of donor concentration for unintentionally doped 4H-SiC epilayers [19].

Figure 18. N donor concentration vs. N_2 flow rate in epitaxial growth of 6H-SiC.

temperature, and reaches up to 6050 cm^2/V-s for 6H-SiC and 11000 cm^2/V-s for 4H-SiC at 77 K, as shown in Fig. 16 [19]. The increasing mobility at low temperature reflects the low compensation in the epilayer.

6. IN SITU DOPING OF IMPURITIES

6.1. Donor doping

The doping efficiency of nitrogen (N) strongly depends on the C/Si ratio during CVD growth (site-competition epitaxy) [71]. The higher C/Si ratio leads to the lower N concentration, which can be explained by the higher C-atom coverage. This coverage on a growing surface prevents the incorporation of N atoms into C sites. Figure 17 shows the C/Si ratio dependence of background doping concentration from capacitance-voltage characteristics for unintentionally doped epilayers [19]. For the C/Si ratio of 2, no significant difference was observed between epilayers on the (0001)Si and (000$\bar{1}$)C faces. By increasing the C/Si ratio, the background doping level can be drastically reduced on the (0001)Si face. The lowest value was in the range of $5 \times 10^{13} \sim 2 \times 10^{14}$

cm^{-3}. On the (000$\bar{1}$)C face, however, the background doping concentration is not sensitive to the C/Si ratio.

In situ n-type doping can be easily achieved by doping N_2 during epitaxial growth. As shown in Fig. 18, the donor concentration is proportional to the N_2 flow rate in the wide range on both (0001)Si and (000$\bar{1}$)C faces of 6H-SiC [72].

Relatively high ionization energies of N donors in SiC (80~144 meV in 6H-SiC [73]) cause incomplete activation at room temperature. The activation ratio n/N_d (n: electron concentration, N_d: donor concentration) can be theoretically estimated to be 0.4~0.7 at room temperature. Figure 19 shows the electron mobility at room temperature vs. the carrier concentration of 6H-SiC and 4H-SiC epilayers [19]. Although 4H-SiC exhibits twice the electron mobility of 6H-SiC for lightly doped layers, the difference seems to be small for heavily doped layers in agreement with a previous report [74]. For very low-doped epilayers grown on (0001)Si with a C/Si ratio of 4~5, high electron mobilities of 431 cm^2/V-s ($n = 2 \times 10^{14}$ cm^{-3}) for 6H-SiC and 851 cm^2/V-s ($n = 6 \times 10^{13}$ cm^{-3}) for 4H-SiC were obtained at room temperature. For device applications, 4H-SiC is much
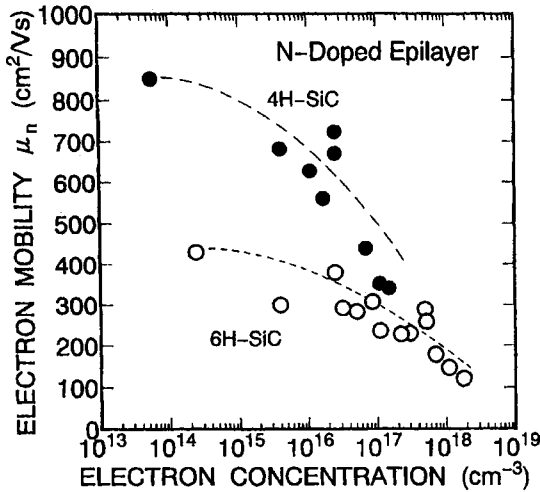
Figure 19. Electron mobility vs. carrier concentration of n-type 6H-SiC and 4H-SiC [19].

Figure 20. Al acceptor concentration vs. TMA flow rate in epitaxial growth of 6H-SiC [19].

more attractive owing to its higher electron mobility and smaller anisotropy in mobility [74,75].

6.2. Acceptor doping

The addition of small amount of trimethylaluminum (TMA: $Al(CH_3)_3$) is effective for *in situ* p-type doping. Although most Al-doped epilayers showed very smooth surfaces, pits and hillocks were observed in heavily doped (Al concentration $\geq 10^{19}$ cm^{-3}) samples grown on $(000\bar{1})C$ faces. The supply of TMA causes the shift of growth conditions toward C-rich due to the decomposition of CH_3 species from TMA. The surface migration of chemical species is suppressed and the nucleation is enhanced on terraces under C-rich growth conditions [58]. Besides, $(000\bar{1})C$ faces easily grow by two-dimensional nucleation due to its low critical supersaturation ratio as shown in Fig. 8. This may be the cause for the surface roughening of heavily doped epilayers. The aluminum (Al) acceptor concentration from capacitance-voltage measurements vs. the TMA flow rate for 6H-SiC is shown in Fig. 20 [19]. The doping efficiency is much higher on (0001)Si faces than that on $(000\bar{1})C$ faces by a factor of 10~80. On (0001)Si faces, the acceptor concen-

tration increases superlinearly with TMA supply, which may be caused by the increased effective C/Si ratio under high TMA flow.

Because of high ionization energy of Al acceptors (~250 meV) [70], the activation ratio p/N_a (p: hole concentration, N_a: acceptor concentration) is as low as 10^{-2}~10^{-1} at room temperature. However, a very high hole concentration of 4~6×10^{19} cm^{-3} could be achieved for heavily doped epilayers (Al concentration is in the mid 10^{20} cm^{-3}). The lowest p-type resistivity was 0.042 Ω-cm for 6H-SiC and 0.025 Ω-cm for 4H-SiC, which were obtained on (0001)Si faces. Figure 21 shows the hole mobility vs. the carrier concentration of Al-doped 6H-SiC and 4H-SiC epilayers at room temperature [19]. The hole mobility is 67 cm^2/V-s at 2 × 10^{16} cm^{-3}, and 6 cm^2/V-s at 1 × 10^{19} cm^{-3} for 6H-SiC, and that of 4H-SiC seems to be higher than that of 6H-SiC at the same hole concentration.

Another hopeful acceptor boron (B) can be easily doped using B_2H_6 gas. Although B doping did not affect surface morphology, the growth rate was reduced by 20~30% in epitaxial growth. The B acceptor concentration increases proportionally with the B_2H_6 flow rate. B-doped samples ex-
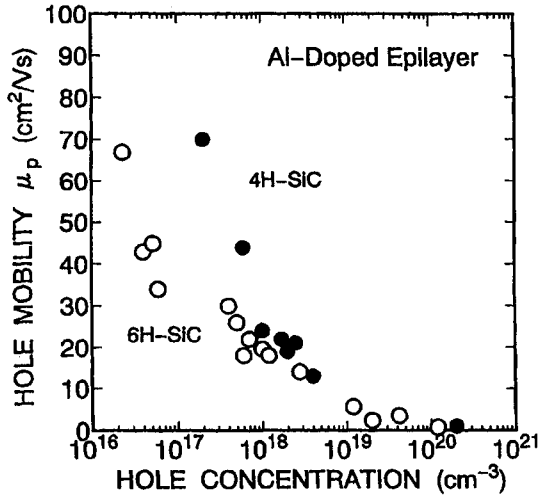
Figure 21. Hole mobility vs. carrier concentration of Al-doped 6H-SiC and 4H-SiC [19].

hibit high resistivity, which is ascribed to the high ionization energy of B acceptor (300~390 meV) [76,77]. The resistivity at room temperature is as high as 520 Ω-cm for a 6H-SiC layer with a boron concentration of 1×10^{17} cm^{-3}. B-doped layers are known to contain so-called D centers locating at 0.6 eV above the valence band, which act as donor-like hole traps [77].

7. RECENT PROGRESS

7.1. Practical epitaxial growth

A horizontal hot-wall reactor has been proposed for the growth of thick epilayers with low doping and good morphology [78]. Growth on substrates of 35 mm and 51 mm in diameter has been tried. The epilayers grown at around 1600°C were n-type, and the background doping was in the low 10^{14} cm^{-3} range. The uniformity in thickness and doping was less than 4% and 7~11%, respectively, for epilayers with a 50-μm thickness grown with a rate of 3~5 μm/h. Epitaxial layers thicker than 150 μm have been grown in a hot-wall type reactor [79]. A vertical hot-wall reactor has been applied for thick epitaxial growth [80]. Mirror-like 4H-SiC epitaxial layers

with undoped level of 5 ~ 7×10^{15} cm^{-3} were obtained at 1750~1900°C with growth rates of 10~30 μm/h.

A planetary reactor for multi-wafer SiC epitaxial growth has been designed for atmospheric- and reduced-pressure operation at temperatures exceeding 1600°C [81]. Epitaxial layers with specular surfaces have been grown at growth rates of 3~5 μm/h, and the thickest layer was 42 μm. The uniformity in thickness and doping was 4% and 7% on wafers with 35-mm diameter, and a thickness uniformity of 3% was on the 50-mm diameter wafer. The lowest unintentional doping was 1×10^{15} cm^{-3} with an electron mobility of 1000 cm^2/V-s.

A vertical cold-wall reactor has been used to grow SiC with high purity and crystal quality [82]. A bell-jar-like graphite susceptor was heated up to 1600°C using radio-frequency induction heating with a flat pancake coil. The susceptor and sample holder are usually rotated with a speed of 800 rpm to stabilize the gas flow. The thickness uniformity was within 5% across the 35 mm diameter, and the background doping level was below 10^{14} cm^{-3} (p-type) with a growth rate of 5 ~ 6μm/h.

A new method using high-temperature CVD to obtain thick epitaxial layers with a very high growth rate has been demonstrated [83]. Growth rates of 10~500 μm/h were demonstrated at temperatures between 1800~2300°C.

7.2. Epitaxial growth on (11\=20) face

Off-axis SiC {0001} wafers with several degrees have been employed for high-quality homoepitaxial growth. SiC {0001} wafers, however, contain micropipes along the c-axis, ⟨0001⟩, which tend to be replicated from substrates to epitaxial layers. To avoid extending the micropipes into epilayers, SiC homoepitaxial growth on (11\=20) and (1\=100) faces is promising because those faces are parallel to ⟨0001⟩ and do not have micropipes in their plane. In the epitaxial growth on (11\=20) and (1\=100) faces, the presence of misfit attributed to the doping difference between the epitaxial layer and the substrate caused a wide FWHM in X-ray diffraction peaks [84]. In the epitaxial growth on the (11\=20) face, the C/Si ratio dependence of

impurity incorporation is between the cases on (0001)Si and (000$\bar{1}$)C faces [85].

Homoepitaxial growth on 4H-SiC substrates with (11$\bar{2}$0) faces has been successfully carried out [86]. An extremely smooth surface with a surface roughness of 0.18 nm can be obtained. On the (11$\bar{2}$0) face, specular surfaces were obtained without macro steps (step bunching), large triangular defects, and carrot-like defects, which are quite often observed in epilayers on the (0001)Si face. Typical surface defects are small shallow depressions. The surface defect area, defined by the ratio of defective area to the total area, is as small as 0.010% compared with 0.16% for epilayers on the (0001)Si face with an off-angle of 8°. The epitaxial growth on the (11$\bar{2}$0) face becomes very important, because the small channel mobility in the (0001)Si face of 4H-SiC can be dramatically improved by using the (11$\bar{2}$0) face [87, 88].

8. SUMMARY

Step-controlled epitaxial growth of SiC on off-axis SiC {0001} substrates was reviewed. Step-flow growth is essential to realize polytype replication in epilayers at the rather low temperatures of around 1500°C without 3C-SiC inclusions. Homoepitaxial CVD growth was achieved at a temperature as low as 1200°C. The introduction of a substrate off-angle induces the change of rate-determining process from surface-reaction control to diffusion control. Critical growth conditions where the growth mode changes from step-flow to two-dimensional nucleation were predicted as a function of growth temperature, growth rate, and substrate off-angle, by using a model describing SiC growth on vicinal SiC {0001}. Step structures of epilayer surfaces depended on the substrate polarity as well as the polytypes. Dominant step heights corresponded to the half or full unit cell of SiC polytypes. Short-time growth experiments revealed that the nucleation rate is much higher on (000$\bar{1}$)C faces, and the lateral growth rate of macrosteps was faster than the vertical growth rate by three orders of magnitude.

The background doping level of epilayers could be reduced to less than 1×10^{14} cm^{-3} by the growth under C-rich conditions, by which very high electron mobilities of 431 cm^2/V-s for 6H-SiC and 851 cm^2/V-s for 4H-SiC were obtained. Excellent doping control has been obtained by *in situ* doping of a nitrogen donor and aluminum/boron acceptors. Polarity dependencies of epitaxial growth and impurity doping were also discussed.

Recent progresses including practical epitaxial growth and epiaxial growth on (11$\bar{2}$0) faces were presented.

ACKNOWLEDGMENTS

REFERENCES

1. W. von Muench and I. Pfaffeneder, J. Appl. Phys., 48 (1977) 4831.
2. D. W. Feldman, J. H. Parker Jr., and W. J. Choyke, L. Patrick, Phys. Rev., 173 (1968) 787.
3. W. von Muench and E. Pettenpaul, J. Appl. Phys., 48 (1977) 4823.
4. G. A. Slack, J. Appl. Phys., 35 (1964) 3460.
5. M. Bhatnagar and B. J. Baliga, IEEE Trans. Electron Devices, ED-40 (1993) 645.
6. Yu. M. Tairov and V. F. Tsvetkov, J. Crystal Growth, 52 (1981) 146.
7. *for example*, K. Koga and T. Yamaguchi, Prog. Crystal Growth and Charact., 23 (1991) 127.
8. V. J. Jennings, A. Sommer, and H. Chang, J. Electrochem. Soc., 113 (1966) 728.
9. W. von Muench and I. Phaffeneder, Thin Solid Films, 31 (1976) 39.
10. S. Yoshida, E. Sakuma, H. Okumura, S. Misawa, and K. Endo, J. Appl. Phys., 62 (1987) 303.
11. N. Kuroda, K. Shibahara, W. S. Yoo, S. Nishino, and H. Matsunami, *Ext. Abstr. the*

34th Spring Meeting of the Japan Society of Applied Physics and Related Societies(Tokyo, 1987), p.135 (in Japanese).

12. N. Kuroda, K. Shibahara, W. S. Yoo, S. Nishino, and H. Matsunami, *Ext. Abstr. 19th Conf. on Solid State Devices and Materials* (Tokyo, 1987), p.227.

13. H. S. Kong, J. T. Glass, and R. F. Davis, J. Appl. Phys., 64 (1988) 2672.

14. J. A. Powell, D. J. Larkin, L. G. Matus, W. J. Choyke, J. L. Bradshaw, L. Henderson, M. Yoganathan, J. Yang, and P. Pirouz, Appl. Phys. Lett., 56 (1990) 1442.

15. S. Karmann, W. Suttrop, A. Schöner, M. Schadt, C. Haberstroh, F. Engelbrecht, R. Helbig, G. Pensl, R. A. Stein, and S. Leibenzeder, J. Appl. Phys., 72 (1992) 5437.

16. O. Kordina, A. Henry, C. Hallin, R. C. Glass, A. O. Konstantinov, C. Hemmingsson, N. T. Son, and E. Janzén, Mat. Res. Soc. Sympo. Proc., 339 (1994) 405.

17. A. A. Burk, Jr, D. L. Barrett, H. M. Hobgood, R. R. Siergiej, T. T. Braggins, R. C. Clarke, G. W. Eldridge, C. D. Brandt, D. J. Larkin, J. A. Powell, and W. J. Choyke, *Silicon Carbide and Related Materials* (Inst. of Phys., Bristol, 1994), p.29.

18. R. Rupp, P. Lanig, J. Völkel, and D. Stephani, J. Crystal Growth, 146 (1995) 37.

19. H. Matsunami and T. Kimoto, Materials Science and Engineering, R20 (1997) 125.

20. H. Matsunami, Mater. Sci. Forum, 338-342 (2000) 125.

21. W. von Muench and I. Pfaffeneder, J. Electrochem. Soc., 122 (1975) 642.

22. A. Suzuki, H. Ashida, N. Furui, K. Mameno, and H. Matsunami, Jpn. J. Appl. Phys., 21 (1982) 579.

23. J. W. Matthews, Ed., *Epitaxial Growth, part B* (Academic Press, New York, 1975), Chap.5.

24. T. Ueda, H. Nishino, and H. Matsunami, J. Crystal Growth, 104 (1990) 695.

25. V. Heine, C. Cheng, and R. J. Needs, J. Am. Ceram. Soc., 74 (1991) 2630.

26. W. S. Yoo and H. Matsunami, *Amorphous and Crystallline Silicon Carbide IV*, (Springer-Verlag, Berlin, 1992), p.66.

27. Yu. M. Tairov, V. F. Tsvetkov, S. K. Lilov, and G. K. Safaraliev, J. Crystal Growth, 36 (1976) 147.

28. Y. Matsushita, T. Nakata, T. Uetani, T. Yamaguchi, and T. Niina, Jpn. J. Appl. Phys., 29 (1990) L343.

29. S. Tanaka, R. S. Kern, and R. F. Davis, Appl. Phys. Lett, 65 (1994). 2851

30. T. Kimoto, H. Nishino, W. S. Yoo, and H. Matsunami, J. Appl. Phys., 73 (1993) 726.

31. B. Wessels, H. C. Gatos and A. F. Witt, *Silicon Carbide 1973*, (University of South Carolina Press, Columbia, 1974), p.25.

32. M. D. Allendorf and R. J. Kee, J. Electrochem. Soc., 138 (1991) 841.

33. C. D. Stinespring and J. C. Wormhoudt, J. Crystal Growth, 87 (1988) 481.

34. C. Hallin, A. O. Konstantinov, O. Kordina, E. Janzén, *Silicon Carbide and Related Materials 1995* (Inst. of Physics, Bristol, 1996), p.85.

35. A. A. Burk, Jr. and L. B. Rowland, J. Crystal Growth, 167 (1996) 586.

36. H. S. Kong, J. T. Glass, and R. F. Davis, J. Mater. Res., 4 (1989) 204.

37. S. Karmann, C. Haberstroh, F. Engelbrecht, W. Suttrop, A. Schöner, M. Schadt, R. Helbig, G. Pensl, R. A. Stein, and S. Leibenzeder, Physica B, 185 (1993) 75.

38. F. C. Eversteyn, P. J. W. Severin, C. H. J. v.d.Brekel, and H. L. Peek, J. Electrochem. Soc., 117 (1970) 925.

39. *CRC Handbook of Chemistry and Physics*, R. C. Weast Ed. (CRC Press, Boca Raton, 1975).

40. W. K. Burton, N. Cabrera, and F. C. Frank, Philos. Trans. Roy. Soc. London, A243 (1951) 299.

41. J. O. Hirschfelder, F. Curties, and R. B. Bird, *Molecular Theory of Gases and Liquids* (John Wiley & Sons, New York, 1954).

42. T. Kimoto and H. Matsunami, J. Appl. Phys., 75 (1994) 850.

43. J. P. Hirth and G. M. Pound, *Condensation and Evaporation, Nucleation and Growth Kinetics* (Pergamon Press, Oxford, 1963), Chap.D.

44. E. Pearson, T. Takai, T. Halicioglu, and W. A. Tiller, J. Crystal Growth, 70 (1984) 33.

45. A. J. Steckl, M. D. Roth, J. A. Powell, and D. J. Larkin, Appl. Phys. Lett., 62 (1993) 2545.

46. S. Tyc, *Silicon Carbide and Related Materials* (Inst. of Physics, Bristol, 1994), p.333.

47. M. A. Kulakov, P. Heuell, V. F. Tsvetkov, and B. Bullemer, Surf. Sci., 315 (1994) 248.

48. J. A. Powell, D. J. Larkin, and P. B. Abel, J. Electron. Mater., 24 (1995) 295.

49. S. Tanaka, R. C. Kern, R. F. Davis, J. F. Wendelken, and J. Wu, Surf. Sci., 350 (1996) 247.

50. T. Kimoto, A. Itoh, and H. Matsunami, Appl. Phys. Lett., 66 (1995) 3645.

51. C. Herring, Phys. Rev., 82 (1951) 87.

52. W. A. Tiller, *The Science of Crystallization: Microscopic Interfacial Phenomena* (Cambridge Univ. Press, Cambridge, 1991), Chap.2.

53. V. F. Tsvetkov, S. T. Allen, H. S. Kong, and C. H. Carter, Jr., *Silicon Carbide and Related Materials 1995* (Inst. of Physics, Bristol, 1996), p.17.

54. T. Kimoto, Ph.D. dissertation, Kyoto University, 1995.

55. B. A. Joyce, R. R. Bradley, and G. R. Booker, Phil. Mag., 15 (1967) 1167.

56. J. Nishizawa, T. Terasaki, and M. Shimbo, J. Crystal Growth, 17 (1972) 241.

57. T. Kimoto and H. Matsunami, J. Appl. Phys., 76 (1994) 7322.

58. T. Kimoto and H. Matsunami, J. Appl. Phys., 78 (1995) 3132.

59. K. Koga, Y. Fujikawa, Y. Ueda, and T. Yamaguchi, *Amorphous and Crystalline Silicon Carbide IV, Springer Proc. in Physics 71* (Springer, Berlin, 1992), p.96.

60. R. A. Stein, Physica B, 185 (1993) 211.

61. H. M. Hobgood, D. L. Barrett, J. P. McHugh, R. C. Clarke, S. Sriram, A. A. Burk, J. Greggi, C. D. Brandt, R. H. Hopkins, and W. J. Choyke, J. Crystal Growth, 137 (1994) 181.

62. J. Takahashi, M. Kanaya, and Y. Fujiwara, J. Crystal Growth, 135 (1994) 61.

63. S. Wang, M. Dudley, C. H. Carter, Jr., and H. S. Kong, Mat. Res. Soc. Sympo. Proc., 339 (1994) 735.

64. A. W. C. van Kemenade and S. H. Hagen, Solid State Commun., 14 (1974) 1331.

65. K. M. Lee, L. S. Dang, G. D. Watkins, and W. J. Choyke, Phys. Rev., B32 (1985) 2273.

66. W. J. Choyke and L. Patrick, *Silicon Carbide 1973*, (Univ. of South Carolina Press, Columbia, 1974), p.261.

67. L. L. Clemen, R. P. Devaty, M. F. MacMillan, M. Yoganathan, W. J. Choyke D. J. Larkin, J. A. Powell, J. A. Edmond, and H. S. Kong, Appl. Phys. Lett., 62 (1993) 2953.

68. S. M. Sze, *Physics of Semiconductor Devices* (John Wiley & Sons, New York, 1981).

69. W. Choyke and L. Patrick, Phys. Rev., 127 (1962) 1868.

70. M. Ikeda, H. Matsunami, and T. Tanaka, Phys. Rev., B22 (1980) 2842.

71. D. J. Larkin, P. G. Neudeck, J. A. Powell, and L. G. Matus, Appl. Phys. Lett., 65 (1994) 1659.

72. T. Kimoto, A. Itoh, and H. Matsunami, Appl. Phys. Lett., 67 (1995) 2385.

73. W. Suttrop, G. Pensl, W. J. Choyke, R. Stein, and S. Leibenzeder, J. Appl. Phys., 72 (1992) 3708.

74. W. J. Schaffer, G. H. Negley, K. G. Irvine, and J. W. Palmour, Mat. Res. Soc. Sympo. Proc., 339 (1994) 595.

75. M. Schadt, G. Pensl, R. P. Devaty, W. J. Choyke, R. Stein, and D. Stephani, Appl. Phys. Lett., 65 (1994) 3120.

76. Yu. A. Vodakov, N. Zhumaev, B. P. Zverev, G. A. Lomakina, E. N. Mokhov, V. G. Oding, V. V. Semenov, and Yu. F. Simakhin, Sov. Phys. Semicond., 11 (1977) 214.

77. W. Suttrop, G. Pensl, and P. Lanig, Appl. Phys., A51 (1990) 231.

78. O. Kordina, A. Henry, E. Janzen, and C. H. Carter,Jr., Mater. Sci. Forum, 264-268 (1998) 97.

79. O. Kordina, K. Irvine, J. Sumakeris, H. S. Kong, M. J. Paisley, and C. H. Carter, Jr., Mater. Sci. Forum, 264-268 (1998) 107.

80. A. Ellison, J. Zhang, J. Peterson, A. Henry, P. Bergman, Y. Makarov, S. Elvstrom, A. Ve-

302

hanen, and E. Janzen, Abstracts of EC-SCRM'98 (1998) 25.

81. A. A. Burk, Jr., M. J. O'Loughlin, and S. S. Mani, Mater. Sci. Forum, 264-268 (1998) 83.

82. R. Rupp, A. Wiedenhofer, P. Friedrichs, D. Peters, A. Schorner, and D. Stephani, Mater. Sci. Forum, 264-268 (1998) 89.

83. O. Kordina, C. Hallin, A. Ellison, A. S. Bakin, I. G. Ivanov, A. Henry, R. Yakimova, M. Tuominen, A. Vehanen, and E. Janzen, Appl. Phys. Lett., 69 (1996) 1456.

84. C. Hallin, A. Ellison, I. G. Ivanov, A. Henry, N. T. Son, and E. Janzen, Mater. Sci. Forum, 264-268 (1998) 123.

85. T. Yamamoto, T. Kimoto and H. Matsunami Mater. Sci. Forum, 264-268 (1998) 111.

86. T. Kimoto, T. Yamamoto, Z. Y. Chen, H. Yano, and H. Matsunami, Mater. Sci. Forum, 338-342 (2000) 189.

87. H. Yano, T. Hirao, T. Kimoto, H. Matsunami, K. Asano, and Y. Sugawara, IEEE Electron Device Lett., 20 (1999) 611.

88. H. Yano, T. Hirao, T. Kimoto, H. Matsunami, K. Asano, and Y. Sugawara, Mater. Sci. Forum, 338-342 (2000) 1105.

Crystal growth and characterization of magnetic semiconductors

Katsuaki Sato

Department of Applied Physics, Tokyo University of Agriculture and Technology, Koganei, Tokyo 184-8588, Japan

Magnetic semiconductors, in which spin- and charge-dependent properties of electrons coexist, has been attracting attention as the possible next-generation spin-electronics materials. This article provides a historical review of studies on crystal growth and characterization of magnetic semiconductors.

1. INTRODUCTION

Spintronics or electronics using spin-related phenomena has been attracting attention because of its potential applicability to new functional devices combining transport and magnetic properties. Magnetic semiconductors and ferromagnet/semiconductor hybrid structures are now the most important topics of investigation in the field of new functional semiconductor devices.

There is a long history of research on this category of materials. The first-generation materials are europium chalcogenides [1] and chalcogenides of chromium with spinel-type crystal structures [2], which were studied intensively in the late 60's and early 70's. Interesting physical properties of magnetic semiconductors, such as magnetic red shift of the absorption edge and huge negative magneto-resistance (MR) around the Curie temperature, were discovered at that time. However, researcher lost interest in these materials because their low Curie temperatures were far below the room temperature and because growth of good-quality single crystals was very difficult.

The second-generation materials are II-VI-based diluted magnetic semiconductors (DMSs), [3] among which $Cd_{1-x}Mn_xTe$ was the focus of most attention due to its capability to accommodate a high percentage of Mn atoms (as high as 77%) and its appropriate energy gap for optical application. The magnetic properties of most of these materials are either paramagnetic or spin-glass. Although the controllability of transport properties is relatively poor, the material shows a good optical property that led to its application to optical isolators.

The third-generation materials are III-V-based diluted semiconductors, in which magnetic properties have been found to be strongly dependent on the carrier concentration in the material. [4] This series of materials can only be produced by using an MBE technique with very low substrate temperatures. Since III-V compound semiconductors are widely used in electronic devices, the III-V-based DMSs are inherently capable of device-integration.

Recently, ferromagnetism in some DMS materials has been predicted by *ab-initio* calculation. A few attempts to obtain room-temperature ferromagnetism have also been reported.

Magnetic semiconductors have been discussed from the viewpoint of crystal growth in only a few reports. This paper is intended to provide a comprehensive review on crystal growth and characterization of magnetic semiconductors.

* E-mail: satokats@cc.tuat.ac.jp

2. FIRST-GENERATION MAGNETIC SEMICONDUCTORS

2.1. Chalcogenide spinels

Multinary chalcogenides with spinel structures, such as $CdCr_2Se_4$, are the most extensively studied species of magnetic semiconductor. The spinel structure belongs to the space group Oh^7 (Fd3m) with an fcc lattice system. The unit cell contains 8 formula units of AB_2X_4-type compound, where A and B are metal atoms and X is a chalcogen atom (S, Se, Te). Chemical vapor transport using halogen or halides as transporting agents was the most widely used technique for bulk crystal growth of chalcogenide spinels. The physical property of these chalcogenides and crystal growth techniques are listed in Table 1.

Exchange interaction between conduction electron spin and local magnetic moment results in various physical properties characteristic of magnetic semiconductors, such as the magnetic red shift of the band gap and the huge negative magneto-resistance in the vicinity of T_C (Curie temperature). Fig. 1 shows the temperature dependence of the absorption edge in $CdCr_2Se_4$. Below the Curie temperature, the absorption edge shifts to lower energies as the temperature decreases. The temperature dependence curve of the absorption edge has been explained in terms of short-range correlation in a localized spin system.

Table 1
Physical properties of spinel chalcogenides

Compounds	Magnetic order	Band gap (eV)	Transporting agent; Growth temperatures (°C)	Lattice parameter (nm)
$ZnCr_2S_4$	AF, T_N=13K		Open tube	0.9974
			Liquid transport $CrCl_3$+ZnS	
$CdCr_2S_4$	FM, T_C=86K	1.8(0K)	Open tube Cd+$CrCl_3$+S_2	1.0244
		1.57(RT)	H+Cl, 1100→1030,	
			Cl_2, 825→775; 825→775; 1000→750	
			$CrCl_3$, 950→900, 1000→950, 1000→750	
$HgCr_2S_4$	Heli, T_N=40K	0.98(0K)	Cl_2,	1.0206
		1.42(RT)	900→850	
$FeCr_2S_4$	Ferri, T_C=177K		Cl_2, H+Cl, 950→875	0.997
			Open tube	
			$CrCl_3$ on FeS	
			HCl, 800→725	
$MnCr_2S_4$	Ferri, T_C=71K		$AlCl_3$, 1000→900	1.0045
$CoCr_2S_4$	Ferri, T_C=22 K		NH_4Cl, 1150→1000	0.990
			H+Cl, 1030→940	
			Liq. transport $CrCl_3$+CoS 1070→1020	
$CuCr_2S_4$	FM, T_C=400K	0	HCl, 800→725	0.9629
$ZnCr_2Se_4$	Spiral AF, T_N=23K	1.1(0K) 1.29(RT)		1.0443
$CdCr_2Se_4$	FM, T_C=130K	1.15(0K) 1.32(RT)	Cl_2, 825→780 CdCl_2, CdI_2 $CrCl_3$, 1000→750, 800→700, 800→750	1.0755
$HgCr_2Se_4$	FM, T_C=110K	0.32(0K) 0.84(RT)	$CrCl_3$, 745→670, 700→670 $AlCl_3$, Al+Cl, 650→625	1.0753
$CuCr_2Se_4$	FM, T_C=426K	0	$AlCl_3$	1.0357

Table 2

Physical properties of the divalent europium chalcogenides

Compound	Color in reflection	Melting point (°C)	Lattice constant (Å)	Energy gap, E_g (eV)	Magnetism, Transition temperature Tc(K)	Curie Weiss Temperature Θ_p (K)
EuO	Magenta	2015±8	5.144	1.12	F, 68	76
EuS	Gold	>2320	5.968	1.65	F,16.2	17
EuSe	Green	>2320	6.195	1.78	Meta	4.6
EuTe	Black	>2250	6.598	1.06	AF	-6

2.2. Europium chalcogenides

Europium mono-chalcogenides crystallize in an NaCl-type fcc structure. They belong to the space group Oh[5]-Fm3m. The magnetic properties of EuO and EuS are ferromagnetic with Curie temperature of 68K and 16.2K, respectively. EuSe is metamagnetic and EuTe antiferromagentic.

The physical properties of europium chalcogenides are summarized in Table 2. Since europium chalcogenides are quite refractory materials with melting points higher than 2000°C, growth from the melt is quite difficult, although not impossible. Indeed, Nakayama et al. grew EuS single crystals by a normal freezing technique using a high-pressure furnace, in which the temperature was raised above 2500 °C and vaporization of carbon heater was suppressed by Ar gas with a pressure of 300-atm.[5] Good-quality single crystals of several millimeter in size were cleaved out from the ingot. They showed a gold metallic luster suitable for optical measurements.

Growth of EuO crystals from a metal-rich solution was carried out by Schafer. [6] The reaction of sesquioxide (Eu_2O_3) and Eu metal proved to be the best method for producing pure EuO. However, the final product obtained by this method contains some unreacted Eu_2O_3. In addition, the large difference between particle sizes of the metal and the oxide often results in an inhomogeneous product in which metal particles are coated by inert Eu_2O_3. These difficulties can be overcome by using a quite large excess of Eu metal, i.e., of the order of 1.5 to 2 times the stoichiometric quantities.
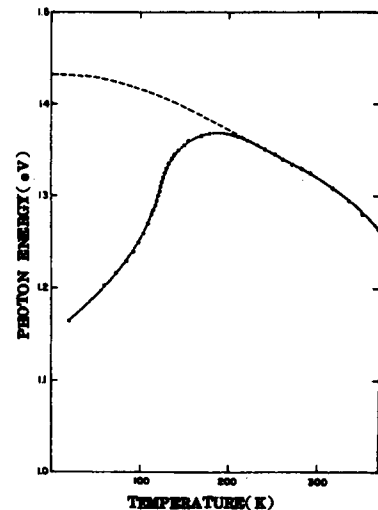
Figure 1. Temperature dependence of the energy gap (Eg) in $CdCr_2Se_4$ crystal. The dotted curve is a plot of the temperature-dependence of Eg expected for a conventional non-magnetic semiconductor extrapolated from the high-temperature part of the curve.

A mixture of Eu and Eu_2O_3 is pressed into slugs, which are wrapped in tantalum foil. The slugs are then slowly heated to 800°C in a closed evacuated system for 16 h. The system was evacuated down to 10^{-7} atm and heated to 1100°C for several hours, and all of the excess metal is distilled off due to the high vacuum.

Dimmock et al. reported the growth of single crystals by a solution method. They grew EuO in a welded tungsten crucible from a starting material consisting of Eu metal and Eu_2O_3 with a large metal excess. The crucible was heated to 2185°C and then cooled slowly
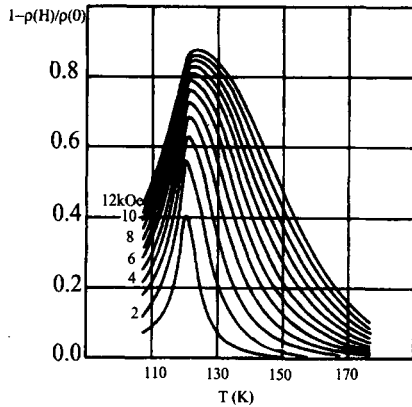
Figure 2. Negative magneto-resistance observed in $CdCr_2Se_4$ for different values of applied field.

at an initial rate of approximately 3.3°C per hour. Samples up to several millimeters in size could be cleaved from single crystalline sections. [7]

Kaldis described crystal growth of europium chalcogenides from the vapor phase. He used chemical transport for growth of EuSe and EuTe and sublimation for EuS and EuTe.[8] The vapor transport data are summarized in Table 3.

Thin films of EuS and EuO have been prepared by vacuum evaporation. Suits et al. used an EuO film as a magneto-optical recording material. [9] However, to keep the film in a ferromagnetic state, the MO-drive had to be immersed in liquid nitrogen. Efforts were therefore made to increase T_c. For this purpose, Fe-doped and Eu_2O_3-doped EuO films were examined. The films were deposited by a three-source simultaneous evaporation of Eu_2O_3, Eu metal and Fe metal. The substrate temperature, T_s, was 150°C and the vacuum during deposition was about 10^{-6} Torr. Film thickness was about 200 nm. The Fe content was 2 wt%. The starting ratio was 0.73 indicating that these films had been heavily doped with excess Eu, in addition to having been doped with Fe. By optimizing the temperature of annealing after deposition, a Curie temperature as high as 150 K was obtained. [10] Mitani et al. prepared thin films by vacuum evaporation on NaCl substrates. [11] Iwata et

al. succeeded in epitaxial growth of an EuO film on an MgO substrate using the MBE technique. [12] The epitaxial EuO film was employed for imaging intrusion of magnetic flux into a high-temperature superconductor (HTSC).

Table 3
Crystal growth data of vapor transport for europium chalcogenides

	T_1 (°C)	T_2 (°C)	ΔT (°C)	I_2 (mg/ cm³)	Rate (mg/ h)	Size (mm)
EuTe+ I_2	1722	1627	95	1.35	18.6	9×8×3
EuTe	2000	1857	143	-	14.7	5×4×2
EuSe +I_2	1687	1619	68	1.0	9.3	4×4×3
EuS	2050	1950	96	-	16	2×2×2

3. II-VI-BASED DILUTED MAGNETIC SEMICONDUCTORS

3.1. General description [3]

Semiconductor alloys, of which the lattice is made up in part of substitutional magnetic atoms, are called diluted magnetic semiconductors (DMS). The most extensively studied and most thoroughly understood DMS species are $A^{IIb}_{1-x}Mn_xB^{VI}$ alloys in which a fraction of the group IIb sublattice is replaced at random by Mn. The lattice constant and band parameters of the alloy system can be "tuned" by varying the composition of Mn. An exchange interaction between the sp-band electrons of the alloy and the localized d electrons associated with Mn^{2+} results in a strongly enhanced g-value for Zeeman splitting of electronic levels, which in turn causes a strong Faraday effect near the band gap of the DMS material. The large magneto-optical effects of $Cd_{1-x}Mn_xTe$ and $Cd_{1-x-y}Hg_xMn_yTe$ are already in practical use as a Faraday rotator in optical isolators for the wavelength region in which magnetic garnet crystals are not applicable

due to presence of strong absorption. [13] The magnetic properties of most II-VI-based semiconductors are either paramagnetic or spin-glass state, properties being not very attractive for practical applications. Recently, development in novel doping technology for II-VI semiconductors has enabled to obtain p-type II-VI-based DMS materials. Theory predicts ferromagnetism in some of these DMS materials if they are heavily-doped p-type. [14]

Table 4
Crystal structure and range of composition

Material	Crystal structure	Range of Composition
$Zn_{1-x}Mn_xS$	ZB	$0<x<0.10$
	WZ	$0.10<x\leq0.45$
$Zn_{1-x}Mn_xSe$	ZB	$0<x\leq0.30$
	WZ	$0.30<x\leq0.57$
$Zn_{1-x}Mn_xTe$	ZB	$0<x\leq0.86$
$Cd_{1-x}Mn_xS$	WZ	$0<x\leq0.45$
$Cd_{1-x}Mn_xSe$	WZ	$0<x\leq0.50$
$Cd_{1-x}Mn_xTe$	ZB	$0<x\leq0.77$
$Hg_{1-x}Mn_xS$	ZB	$0<x\leq0.37$
$Hg_{1-x}Mn_xSe$	ZB	$0<x\leq0.38$
$Hg_{1-x}Mn_xTe$	ZB	$0<x\leq0.75$

3.2. Crystal structure and composition

Crystal structures and miscibility ranges of components of ternary alloys are listed in Table 4.

Crystal structures are either of zinc-blende (ZB) or wurtzite (WZ) type. Under thermal equilibrium conditions, $Cd_{1-x}Mn_xTe$ forms an alloy of ZB structure with x up to 0.77, while $Zn_{1-x}Mn_xSe$ crystallizes in ZB structure for $x\leq0.30$ and WZ structure for $0.30<x<0.57$. The ZB structure of the host II-VI compound survives for x as large as 0.86 in $Zn_{1-x}Mn_xTe$.

Fig. 3 shows plots of lattice constants of telluride alloys as a function of Mn mole fraction. [15] The lattice parameter a can be expressed as a function of the Mn fraction x in the form

$$a = (1 - x)a_{II-VI} + xa_{Mn-VI},$$

where a_{II-VI} is the lattice parameter of the host II-VI compound and a_{Mn-VI} is that of the "hy-
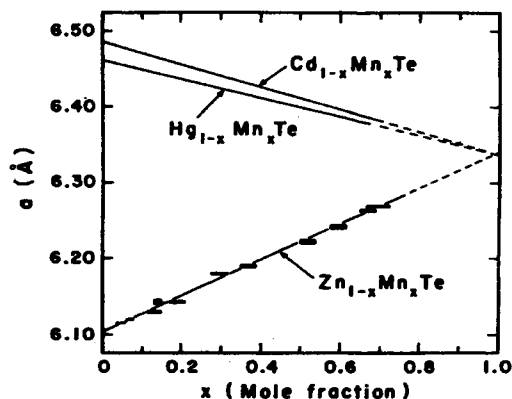
Figure 3. Lattice parameter as a function of Mn mole fraction for telluride DMS. Note that extrapolated lines converge to a single lattice constant (6.334Å), which determines the lattice parameter for the ZB phase of MnTe.
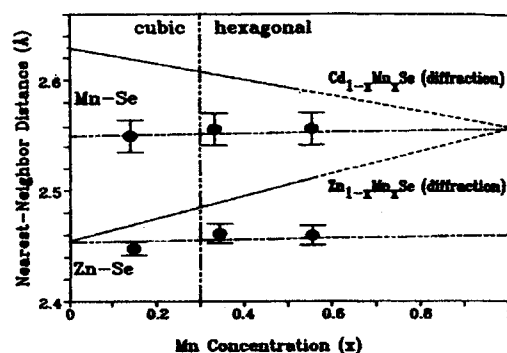
Figure 4. Actual Mn-Se and Zn-Se bond length as a function of Mn fraction determined by EXAFS. Note that either bond length has no observable change at $x\approx0.3$, where $Zn_{1-x}Mn_xSe$ transforms from ZB to WZ.

pothetical" ZB Mn-VI compound. Note that MnS and MnSe crystallize in a cubic rock salt-type structure, while MnTe crystallizes in a hexagonal NiAs-type structure. Although the macroscopic lattice parameter determined by XRD (X-ray diffraction) follows Vegard's law, the local bond length obtained by EXAFS (extended x-ray absorption fine structure) remains practically constant throughout the entire range of compositions studied, as shown in Fig. 4 for the Mn-Se bond in $Zn_{1-x}Mn_xSe$ alloys. [16]

3.3. Bulk crystal growth

Bulk single crystals of $Cd_{1-x}Mn_xTe$ for optical isolators were grown by the Bridgman technique. Elements of Cd, Mn and Te were sealed in vacuum into a quartz crucible. The crucible was lowered in a furnace to crystallize the melt through the melting point of 1100°C with a growth rate of 4 mm/h. Optical transmission of crystals often becomes poor due to twinning defects generated by the phase change from a high-temperature phase (WZ) to a low-temperature phase (ZB) during the cooling process. To obtain twin-free single crystals, Te-excess melt-composition is recommended.

The optical energy gap of a II-VI-based DMS depends linearly on the Mn concentration as shown in Fig. 5. Since a large Faraday effect of II-VI-based DMS materials is observed at photon energies close to the band gap, precise control of the composition ratio in a multinary system is required. For this purpose, recrystallization [17] and zone-melting [18] techniques of polycrystalline ingot were proposed. In order to obtain improved compositional homogeneity of a polycrystalline ingot, Onodera et al. used an quench-and-anneal technique under pressure.

3.4. MBE growth

Nonmagnetic II-VI thin films can be easily obtained by MBE, HWE and PLD techniques. In MBE growth, the condition for obtaining compounds stably can be adjusted, thus avoiding condensation of elements, since the vapor pressure of group II and group VI elements is relatively high compared with that of II-VI compounds. This condition is realized in CdTe for substrate temperatures between 230 and 360°C for a wide range of Cd/Te ratios. RHEED observations have revealed that the atomic arrangements of growth surfaces in Cd-excess and Te-excess conditions are different: For [001] growth, a (2×1) RHEED pattern appears under a Te-excess condition, whereas c(2×2) and (2×1) patterns appear under a Cd-excess condition, suggesting half-atomic coverage of the Cd layer on the Cd-stabilized surface. [19]
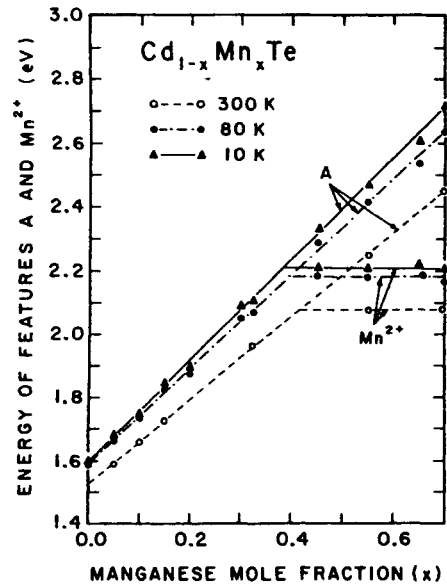
Figure 5. The optical energy gap (determined from the free exciton transition marked "A" in the figure) as a function of Mn concentration x for $Cd_{1-x}Mn_xTe$ for three temperatures.

Epitaxial growth of II-VI-based DMS using the MBE technique has been studied from mid 1980's. [20] MBE growth of $Cd_{1-x}Mn_xTe$ can easily be conducted in the same way as that of CdTe using a simultaneous supply of Mn and Cd fluxes. The addition of Mn flux does not change the growth rate under a Cd-excess condition, whereas the growth rate is increased under a Te-excess condition. The Mn composition is determined by the flux ratio of Mn/Te in the former case and by Mn/(Cd+Mn) in the latter case. [21]

Growth of high-quality $Cd_{1-x}Mn_xTe$ thin films for wave-guide-type optical isolators was achieved on a GaAs substrate by the appropriate use of ZnTe and CdTe buffer layers. [22] The transmission loss parallel to the film plane was so greatly reduced that TE-TM conversion was successfully realized using an epitaxial film.

3.5. Heterostructures and superlattices

Heterostructures consisting of DMS layers are expected to display many interesting and novel phenomena originating from the modulation of the exchange interaction, in addition

to the quantum confinement due to the modulation of the band gap. CdTe/Cd$_{1-x}$Mn$_x$Te multiple quantum wells (MQWs) and superlattices are typical of such systems, where carriers are confined in the non-magnetic CdTe layer and their confinement potential depends strongly on the spin-states of both carriers and Mn^{2+} ions in the barrier layer. This leads to magnetic field tuning of the electronic states, including the drastic phenomenon of field-induced type I→type II transition. [23]

MQW structures of DMS have been prepared on GaAs (100) substrates by MBE, HWE and ICB techniques. Kuroda et al. described MBE growth as follows: The growth was initiated by a thin buffer layer (~2000Å) followed by a Cd$_{1-x}$Mn$_x$Te cladding layer (~8000Å) on the (100) surface of a GaAs substrate, and then an MQW structure consisting of 50 cycles of CdTe and Cd$_{1-x}$Mn$_x$Te layers was grown and finally capped by a Cd$_{1-x}$Mn$_x$Te layer of 2000 Å. [24]

3.6. Carrier-induced ferromagnetism in II-VI-based DMS materials

Theoreticians predict carrier-induced ferromagnetism in DMSs either by the Ruderman-Kittel-Kasuya-Yoshida (RKKY) mechanism or by its continuous-medium limit, the Zener model. [25]
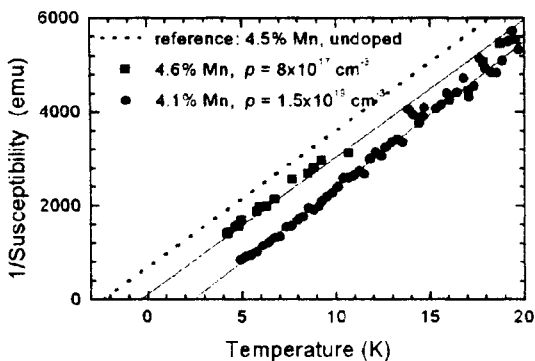
Figure 6. Inverse magnetic susceptibility (squares) for two p-Zn$_{1-x}$Mn$_x$Te samples with similar Mn content x~0.045 but different hole concentration. Solid lines are linear fit, which serve us to determine the Curie Weiss temperature T_{CW}. The dotted line presents the dependence expected for an undoped sample with a similar Mn content.

Recent progress in nitrogen doping of Zn$_{1-x}$Mn$_x$Te by MBE has made it possible to grow II-VI DMSs, in which the kinetic energy of holes is high enough to over-compensate electrostatic and magnetic disorder, so that the metal phase exists down to the millikelvin temperature range. In such a sample, the doping-induced ferromagnetic ordering has been realized. According to Ferrand, [26] Zn$_{1-x}$Mn$_x$Te:N layers were grown by MBE on an 800-μm-thick (001) Cd0.96Zn0.04Te substrate, on which a 300-nm-thick CdTe and a 200-nm-thick ZnTe buffer layer were deposited. The thickness of the ZnTe layer is larger than the critical thickness of the ZnTe/CdTe system. The Zn$_{1-x}$Mn$_x$Te layer was deposited either by using stoichiometric Zn/Te flux from a ZnTe load and simply adding an Mn flux during growth or by using a Zn-rich flux, resulting in a rather rough surface or a smooth surface, respectively. An electron cyclotron resonance (ECR) plasma cell was used for the doping of atomic nitrogen. A hole concentration as high as 1.2×10^{20} cm^{-3} was obtained in the case of ZnTe and Zn0.981Mn0.019Te epilayers. The magnetic properties of the Zn0.981Mn0.019Te film were ferromagnetic with a T_{CW} (magnetic transition temperature determined by the Curie-Weiss law) of 1.4 K. The value of T_{CW} depends on the concentration of holes. As shown in Fig. 5, inverse magnetic susceptibility is different for two p-type Zn$_{1-x}$Mn$_x$Te samples with similar Mn concentrations (x~0.045) but different hole concentrations. A doping-induced positive shift of T_{CW} is clearly visible, as instead of T_{CW} =-2.3K estimated for x~0.045, the observed values of TCW are 2.4 K and -0.4 K for p=1.5×10^{19}cm^{-3} and p=7×10^{17} cm^{-3}, respectively.

4. III-V-BASED DILUTED MAGNETIC SEMICONDUCTORS

4.1. General description [4]

Although great efforts have been made to obtain highly conductive p- and n-type materials, realization of ferromagnetism in a II-VI-based DMS with a reasonable value of T_C is still difficult, and the material is therefore

not so attractive as a material for spintronic devices. III-V compounds are more suitable than II-VI compounds for electronic applications because they possess better controllability of electrical transporting properties. This fact prompted scientists to study III-V-based DMS materials. In 1989, Munekata et al. made the first attempt to fabricate a III-V-based DMS on $In_{1-x}Mn_xAs$ semiconductor alloy by means of a low-temperature MBE method. [27] Their work gained further attention when they found hole-induced ferromagnetism in p-type $In_{1-x}Mn_xAs$. [28] The Curie temperature was found to increase with hole concentration to give the highest Tc of 110 K in $Ga_{1-x}Mn_xAs$ with x=0.05. [29] Owing to the advanced technology for fabrication and manipulation of good quality films and multiplayer structures, a number of interesting studies have been conducted in fundamental physics based on device-structures; for example, spin-injection LED and gate-voltage controlled ferromagnetism. [30]

4.2. Structure and growth technique

Generally speaking, the equilibrium solubility of transition elements in III-V compounds is as low as $10^{17}cm^{-3}$, which is far below the level (hopefully of the order of a few %) required to show (if any) a magnetic ordering effect. Vapor growth methods such as MBE seem to enable the doping of magnetic ions in excess of the thermodynamic solubility limit, since the growth method is not the equilibrium condition used for determining the solubility. Nevertheless, incorporation of a high concentration of transition atoms by MBE often leads to roughened surfaces due to surface segregation of the impurity that occurs during growth. Low-temperature MBE (LT-MBE) deposition is considered to be effective for preventing the segregation effect, because the staying time of impurity atoms on the surface is prolonged. Low temperature growth also has the merit of preventing reaction between the magnetic ions and the host elements. However, a very low growth temperature suppresses epitaxial growth, resulting in the formation of polycrystalline films. Therefore, the optimum growth temperature should be determined so as to satisfy the following two conditions: suppression of surface segregation and capability of epitaxial growth. The LT-MBE method with a substrate temperature T_s of even less than 300°C ha given good results for the growth of $In_{1-x}Mn_xAs$ and $Ga_{1-x}Mn_xAs$.

4.3. Growth of $In_{1-x}Mn_xAs$ [31]

LT-MBE growth of $In_{1-x}Mn_xAs$ is carried out either directly on a GaAs (100) substrate that has a 7% lattice mismatch or on the buffer layer of InAs or $Al_{1-x}Ga_xSb$ that lattice-matches with the epilayer. The substrate temperature used is 200-300°C. If $In_{1-x}Mn_xAs$ layers are grown directly on a GaAs substrate with a thickness larger than 1 μm, both n-type and p-type conduction occurs in the obtained layer. The conduction type depends on x and T_s; p-type conduction occurs at Ts>275°C for x<0.03, while n-type conduction occurs at Ts <275. When Ts is below 200°C, polycrystalline growth occurs. On the other hand, thin layers (<30 nm) of $In_{1-x}Mn_xAs$ grown pseudomorphically on thick buffer layers are p-type at T_s of less than 270°C for x>0.1. In this Mn concentration region, higher Ts results in the segregation of NiAs-type MnAs compounds.

4.4. Growth of $Ga_{1-x}Mn_xAs$

Ohno succeeded in LT-MBE growth of $Ga_{1-x}Mn_xAs$ using solid sources of the constituent elements. Epitaxial films of $Ga_{1-x}Mn_xAs$ were grown on semi-insulating (001) GaAs substrates at a growth rate of 0.6 μm/h. Reflection high-energy electron diffraction (RHEED) patterns were used to monitor the surface reconstruction during growth. Appropriate buffer layers such as GaAs or $Al_{1-x}Ga_xAs$ were deposited prior to growth of $Ga_{1-x}Mn_xAs$. In order to control the strain introduced by the lattice mismatch, a strain-relaxed $In_{1-x}Ga_xAs$ buffer layer with a lattice parameter value larger than the subsequent $Ga_{1-x}Mn_xAs$ layer can be employed.
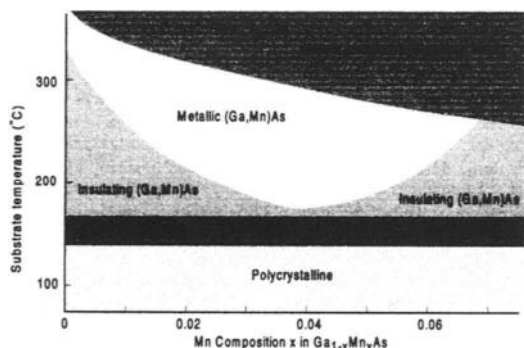
Figure 7 Phase Diagram showing the relation between growth parameters and the properties of $Ga_{1-x}Mn_xAs$ grown by MBE

Fig. 7 shows a schematic phase-diagram of MBE growth of the GaAs-MnAs system. [32] In the case of a high Mn-content x exceeding a critical value or in the case of a high substrate temperature, a spotty RHEED pattern due to MnAs with a hexagonal NiAs structure appears (heavily shaded region in Fig. 7). The MnAs is a metallic ferromagnet with T_C=310 K. The spotty RHEED pattern suggests an island
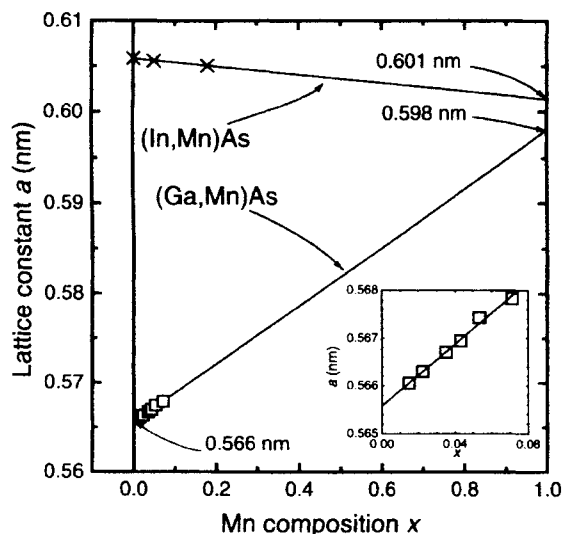
growth of MnAs. Preferable growth of a ferromagnetic $Ga_{1-x}Mn_xAs$ layer occurs in the case of a substrate temperature of ~250 and Mn content of $1\% \leq x \leq 7\%$. The maximum value of the Mn fraction that provides homogeneous incorporation is 7-8%, above which MnAs crystals precipitate on the surface. The electrical properties of the obtained $Ga_{1-x}Mn_xAs$ vary depending on the composition and temperature. Films are insulating when prepared under the condition shown by the lightly shaded region in Fig. 7, while they are metallic for the white region.

In the case of extremely low temperatures (T_s<140 °C), the RHEED pattern becomes spotty again, showing island growth, which in turn results in a ring-like RHEED pattern, suggesting the formation of a polycrystalline material.

4.5. Structural and magnetic characterization

The lattice parameter a of $In_{1-x}Mn_xAs$ and that of $Ga_{1-x}Mn_xAs$ are plotted as a function of Mn content x in Fig. 8. [33] The lattice parameter of $Ga_{1-x}Mn_xAs$ was determined from the X-ray diffraction (XRD) assuming that $Ga_{1-x}Mn_xAs$ is fully strained and has the same elastic constant with a GaAs underlayer, while that of $In_{1-x}Mn_xAs$ was determined under the

Figure 8. Relaxed lattice constant a versus Mn composition x in $Ga_{1-x}Mn_xAs$ films. a was calculated from the X-ray diffraction peaks at room temperature, assuming that Ga1-xMnxAs layers are fully strained and that $Ga_{1-x}Mn_xAs$ has the same elastic constant as GaAs.
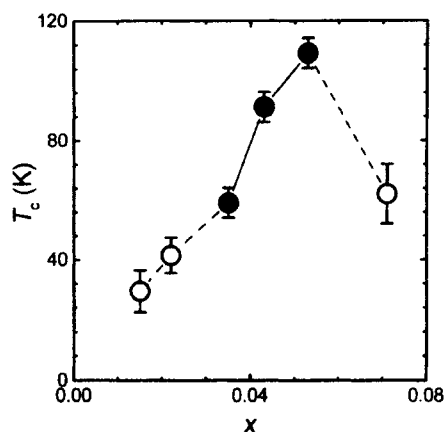
Figure 9. Ferromagnetic transition temperature TC determined from magnetotransport measurements as a function of Mn composition x. Closed circles show metalliv samples,whereas open circles show insulating samples.

assumption that the lattice is fully relaxed. As seen in the figure, the lattice parameter seems to obey Vegard's law. The extrapolated lattice parameter of hypothetical ZB-MnAs determined from the $Ga_{1-x}Mn_xAs$ system (a=0.598 nm) is in good agreement with that determined from $In_{1-x}Mn_xAs$ (a=0.601 nm). From the temperature dependence curve of the saturation magnetization determined from transport measurements in $Ga_{1-x}Mn_xAs$ with x=0.053, Tc is determined to be 110 K. Tc is plotted as a function of Mn content x in Fig. 9. For x<0.07, linear relationship between Tc and x holds as expressed by Tc=2000x±10 K. The decrease is attributed to the beginning of partial segregation of the MnAs phase. [4]

4.6. Attempts to increase Mn-concentration

Recently an attempt to exceed the critical value (~7%) of Mn content in MBE growth without deterioration of crystallinity was made by Misawa et al. using an alternating supply of Mn and GaAs fluxes instead of a simultaneous supply as used in the previous studies.[34] Growth of $[GaAs(xML)/Mn(yML)]_N$ stacked layers was conducted on a GaAs (001) buffer layer of 100 nm in thickness. The substrate temperature was fixed at 230°C, which provided the best surface morphology in preliminary experiments. Control of shutters for individual K-cells supplied GaAs and Mn fluxes alternately. Surface reconstruction during growth was monitored by the RHEED pattern. The surface of the buffer GaAs showed a c(4×4) pattern, which changed to (1×1) with reduction of diffraction intensity after Mn (02 to 1.1 ML) deposition. After the second GaAs layer (9 ML in thickness) had been deposited, the intensity recovered without a drastic change in the RHEED streak pattern. With repeated deposition, the streaky RHEED pattern persisted even after 350 periods of the sequence, although the surface reconstruction pattern changed to (1×2) as in simultaneous deposition. No segregation of extraneous phases was observed. Fig. 10 shows an XRD pattern of a $[GaAs (9ML)/Mn(1.02ML)]_{100}$ layer, in which diffraction lines associated with $Ga_{1-x}Mn_xAs$ can be ob-
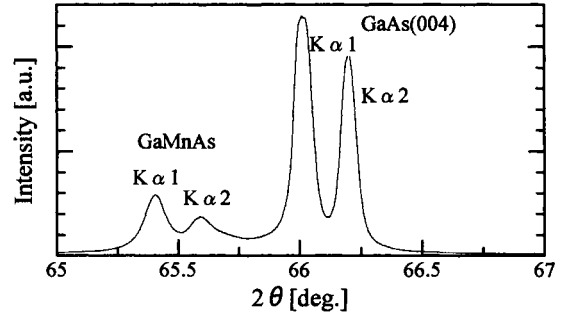

Figure 10. XRD pattern of the $Ga_{1-x}Mn_xAs$ epilayer prepared by MBE technique with alternating deposition of 100 periods of GaAs (9ML)/Mn (1.02ML) stacks

served with a distinct separation from that of the GaAs (001) substrate. The Mn concentration x was determined to be 14% using the relationship between a and x shown in Fig. 8. Surface morphology observed by SEM was quite homogeneous. Based on these results, we postulate that the monolayer or sub-monolayer of Mn acts as a surfactant for the layer-by-layer growth of GaAs, suppressing the formation of three-dimensional island growth, and is finally incorporated into the lattice to form $Ga_{1-x}Mn_xAs$ with a high concentration of Mn. Magnetic measurements in these DMS films are underway.

4.7. Heterostructure devices of III-V DMS

The use of advanced technology for hetero-epitaxy has enabled realization of spin-injection devices employing III-V DMS. Recently, Ohno et al. succeeded in fabricating an LED structure having high spin-injection efficiency. [35] Fig. 11 gives a schematic illustration of the structure. The $Ga_{1-x}Mn_xAs$ DMS is used as a spin-injector electrode. Spin-polarized electrons and unpolarized holes recombine at the $In_{1-x}Ga_xAs$ multi-quantum well (MQW) active layer, from which circularly polarized light is emitted depending on the magnetization of the DMS layer. The degree of polarization ΔP of the light emitted from the diode is plotted as a function of the applied magnetic field B. A considerable degree of polarization appears with a clear hysteresis loop in the magnetic-field dependence, dem-
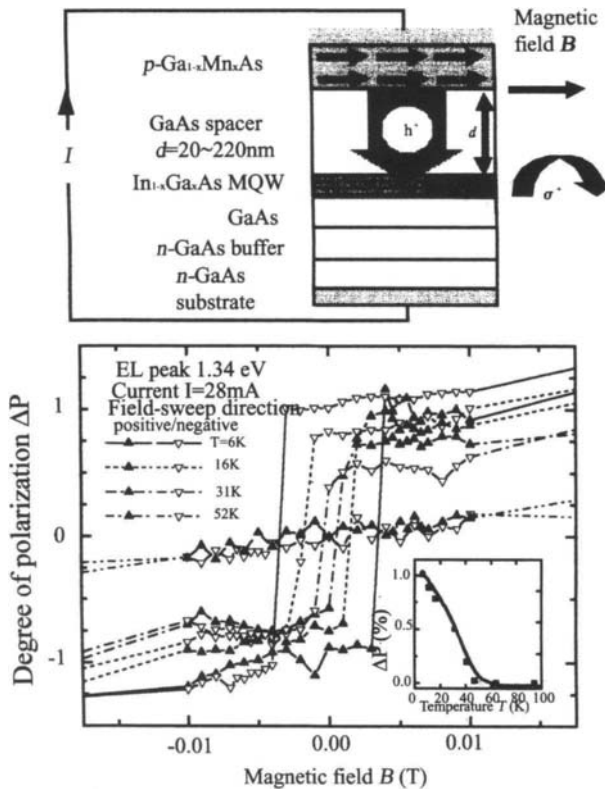
Figure 12. Gate-voltage dependence of Hall resistance in $In_{1-x}Mn_xAs$ FET.

Figure 11. Electrical spin injection in pn-junction LED using ferromagnetic semiconductor $Ga_{1-x}Mn_xAs$.

onstrating the occurrence of electrical spin injection and spin transport.

A field effect transistor (FET) structure was fabricated with a thin (5 nm) $In_{0.97}Mn_{0.03}As$ layer as a channel on which a gate is formed using a polyimide as a gate insulator. By controlling the hole concentration by application of the gate voltage, carrier-induced ferromagnetism can be controlled. The change in magnetization in the channel was monitored by the Hall effect.

Fig. 12 shows Hall resistance as a function of applied field near the Curie temperature (22.5 K). The hysteresis curve dramatically changed with gate voltage. Application of negative voltage increases the magnetization as well as the in-plane anisotropy, while application of positive voltage completely reduces the coercivity.

Recently, a ferromagnetic tunnel junction device using $Ga_{1-x}Mn_xAs/AlAs/Ga_{1-x}Mn_xAs$ hetero-epitaxial structure was fabricated by Tanaka's group and was found to show the magneto-resistance as large as 72% as shown in Fig. 13. [36] An AlAs has thus been demonstrated to be suitable for a tunnel barrier.

5. FERROMAGNET/SEMICONDUCTOR HYBRIDS [37]

Ferromagnet/semiconductor hybrid structures have been attracting attention since room-temperature operation of spintronic devices is only possible at present by using such a structure. Tanaka has been investigating on epitaxial growth of various kinds of ferromagnetic films and superlattices on semiconductor substrates.

MBE growth of MnGa (tetragonal with a CuAu-type ordered structure) on GaAs was conducted using an ultrathin amorphous MnGa template (0.9 nm in thickness) deposited at a very low temperature (20-40°C). The amorphous template was heated up to 200-250°C to form a single crystalline template by solid-state epitaxy. Mn and Ga were subsequently deposited at a relatively low temperature in the range of 150-200°C and at a growth rate of 0.05 μm/h. Transmission electron microscope (TEM) observations re-
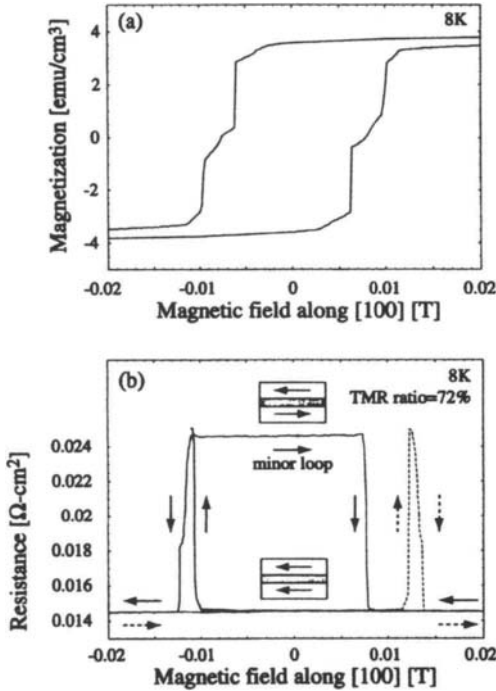
Figure 13. Tunnel magneto-resistance in $Ga_{1-x}Mn_xAs/$ AlAs/ $Ga_{1-x}Mn_xAs$ trilayer structure at 8K.
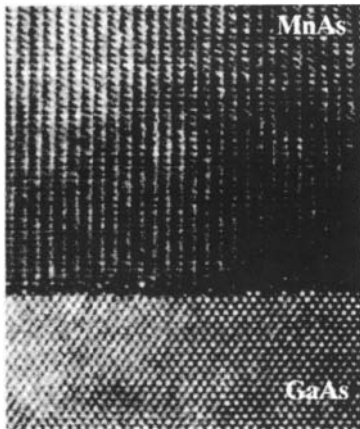
MnAs($\bar{1}100$) / GaAs(001) heterostructure

Figure 14. Cross-sectional TEM image of the MnAs/GaAs heterostructures of type-A.

vealed good epitaxial growth with a smooth and abrupt interface. The c parameter of the MnGa epilayer is significantly reduced from
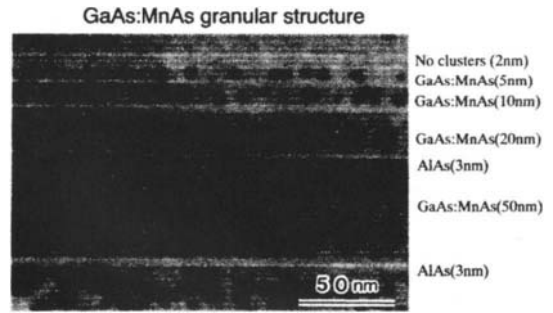
GaAs:MnAs granular structure

Figure 15. Cross-sectional TEM image of the GaAs:MnAs granular layers with different values of layer thickness.

that of bulk material.

Growth of NiAs-type hexagonal MnAs layers on GaAs substrates has been extensively studied by many researchers. MnAs can be deposited at 200-250°C using MBE on a GaAs (001) buffer layer of 100 nm in thickness prepared at 580°C. Tanaka showed that MBE-grown ferromagnetic MnAs thin films have two types of epitaxial orientation depending on the growth conditions. [37] In type-A, the MnAs growth plane of ($\bar{1}100$) is parallel to the GaAs (001) surface with the epitaxial relationship MnAs [11$\bar{2}$0],[0001] // GaAs [110],[$\bar{1}$10] . Fig. 14 shows a
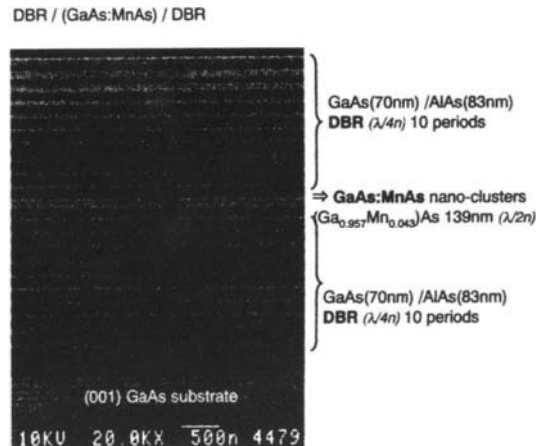
DBR / (GaAs:MnAs) / DBR

Figure 16. Photonic crystal structure consisting of GaAs:MnAs nanocluster layer sandwiched by GaAs/AlAs DBR.

cross-sectional TEM image of a type-A film. On the other hand, in type-B, $(\bar{1}101)$ is parallel to GaAs (001) with MnAs $[11\bar{2}0],[1\bar{1}02]$ //GaAs $[\bar{1}10],[110]$. A type-A film appears on the surface with disordered c(4×4), while type-B appears on a 1-ML-thick Mn layer deposited on a clean c(4×4) GaAs surface. Magnetic anisotropy differs between the two types of growth mode in relation to the substrate orientation.

Hybrid structures with granular MnAs clusters buried in GaAs have been attracting attention as possible new functional materials for spintronics and photo spinics. The granular material can be fabricated by annealing $In_{1-x}Mn_xAs$ and $Ga_{1-x}Mn_xAs$ layers at a temperature above 500°C. Segregation of nano-cluster occurs since $In_{1-x}Mn_xAs$ and $Ga_{1-x}Mn_xAs$ layers are thermodynamically in a non-equilibrium state. Fig. 15 shows cross-sectional TEM images of GaAs:MnAs granular structures for different layer thicknesses. The granular films are super-paramagnetic and have recently been found to show a large magneto-optical effect [38] and magneto- resistance. [39] Tanaka fabricated a photonic device, shown in Fig. 16, using a granular layer sandwiched by two distributed Bragg reflectors (DBRs) consisting
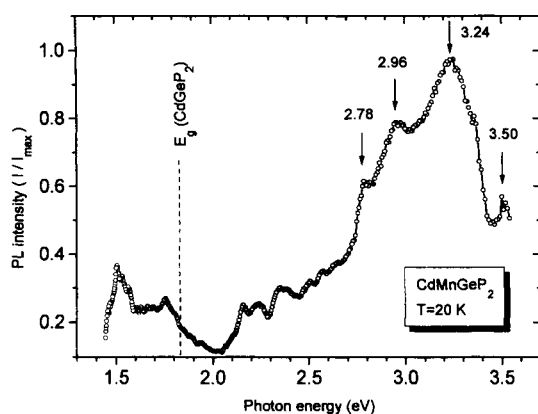

Figure 17. Photoluminescence spectrum of Mn-diffused layer on $CdGeP_2$. The excitation source is a He-Cd laser with λ=325nm. Measurement was carried out at 298K.

of GaAs/AlAs multilayers. Enhancement of the Faraday effect was observed, although the Faraday rotation is still insufficient for practical use. [40]

6. CHALCOPYRITE-TYPE MAGNETIC SEMICONDUCTORS

Ternary semiconductors of I-III-VI$_2$ and II-IV-V$_2$ types are ternary analogs for II-VI and III-V compounds, respectively. They crystallize in a chalcopyrite structure that belongs to the tetragonal crystal system, the unit cell of which is formed by stacking two ZB cell with an ordered arrangement of two cation atoms. The structure is the same as that of the natural mineral $CuFeS_2$, from which the name "chalcopyrite" is derived.

A number of studies have been done to substitute cations of I-III-VI$_2$ semiconductors by transition elements. The solubility of transition elements in the I-III-VI$_2$ compounds is rather poor except for $CuAl_{1-x}Fe_xS_2$ and $CuGa_{1-x}Fe_xS_2$, in which solid-solution is formed for the entire range of x. The magnetic properties of transition element-doped I-III-VI$_2$ semiconductors are paramagnetic or antiferromagnetic.

No systematic studies have been carried out to dope II-IV-V$_2$ compounds with transition elements. Recently, Medvedkin et al. succeeded in incorporating a high concentration of Mn atoms in a $CdGeP_2$ crystal and reported that the Mn-doped sample shows ferromagnetic behavior at room temperature. [41]

A thin Mn layer of about 30 nm in thickness was deposited at 180°C on a $CdGeP_2$ crystal in an MBE chamber using a Knudsen cell for evaporation of the Mn source, and this was followed by thermal treatment at about 500°C for 30 minutes. Prior to the deposition, the surface of the single crystal was etched with bromine-methanol solution. This treatment has been proved to be quite effective for removal of surface contamination and damage in II-VI-V$_2$ semiconductors. The RHEED pattern of the chalcopyrite structure was recovered after annealing.

Crystallographic analysis was carried out using a Rigaku RAD-IIC diffractometer, and

lattice constant change was investigated in detail using a RIGAKU RAD-B with an InP crystal monochromator.

Taking into account the ionic radii of Mn, Cd and Ge, we assume most of the Mn occupies the divalent Cd site. The Mn/Cd ratio analyzed by EDX at the surface reaches 53.4% and drops rapidly with depth, the values being 12.7% at 0.6 μm and 0.9% at 2.5 μm. The average Mn/Cd ratio was calculated to be 20% for an effective thickness of 0.5 μm.

It was found that the crystal structure of the grown layer does not strongly differ from the substrate CdGeP₂, except for the topmost surface, in which a texture formation was confirmed. Detailed crystallographic analysis was carried out, and it was revealed that the lattice

Figure 19. Temperature dependence of magnetization in the Mn-diffused $CdGeP_2$ layer.

Figure 18. Magnetization curves of Mn-diffused $CdGeP_2$ layer. (a) before and (b) after correction for diamagnetic part and demagnetization field.

parameter of an Mn-diffused layer is 0.8% smaller than that of the host semiconductor.

The photoluminescence (PL) spectrum is given in Fig. 17, showing a broad emission band between 1.6 and 3.6 eV with a peak at 3.24 eV, which suggests that the new material $CdMnGeP_2$ grown on the crystal surface of a $CdGeP_2$ single crystal is also a semiconductor with an enlarged energy gap E_g relative to the gap (1.83 eV) of the host semiconductor.

Magnetization was measured using a vibrating sample magnetometer (VSM) with a temperature-controlling attachment. Fig. 18(a) shows raw data of the magnetic hysteresis curves of the $CdGeP_2$:Mn system measured at room temperature (298 K). Straight lines represent in-plane magnetization, and dots represent perpendicular magnetization. The curves are clearly composed of diamagnetic and ferromagnetic components. The former may be attributed to the host substrate and the latter to the new magnetic semiconductor layer. Applying suitable corrections for diamagnetism and demagnetization fields, ferromagnetic hysteresis curves are obtained, as shown in Fig. 18(b). The ferromagnetic component shows a well-defined hysteresis loop with a saturation field H_s of about 3 kOe and coercivity H_c of about 0.4 kOe. The saturation magnetization at room temperature was 3.5×10^{-4} emu. Assuming that deposited Mn of

30 nm in thickness on a 3×5 mm² surface area was completely incorporated into the host semiconductor, the magnetization per atomic unit was estimated to be 0.956×10^{-20} emu/atom, from which the gS value was determined to be 1.03 μ_B. ($S \sim 1/2$)

Fig. 19 shows a plot of remanent magnetization as a function of temperature. The curve was simulated by molecular field theory. The best fit was obtained using the Brillouin function with $S=1/2$. This fact is consistent with the spin value estimated from the saturation magnetization. The Curie temperature estimated by fitting was 320 K. [42]
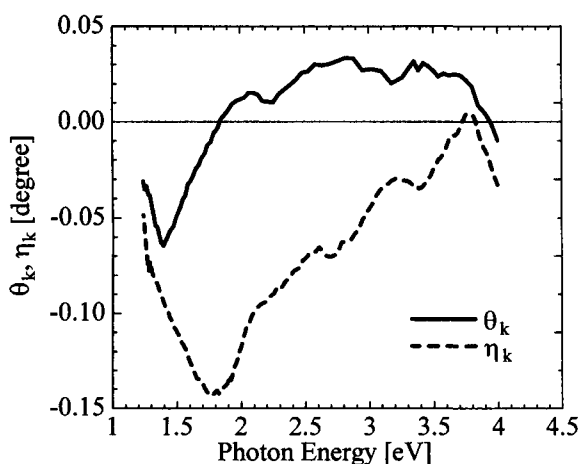
Figure 20. Spectra of magneto-optical Kerr rotation (straight line) and Kerr ellipticity (dashed line) of the Mn-diffused layer measured at room temperature.

Spectra of polar magneto-optical Kerr rotation and ellipticity measured at room temperature are shown in Fig. 20. The Kerr rotation is relatively small and is subjected to a negative peak at 1.4 eV, a zero-crossing behavior at 1.8 eV, and a few peaks between 2 and 4 eV. The peak Kerr rotation value was only 0.065 deg. On the other hand, Kerr ellipticity has a distinct negative peak at 1.75 eV and gradually approaches zero towards higher energies up to 3.8 eV, where a small positive peak appears. The energy at which the ellipticity shows a maximum coincides with the value of the energy gap of the host crystal.

Specific Faraday rotation was estimated from the Kerr effect and was determined to have a value as large 5.2×10^4 deg/cm, which is comparable to the Faraday rotation of Bi-substituted magnetic garnet. [43]

For detailed analysis of structural characterization and explanation of electronic origin of ferromagnetism, further studies on growth of well-defined single-phase crystal and characterization of electrical transport properties are needed.

7. ZnO-BASED MAGNETIC SEMICONDUCTORS

Recently, the results of theoretical studies based on first-principle calculation have indicated that high T_c ferromagnetism can be realized in an Mn-doped ZnO magnetic semiconductor either by RKKY [25] or double-exchange [44] mechanism. However, the Mn-doped ZnO prepared using laser-MBE by Fukumura et al. did not show ferromagnetism. [45] Since ferromagnetic interaction is mediated by a high concentration of holes, heavy p-type doping is necessary. To solve the difficulty in preparation of p-type ZnO, a co-doping process has been proposed. [46] Recently, Tabata et al. succeeded in preparation of heavily Co-doped ZnO films and found that Curie temperature is as high as 290-380 K. [47] Further studies are necessary to elucidate the mechanism of high Tc ferromagnetism in ZnO:Co.

8. CONCLUSIONS

Various kinds of magnetic semiconductors have been studied extensively to determine their potential as next-generation spintronics materials. Developments in epitaxial growth techniques and in doping technology have enabled us to pioneer on a new paradigm of magnetic semiconductors. Some prototype experiments show a high potential of new functional devices using spin-dependent phenomena of magnetic semiconductors.

Room-temperature ferromagnetism has been strongly required for practical applications of magnetic semiconductors. Recently,

stabilization at room temperature of carrier-induced ferromagnetism relative to anti-ferromagnetism has been predicted by theoretical studies. A number of experimental attempts have been made to bring the theoretical prediction into reality, and recently several groups have announced observation of room-temperature ferromagnetism in different materials, although further studies are needed to establish the electronic origin of ferromagnetism and their preparation techniques.

ACKNOWLEDGMENTS

REFERENCES

1. R. P. Van Stapele, *Ferromagnetic materials, vol. 3,* ed. E. P. Wohlfarth (North-Holland Publ. Co, Amsterdam, 1982) Chap.8.
2. T. Kasuya, A. Yanase, Rev. Mod. Phys. 40 (1968) 684
3. J. K. Furdyna, J. Appl. Phys. 64 (1988) R29
4. H. Ohno, J. Magn. Magn. Mater. 200 (1999) 110
5. T. Nakayama and T. Teranishi, Oyobutsuri (Bulletin of Jpn. Soc. Appl. Phys.) 39 (1970) 492
6. M.W. Schafer, J. Appl. Phys. 36 (1965) 1145
7. J.O. Dimmock, C.E. Hurwitz and T.B. Reed, Appl. Phys. Lett. 142 (1969) 49
8. E. Kaldis, J. Cryst. Growth 3-4 (1968) 146
9. J.C. Suits, IEEE Trans. Mag. MAG-8 (1972) 421
10. J.C. Suits, K. Lee, H.F. Winters, P.B.P. Phipps and D.F. Kyser, J. Appl. Phys. 42 (1971) 1777
11. T. Mitani, M. Ishibashi and T. Koda, J. Phys. Soc. Jpn. 38 (1975) 731
12. N. Iwata, G. Pindoria, T. Morishita and K. Kohn, J. Phys. Soc. Jpn. 69 (2000) 230
13. K. Onodera, T. Matsumoto and M. Kimura, Electron. Lett. 39 (1994) 1954
14. D. Ferrand, J. Cibert, C. Bourgognon, S. Tatarenko, A. Wasiela, G. Fishman, A. Bonanni, H. Sitter, S. Kolesnik, J. Jaroszynski, A. Barcz and T. Dietl, J. Cryst. Growth 214/215 (2000) 387
15. J. K. Furdyna, W. Giriat, D. Mitchell and G. Sproule, J. Solid State Chem. 46, (1983) 349
16. B. A. Bunker, W.-F. Pong, V. Debska, D. R. Yoder-Short and J. K. Furdyna, Diluted Magnetic (Semimagnetic) Semiconductors, ed. by R. L. Aggarwal, J. K. Furdyna, and S. von Molnár (Mater. Res. Soc., Pittsburg, PA, 1987) Vol. 89, p. 231
17. K. Onodera and H. Ohba, Cryst. Res. Technol. 31, Suppl. (1996) 29
18. K. Onodera, *Extended Abstracts of the 55th Fall Meeting of the Japan Society of Applied Physics*, p.240 (1993)
19. S. Tatarenko, F. Bassani, J. C. Klein, K. Saminadayer, J. Cibert and V. H. Etgens, J. Vac. Sci. Technol. A12 (1994) 140
20. S. Datta, J. K. Furdyna, and R. L. Gunshor, Superlattices and Microstructures 1 (1985) 327
21. C. Bodin, J. Cibert, W. Grieshaber, Le Si Dang, F. Marcenat, A. Wasiela, P. H. Jouneau, G. Feuiiet, D. Herve and E. Molva, J. Appl. Phys. 77 (1995) 1069
22. K. Ando, W. Zaets and K. Watanabe, Mater. Res. Soc. Symp. Proc. 517 (1999) 625
23. X. Liu, A. Petrou, J. Warnock, B. T. Jonker, G. A. Prinz and J. J. Kretz, Phys. Rev. Lett. 63 (1989) 2280
24. S. Kuroda, Y. Shirai, K. Kojima, K. Uchida, N. Miura and K. Takita, Jpn. J. Appl. Phys. 32, Suppl. 32-3 (1993) 364.
25. T. Dietle, A. Haury and Y. Merle d'Aubigné, Phys. Rev. B55 (1997) R3347; T. Dietle, J. Cibert, D. Ferrand and Y. Merle d'Aubigné, Mater. Sci. Eng. B63 (1999) 103; T. Dietl, H. Ohno, F. Matsukura, J. Cibert and D. Fer-

rand, Science 287 (2000) 1019

26. D. Ferrand, J. Cibert, A. Wasiela, C. Bourgognon, S. Tatarenko and G. Fishman, Phys. Rev. B63 (2001) 085201

27. H. Munekata, H. Ohno, S. von Molnár, A. Segmüller, L.L. Chang, L. Esaki, Phys. Rev. Lett. 63 (1989) 1849.

28. H. Ohno, H. Munekata, T. Penny, S. von Molnár, L.L. Chang, Phys. Rev. Lett. 68 (1992) 2664.

29. A. Shen, H. Ohno, F. Matsukawa, Y. Sugawara, Y. Ohno, N. Akiba, T. Kuroiwa, Jpn. J. Appl. Phys. 36, Part 2 (1997) L73.

30. Y. Ohno, D. K. Young, B. Beschoten, F. Matsukura, H. Ohno and D. D. Awschalom, Nature 402 (1999) 790

31. H. Ohno, H.Munekata, S. von Molnár, A. Harwit, A. Segmüller, L.L. Chang, J. Vac. Sci. Technol. B8 (1990) 176

32. A. Shen, H. Ohno, F. Matsukura, Y. Sugawara, N. Akiba, T. Kuroiwa, A. Oiwa, A. Endo, S. Katsumoto and Y. Iye, J. Cryst. Growth 175/176 (1997) 1069

33. H. Ohno, A. Shen, F. Matsukura, A. Oiwa, A. Endo, S. Katsumoto and Y. Iye, Appl. Phys. Lett. 69 (1996) 363

34. R. Misawa, T. Morishita and K. Sato, J. Jpn. Assoc. Cryst. Growth 27 (2000) 34 (In Japanese)

35. Y. Ohno, D.K. Young, B. Beschoten, F. Matsukura, H. Ohno and D.D. Awschalom, Nature 402 (1999) 790

36. Y. Higo, H. Shimizu and M. Tanaka, *Digest Int. Conf. Phys. & Appl. Spin-Related Phenomena in Semiconductors, NL-2*, Sendai, Sept. 2000.

37. M. Tanaka, Mater. Sci. & Eng. B31(1995) 117

38. H. Akinaga, S. Miyanishi, K. Tanaka, W. van Roy and K. Onodera, Appl. Phys. Lett. 76 (2000) 97

39. H. Akinaga, M. Mizuguchi, K. Ono and M. Oshima: Appl. Phys. Lett. 76 (2000) 357

40. H. Shimizu and M. Tanaka: *Extended Abstracts, 8th MMM-Intermag Conference, FE04*, San Antonio, Jan. 2000 p.358

41. G.A. Medvedkin, T. Ishibashi, T. Nishi, K. Hayata, Y. Hasegawa and K. Sato: Jpn. J. Appl. Phys. 39, Part 2 (2000) L949

42. K. Sato, G. A. Medvedkin, T. Ishibashi, T. Nishi, R. Misawa, K. Yonemitsu, and K. Hirose: J. Magn. Soc. Jpn. 25 (2001) No. 4-2 (in press). (In Japanese)

43. K. Sato, G.A. Medvedkin, K. Hayata, Y. Hasegawa, T. Nishi, R. Misawa and T. Ishibashi, J. Magn. Soc. Jpn. 25 (2001) 283

44. K. Sato and H. Katayama-Yoshida, Jpn. J. Appl. Phys. 39 Part 2 (2000) L555.

45. T. Fukumura, Zhengwu Jin, A. Ohtomo, H. Koinuma and M. Kawasaki, Appl. Phys. Lett. 75, 3366 (1999)

46. T. Yamamoto and H. Katayama-Yoshida, Jpn. J. Appl. Phys.38, Part 2 (1999) L166

47. H.Tabata, K. Ueda and T. Kawai, *Proc. Int. Conf. Physics and Application of Spin-Related Phenomena in Semiconductors, Sendai, September 2000.* (to be published)

X-Ray Characterization of Epitaxial Layers

Y. Takeda and M. Tabuchi

Department of Materials Science and Engineering, Graduate School of Engineering, Nagoya University Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan

In this chapter, two techniques of X-ray characterization for epitaxial layers, especially heteroepitaxial layers, are described. Major interests in the characterization of heteroepitaxial layers are 1) the lattice-distortion due to lattice-mismatching that is commonly observed in heteroeptaxy, 2) interface abruptness, and 3) thickness and 4) composition of embedded layers. To investigate those layer structures, a regular θ–2θ diffractometer measurement and a crystal truncation rod scattering technique are used for several examples.

1. INTRODUCTION

X-ray diffraction is a traditional and very powerful technique that is indispensable for crystal structure analysis. There are many textbooks on X-ray diffraction and X-ray crystallography, which are well-written and worldwide bestsellers. Using X-rays, there are many other characterization techniques such as X-ray fluorescence measurements, EXAFS and XANES measurements, EPMA, XPS, small angle scattering and so on. The crystals can be perfect single crystals, thin epitaxial layers, polycrystals, powders, oriented, and misoriented.

In this chapter, we would like to focus only on two techniques, i.e., 1) the X-ray diffraction for characterizing the lattice-distortion due to lattice-mismatching that is quite often a problem in the heteroepitaxy of single crystals and 2) the X-ray crystal truncation rod scattering for characterizing an extremely thin heteroepitaxial layers embedded (quantum well layers as thin as several Å) and interface structures. We would further like to discuss those characterizations in conjunction with growth processes. Several examples studied by us will be taken and characterization results will be demonstrated.

2. LATTICE DISTORTION AND SYMMETRIC/ASYMMETRIC REFLECTIONS

2.1. Coherent growth and symmetric reflections

Fig. 1 (a) schematically illustrates an epitaxial
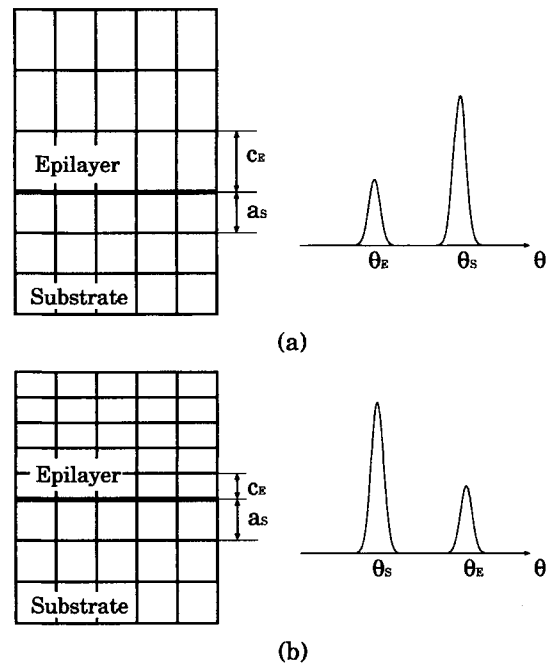

(a)

(b)

Figure 1. (a) An epitaxial layer with a larger lattice parameter and (b) an epitaxial layer with a smaller lattice parameter are grown coherently on a (001) substrate surface. All the lattices are originally cubic. The insets are diffraction peaks from the epitaxial layer (at θ_E) and the substrate (at θ_S).

layer with a larger (cubic) lattice parameter grown coherently on a (001) substrate surface. Due to the coherency of the lattice parameters in plane, the cubic lattice is tetragonally distorted with the perpendicular lattice parameter elongated. In this case the diffraction peaks from the epitaxial layer (at θ_E) and the substrate (at θ_S) are located as shown in the inset when a regular $\theta-2\theta$ diffractometer is used for the symmetric reflections from (00l) planes (see Fig. 3 (a)).

. In the case of a smaller (cubic) lattice parameter in the coherent epitaxial layer, the lattice is under a tensile stress as in Fig. 1 (b) and the diffraction peaks are observed as shown in the inset. When the Bragg law,

$$2d\sin\theta = \lambda \qquad (1)$$

is applied to Figs. 1 (a) and (b), the lattice parameter c_E is easily obtained from the plane spacing d_E as $d_E = c_E/\sqrt{4^2}$ in the case of (004) plane. However, we do not know whether the layer with the lattice parameter a_E is partly strained (Fig. 2 (a)), coherently grown (b), or strain-free (c).
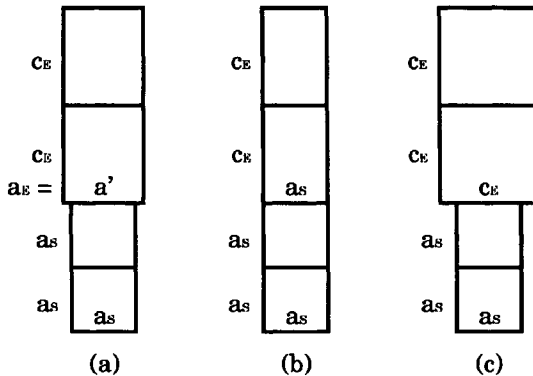

Figure 2. Relationships between the substrate lattice and the epitaxial layer lattice. The cubic lattice is partly strained (a), coherent (b), or strain-free (c).

We want to distinguish those three cases (a), (b) and (c), i.e., to measure the lattice parameters in plane ($a_E = a'$, a_S or c_E). To do this, asymmetric reflections are measured. Since the lattice planes nonparallel to the growth direction (here it is [001]) contain both information of lattice parameters parallel and perpendicular to the growth direction, the in-plane parameter a_E should be deduced from the asymmetric reflections. a_S and then θ_S is assumed to be known.

2.2. Asymmetric reflections

In Fig. 3, symmetric reflections (a) and asymmetric reflections ((b) and (c)) from the epitaxial layer and the substrate are shown. The relationships among the measured angles at the Bragg condition are

$$D_1 = \theta_S - \theta_E - \phi \qquad (2)$$
$$D_2 = \theta_S - \theta_E + \phi \qquad (3)$$

and those peak positions are illustrated in the insets. From Eqs. (2) and (3),

$$\phi = (D_2 - D_1)/2 \qquad (4)$$
$$\theta_S - \theta_E = (D_2 + D_1)/2 \equiv \Delta\theta. \qquad (5)$$

Under the condition that θ_S (or θ_E)$>>\Delta\theta$, in other words, if the difference between a_S and c_E is small enough, the equation

$$\Delta d/d \approx -(\cot\theta)\,\Delta\theta \qquad (6)$$

is obtained from Eq. (1).

In a general case such as Fig. 2 (a), let $c_E = a_S + \Delta a^\perp$ and $a_E = a_S + \Delta a''$. For the (004) reflection, Eq. (6) is easily deduced as

$$\Delta d_{004}/d_{S004} = \Delta a^\perp/a_S \approx -(\cot\theta_{S004})\,\Delta\theta_{004} \qquad (7)$$

and is measured by symmetric reflections.

Then, let us consider the (hkl) plane [1]. The plane spacings are

$$d_{Shkl} = a_S/\sqrt{(h^2 + k^2 + l^2)} \qquad (8)$$

for a cubic lattice (the substrate crystal is cubic in our example) and

$$d_{Ehkl} = 1/\sqrt{\{(h^2 + k^2)/a_E^2 + l^2/c_E^2\}} \qquad (9)$$

for a tetragonal lattice (the epitaxial layer is tetragonally distorted in our example). Using Eqs. (8) and (9), $\Delta d_{hkl}/d_{Shkl} \equiv (d_{Ehkl} - d_{Shkl})/d_{Shkl}$ is obtained as

$$\Delta d_{hkl}/d_{Shkl} \approx \{(h^2 + k^2)\Delta a''/a_S + l^2\Delta a^\perp/a_S\}/(h^2 + k^2 + l^2) \qquad (10)$$

Since $\Delta a^\perp/a_S$ is obtained from Eq. (7) and $\Delta d_{hkl}/d_{Shkl}$ is measured by the asymmetric reflection for the (hkl) planes and is equal to $-(\cot\theta_{Shkl})\,\Delta\theta_{hkl}$, the remaining $\Delta a''/a_S$ is calculated from Eq. (10). Thus, the lattice parameters of the distorted epitaxial layer
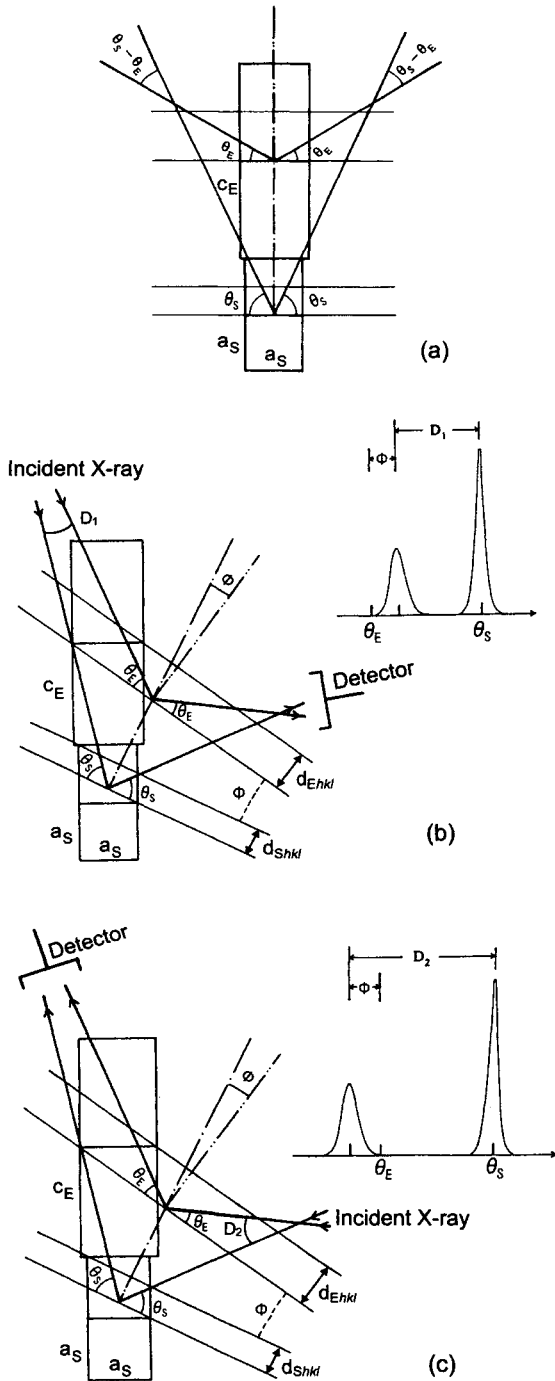
Figure 3. Symmetric reflections (a) and asymmetric reflections ((b) and (c)) from the epitaxial layer and the substrate. In (b) the incident X-ray is from the left-hand side and in (c) it is from the right-hand side. ϕ is the plane angle.

are obtained as $a_E = a_S + \Delta a^{//}$ and $c_E = a_S + \Delta a^{\perp}$.

Examples of the symmetric and asymmetric reflections from (004) and (115) planes are shown in Fig. 4. The epitaxial layer is $Al_xGa_{1-x}Sb$ with the thickness about 2 μm grown by liquid phase epitaxy on a [001]-oriented GaSb substrate. The crystal structure of $Al_xGa_{1-x}Sb$ and GaSb is the zincblende. a_S of GaSb is known as 6.096Å. The wavelength of X-ray is 1.5405Å of Cu $K_{\alpha 1}$. Using those measured data of $\Delta\theta_{004}$, D_1 and D_2 and from Eqs. (5), (7) and (10),

$$\Delta a^{\perp}/a_S = +3.418 \times 10^{-3}$$
$$\Delta a^{//}/a_S = +4.3 \times 10^{-5}$$

are calculated. These results indicate that the $Al_xGa_{1-x}Sb$ epitaxial layer is under a compressive strain as in Fig. 1 (a). The lattice is elongated to the growth direction and the lattice in plane is
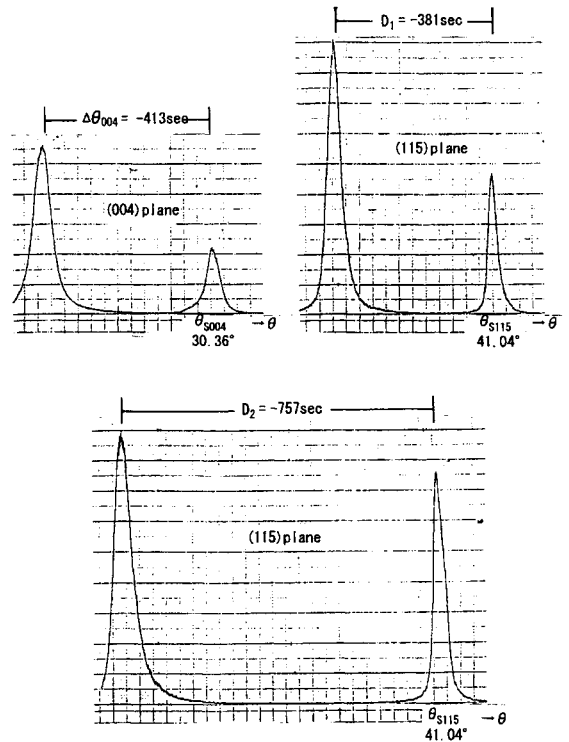

Figure 4. Examples of the symmetric and asymmetric reflections from (004) and (115) planes. The sample is an $Al_xGa_{1-x}Sb$ layer with the thickness about 2 μm grown by liquid phase epitaxy on a [001]-oriented GaSb substrate.

closely matched with that of the substrate.

A series of data are plotted in Fig. 5. The lattice parameter along the growth direction increases linearly with the increase of the Al atom fraction in liquid χ^l_{Al}. However, at $\chi^l_{Al}=0.012$ $\Delta a^{\perp}/a_S$ shrinks and $\Delta a^{///}/a_S$ increases. This is due to the strain relaxation of the lattice by the introduction of misfit dislocations, i.e., the layer is beyond the critical thickness.
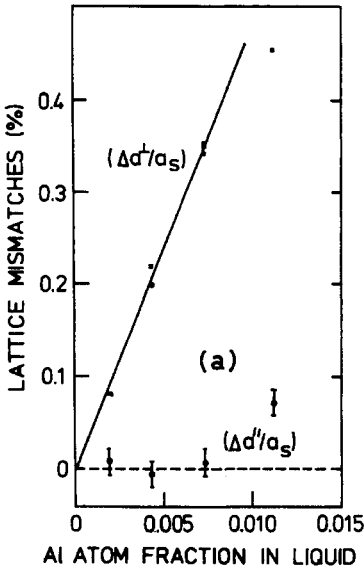

Figure 5. Data of $\Delta a^{\perp}/a_S$ and $\Delta a^{///}/a_S$ are plotted as functions of Al atom fraction in liquid χ^l_{Al}.

3. ATOMIC LAYER CHARACTERIZATION BY X-RAY CRYSTAL TRUNCATION ROD SCATTERING

In this section, X-ray crystal truncation rod (CTR) scattering measurement is described. By this technique, we can reveal the layer thickness, composition, lattice parameters, interface roughness of embedded layers, and the surface roughness to the monolayer scale nondestructively.

3.1. X-ray CTR scattering

X-ray CTR scattering is a rod that appears around a Bragg diffraction spot in reciprocal space (k-space) [2]. It is caused by the abrupt truncation of a crystal at the surface. This can be easily illustrated using the Laue equation;

$$L = \frac{\sin^2(\pi N_1 h)}{\sin^2(\pi h)} \frac{\sin^2(\pi N_2 k)}{\sin^2(\pi k)} \frac{\sin^2(\pi N_3 l)}{\sin^2(\pi l)} \quad (11)$$

where hkl are the Mirror indices. When h and k are fixed and only l is variable, the Laue equation is written as

$$L = \text{Const.} \frac{\sin^2(\pi N_3 l)}{\sin^2(\pi l)} \quad (12)$$

The function $\sin^2(\pi N_3 l)/\sin^2(\pi l)$ with $N_3=10$ around the index $l=2$ is shown in Fig. 6 (a) with the axes in a linear scale. With $N_3=1000$, the function is plotted in Fig. 6 (b) again in a linear scale. It looks zero except at $l=2$. However, in a logarithmic scale, a broad tail is observed around $l=2$. The envelope shape is expressed as $\text{Const.}/\sin^2(\pi l)$. When this wide distribution of the X-ray intensity is plotted in k-space, it looks like a rod or needle that extends perpendicular to the $(00l)$ plane.

A calculated spectrum, considering the atomic form factors, structure factors and others, of an InP wafer is shown in Fig. 7 (a) where a broad tail around the 002 Bragg peak is displayed. Because of this tail, perturbations in the crystal structures, if any, can be superimposed on the spectrum and becomes observable [2–6].

A simulated spectrum for InP with such perturbations by insertion of 10 monolayers (MLs) $In_{0.53}Ga_{0.47}As$ is shown in Fig. 7 (b). The shape is largely modulated due to the X-ray scattering and X-ray interference caused by the crystal truncation and the heterostructure.

In the spectrum, the surface roughness affects the damping of the tail, the layer thickness and the difference in lattice parameters in the case of a heterostructure change the period and the phase of oscillation, respectively, and the interface roughness affects the amplitude of the oscillation. Since X-ray scattering is sensitive to atom species, the spectrum is further modulated depending on the constituent atoms and their distributions. The modulations in the spectrum are especially sensitive to the layer thickness in a monolayer or even sub-monolayer scale since the X-ray wavelength (~ 1 Å) is comparable to the lattice spacing ($2\sim 3$ Å) and monochromatic [7–13].

The CTR scattering spectrum is also sensitive to the crystal structure such as the zincblende structure and the rocksalt structure of the embedded layer even of a monolayer [14–17].
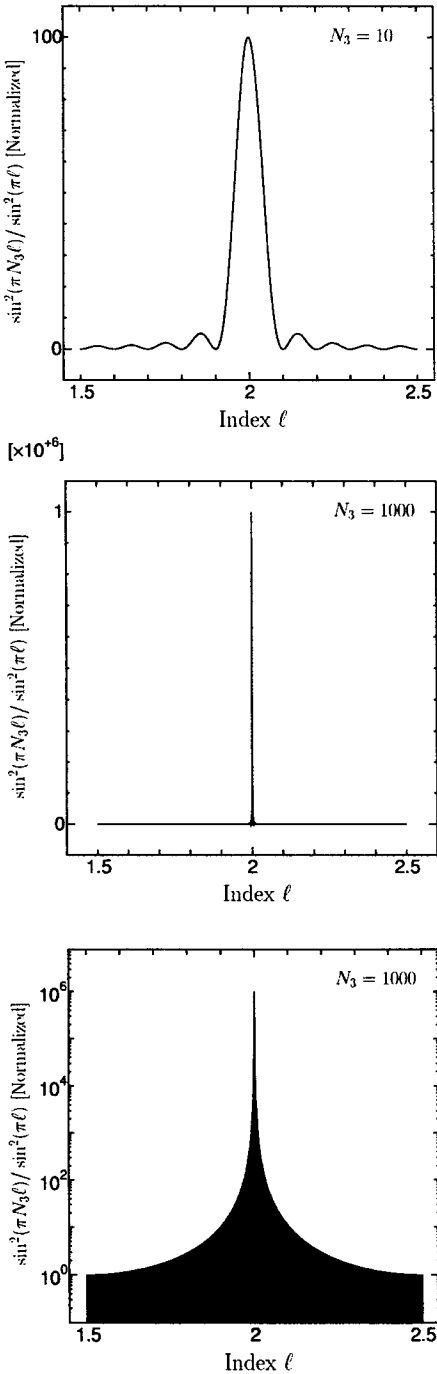
Figure 6. (a) The Laue equation $\sin^2(\pi N_3 l)/\sin^2(\pi l)$ with $N_3=10$ around $l=2$ is plotted in a linear scale of the axes. (b) The Laue equation with $N_3=1000$ in a linear scale and (c) in a logarithmic scale of the vertical axes. In (c) a widely spread tail is observed, which looks like a rod in k-space.
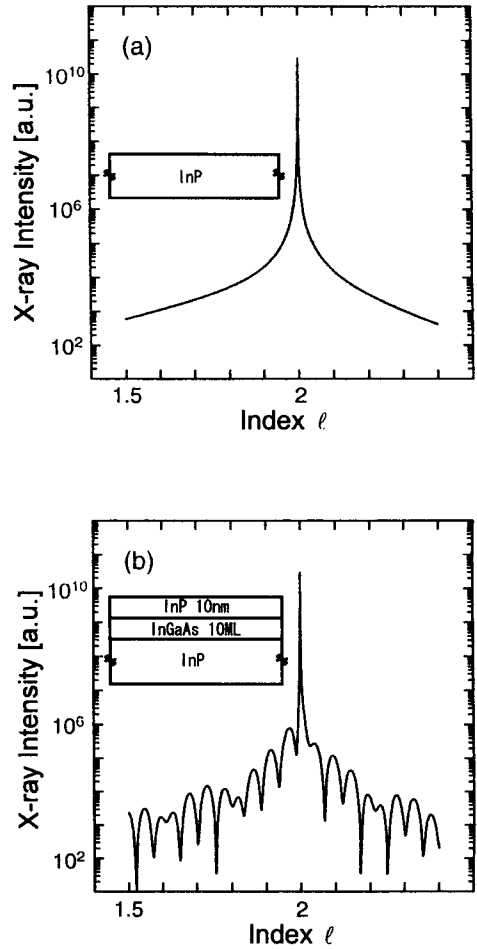

Figure 7. (a) shows a simulated CTR spectrum from InP wafer with a smooth surface, and (b) shows a simulated spectrum for InP with an embedded 10 ML $In_{0.53}Ga_{0.47}As$. Schematic structures are shown in the insets.

3.2. X-ray CTR scattering measurement

The X-ray CTR scattering measurement was conducted using synchrotron radiation (SR) at the Photon Factory in the High Energy Accelerator Research Organization at Tsukuba. The X-ray wavelength was set mostly at 1.600 Å where the SR intensity was highest and the wavelength was far enough from the absorption edges of Ga, In, As, P and Er, and the effects of anomalous dispersion were avoided. The diffraction spots and the CTR scattering were recorded using a Weissenberg camera on an imaging plate (IP). The recorded

patterns on the IP were read out optically and stored as digital data. A schematic drawing of the measurement system is shown in Fig. 8 and the configuration of the X-rays, the sample, and the IP is shown in Fig. 9.
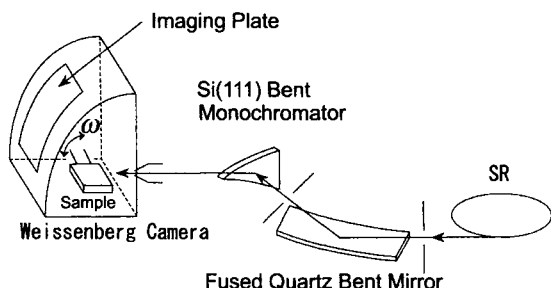

Figure 8. Schematic drawing of CTR measurement system using synchrotron radiation. Monochromatized X-ray by Si (111) bent crystal is collimated to a 0.1 or 0.4 mm spot on sample surface.
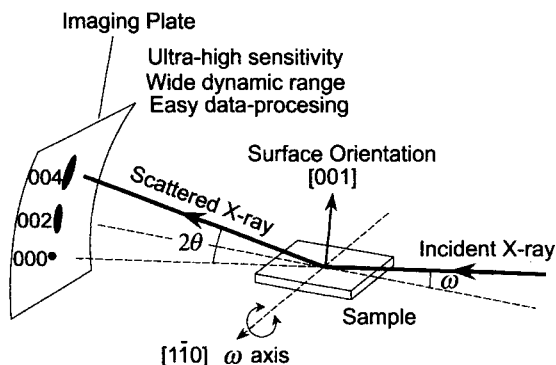

Figure 9. Configuration of sample and imaging plate with respect to incident X-ray beam. Scattered X-ray is recorded on the imaging plate.

The diffraction spot and CTR spectrum around the 002 or 004 Bragg peak were used. By subtracting the background X-ray diffuse scattering from the measured X-ray intensity spectra around the Bragg point, we obtained the CTR spectra. It is essential to subtract the background diffuse scattering to obtain correct CTR spectra for a quantitative analysis. The strong X-ray beam from SR is necessary to obtain good-quality signals in the tail part of the spectra where most of the CTR signals are superimposed. The X-ray CTR scattering intensity varies in a wide range from 10^2

to 10^7. An example of the CTR spectra recorded on the IP is shown in Fig. 10. Around the Bragg point a round area due to diffuse scattering is seen and the horizontal line extending both sides of the Bragg point is the CTR scattering. The modulation of intensity is clearly seen on the rod. The CTR spectrum that is plotted after subtracting the diffuse scattering in Fig. 10 is shown in Fig. 11.

Index l that is the abscissa of Fig. 11 means the index in k-space. We measured the X-ray CTR scattering along the $[00l]$ direction that is normal to the surface of the samples.
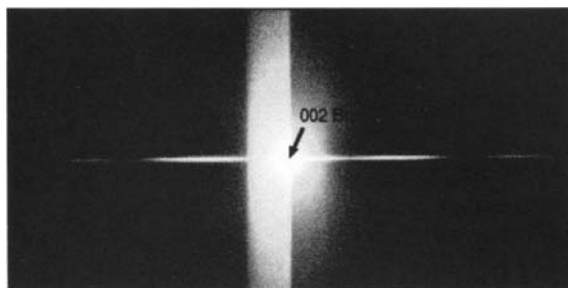

Figure 10. CTR spectrum of InP/InGaAs(3ML)/InP recorded on imaging plate. Horizontal line is the CTR of which intensity is modulated. Vertical thick line is due to saturation of photomultiplier of the IP reader and round white area is due to diffuse scattering.
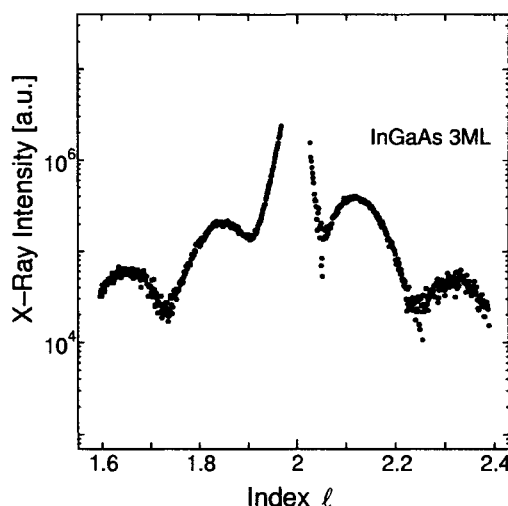

Figure 11. CTR spectrum that is plotted after subtracting the diffuse scattering in Fig. 10. l is the index of $[00l]$ direction.

3.3. Model structures

To analyze the CTR scattering data, we need a model structure to generate a CTR spectrum with several fitting parameters. We assumed model structures shown in Figs. 12 and 13.
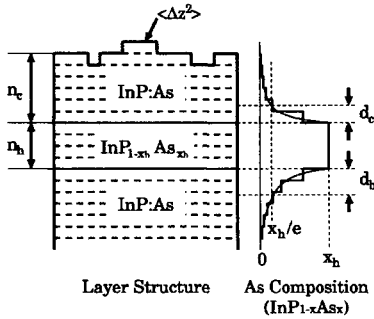

Figure 12. Model structure for InP/InPAs/InP. A space between two horizontal lines corresponds to 1 ML. As distribution is step-wise though the distribution function assumed is $x=x_h\exp(-n/d)$.
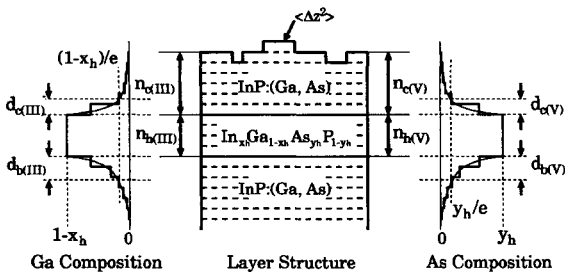

Figure 13. Model structure for InP/InGaAs/InP. Parameters are assigned separately for group-III and group-V atoms.

The model structure in Fig. 12 is for the case of AsH$_3$-exposed and capped InP, i.e., InP/InPAs/InP and is slightly modified for InP/InErP/InP. This model contains parameters such as n_c, n_h, x_h, d_c, d_b, $<\Delta z^2>$, and c/a. Here, n_c and n_h are the thicknesses of the cap layer and the heterolayer, respectively. x_h denotes the As composition of the heterolayer that is formed by the As and P atom exchange during the AsH$_3$-exposure period. d_c and d_b denote the distributions of As atoms in the InP cap layer and in the InP buffer layer, respectively. In the model, the As distribution is assumed to be represented by the formula

$$x=x_h\exp(-n/d) \qquad (13)$$

where n in units of ML is the distance from the upper or lower interface, and x is the As composition of InP$_{1-x}$As$_x$ in the layer at the distance n [ML] from the interface. d is equal to d_c in the cap layer and equal to d_b in the buffer layer. $<\Delta z^2>$ is the mean square deviation that represents the roughness of the surface. c/a is the ratio of the lattice constants normal (c) and parallel (a) to the surface. Coherent growth is assumed and then a is constant. When the ratio is not equal to 1, there is a tetragonal distortion of the lattice in the InP$_{1-x}$As$_x$ layer. The c/a value is dependent on the composition. In the list of c/a, the value at the composition x=1 (of InAs in the case of InP$_{1-x}$As$_x$) is shown for simplicity where a linear dependence on composition is assumed.

The model structure in Fig. 13 is for the case of InP/InGaAs/InP single quantum-well. Parameters contained in this model are basically the same as in Fig. 12, but parameters are assigned separately for group-III atoms (Ga and In) and group-V atoms (P and As) with subscripts III and V to each parameter.

For the CTR spectrum calculation, kinetic theory was used [18]. Comparing the calculated spectrum and the measured data, we determined the best-fit values of the above parameters at the lowest R-factor (residual error ratio).

3.4. R-factor and fitting accuracy

A value that is calculated as

$$\frac{\sum |(\text{Measured value}) - (\text{Calculated value})|}{(\text{Number of measured points})} \qquad (14)$$

is usually called as the "R-factor". We used the R-factor to evaluate the curve-fitting quality realized by a set of parameters. When the R-factor is less than 0.02, we considered that the calculated curve fitted well to the measured data.

The accuracy of each parameter determined was different. Some of the parameters were very accurate since they affect largely the X-ray CTR spectrum. For example, the parameters related to the layer thicknesses (n_c and n_h) were very accurate. If one of these parameters change only 1 ML, obtained spectrum is utterly different. The parameters related to the composition of the heterostructures (x_h) and the parameter c/a were also accurate. If one of these parameters change by

1~2%, the R-factor increases about 10%. On the other hand, the parameters related to the distributions of atoms (d_c and d_b) were not so accurate since they did not affect the spectrum so seriously. Even if one of these parameters changes by 20%, the R-factor increases only about 5%.

3.5. Sample Preparation

In the OMVPE, AsH_3 and PH_3 were used as group-V sources and TMIn (trimethylindium) and TMGa (trimethylgallium) as the group-III sources. (001)-oriented InP was used as the substrate. During the growth and exposure, the reactor pressure was kept at 76 Torr and the growth temperature at 600°C. These growth conditions were previously found as the optimum conditions for obtaining high-quality $InP/In_{0.53}Ga_{0.47}As/InP$ quantum-wells [19]. A typical example of gas flow sequence for growth of InP/InGaAs/InP single quantum-well structure is shown in Fig. 14. AsH_3 and PH_3 are switched twice to form the double heterostructure. The interface and/or quantum-wells were capped with 20~150 Å of an InP layer all in the continuous gas flow sequence.

We investigated the As/P atom exchange at the interfaces, the Ga and As distributions in the quantum-well, and Er distribution in δ-doped InP by the X-ray CTR scattering.

Specific procedure for sample preparation is described in each section.

3.6. AsH_3-Exposed InP/InPAs/InP

The experiments of AsH_3-exposure to InP were conducted to study the effect of the purge sequence, which appears in the growth sequence of InP/InGaAs/InP structures (Fig. 14) [7–9]. The purge sequence is unavoidable to grow heteroepitaxial layers, which consist of different group-V atoms, and is considered to affect largely the structure at the interface.

The samples were prepared by exposing the grown InP surface to AsH_3 and capped by a 20 Å InP layer in one continuous gas sequence as shown in Fig. 15 and the schematic layer structure of the samples is shown in Fig. 16. (001)-oriented InP was used as the substrate. The exposure time was varied from 0.5 to 30 s. During the growth and exposure, the reactor pressure was kept at 76 Torr and the growth temperature at 600°C.

3.6.1. X-ray CTR spectra
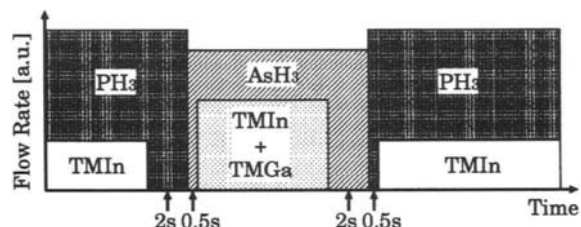
All the measured X-ray CTR spectra for the

Figure 14. Gas flow sequence for growth of InP/InGaAs/InP single quantum-well structure. The carrier gas hydrogen is supplied continuously but not shown here and hereafter in the similar figures.
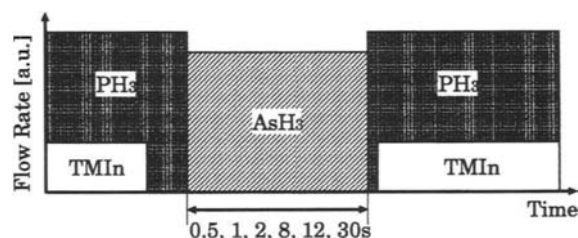

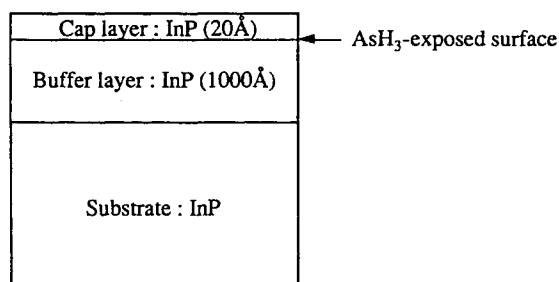Figure 15. Gas flow sequence to study AsH_3-exposure effect.

Figure 16. Structure of measured samples.

AsH_3-exposed samples are shown in Fig. 17. The shoulders observed in the figure at l around 2.1~2.2 reflect information of As atoms contained in InP since this part changes successively with increasing exposure time.

The model structure used is the same as in Fig. 12. Figure 18 shows an example of curve fitting. Data points close to the 002 Bragg point were omitted from curve fitting since the kinetic theory is not correct when the diffraction is too strong.

Table 1
Values of fitting parameters that give the best-fit to measured CTR spectra of AsH_3-exposed samples.

	0.5s	1s	2s	8s	12s	30s
n_c (ML)	7 (20.5Å)	7	7	7	8 (23.5Å)	7
n_h (ML)	1	0	0	1	0	1
x_h (ML)	0.291	0.275	0.294	0.279	0.318	0.210
d_c (ML)	0.525	1.02	1.41	3.10	2.15	3.10
d_b (ML)	0.442	0.393	0.497	1.40	0.791	0.720
$<\Delta z^2>$(ML^2)	0.088	0.084	0.155	0.269	0.00375	0.212
c/a	1.087	1.091	1.078	1.061	1.087	1.073
R-factor	0.0145	0.00905	0.00927	0.0205	0.0133	0.0157

The parameters for the heterostructures obtained at the best fit are listed in Table 1 together with the lowest R-factors at each best-fit.

3.6.2. Results and discussion
a. Layer thicknesses, surface roughness and lattice distortion: In Table 1, most of the n_c values (the InP cap layer thicknesses) are 7 ML (20.5 Å) which is quite close to the designed layer thickness of 20Å. The n_h values are 1 or 0. The difference is not very meaningful in the analysis of the As distribution. It slightly changes the shape near the peak As composition.

The $<\Delta z^2>$ values are much less than 1 ML^2. This means that even if there are steps of 1-ML height, the smooth terraces predominate dips or plateaus on the surface of the InP cap layer. The c/a values are between 1.06 and 1.09. c/a in the coherently grown InAs between InP layers is calculated to be 1.068 using a simple elastic model
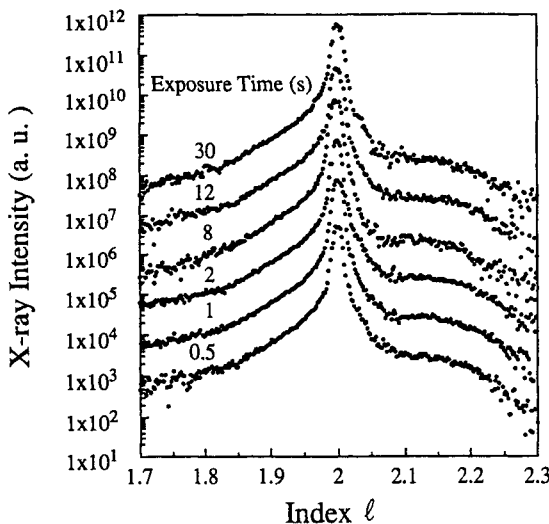

Figure 17. Measured CTR spectra. Exposure time is shown at the left ends of each spectrum. Data for 0.5 s-exposure is plotted at the measured scale and others are shifted upward by one order of magnitude for clarity.
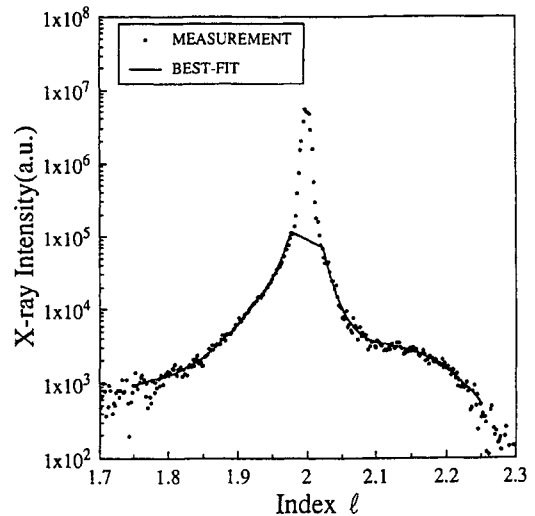

Figure 18. An example of curve fitting. Data points around 002 Bragg point were not used for curve fitting since the kinetic theory that was used for calculation is not correct when diffraction is too strong.

[20] with stiffnesses C_{11} (=8.651011[dyne·cm^{-2}]) and C_{12}(=4.851011 [dyne·cm^{-2}]) of InAs. The obtained c/a values at the best fit are quite close to the calculated one.

b. As atom distribution: The important parameters for discussion of atom exchange are x_h, d_c and d_b which describe the amount and the distribution of As atoms due to the AsH$_3$-exposure of the InP surface. The distributions of As atoms calculated using these three parameters and the cap layer thicknesses (n_c) are shown in Fig. 19. Considerable amounts of As exist in each sample.

In Fig. 19, two features are readily observed. Firstly, the As compositions at each peak are almost the same, *i.e.*, ~0.3 in all the AsH$_3$-exposed samples in spite of very different exposure times. There must be some mechanism that stops the exchange at that composition, *e.g.*, the formation of a surface reconstructed structure. Secondly, the distributions of As atoms in the cap layer and in the buffer layer are very different. In order to make this difference distinct, As amounts, which are equal to the integral of As composition with the layer thickness in units of ML, in the cap layer (O) and in the buffer layer (•) are plotted as functions of the exposure time in Fig. 20. In the figure, the bars indicate ambiguity of values at the R-factor of ±1%. Here, we defined the layer whose As composition was at maximum was the top of the buffer layer. From Fig. 20, it is clear that with increasing exposure time, the amounts of As atoms taken in the cap layer increase steeply for the first few seconds and appears to saturate after about 10 s. On the contrary, the amounts of As atoms in the buffer layer are about 0.3 ML and independent of the exposure time.

3.6.3. Summary

X-ray CTR scattering measurement was conducted for samples prepared by exposing the grown InP surface to AsH$_3$ and capped by InP to reveal the As/P exchange process that is unavoidable during the growth of the heterointerface. By the analysis of the CTR spectra for samples with different exposure times, we obtained the As atom distribution profiles in the samples.

It was found that considerable amounts of As atoms existed in all the samples though the InP surface was only exposed to AsH$_3$, and that the P atoms on the surface were quite quickly exchanged by the As atoms. The exchange stopped at the As composition of 0.3, probably due to formation of a

surface reconstructed structure.

The long tail of As atom distribution in the cap layer was probably due to the memory effect of the As on the reactor wall.
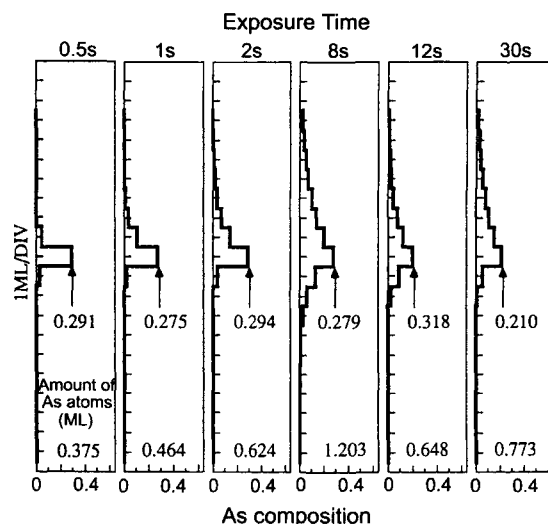
Since the long tail of As distribution in the cap

Figure 19. Profiles of As atom distribution in each AsH$_3$-exposed sample. As compositions at peak of AsH$_3$-exposed surfaces are almost the same, i.e., ~0.3. From the area of the profile the amount of As atoms is calculated and shown in the bottom of each profile.
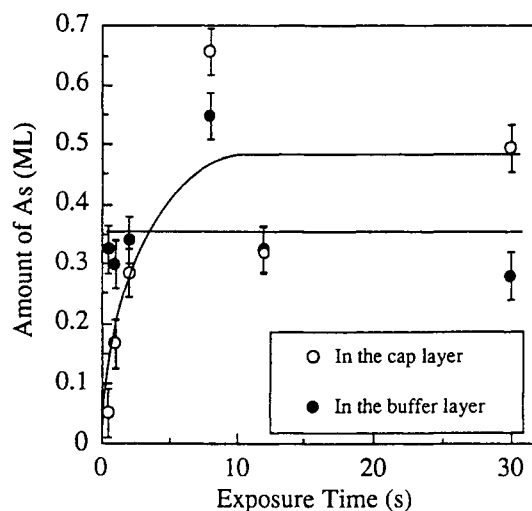

Figure 20. Amounts of As in the cap layer (O) and in the buffer layer (•) as functions of exposure time.

layer is a serious obstacle to obtain a sharp heterointerface, a proper gas sequence is needed. Based on the present analysis, even the exposure time of 0.5 s is long enough to cause the As distribution into a few MLs.

3.7. InP/InGaAs/InP

3.7.1. Ga and As distributions in InP/InGaAs/InP

Here, we will show the distributions of both Ga and As in the OMVPE-grown InP/InGaAs/InP single quantum-wells [10, 12, 13]. CTR spectra from InP/InGaAs(m ML)/InP with m=1, 2, 3, 4 and 5 are shown in Fig. 21. Each spectrum is shifted upward by one order of magnitude for clarity. Oscillatory peaks on both sides are mainly due to interference of X-ray in the single 20 Å InP cap layer and the single InGaAs-well layer. It does demonstrate that the spectra change clearly by each sample and that the CTR scattering has a resolution of 1 ML.

For simplicity, InP/InGaAs(3ML)/InP single quantum-well (denoted as InGaAs-3ML) and InP/InGaAs(5ML)/InP single quantum-well (denoted as InGaAs-5ML) samples were taken for analysis and discussion of the distribution profiles. The samples were prepared by a low-pressure OMVPE on InP (001) substrates. The growth temperature was 600°C. The gas flow sequence was shown in Fig. 14. The PH_3 and AsH_3 flow rates were 20 sccm and 2 sccm, respectively. InP buffer layer was 700 Å and InP cap layer was 20 Å.

3.7.2. Model structure

We assumed the model structure as shown in Fig. 13. This model contains parameters such as n_{cGa}, n_{hGa}, x_{hGa}, d_{cGa} and d_{bGa} for Ga atom, and n_{cAs}, n_{hAs}, x_{hAs}, d_{cAs} and d_{bAs} for As atom. n_c and n_h are the thicknesses of the cap layer and of the well layer, respectively, and defined separately for Ga and As atoms. x_{hGa} and x_{hAs} denote the Ga and As compositions in the well layer, respectively, and needless to say $x_{hGa}+x_{hIn}=1$ and $x_{hAs}+x_{hP}=1$ from stoichiometry requirement. d_c and d_b are the extension of As or Ga atoms into the InP cap layer and into the InP buffer layer, respectively. The Ga and As distributions are assumed to have the formula

$$x_{(Ga, As)}=x_{h(Ga, As)}\exp(-n/d) \qquad (15)$$

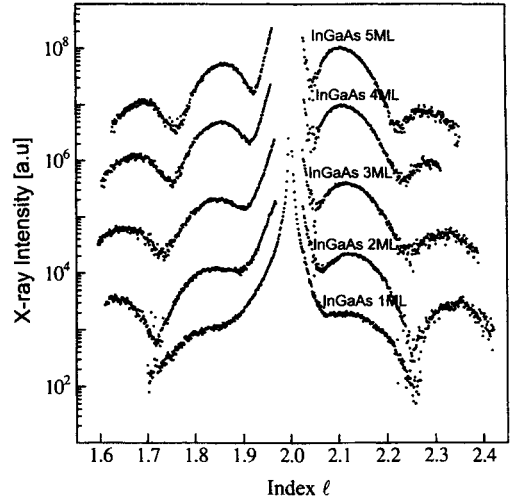where n [ML] is a distance from the upper or lower

Figure 21. CTR spectra from InP/InGaAs(1, 2, 3, 4 and 5ML)/ InP samples. The spectra change clearly by each sample from bottom to top, indicating that the CTR spectrum has a resolution of 1 ML.

interface. d is d_{cGa} or d_{cAs} in the cap layer, and d_{bGa} or d_{bAs} in the buffer layer. The InGaAs well layer is assumed to grow coherently with InP and thus the lattice-distortion, if any, is expressed as c/a where c is the lattice parameter of InGaAs normal to the surface and a is that parallel to the surface and equal to that of InP. This model contains as many as 12 parameters to fit the model curve to one experimental curve and seems to have too much freedom of fitting. However, the CTR spectrum has the amplitude, phase shift, period, and decay as independent physical variables, and also the scattering factors for Ga atom and As atom (at the same time for In atom and P atom) are different. From the simulation and previous experiments, we know that in InP the Ga atoms tend to enhance the shoulder (amplitude) at the lower index and the As atoms enhance the shoulder at the higher index. The best-fit parameters at the lowest R-factor converged to quite a narrow range of the parameter values.

3.7.3. Results and discussion

In Fig. 22, obtained Ga and As profiles for InP/InGaAs/InP with the designed InGaAs well thickness of 3 ML and 5 ML (designated as InGaAs-

3ML and InGaAs-5ML, respectively) are shown. The vertical axis is in units of ML and 1 ML is about 2.9 Å for InP (001) orientation. First of all, Fig. 22 clearly shows the Ga and As distributions to 1-ML resolution and the widths of the well layers are almost equal to those designed. Total amount of As in InGaAs-3ML is calculated to be 2.8 ML and that in InGaAs-5ML to be 4.2 ML. Total amount of Ga in InGaAs-3ML is calculated to be 1.0 ML and that in InGaAs-5ML to be 1.8 ML. The total As amounts are quite close to those designed, but the total Ga amounts are by 20-30% less than those for the lattice-matching composition (0.47×3ML for 3 ML and 0.47×5ML for 5 ML).

As in the case of AsH_3-purged InP, the lower interface of InGaAs/InP is somewhat graded because there is atom exchange during the change of group-V source gas from PH_3 to AsH_3. Since this atom exchange takes place during the same period of the source gas change for InGaAs-3ML and 5ML, the areas of the graded composition in both lower interfaces are quite similar. It is also natural that there is a composition grading at the upper interface of InP/InGaAs. It is understood that the tail is longer in InGaAs-5ML since more amounts of As source gases remain on the reactor wall. It should be noted, however, the X-ray does not have a high spatial resolution and the interface roughness is also represented as the interface grading, in both top and lower interfaces, if the roughness is smaller in area than the X-ray coherent length (~1 μm).

As far as the CTR analysis is concerned, the thickness difference is very sensitively reflected to the modulation period of the spectra and other combinations of n_{cGa} and n_{cAs} failed in fitting to the CTR data points. An example of the effect of the layer thickness n_{hAs} on the spectral change is shown in Fig. 23. Since there are no other techniques to reveal the profiles to this resolution, it is not possible to make a further comparative evaluation of the accuracy. We have been using the resultant total amount of As in the layer as a test of the accuracy. It is because the amount of As at this concentration can be measured by fluorescence X-ray using SR as the excitation source [7, 11] and also because the total amount of As can be accurately controlled by the AsH_3 flow rate and flow duration by mass flow controller even though the distribution profile may not be controlled perfectly as expected by the gas switching sequence.

3.7.4. Summary

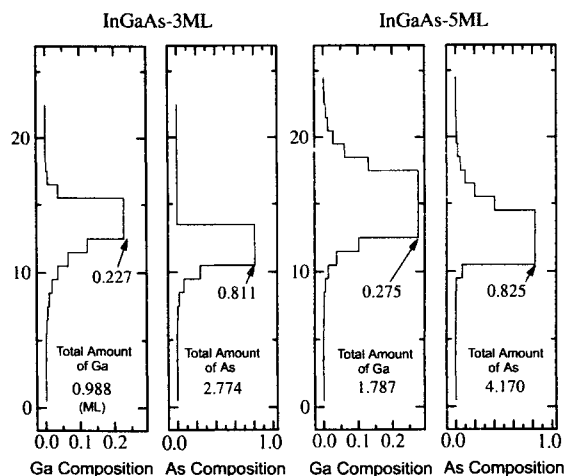The distributions of Ga and As atoms in the OMVPE-grown InP/InGaAs/InP single quantum-

Figure 22. Profiles of Ga and As atom distributions in InGaAs-3ML and InGaAs-5ML samples. Well thicknesses are close to those designed, but longer tails to the cap layer are observed in InGaAs-5ML.
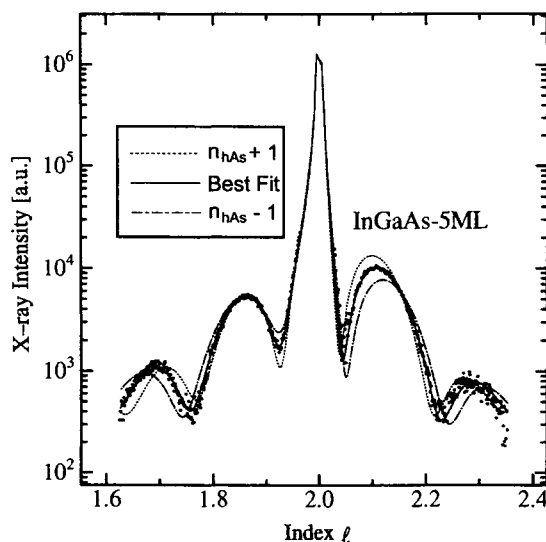

Figure 23. Effect of layer thickness n_{hAs} on the spectral change. Dotted line is the curve calculated with n_{hAs} 1 ML added to the best-fit value and dash-dotted line with n_{hAs} 1 ML subtracted from the best-fit value. Only 1-ML difference affects the spectral shape largely. Since the wavelength of the X-ray is ~1 Å, it is a good measure for the lattice spacings.

wells with 3 ML and 5 ML InGaAs-well were investigated by the modulation of X-ray CTR scattering caused by X-ray interference in the heterostructures, which results in CTR spectra. The CTR spectrum extending around 002 Bragg point was taken. By curve fitting of theoretical CTR spectrum to the experimental data, the Ga and As distributions were obtained to 1-ML resolution. Graded heterointerfaces at InP/InGaAs and InGaAs/InP extending into several monolayers were revealed. The widths of the well layers were almost equal to those designed. The total As amount was quite close to the designed one, but the total Ga amount was by 20-30% less than that for the lattice-matching composition.

3.8. InP/InErP/InP

There is a considerable interest in achieving good quality heterostructures of semimetallic compounds (RE-V) between rare earth (RE) elements and group-V elements on III-V semiconductors, which may lead to new electronic devices such as resonant tunneling devices [21]. III-V compound semiconductors themselves have been applied to various electronic and optical devices using heteroepitaxial techniques. To add new functions to these III-V semiconductor devices by utilizing the semimetals of RE-V compounds, a technique to grow RE-doped layers epitaxially on III-V semiconductors and a technique to characterize the structures are necessary.

From this point of view, we have been studying δ-doping of Er in InP [14–17]. For δ-doping, both of an Er doping process itself and a selective layer doping technique are required. At the same time, it is necessary to establish a characterization technique that can evaluate the thickness, the composition, the crystal structure, and the crystalline quality of the extremely thins layers.

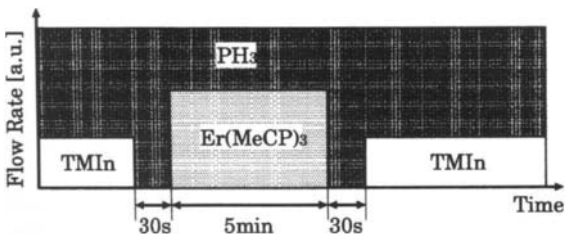In this section, the X-ray CTR scattering and

interference measurements are applied to study the Er δ-doped InP. The layer thicknesses and the distribution of Er in the δ-doped layer are clearly obtained, as is usual, by the X-ray CTR scattering analysis. In addition, it is demonstrated that the crystal structure of the δ-doped layer is determined.

3.8.1. Sample preparation

The Er δ-doped InP sample was grown by OMVPE with a vertical quartz reactor at 76 Torr using TMIn (trimethylindium), TBP (tertiarybutylphosphine), and Er(MeCp)₃ (tris(methylcyclopentadienyl)erbium) as source materials. Er(MeCp)₃ was maintained at 100°C and introduced into the reactor by H_2 flow of 125 sccm. The time sequence of source gases to grow the Er δ-doped layer in InP is shown in Fig. 24. The growth temperature was 530°C. Undoped InP buffer layer of 1000 Å in thickness was grown on Fe-doped InP (001) substrate. The TMIn supply was stopped to suspend the growth of InP, and the Er source was supplied for 5 min for δ-doping of Er on InP. Finally, a 100Å-thick undoped InP cap layer was grown. Before and after supplying the Er source, purge periods of 30 s were inserted to ensure the source gas change. TBP was continuously supplied during these steps.

3.8.2. Model structure

The model of the layer structure is schematically shown in Fig. 25. This model contains seven parameters, i.e., n_c, n_h, d_c, d_b, x_h, $<\Delta z^2>$, and c/a similar to the model shown in Fig. 12. The main difference is that the layer is composed of InErP

Figure 24. Gas flow sequence for growth of Er δ-doped InP.
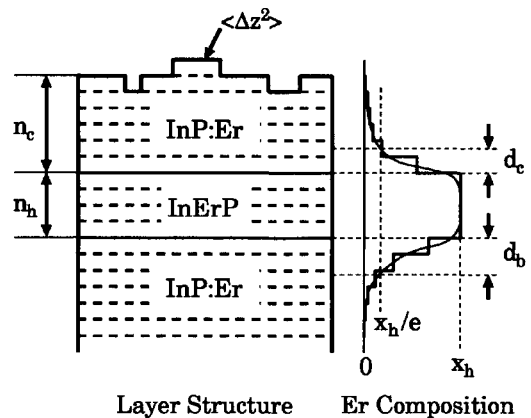

Layer Structure   Er Composition

Figure 25. Schematic drawing of the model structure and Er distribution.

instead of InPAs and the crystal structure is not necessarily the zincblende. As will be described later, Er atoms were assumed to be either substitutional or interstitial impurity in InP matrix. When the Er atoms are included in a layer as substitutional or interstitial impurity, the composition of the layer should be represented as $In_{1-x}Er_xP$ or $InEr_xP$, respectively.

d_c and d_b denote the extension of the Er atoms into the InP cap layer and into the InP buffer layer, respectively. The Er distribution is assumed to have the formula

$$x = x_h \exp\{-(n/d)^2\} \qquad (16)$$

where n [ML] is a distance from the upper or the lower interface. x_h and x are the peak Er composition in the δ-doped layer and in the layer at the distance n [ML] from the interface, respectively. In the equation, d is d_c or d_b.

In order to investigate the Er atom position in the InErP layer, four models of crystal structure were assumed for the InErP layer. Those crystal structures are schematically shown in Fig. 26. In Fig. 26 (a), Er atoms were assumed to form the rocksalt structure ErP. In Fig. 26 (b), Er atoms were assumed to substitute the In-site. In Figs. 26 (c) and (d), Er atoms were assumed to be on the Td (tetrahedral)-site with four P nearest neighbors and on the Td-site with four In nearest neighbors, respectively. Er composition x was varied in each monolayer in the curve fitting.

3.8.3. Results and discussion

In Fig. 27, the fitting results of the four curves calculated with four different positions of Er atoms to the measured CTR spectrum around the 002 Bragg point are shown by white lines. Each pair of curves and data was shifted by two orders of magnitude to avoid overlapping. The bottom pair is shown at the measured intensity. In the process of the curve fitting, data of index greater than 2.0 were given priority, in other words, right-hand side of the CTR spectrum was fitted at first. Data near the Bragg peak ($1.98 \le l \le 2.02$) were excluded from the fitting as usual since the intensity may be too strong for kinetic theory of the scattering. As shown in Fig. 27, measured CTR spectrum was fitted very well to the curve that was generated for the rocksalt structure ErP, but other curves for different crystal structures were unable to fit to the whole spectrum. EXAFS measurement of Er
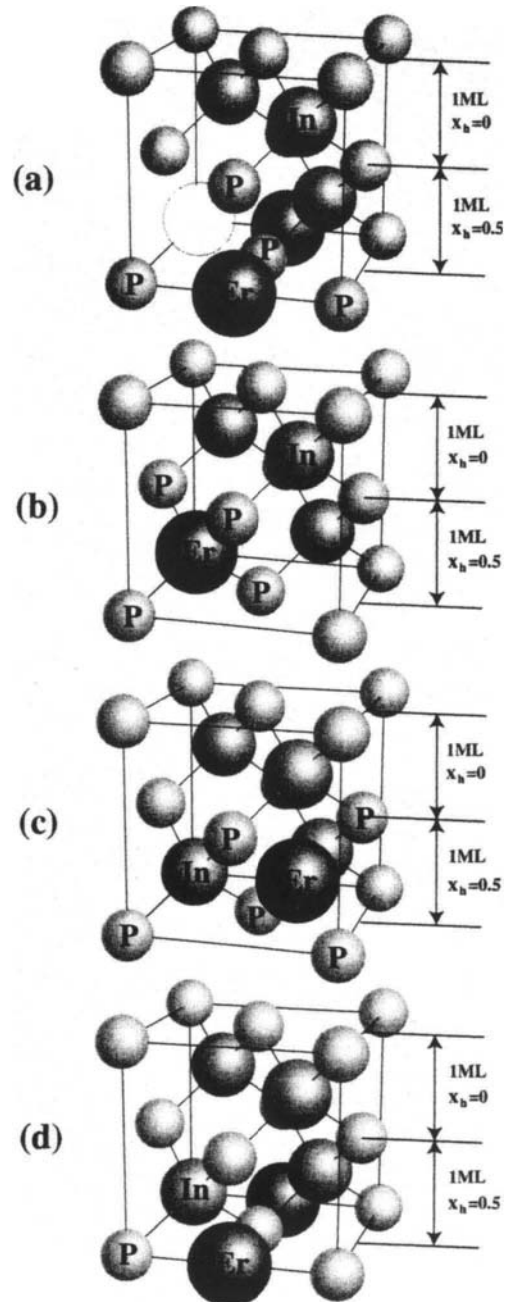

Figure 26. Schematic drawing of model crystal structures. In (a), Er atoms were assumed to form the rocksalt structure ErP. In (b), Er atoms were assumed to substitute the In site. In (c) and (d), Er atoms were assumed to be on Td (tetrahedral)-site with four P nearest neighbors and on the Td-site with four In nearest neighbors, respectively.

atoms also revealed that the Er atoms in the same δ-doped samples form the rocksalt structure [22]. It was reported that Er atoms formed rocksalt structure ErAs in MBE-grown GaAs [23, 24].

Values of the parameters at the best fit are listed in Table 2. The obtained distribution of the Er atoms is shown in Fig. 28 in the monolayer scale. The thickness of the InP cap layer (n_c) is nearly equal to the designed value (100 Å). When the thickness of the InErP layer n_h as the full width at half maximum (FWHM) in Fig. 28 is taken into the cap layer thickness, (n_c+n_h) is quite close to 100 Å. The value of $<\Delta z^2>$ was 0.104 ML^2. It means that the top surface of the InP cap layer was quite flat. The value of c/a was 0.983 which is close to the calculated value 0.985 with the lattice parameters of InP (5.8694 Å) and ErP (5.606 Å), respectively, and the assumed Poisson's ratio of ErP (1/3, to our

Table 2
Values of fitting parameters at the best fit to the measured data. Er atoms form rocksalt structure ErP in the δ-doped layer.

n_c	$<\Delta z^2>$	c/a	Total Er atoms
30 ML(88 Å)	0.104 ML^2	0.983	0.171 ML

knowledge there is no data of it) [25].

The FWHM of the Er distribution was 5 ML (15 Å). The Er distribution profile is almost symmetric. The Er peak composition was 0.032 and the total amount of the Er atoms was 0.171 ML. This amount is quite close to 0.2 ML, which was obtained from the Rutherford backscattering measurement on the same sample [14].

3.8.4. Summary

The sample of Er δ-doped InP was prepared by OMVPE. The δ-doped layer was formed by exposing InP surface to Er source gas for 5 min at 530°C. Layer structure of the Er δ-doped InP was analyzed by the X-ray CTR scattering measurement and X-ray interference using synchrotron radiation. From the result of the measurement and computer
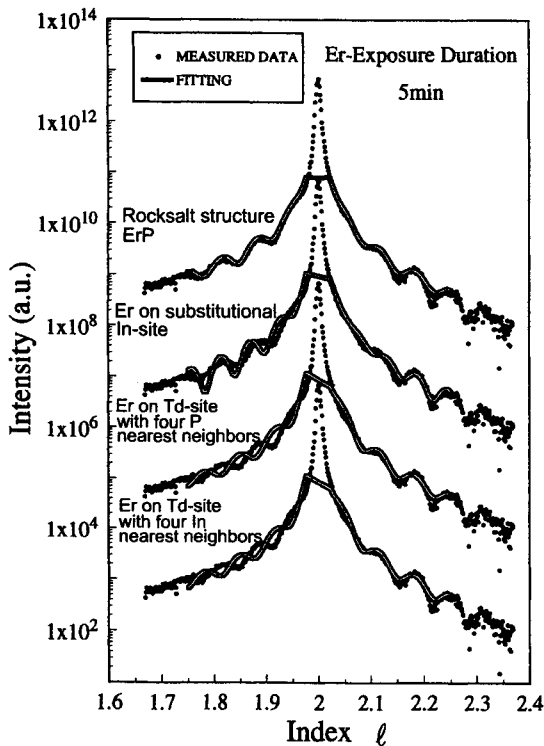

Figure 27. Solid circles show the measured CTR spectrum around the 002 Bragg point. White lines are the best-fit curves calculated with each model of the crystal structures in Fig. 26. Each pair of curves and data was shifted by two orders of magnitude to avoid overlapping. Data near the Bragg peak (1.98<ℓ<2.02) were excluded from the fitting.
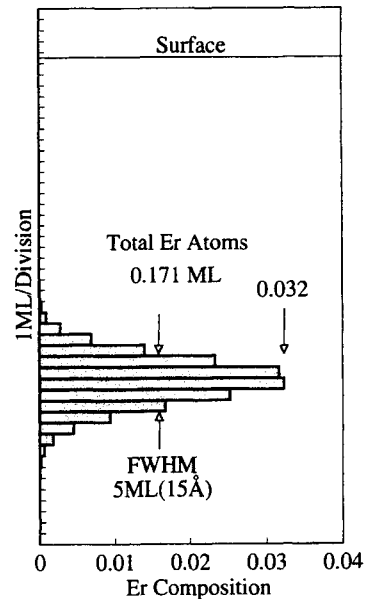

Figure 28. Obtained Er distribution. The peak Er composition is 0.032 and the FWHM is 5 ML (15 Å). The amount of Er is calculated as the area of dark region. It is 0.171 ML.

simulation, we revealed that the Er atoms δ-doped in InP formed the rocksalt structure ErP. The Er distribution was resolved in the order of 1 ML. Er atoms in the Er δ-doped InP layer were found to be confined in 5 ML (15 Å).

4. CONCLUSIONS

In the first half, use of asymmetric reflections to investigate the lattice-distortion in the heteroepitaxial layer due to a lattice mismatching was described. Useful equations to analyze those distortions from regular θ-2θ diffractometer measurements were introduced. Those equations were applied to an example of $Al_xGa_{1-x}Sb$ layers on GaSb substrates and tetragonal distortion of the lattice was analyzed.

In the second half, X-ray CTR scattering and interference measurements were conducted for a number of different samples prepared under various growth conditions and procedures, using synchrotron radiation (SR) as the X-ray source and the imaging plate (IP) as the two-dimensional detector. By the analysis of the CTR spectra for the samples, we obtained the atom distribution profiles in the samples.

In the InP/InPAs/InP samples made by AsH_3-exposure to InP surface, it was found that considerable amounts of As atoms existed in all the samples though the InP surface was only exposed to AsH_3, and that the P atoms on the surface are quite quickly exchanged by the As atoms. The exchange stops at the As composition of ~0.3, probably due to formation of a surface reconstructed structure.

The long tail of As atom distribution in the cap layer was probably due to the memory effect of the As on the reactor wall.

The distributions of Ga and As atoms in the OMVPE-grown InP/InGaAs/InP single quantum-wells with 3 ML and 5 ML InGaAs-well were investigated by the X-ray CTR scattering. Graded heterointerfaces at InP/InGaAs and InGaAs/InP extending into several monolayers were revealed. The widths of the well layers were almost equal to those designed. The total As amount was quite close to the designed one, but the total Ga amount was by 20-30% less than that for the lattice-matching composition.

The sample of Er δ-doped InP was prepared by OMVPE. The δ-doped layer was formed by exposing InP surface to Er source gas for 5 min at 530°C. From the result of the CTR scattering measurement and computer simulation, we revealed that the Er atoms δ-doped in InP formed the rocksalt structure ErP. The Er distribution was resolved in the order of 1 ML. Er atoms in the Er δ-doped InP layer were found to be confined in 5 ML (15 Å).

Atom distributions at the heterointerfaces and in the ultra-thin quantum-wells were revealed in the atomic scale by the X-ray CTR scattering and interference measurement using synchrotron radiation and imaging plate, which cannot be done by any other techniques. Those revealed structures were close to the designed one in some cases, but in most cases are very different from those expected, and were compared with the results obtained by other indirect techniques and discussed. It was concluded that in all cases the X-ray CTR scattering and interference measurement is a very powerful technique to investigate atomic scale structures at the heterointerfaces and quantum wells.

REFERENCES

1. Any standard textbooks on X-ray diffraction and X-ray crystallography. For example, "X-ray Diffraction" by B.E. Warren (Dover Publications, New York, 1990),
2. I.K. Robinson, Phys. Rev. B 33 (1986) 3830.
3. J. Harada, N. Kashiwagura, M. Sakata and H. Miyatake, J. Appl. Phys. 62 (1987) 4159.

336

4. I.K. Robinson, R.T. Tung, R. Feidenhans'l, Phys. Rev. B 38 (1988) 3632.
5. I.K. Robinson and D.J. Tweet, Rep. Prog. Phys. 55 (1992) 599.
6. T. Shimura and J. Harada, J. Appl. Cryst. 26 (1993) 151.
7. Y. Takeda, Y. Sakuraba, K. Fujibayashi, M. Tabuchi, T. Kumamoto, I. Takahashi, J. Harada and H. Kamei, Appl. Phys. Lett. 66 (1995) 332.
8. M. Tabuchi, Y. Takeda, Y. Sakuraba, T. Kumamoto, K. Fujibayashi, I. Takahashi, J. Harada and H. Kamei, J. Cryst. Growth 146 (1995) 148.
9. M. Tabuchi, N. Yamada, K. Fujibayashi, Y. Takeda and H. Kamei, J. Electron. Mat. 25 (1996) 671.
10. M. Tabuchi, K. Fujibayashi, N. Yamada, K. Hagiwara, A. Kobashi, H. Kamei and Y. Takeda, Inst. Phys. Conf. Ser. 145 (1996) 227.
11. M. Tabuchi, K. Fujibayashi, N. Yamada, Y. Takeda and K. Kamei, J. Appl. Phys. 81 (1997) 112.
12. Y. Takeda, M. Tabuchi, H. Hamamatsu and S. Ichiki, 14th Inter. Symp. Chemical Vapor Deposition' CVD XIV and EUROCVD 11 (1997) 2451.
13. M. Tabuchi, K. Fujibayashi, N. Yamada, K. Hagiwara, A. Kobashi, T. Iguchi, H. Kamei and Y. Takeda, J. Cryst. Growth 186 (1998) 48.
14. Y. Fujiwara, N. Matsubara, J. Yuhara, M. Tabuchi, K. Fujita, N. Yamada, Y. Nonogaki, Y.

Takeda and K. Morita, Inst. Phys. Conf. Ser. 145 (1996) 149.
15. M. Tabuchi, D. Kawamura, K. Fujita, N. Matsubara, N. Yamada, H. Ofuchi, S. Ichiki, Y. Fujiwara and Y. Takeda, Rare-Earth Doped Semiconductors II (MRS, Pittsburgh, 1996) 150.
16. K. Fujita, J. Tsuchiya, S. Ichiki, H. Hamamatsu, N. Matsunmoto, M. Tabuchi, Y. Fujiwara and Y. Takeda, Appl. Sur. Sci. 117/118 (1997) 785.
17. Y. Takeda, K. Fujita, N. Matsubara, N. Yamada, S. Ichiki, M. Tabuchi and Y. Fujiwara, J. Appl. Phys. 82 (1997) 635.
18. J. Harada, Acta Cryst. A48 (1992) 764.
19. H. Kamei and H. Hayashi, J. Cryst. Growth 107 (1991) 567.
20. E. Estop, A. Izrael and M. Sauvage, Acta Cryst. A32 (1976) 627.
21. N. Tabatabaie, T. Sands, J. P. Harbison, H. L. Gilchrist, and V. G. Keramidas, Appl. Phys. Lett. 53 (1988) 2528.
22. H. Ofuchi, J. Tsuchiya, S. Ichiki, H. Hamamatsu, N. Matsumoto, M. Tabuchi, Y. Fujiwara and Y. Takeda, Appl. Sur. Sci. 117/118 (1997) 781.
23. N. Jourdan, H. Yamaguchi, and Y. Horikoshi, Jpn. J. Appl. Phys. 32 (1993) L1784.
24. J. D. Ralston, J. Wagner, F. Fuchs, P. Hiesinger, and J. Schamalzlin, J. Cryst. Growth 111 (1991) 989.
25. K.-H. Hellwedge and A. M. Hellwedge, LANDOLT-BORNSTEIN 20, 546.

Principles and applications of optical crystals; their stoichiometry study

Shintaro MIYAZAWA

SHINKOSHA Co. Ltd.,

2-4-1 Kosugaya, Sakae, Yokohama
Kanagawa 247-0007, Japan

Optical crystals, which have a transparent range from the near-ultraviolet to infrared regions, are very attractive for diverse optical devices.   In particular, "active" crystals, which have various optical functions, are getting popular.   In this lecture, light propagation in crystals is discussed briefly and it is pointed out that light propagation is affected by the crystallinity and optical inhomogeneities associated with compositional fluctuations. How to obtain optically homogeneous single crystals is discussed in detail using the phase relations of $LiTaO_3$ and $Bi_{12}TiO_{20}$ as examples, and their stoichiometry is discussed from the viewpoint of crystal growth.   For optical devices, compositionally homogeneous single crystals are in high demand.   To obtain such crystals, the "congruent" composition should be established, and the phase diagram should be. reexamined in some useful optical crystals

1. Introduction

A large number of oxide and fluoride crystals have been synthesized in the past three decades, and they are now called "optical crystals".   The conventional definition of optical crystal; transparent to visible light or light of wavelength (380~780 nm). Because of this old convention, semiconductor crystals are not generally called optical crystals. However, modern optical detectors detect light of wavelength far outside of the 380~780 nm range. Moreover, high-intensity light sources such as lasers are now available with wavelengths significantly longer or shorter than the visible.   Therefore, it is well accepted to redefine optical materials to include all crystalline materials that have a transparent region within the electromagnetic wavelength range from ~150 nm (ultraviolet) to ~ 50 μ m (infrared) the ultraviolet (from ~150nm) to infrared (up to ~50 μ m).

At present, a large portion of research on optical crystals has advanced, and many useful materials have been developed artificially in singles using various crystal growth technologies.   Single crystals are grown successfully from a melt by the Czochralski (Cz), Bridgmann, Kyropoulas, TSSG-pulling and Floating-zone (FZ) methods.   However,

these mrthods can grown crystals of various quality, which is sometimes a problem.

This lecture will focus on how to grow single crystals of high optical homogeneity by reviewing the fundamental issues of optical crystals, and studies on establishing their phase diagrams.   For a case study, the stoichiometry of electrooptic $LiTaO_3$ and $Bi_{12}TiO_{20}$ is described in detail as typical optical crystals.   These crystal have now remained popular by tremendous innovations such as optical WDM (Wavelength Division Multiplexing) technology, mobile communication systems and photorefractive memories.

2. What is the optical crystal ?

Classification of crystalline materials based on optical properties was well established in optical mineralogy, and here we use the basic framework of classification of optical crystals [1].

2.1. Classification of optical crystals

All optical crystals are classified into two kinds: *Isotropic* crystals: In these crystals, the speed of monochromatic light is independent of the bivration direction.   In an isotropic medium, the bivration direction of a light ray is always perpendicular to the
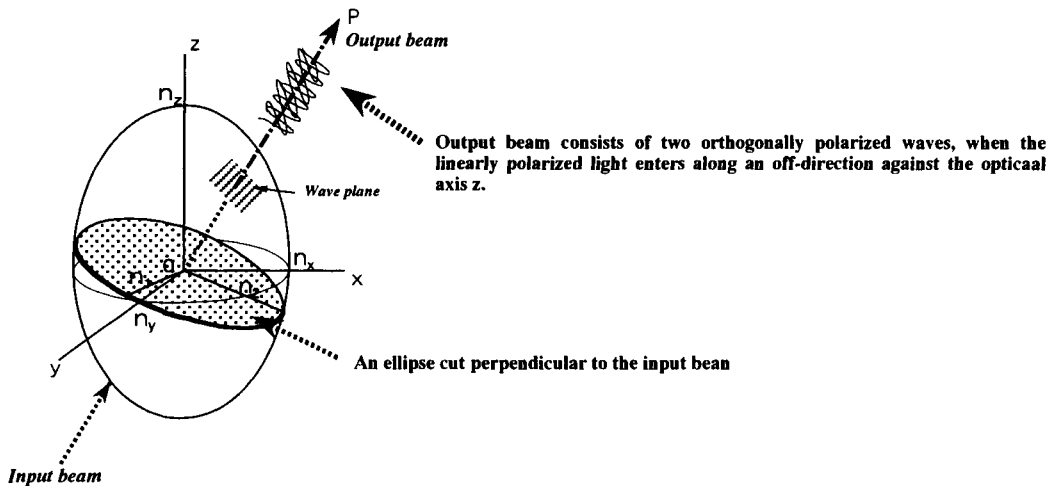
338

Figure 1 Relation between the optical indicatrix and light propagation

ray path direction. Only crystals with a cubic symmetry and all amorphous materials such as glasses are isotropic.

Anisotropic crystals : In these crystals , a light ray travels with different speeds for different vibration directions, and the angle between the vibration directions and ray path can differ from 90 $^\circ$. Moreover, the refractive index varies according to the vibration direction of the light. Thus, the optical indicatrix is an ellipsoid, as shown in Fig.1, instead of a sphere as it is for isotropic crystals. Depending on the geometry of the ellipsoid, it is necessary to divide the crystals further into two sub-categories, that depend on their crystal symmetry.

Crystals with tetragonal, hexagonal, or trigonal (or orthorhombic) symmetry exhibit a unique refractive index when a light vibrates parallel to the crystallographic c-axis (denoted as the extraordinary ray; n_e). For light vibrating at 90° to the c-axis, the refractive index is the same in all 360° directions (denoted as the ordinary ray; n_o). Crystals with this property are usually called "uniaxial crystals".

Crystals with orthorhombic, monoclinic, or triclinic symmetry posses three significant refractive indices, commonly symbolized as α, β, and γ in the order from smallest to largest. The shape of the indicatrix is a three-dimensional ellipsoid with all central sections being ellipses, except for two. These later two are circular sections of radius β. Crystals with these types of optical properties are called "biaxial crystals".

2.2. Category of optical crystals

There are passive and active optical crystals. "Passive" crystals are usually optically isotropic, whereas "active" crystals are usually single crystals with optically anisotropic properties that have many attractive optical functions when used with a coherent light.

Optically isotropic crystals are used most frequently for windows and lenses. A practical example is the cubic and isotropic crystal GGG ($Gd_5Ga_3O_{12}$), which is used now as a substrate for Faraday rotating devices. However, anisotropic crystals are also used as window. One is their use as a precisely oriented uniaxial single crystal such as sapphire (α-Al_2O_3) along the optical axis. The choice of sapphire is now very attractive because of its high mechanical strength and high thermal conductivity. Sapphire is also used widely as a substrate for recently developed GaN-related micro-/opto-electronic devices. It is grown by the Cz, EFG (Edge-defined, Film-fed Growth) and Verneuil techniques. Use as a "substrate" is another functions of optical crystals.

Anisotropic single crystals are widely used for many valuable optical applications. In fact, it is precisely the anisotropic properties that are needed for applications in the WDM optical information system. Such applications include polarizers, optical waveplates and wedges for which the crystals are "passive". Figure 2 shows how the wedges function in an optical isolator with a Faraday rotator made of a magnetooptic garnet crystal. The
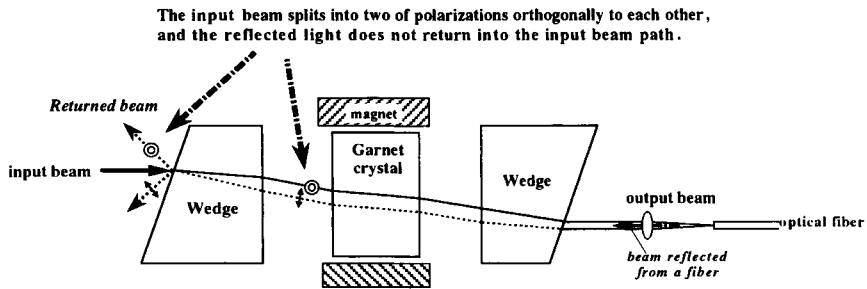
The input beam splits into two of polarizations orthogonally to each other,
and the reflected light does not return into the input beam path.

Figure 2 Principle of optical isolator consisted of birefringent wedges and Faraday
rotating granet crystals.

typical crystal used here for wedges is rutile (TiO$_2$), grown by the Verneuil method. It has a relatively large birefringence (Δ n~0.2), which is very attractive for beam splitting devices through an orthogonal polarization.

Moreover, most optical crystals are used as host crystals for solid-state lasers and optical nonlinear frequency conversion. Although the most widely used solid-state laser host crystal; YAG (Y$_5$Al$_3$O$_{12}$) is cubic and isotropic, the interesting trend is to look for linearly polarized laser sources that are generated

from anisotropic hosts such as YAP (YAlO$_3$), YLF (YLiF$_4$) and others. This is because linearly polarized coherent light is demanded for any nonlinear wavelength conversion such as harmonic generation, sum frequency mixing, and optical parametric oscillation. In nonlinear frequency conversion, all the optical crystals used at present must be highly anisotropic, non-centrosymmmetric, and should have a wide optical transparent region. Anisotropic crystals have many potential optical functions when used with a coherent light.

Consider the propagation of the light in an isotropic medium ;

from the Maxwell's equation

$$\frac{\partial^2 E}{\partial z^2} = \frac{\varepsilon \mu}{c^2} \cdot \frac{\partial^2 E}{\partial t^2}$$

, we deduce the following principal equation expressing the propagation of the light of electric field E;

$E = E_0 \exp[\, i\, (\omega t - kz)\,]$

$= E_0 \exp[\, i\omega(t - \frac{k}{\omega} z)]$

$\frac{k}{\omega} = \frac{1}{\sqrt{\varepsilon \mu}} = \frac{1}{\sqrt{\varepsilon_0 \mu_0}} \cdot \frac{1}{\sqrt{\varepsilon / \varepsilon_0}} = \frac{c}{N}$

$= E_0 \exp[\, i\omega(t - \frac{c}{N} z)]$

$N = n - i \kappa$

$= E_0 \exp[\, i\omega(t - \frac{n}{c} z)] \times \exp(- \frac{\omega \kappa}{c} z)$

$\underbrace{\qquad\qquad\qquad}_{phase\ term\ of\ light} \quad \underbrace{\qquad\qquad}_{amplitude\ term\ of\ light}$

Figure 3 Theoretical approach to the propagation of light in solid

3. Functions of optical crystals in applications

To understand the optical functions of optical crystals, let us consider the behavior of a light wave in a crystal. The crystal gives is a medium where the light wave interacts with external signals. For simplicity, we here consider the propagation of light in an optically isotropic medium. Figure 3 summarizes a theoretical approach to describe how light propagates in an isotropic medium. E_0 is an amplitude of a input light wave, ω is an angular frequency ($=2\pi f$, f; frequency), k is a propagation constant ($=2\pi/\lambda$, usually called the wave number), t is time, and z is the propagating length. ω/k is then the light's phase velocity in a medium as described in Fig.1. In this figure, ε is the dielectric constant, μ is the magnetic permeability (subscript 0 means in vacuum), c is the light velocity in vacuum, and N is a complex number composed of a refractive index n and an extinction number k, as $N=n-ik$.

external signals through changes of n, k, or z. These external signals are electric, magnetic, acoustic, optical, and strain fields. A refractive index of crystalline solids is changed by several parameters, as is summarized in Fig.4, by means of tensor expressions. Strain is also one of the main causes for a refractive index change, and hence compositional variations and fluctuations in a medium cause varying amounts of strains.

Figure 5 demonstrates schematically optical functions of active optical crystals, where the input light is modulated by signals applied external to the crystal. For example, an applied electric field provides a change in refractive index, allowing a change in the phase of light. This phenomenon, called an electrooptic (EO) effect, is used as an optical modulator/switch. The change of the parameters n, k or z, as is described in Fig1, produces the different functions shown in Fig.2, giving several
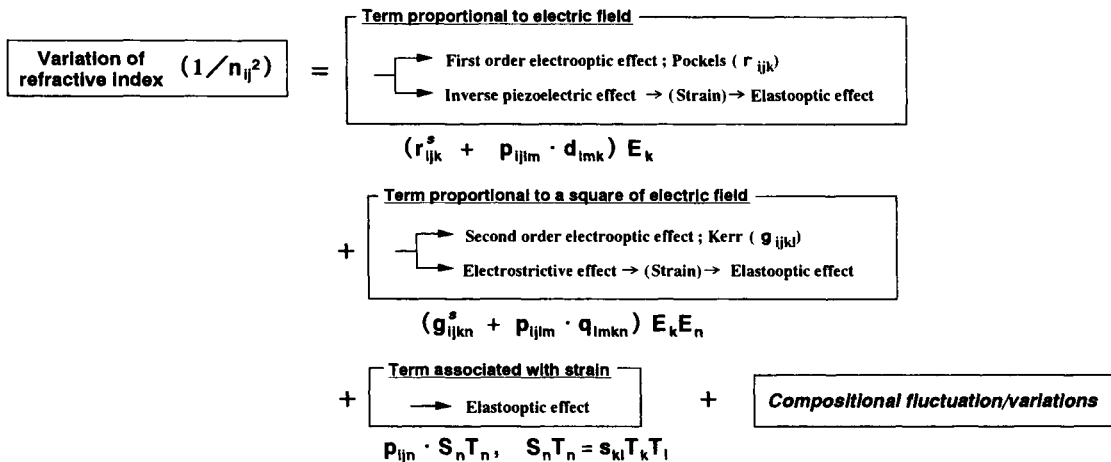

Figure 4 Parameters giving rise to a change of refractive index

Because the electric field E_0 of light from the boundary conditions when the light of E_0 enters into a medium, the final equation indicates that the propagation of a light in a medium is influenced by either n or k of the medium under consideration. The sign of n is always positive, whereas that of k is either negative or positive. Therefore, the first term of the final equation in Fig.3 gives the "phase" term and the second gives the "amplitude" term of the light. As a brief summary, the propagation mode of light in a medium, such as the phase, amplitude, and path, can be modulated by applying

important phenomena such as an electrooptic (EO), acoustooptic (AO), magnetooptic (MO), nonlinear optical (NLO), photorefractive (PR) effects, and lasing. These effects have been well investigated with many active crystals so far and practical optical devices are now developed. More recently, monolithically integrated devices that combine two effects are in progress. One example is the self-doubling laser $Nd:YVO_4$, pumped of laser diode (LD), where about $1\mu m$ emission of Nd ions in host is frequency-doubled ($\sim 0.5\mu m$) by means of NLO property of YVO_4 host crystal.
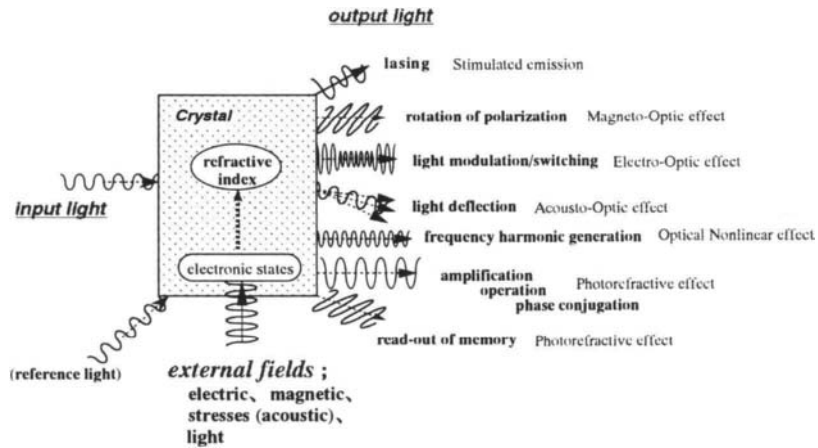
Figure 5 Various optical functiions of "active" optical crystals. The propagation of light in a crytal is modulated by changes of either phase (n), amplitude (k), or strain (z), which are induced by applied external fields such as electric field, magnetic field, optical pumping, and stresses.

One of the most important points, in turn, is that an inhomogeneity of a refractive index n in a medium should cause the propagation mode to fluctuate, which makes the device characteristics difficult to control. Therefore, optical crystals with high homogeneity of a refractive index n and a uniform concentration of dopants (k) are strictly required for device applications.

Most compound crystals have optical refractive indices that strongly correlate the composition ratio of their constituents. Therefore, it is very important in most optical applications to use crystal growth methods that produce single crystals with highly homogeneous composition and uniform dopant concentrations.

4. Basic study for the growth of optical crystals

We will focus on how to establish growth technology for obtaining single crystals with high optical homogeneity. The approach described here
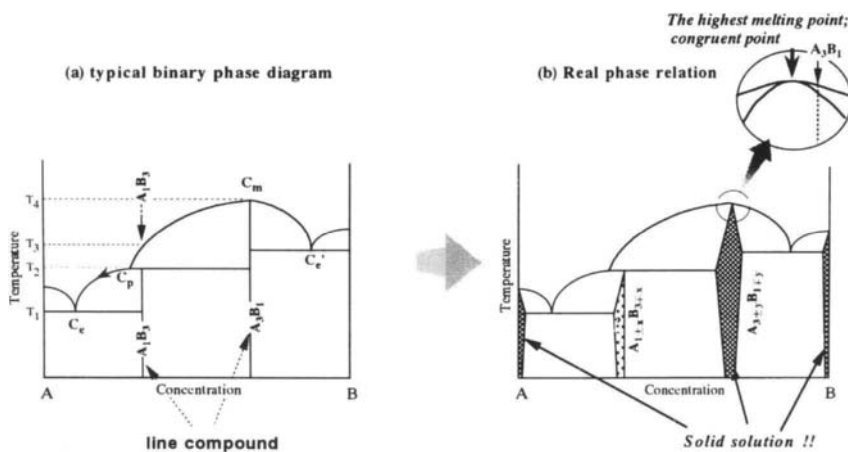

Figure 6 General illustrations of binary phase diagram (a) and real diagram (b).

requires understanding the phase diagram of a compound to be grown in single crystalline form. Figure 6(a) shows an illustration of general binary phase diagram, where end members A, B and compounds A_1B_3 and A_3B_1 are given as a line compound. However, recent crystal growth experiments have shown that each line compound has a solid solution region, as illustrated in (b). Even for mono-oxides, M-O_2 (M:metal ion) such as rutile (TiO_2), the binary phase relation has been reexamined [2].

In Fig.6(b), a very important point to be stressed is that the composition with the highest melting temperature (sometimes called the "congruent" point) does not coincide necessarily with a stoichiometric line compound, as is indicated in the insert of the figure. A more general definition of "congruent" is the crystal composition that equals the melt composition from which the crystal is grown. From the principles of crystal growth, the composition of a crystal grown from an A_3B_1 melt changes gradually along the growth direction, according to the "normal freezing" process along a solidus line. On the other hand, compositionally homogeneous single crystals can be grown from a melt of the congruent composition. Therefore, a minute phase relation close to a compound to be grown should be established for obtaining optically homogeneous crystals, because the refractive index is sometimes correlated tightly with the composition ratio of constituents, i.e., its *stoichiometry*.

As examples, we discuss how the phase relations of electrooptic $LiTaO_3$ and $Bi_{12}TiO_{20}$ single crystals [3] are established so as to grow them with high optical homogeneity. Recently, the WDM optical fiber communication system has captured worldwide attention; this system uses electrooptic crystals as an external high bit-rate optical modulator in the form of optical waveguide. A ferroelectric $LiNbO_3$ is now a standard crystal for this purpose, but its isostructural $LiTaO_3$ will be the next candidate to use in several related types of devices.

4.1. Determination of a congruent composition of $LiTaO_3$

The ferroelectric Curie temperature T_c of $LiTaO_3$ reported earlier has varied considerably from report to report, for instance, 665 °C [4], 660 ± 10 °C [5], 630°C[6], and 618°C[7]. In our experiments, the T_c of the single crystals grown from the

stoichiometric melt (Ta_2O_5/Li_2O=50/50 in molar ratio) was measured to be about 620°C, whereas that of a calcined ceramics was about 660°C. This discrepancy in T_c is probably caused by the compositional variations, because the stoichiometric melt does not provide a stoichiometric single crystal.

Then, to determine the congruently melting composition in the binary Li_2O-Ta_2O_5 system so as to obtain single crystals with high optical homogeneity, preliminary studies were undertaken as follows:

(1) Comparison of X-ray powder diffraction patterns of calcined ceramics with stoichiometric composition with that of a bulk poly-crystal prepared by melting the stoichiometric ceramics. If the stoichiometric and congruently melting compositions are the same, they would have the same diffraction patterns.

(2) Comparison among the X-ray powder diffraction patterns of ceramics with ratios of Li_2O/Ta_2O_5=1/3, 1/1, 2/1 and 3/1 in molar ratio, and consider the change in peaks corresponding to phases other than $LiTaO_3$.

(3) The congruently melting composition was determined by comparing T_cs of single crystals grown from different melt compositions with those of residual solids solidified in a crucible. The melt composition was varied from $Ta_2O_5/(Li_2O+Ta_2O_5)$ = 49.0 to 54.25mol%.
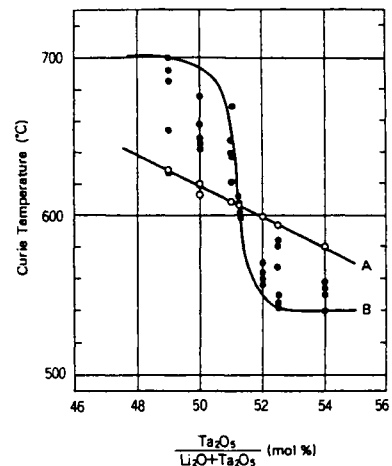

Figure 7   Stoichiometry dependence on The Curie temprature of $LiTaO_3$. Open and closed circles are of single crystals and solidified melts, respectively.

These experiments show the congruently melting composition does not coincide with the stoichiometric one. The congruently melting composition is assumed to be located at Ta_2O_5-rich side from the stoichiometric composition, as shown in Fig.7. The open circles represent the T_c of single crystals, whereas the filled circles are of residual solids. If the melt has thr congruent melting composition, the T_c of a single crystal grown from this melt coincides with that of any crystals picked up from the residual solids. So, we conclude that the congruently melting composition lies close to $Ta_2O_5/(Li_2O+Ta_2O_5) = 51.25mol\%$.[3]

The relatively wide dispersion of T_cs of the residual solids in a crucible can be easily understood when we use a simple binary solid solution system where the equilibrium coefficient k_o is larger than unity. From these experimental results, a tentative phase relation at the vicinity of $LiTaO_3$ can be drawn in Fig.8, where the axial temperature scale is arbitrary. The composition of single crystals grown from the stoichiometric melt contained less than 51.20 mol% Ta_2O_5, thus the distribution coefficient of Ta_2O_5 is determined as 1.02.
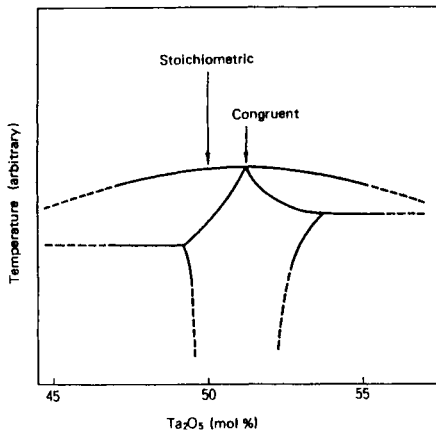

Figure 8 The established phase diagram around the stichiometric $LiTaO_3$

4.2. Evaluation of optical quality

Optical homogeneity of the optical single crystal should be evaluated quantitatively by measuring the dependence of the light extinction ratio (E.R.) on the light beam diameter. Based on the theoretical approaches [8,9], E.R. can be written by

$$E.R. = r^2 (\pi / \lambda)^2 \ \Delta_r n^2 l^2 \qquad (1)$$

, where r is the beam diameter, λ is the wavelength, $\Delta_r n$ is radial variation of refractive index in a beam, and l is sample thickness. Eq.(1) gives us that E.R. is proportional to r^2, when the refractive index variation $\Delta_r n$ is constant.

The measured results for the congruent crystal are shown in Fig.9, in which values of the two different portions in the crystal were plotted. The crystal diameter was about 10 mm. With an increase of the laser beam diameter from 1 mm ϕ up to about 6mm ϕ, E.R. decreases almost linearly from 0.25%
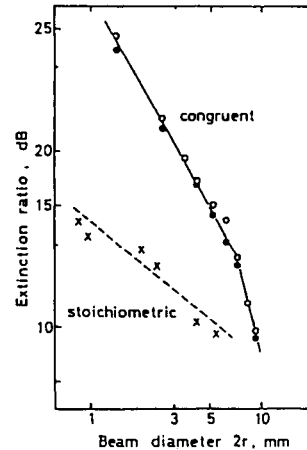

Figure 9 Light extinction ratio as a function of a beam diameter for the congruent $LiTaO_3$ crystal.

to about 4%. This linear tendency clearly obeys to Eq.(1). On the other hand, the E.R. of a crystal grown from the stoichiometric melt does not obey Eq.(1). This indicates the existence of a relatively large birefringence variation in the boule. This homogeneity evaluation was also verified by observing a macroscopic conoscopic pattern.

From the obtained light extinction ratios, a natural birefringence variation in a crystal boule can be estimated to be less than 8.9×10^{-6}/cm from a relation deduced from Eq.(1) as

$$\Delta_r n = (\lambda / \pi l) (r \sqrt{E.R.})^{-1} \qquad (2)$$

. Devices such as light modulators require birefringence variation less than 5×10^{-5}/cm, and then the congruent $LiTaO_3$ is quite applicable to

optical devices.

4.3. Growth-induced inhomogeneity in LiTaO₃

In the growth of single crystals by the Cz method, it is common to change the melt temperature little by little so as to control the crystal shape and dimensions. This temperature change of the melt alters the birefringence during the crystal pulling from the melt.

The distribution coefficient k can be specified as the equilibrium distribution coefficient $k_0 = C_s/C_L$ at the growth rate R=0, and as the effective distribution coefficient $k_{eff} = C_s/C_L$ at the growth in steady state, where C_s and C_L are the composition of the crystal and the melt, respectively. The effective distribution coefficient k_{eff} can be described as $k_{eff} = k^*F$, where $k^* = C_s/C_{L0}$ (C_{L0}; the composition of the melt at the growing interface) and F is dependent on growth rates, turbulence in the melt and properties of the solute. Burton et al., [10] expressed k_{eff} as following:

$$k_{eff} = k^*/\{k^* + (1-k^*)\exp(-\delta R/D)\} \quad (3)$$

, where δ is the diffusion boundary layer thickness of solute, D is the diffusion constant of solute, and R is the growth rate. Since the k_{eff} is a function of R, the birefringence change $d(\Delta n)$ associated with the temperature variation dT can be developed as follows:

$$d(\Delta n)/dT$$

$$= \{d(\Delta n)/dC_s\}\{dC_s/dk_{eff}\}\{dk_{eff}/dR\}\{dR/dT\} \quad (4)$$

Using typical values of oxides for the parameters of described in Eq.(4), we obtained $d(\Delta n)/dT$ to be 1.2 ~ 2.3×10^{-5}/°C.

We directly measured the birefringence change in a single crystal by observing the interference pattern at positions where the melt temperature was abruptly changed. The measured value of $d(\Delta n)/dT$ was 4.5~6.0 × 10⁻⁵/°C , which agrees well with the calculated value in magnitude. As a conclusion, the temperature of the melt should be controlled within the limit of ±0.25°C during the crystal pulling so as to keep the birefringence change below 10⁻⁵. This tolerance seems to be slightly troublesome in conventional Cz-pulling process,

when LiTaO₃ single crystals are grown from the *stoichiometric* melt. On the other hand, the second term on the right side of Eq.(4) becomes zero when growth is from the *congruent* melt. Therefore, we can obtain LiTaO₃ single crystals without any optical inhomogeneity.

More recently, Kitamura et.al. [11] succeeded in growing LiNbO₃ and LiTaO₃ with the stoichiometric composition by means of their originally developed pulling technology, i.e., a double crucible configuration. Surprisingly, their physical and optical properties are more attractive than those of the congruent crystals. This means that the crystals close to their stoichiometric composition will open new opportunities for device applications. The present author has proposed new optical devices using LiTaO₃ with either congruent or stoichiometric compositions [12].

4.4. Reexamination of Bi₁₂TiO₂₀ crystal growth

This section concerns the phase relation in a Bi₂O₃-rich region from the stoichiometric Bi₁₂TiO₂₀ (BTO;14.286 mol.% TiO₂) in the Bi₂O₃-TiO₂ binary system. The reproducible growth of single crystals is required for practical uses , such as an EOS (Electro-Optic Sampling) probe head [13]. The best known sillenide crystal is Bi₁₂SiO₂₀ (BSO), which melts congruently. Its stoichiometry has been well documented [14]. Among sillenides. BTO single crystals have the largest electrooptic constant r_{41}.

4.4.1. Reexamination of the phase diagram

According to the phase diagram reported first by Burton [15], BTO is given as a "line" compound that melts incongruently at 873~875°C at which BTO decomposes into liquid and Bi₄Ti₃O₁₂ associated with a peritectic reaction. Therefore, BTO is usually grown by the TSSG-pulling method from Bi₂O₃-rich solutions of less than about 12 mol.% TiO₂ [16,17]. However, a standard technology for growing BTO single crystals has not been established yet. For this purpose, Burton's phase relation was reexamined experimentally.

For establishing a minute phase diagram, DTA (differential thermal analysis) was carried out in order to reexamine the peritectic and eutectic temperatures. These temperatures were 823±3°C and 855±2°C, respectively, as shown in Fig.10, where Bruton's diagram is given for comparison.

Figure 10 Reexamined phase diagram of Bi_2O_3-TiO_2 hypo-peritectic region by DTA

And to verify the peritectic composition, TSSG-pulling was carried out from 10.25~11.25 mol.% TiO_2 solutions. As a result, a boule grown from a 10.75 mol.% TiO_2 solution consisted of $Bi_4Ti_3O_{12}$ polycrystalline and a monocrystalline BTO phase in the upper and lower parts, respectively. A lattice constant of BTO was 10.1743_9Å. This means that the first crystallized part beneath the seed was of lamellae-type $Bi_4Ti_3O_{12}$ and BTO crystallization after a while resulted from a gradual change in solution composition that passed the peritectic composition along the liquid curve. Therefore, we emphasize that the peritectic composition must lie close to a 10.75 mol.%TiO_2 or less, as indicated by the arrow in Fig.10.

To examine the existence of a solid solution range close to the stoichiometric, mixed binary compounds with 13.3~14.7 mol.%TiO_2 were calcined at $830\pm5°C$, and lattice constants of the BTO phase in each ceramics were evaluated by X-ray diffractometry (XD) by means of the Nelson-Riley approximation method. And single crystals from TiO_2 solutions ranged from 4.0 to 10.75 mol.%TiO_2 were grown, and their lattice constants were characterized by XD.

Figure 11 shows the lattice constants of BTO crystals as a function of the starting composition [18]. The lattice constant decreased monotonically with TiO_2 concentration, but those of crystals grown from a 9.0 mol.% TiO_2 solution show the minimum. This aspect indicates that the solidus curve must be retrograde, regardless of point defects such as vacancies, oxygen deficiencies and/or Bi_{Ti}^{3+}-anticites

[19].

In the case of the TSSG-pulling, the starting solution composition changes gradually along the liquidus line as the crystal grows. Then, the variation of the lattice constant along the crystal length was evaluated by growing long single crystals.

A single crystal about 53 mm long was grown from a 10.10 mol.% TiO_2 solution. Table I summarizes the lattice constant and Bi/Ti atomic ratio variations along the crystal. It can be recognized that the lattice constant did not vary in proportional to the length. The lattice constant did not vary in proportion to the length. This suggests that the lattice constant varied according to the retrograde solidus line through the "turning point".

Figure 11 Stoichiometry dependence on lattice constant of $Bi_{12}TiO_{20}$ single crystals. ◎ is the value of the crystal grown from a solution very close to the peritectic composition.

Table I Lattice constant and Bi/Ti atomic ratio variations along the crystal length of a single crystal grown from 10.10 wt% TiO_2 solution.

Portion	Lattice constant (Å)	Bi/Ti atomic ration
Top	10.1739_7	—
(1)	10.1736_6	12.13_3
(2)	10.1729_8	12.15_6
(3)	10.1736_3	12.29_7
(4)	10.1739_7	12.25_6
Tail	10.1744_3	12.21_5

4.4.2. New phase relation

Together with experimental results and the expected considerations described above, the relevant part of the phase diagram in the hypo-eutectic region from the BTO (14.286 mol.% TiO_2) is possible to draw. Figure 12 shows a renewed phase diagram with a plausible retrograde solid solution range close to the BTO. It appears that the "turning point" of the retrograde solidus curve would lie around 13.80~13.85 mol.% TiO_2. This retrograde characteristic is given in Table1. Single crystals grown from solutions between ~7.5 and 10.10 mol.% TiO_2 have a lattice constant deviation of less than ~1 $\times 10^{-4}$ Å. Consequently, we conclude that starting solutions with a 10.0~10.1 mol.% TiO_2 are very practical for growing relatively homogeneous single

crystals whose lattice constants are uniform within $\pm 1 \times 10^{-4}$ Å over the whole crystal boule, when the solidified fraction is less than 45%.

This is the first report of retrograde solid solution in incongruently melting materials, although the retrograde solidus curve has been well discussed in GaAs [20].

5. Summary

Optical crystals are now widely applied in many optical devices, because they have many functions when used with coherent light. By considering the principle of light propagation modes in a crystalline medium, we found that homogeneous crystals are necessary. For the growth of optical crystals, deep insight of the stoichiometry is a very important issue, and establishing the congruent composition is crucial for obtaining reproducible optical quality for most binary compounds.

The phase relation in the vicinity of $LiTaO_3$ within the Li_2O-Ta_2O_5 system was reviewed as a case study. It was found that the congruently melting composition is located at about 51.25 mol% Ta_2O_5. The optical homogeneity of the congruent crystal was much higher than that grown from the stoichiometric melt. By measuring the compositional dependence of refractive indices, the birefringence change corresponding to the change of melt temperature during the crystal pulling from the stoichiometric melt was determined. The temperature variation of birefringence within the crystal boule was estimated as about 2×10^{-5}/°C.

We have to pay attention to the stoichiometry of crystals in order to find new properties, hence the

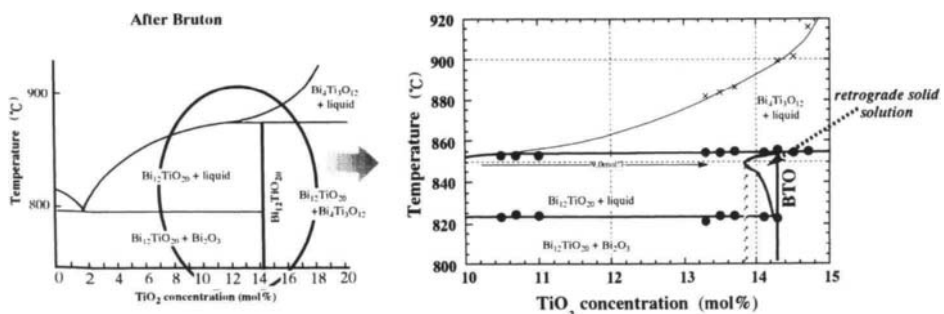**Figure 12** Renewed phase relation of Bi_2O_3-TiO_2 hypo-peritectic region showing the retrograde solid solution of $Bi_{12}TiO_{20}$

study on phase relation with respect to crystal growth technologies is quite important for developing "new" optical crystals. The phase diagram is analogous to a "compass" for growing single crystals with high crystalline quality. Therefore, we have to examine the phase relation, that is the "stoichiometry", of the crystals prior to or during the developing of materials.

How "good" a given crystal material is a complicated function of device types, operational conditions, device density, process parameters, and cost. In microelectronics and photonics, where device size decreases, device density increases and device functions become multiplex, thus the bulk single crystal is the key to the definition of device yield, when material/device process parameters are well controlled. It is then impossible to define the material quality regardless of these factors. Therefore, it is strongly recommended to know the causal relationship between the stoichiometry and material properties so as to grow single crystals with high quality.

References

1. J. F. Nye, Physical Properties of Crystals, Oxford, Great Britain, 1967
2. P. G. Wahlbeck and P. W. Gilles, J. Amer. Ceram. Soc., 49 (1966) 180
3. S. Miyazawa and H. Iwasaki, J.Cryst.Growth, 10 (1971) 276
4. Z. I. Shapiro, S. A. Fedulov and Yu. N. Venevtsev, Fiz.Trevd.Tela., 6 (1964) 316
5. H. J. Levinstein, A. A. Ballman and C. D. Capio, J.Appl.Phys., 7 81966) 4585
6. Z. I. Shpiro, S. A. Fedulov, Yu. N. Venevtsev and L. G. Rigerman, Sov. Phys. Crystallogra., 10 (1969) 725
7. A. M. Glass, Phys.Rev., 172 (1968) 172
8. K. Sugibuchi, H. Tsuya and Y. Fujino, Appl. Phys. Lett., 13 (1968) 107
9. S. Miyazawa, Kougaku Kesshou (in Japanese), Baifukan, Tokyo, 1995
10. J. A. Burton, R. C. Prim and W. P. Slichter, J. Chem.Phys., 21 (1987) 1953
11. Y. Furukawa, K. Kitamura, E. Suzuki and K. Niwa, J. Cryst. Growth 197 (1999) 889
12. S. Miyazawa (unpublished)
13. M. Shinagawa, T. Nagatsuma and S. Miyazawa, IEEE Trans. Instr. & Meas., 47 (1998) 235
14. J. C. Brice, M. J. Hight, O. F. Hill and P. A. C. Whiffin, Philips Tech. Rev., 37 (1977) 250
15. T. M. Bruton, J.Sol.Stat.Chem., 9 (1974) 173
16. D. Rytz, B. A. Wechsler, C. C. Nelson and K. W. Kirby, J.Cryst.Growth, 99 (1990) 864
17. Y. Okano, H. Wada and S. Miyazawa, Jpn.J.Appl. Phys., 30 (1991) L1307
18. S. Miyazawa, Korean Assoc. Cryst. Growth, 9 (1999) 424
19. R. Obershmid, Phys.Sta.Sol.(a), 89 (1985) 263
20. K. Terashima, J. Nishio, A. Okada, S. Washizuka and M. Watanabe, J.Cryst.Growth, 79 (1986) 463

This Page Intentionally Left Blank

PART III

Dynamics of Crystal-liquid Interface

This Page Intentionally Left Blank

Surface X-ray diffraction studies of crystal growth

Elias Vlieg, Marianne Reedijk and Jelena Arsic.

NSR-RIM Dept. of Solid State Chemistry, University of Nijmegen,
Toernooiveld 1, 6525 ED Nijmegen, The Netherlands.

One of the factors determining crystal growth processes is the atomic or molecular structure of the growth interface. For crystals growing in a vacuum environment such structural information is available, but this is not the case for crystals growing from a solution. X-ray diffraction is one of the few techniques that can be applied for this purpose and it is starting to provide information on the structure of both the solid and the liquid side of a growing interface. After a brief introduction of the technique of surface X-ray diffraction we will discuss results on various systems and using various growth methods, including molecular-beam epitaxy and solution growth. By selecting appropriate diffraction spots, one can observe specific structural features like stacking faults or surface roughness. At the solid-liquid interface, both the solid and the solution are found to deviate from their bulk structure.

1. INTRODUCTION

The growth of a crystal occurs at the interface with its growth environment: a vapour, a melt, a solution or even a solid. Many factors determine the actual growth process and the final growth shape (habit) of a crystal, e.g. supersaturation, impurities and temperature. One of the dominant factors in the habit of a crystal is the crystallographic structure: whether a crystal is plate-like, cubic or needle-like is mostly caused by the specific topology and bond strengths of its building blocks. The Hartman-Perdok theory [1], which predicts the crystal growth morphology by calculating which facets are the most stable, assumes that the atomic positions at the surface are the same as in the bulk of a crystal. For crystal surfaces in vacuum, it is well known that this is often not the case: a surface may exhibit relaxation or reconstruction. Such structural rearrangements will lower the free energy of a surface and may thus change the stability of one face with respect to another. In other words, the morphology of a crystal may change due to the detailed structure of the surface. The large amount of knowledge on the structure of surfaces in vacuum has been obtained using a variety of surface-sensitive techniques. Many of these employ electrons or ions and cannot be used outside

vacuum. For this reason, very little is known about the detailed surface structure of crystals grown in other environments like a melt or a solution. Two techniques, scanning-probe microscopy (SPM) and surface X-ray diffraction (SXRD) are now starting to provide interface information with similar detail as common in surface science.

SPM techniques, such as scanning-tunnelling microscopy (STM) and atomic-force microscopy (AFM), are very suitable for non-vacuum applications. Since many crystals are non-conducting, AFM is in practice the most common technique in studies of solution growth [2, 3]. For a full understanding of crystal growth, also a microscopic technique is required, because the local structure on a crystal surface (e.g., steps, defects and impurities) is one of the determining factors in growth. Here, however, we will only discuss SXRD. Being a diffraction technique, SXRD provides an averaged picture of a growth system, but at a much higher resolution than AFM. The two techniques are thus complementary and from a powerful combination for the study of crystal growth.

2. SURFACE X-RAY DIFFRACTION

The technique of surface X-ray diffraction (SXRD) has established itself as a valuable tool for surface crystallography [4]. The main reason why
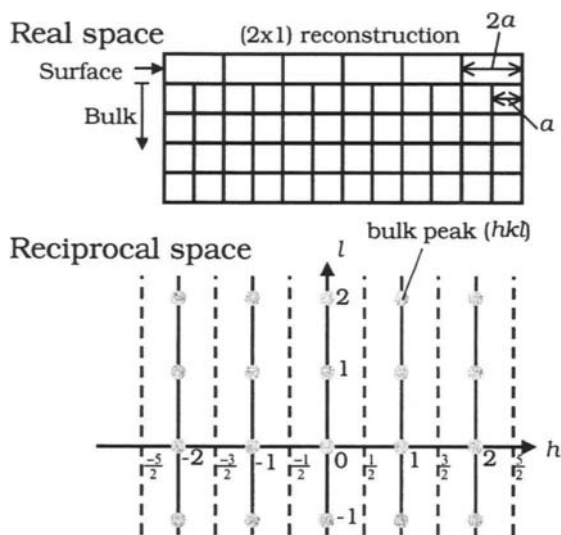
Real space   (2x1) reconstruction   2a

Surface →

Bulk

Reciprocal space   l   bulk peak (hkl)

Fig. 1. Schematic showing the position of surface sensitive information in reciprocal space. The bulk crystal yields only intensity at points (*hkl*) in reciprocal space. The surface is assumed to be (2x1) reconstructed, which gives rise to half-order diffraction rods in reciprocal space (dashed lines). The bulk peaks are connected by tails of weak intensity (crystal truncation rods) that are directed perpendicular to the surface. At positions not too close to a bulk peak, the intensity in such a tail is sensitive to the surface structure.

X-rays, given their dominant role in bulk crystallography, were not used in surface crystallography before the 1980's, was the lack of sufficiently intense X-ray sources. The diffracted intensity from a surface is about a million times less than that of a typical bulk crystal. Only when powerful synchrotron radiation sources became available, X-ray diffraction from a surface became feasible.

The main characteristic of X-rays is their low scattering cross section with matter. This has three main consequences: (1) a straightforward interpretation of the diffracted intensity, because single scattering theory applies, (2) a low yield, hence the need for synchrotron radiation sources and (3) a large penetration depth. The latter property makes X-ray diffraction an ideal tool to study the surfaces of crystals not only in vacuum, but also in gas, liquid or solid environments.

While the large penetration depth of X-rays is a necessity for *in situ* studies in non-vacuum environments, we need to discuss how this is compatible with surface sensitivity. The main trick here is to look at the appropriate places in reciprocal space. The situation is illustrated in figure 1. A bulk crystal has only intensity at sharp spots in reciprocal space, denoted by *integer* values of the diffraction indices (*hkl*). (By convention the index *l* describes the out-of-plane direction.) At these spots, the bulk contribution completely dominates the diffracted signal. For surface sensitive information, these points should thus be avoided. There are two types of locations in reciprocal space where the surface structure can be observed. First of all a surface may be reconstructed, i.e., the unit cell may be a multiple of the bulk unit cell. This larger unit cell in real space gives a smaller unit cell in reciprocal space. Expressed in units of the bulk unit cell, this leads to *fractional*-order reflections, with non-integer values for *h* and/or *k*. Since the surface is non-periodic in the perpendicular direction, one obtains rods of diffracted intensity. The bulk crystal does not contribute to such fractional-order rods, irrespective of the penetration depth, because its scattering is cancelled due to the symmetry of the crystal.

If a surface is not reconstructed, one can still measure surface-sensitive reflections by using a second type of rods. For a crystal that terminates in a flat surface, it turns out that the bulk reflections are connected by tails of diffuse intensity in the direction perpendicular to the surface. These are called crystal-truncation rods (CTR's) [5, 6]. On a crystal truncation rod the intensity varies continuously as function of *l* and reaches a high maximum for integer values of *l* at the bulk peaks. Maximum surface sensitivity is obtained exactly midway between these bulk peaks. Each rod is labelled by the diffraction indices (*h,k*).

The signal from the surface is weak and it is therefore important to minimise the background scattering from the bulk crystal. For this reason, diffraction experiments are typically performed at a small incoming or outgoing angle, because then the penetration depth is reduced. In the limit of very small angles (≤ 0.2°), total reflection occurs and the penetration depth is reduced to the 100 Å range. The main method to obtain surface sensitivity, however, is the selection of the appropriate diffraction rods.

An experimental set-up consists of a sample environment chamber that is coupled to a so-called diffractometer [4]. This is a big manipulator that
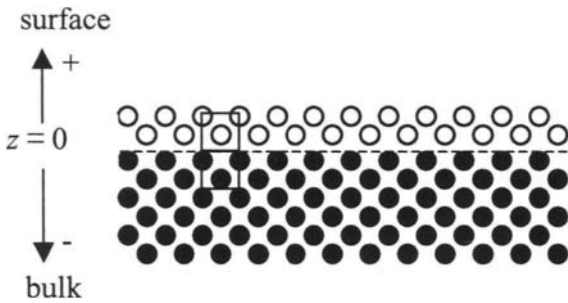
Fig. 2. Schematic of a surface layer on top of a bulk crystal that is assumed to extend to minus infinity. Unit cells (of equal size) for both sides are indicated by outlined squares. In calculating a theoretical structure factor, all atoms that are not at bulk positions are included in the surface unit cell.

allows accurate angular positioning and scanning of both the sample and the X-ray detector. In a typical experiment, one measures the integrated intensity of a slice out of a diffraction rod by rotating the crystal and collecting the scattered photons with a fixed detector. This is called a rocking scan. Using standard procedures [7] this integrated intensity can be converted into the amplitude of the so-called structure factor F_{hkl} for this location in reciprocal space. Structure determination using X-ray diffraction revolves around these structure factors.

The structure factor that one measures is the sum over all bulk and surface contributions from the crystal (fig. 2). The structure factor of a single bulk unit cell is defined as [8]:

$$F_{hkl}^{u} = \sum_{j}^{unit\ cell} f_j e^{-M_j} e^{2\pi i (hx_j + ky_j + lz_j)} , \qquad (1)$$

with f_j the atomic scattering factor of atom j, M_j the Debye-Waller factor that accounts for thermal vibrations and $(x,y,z)_j$ the position of the atom in the unit cell in fractional coordinates (i.e., expressed as fractions of the lattice vectors). In order to calculate the contribution of the bulk crystal to a diffraction rod, we need to sum over all bulk unit cells with the appropriate phase factors, starting at the surface and extending into the crystal:

$$F_{hkl}^{bulk} = \sum_{j=-\infty}^{0} e^{2\pi i l j} e^{\alpha j} F_{hkl}^{u} = \frac{F_{hkl}^{u}}{1 - e^{-2\pi i l} e^{-\alpha}} , \qquad (2)$$

where α is a factor that describes the attenuation of the X-ray beam when it penetrates the bulk crystal. Eq. (2) describes the structure factor of a crystal truncation rod [6]. Figure 3 shows such a rod for the case of a simple cubic lattice. The intensity is very large for integer values of the diffraction index l (i.e., at the bulk reflections) and reaches a minimum value for intermediate l values.

The surface contribution to the structure factor is given by an expression completely analogous to eq. (1), where now the summation is over all the atoms in the surface unit cell. In this surface unit cell we put all atoms that are not at bulk positions. The total structure factor is the interference sum of the bulk and the surface contributions:

$$F_{hkl} = F_{hkl}^{bulk} + F_{hkl}^{surf} . \qquad (3)$$

For a fractional-order rod, $F^{bulk} = 0$ and only the surface contributes to the intensity. The total structure factor is sensitive to surface relaxation, reconstruction and roughness [9]. Figure 3 shows an example of the changes in a crystal-truncation rod when a crystal surface is relaxed. The difference with a non-relaxed surface is only visible away from the bulk spots.

Unfortunately, a measurement does not yield the complete structure factor, but only its amplitude.
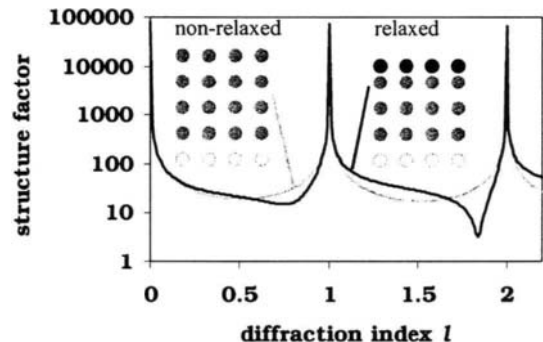

Fig. 3. The variation of the structure factor amplitude as a function of the diffraction index l along a crystal-truncation rod for a simple cubic crystal. For integer values of l one finds diffraction spots from the bulk crystal with very high intensity. In between these spots, the structure factor is surface sensitive. The grey curve is a calculation for a crystal with the surface atoms at the same positions as in the bulk, the black curve corresponds to a 10% inward relaxation of the topmost layer.

This is the famous phase problem in crystallography [8]. Instead of a simple Fourier transform, model calculations and fitting are therefore required to determine the structure [10]. Luckily, also model-independent information can be obtained, because the single scattering character of X-rays makes Fourier methods like the Patterson function and the electron-density difference map applicable [4]. Moreover, the model calculations are simple and fast, so quite complicated systems can be handled in principle.

The *integrated* intensity of a reflection is the quantity necessary to determine a structure factor and thus to obtain crystallographic information. The *profile* of a reflection contains a second important piece of information, because it depends on the long-range order of the surface. Disordered surfaces give broad diffraction profiles, while well-ordered surfaces give narrow ones [9, 11]. Measuring reflection profiles during growth thus yields information about the surface morphology.

3. EPITAXIAL GROWTH

Two common methods to grow an epitaxial film are molecular-beam epitaxy (MBE) and chemical vapour deposition (CVD). The growth process can be accurately followed using diffraction, since the reflected signal is sensitive to the surface roughness. Typically, the specular beam, or (0,0)-rod, is used for this, i.e. the diffracted signal without in-plane momentum transfer. Maximum sensitivity to surface roughness is obtained for points on this rod exactly in between two bulk peaks. In the schematic diagram of figure 1 this would thus be at the (0,0,0.5) reflection. At this point the signal from one layer is exactly out of phase with that of the underlying one (destructive interference). Because there is no in-plane momentum transfer, the specular beam is not sensitive to the in-plane crystalline quality of the deposited layers. This can be seen from eq. (1), because for $h = k = 0$, the structure factor is independent of the x and y coordinates of the atoms.

Figure 4 explains the behaviour of the diffracted signal for the various growth modes that can occur as a function of the substrate temperature. At high substrate temperatures (fig. 4a), the deposited atoms are very mobile and diffuse over the surface until a
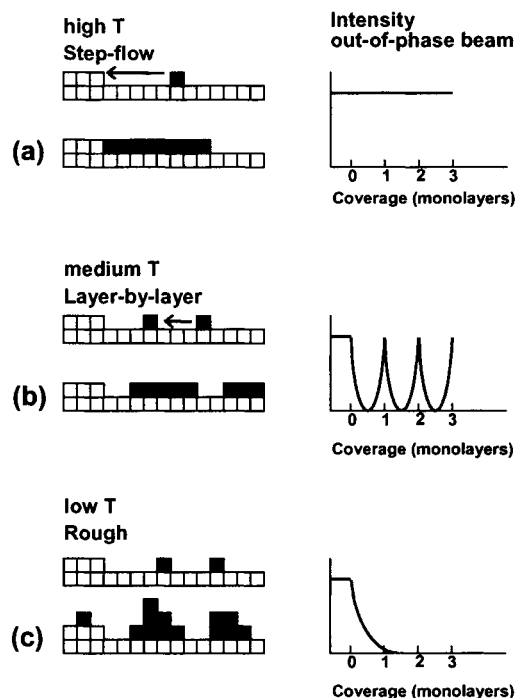
Fig. 4. A schematic diagram of the growth modes during homoepitaxial MBE or CVD growth of a crystal as a function of the substrate temperature. On the right, the corresponding diffracted signal is shown for a reflection for which the scattering amplitudes of consecutive layers are exactly out-of-phase. For decreasing temperatures, the growth modes vary from (a) step-flow, (b) layer-by-layer to (c) rough.

step edge is encountered. They attach to this edge and are incorporated in the growing terrace when more atoms arrive. This regime is called *step-flow* growth. Since the roughness on the surface does not change, the diffracted X-ray intensity remains constant.

For lower substrate temperatures (fig. 4b), the atoms do not reach a step edge, but encounter other deposited atoms with which they form islands ('2-dimensional island nucleation'). Initially the islands grow and the diffracted intensity decreases, because of the destructive interference of the signal from the islands with that of the lower lying layer. At 0.5 monolayer coverage, the diffracted intensity becomes zero. Upon further deposition, the islands start to coalesce and the intensity increases again. If
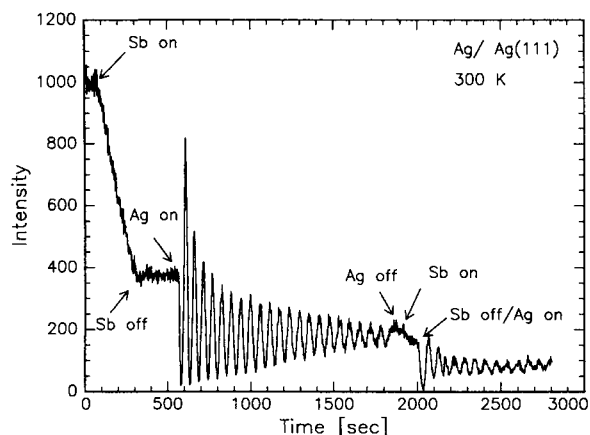
Fig. 5. The intensity of the (0,0,1.5) reflection (specular beam) during successive deposition of Sb (0.2 monolayers) and Ag at room temperature. Without Sb no layer-by-layer growth occurs, thus no oscillations are observed. From ref. [12].

the layer would perfectly close itself before islands are formed on the next level, the intensity would reach the starting level and one speaks of ideal *layer-by-layer* growth. One would observe an oscillatory intensity with constant amplitude and with a period corresponding to the growth of one monolayer. In practise, this ideal behaviour does not occur and damped oscillations are found instead. Finally, at sufficiently low temperatures, the atoms are largely immobile and the surface quickly becomes rough (fig. 4c). During this *rough* growth, no oscillations are observed and the signal decays rapidly.

The behaviour sketched in figure 4 was indeed observed in the first X-ray study of a growing crystal [13], in which Ge was deposited on Ge(111) using MBE. X-ray oscillations with monolayer period were measured, similar to the oscillations found in such growth systems using reflection high-energy electron diffraction (RHEED) [14]. In observing the growth mode or speed, X-rays offer no advantages over RHEED. In fact, X-rays are a far more elaborate and expensive method for this. However, in interpreting the scattered intensity and deriving surface crystallographic information, X-rays have clear benefits. In case of Ge(111) it was, for example, found that at low temperatures, growth does not proceed via the full bilayers of the diamond lattice, but that partial layers occur as well [13].

Epitaxial growth can also be observed when CVD is used as a growth technique [15]. Typical system pressures during growth are 100 mbar, a pressure at which RHEED can no longer be used. A recent example of such experiments is the metal-organic CVD growth of GaN by Stephenson et al. [16].

The behaviour sketched in figure 4 is not always observed. For example, in the case of the homoepitaxial growth of Ag(111) the intensity oscillations characteristic of layer-by-layer growth did not occur [12]. The reason for this was found to be the Ehrlich-Schwoebel barrier (island-edge diffusion barrier) [17]. This barrier prevents most atoms that land on top of an island from descending to the lower-lying terrace. This leads to a quick roughening of the surface. For Ag(111) the situation is completely changed when sub-monolayer amounts of Sb are added as a 'surfactant': then layer-by-layer growth does occur. Figure 5 shows the measured X-ray oscillations. These observations used the specular beam that is insensitive to the lateral ordering of the growing layer.

In order to determine the lateral ordering, a reflection with in-plane momentum transfer has to be measured. Figure 6a shows the intensity of the (0,1,0.3) reflection during Ag deposition after an initial 0.3 monolayer of Sb was deposited. For normal crystal growth, the intensity oscillations should behave as shown in figure 5, with a decaying amplitude never exceeding the initial intensity. The observations using the (0,1,0.3) reflection are very different from this and indicate that the stacking of the top layer is wrong. At a coverage of 1/3 ML Sb, the system reconstructs with the Sb atoms occupying substitutional sites and with a top layer that has the wrong stacking [18]. Figure 6b illustrates the major rearrangements that take place during growth. The Sb segregates to the top layer, while the lower lying Ag atoms return to their correct crystallographic positions. One thus observes effectively a floating stacking fault [19]. At lower Sb coverage, this stacking fault does not occur.

Seeing the floating stacking fault requires a penetration depth of only a few layers. This is always the case with X-rays, but already impossible (or difficult) with many other probes. When stacking faults occur deeper in a crystal, X-rays soon become the only option for *in situ* observations [20].
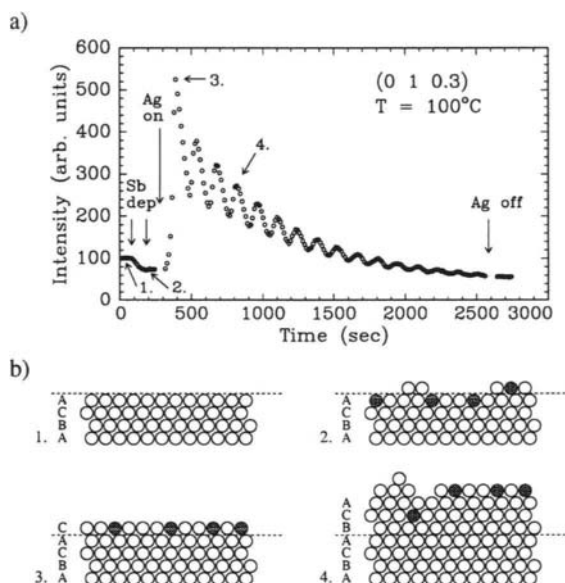
a)

b)

Fig. 6. (a) The intensity of a reflection with an in-plane diffraction component, $(h,k,l) = (0,1,0.3)$, which is sensitive to the in-plane positions of the growing layers. A deposition of 0.3 monolayer Sb is followed by 16 monolayers of Ag at 100°C. The numbers indicate the situations depicted in (b). (b) Schematic side view of the surface during different stages of deposition. Open circles represent Ag atoms and filled circles represent Sb atoms. On the clean Ag(111) surface with the ABC stacking of an fcc crystal (1), 0.3 monolayer of Sb is deposited (2). After a total deposition of 1 monolayer Ag (3), all top layer atoms have the wrong stacking (C instead of B). After continued deposition (4), the surface atoms are partly correctly stacked and partly wrongly stacked. The (buried) starting surface has returned to the correct fcc stacking. From ref. [19].

During growth not only the total roughness is important, but also the lateral ordering of the deposited material. Such ordering may for example occur between the islands during layer-by-layer growth. These islands are typically separated by a distance comparable to the diffusion length of the atoms at the given deposition rate and substrate temperature. If such a preferred island-island distance indeed occurs, the diffraction profile will develop shoulders. By measuring the diffraction profile, this and other types of order can be determined. Fuoss et al. [21] were the first to measure such profiles during vapour-phase epitaxial growth of GaAs. They used rocking scans to obtain the profiles, similar to such measurements during MBE growth [22].

4. SOLUTION GROWTH

In the case of crystal growth from a solution, atomic-scale information about the surface structure is almost completely lacking. This is definitely a 'niche market' for X-ray diffraction. Additional and important information like step heights and roughness can be gained from SPM and optical microscopy, but crystallographic information appears to be the exclusive realm for X-ray diffraction for this type of systems. Both sides of the growth interface are of interest, because both the crystal and the liquid are expected to deviate from their bulk structure.

Measuring data from a crystal immersed in a solution is far more difficult than the corresponding vacuum experiments. The X-rays need to penetrate the liquid, both on the way in and towards the detector. This leads first of all to an attenuation of the intensity. This effect can be minimised by using hard X-rays (typically > 15 keV) and by reducing the travel path through the liquid. The latter is achieved by using a foil to produce a thin film in a reflection geometry, or by using a transmission geometry with small dimensions [23]. A second and more problematic effect of the liquid is that it scatters X-rays more or less isotropically and thus generates a large background. In vacuum, signal-to-background ratios are easily above 10, but in solution this may be a factor 100 to 1000 worse [24], depending on the solution and crystal used.

The growth behaviour in a solution differs considerably from that during the MBE or CVD growth shown in figure 4. This is owing to a considerable difference in the typical supersaturation used. For MBE and CVD this can be several hundred percent, while in solution growth values below 10% are normal. At such low supersaturations, the barrier for 2-dimensional island nucleation is so high that its contribution to the growth is negligible compared with step flow growth. The nicely observable layer-by-layer growth during MBE occurs only at high supersaturations, and is thus not typical for solution growth. Note that what is called 'step flow' in the vacuum growth community, is called 'layer growth' in the solution growth community. This layer growth mode does not lead to strong variations in the diffracted signal. The quantity that is most directly measurable during solution growth is the overall surface roughness, but

few experiments have been done in this direction [24]. At high supersaturation, island nucleation does of course occur. Using AFM, it was for example found that step flow and island nucleation may occur simultaneously [2].

Even when the solution growth process itself has hardly been measured using X-ray diffraction, the (equilibrium) structure of the crystal surface is equally relevant. The first information on the surface structure of a crystal in its growth solution was obtained only five years ago by Chiarello and Sturchio [25] on cleaved calcite $(10\bar{1}4)$ in water. More elaborate experiments have since led to much improved data sets and therefore more detailed information [26].

A crystal with a somewhat complicated unit cell can often terminate in more than one way. The Hartman-Perdok theory can be used to find these possible terminations, but does often not provide simple means to choose the most favourable one. Using X-ray diffraction, Gidalevitz et al. [27] were able to determine which of the two alternative terminations were realised in β-alanine and α-glycine (010) surfaces.

As an example of a solution growth study, we will now discuss KDP (KH_2PO_4, potassium dihydrogen phosphate) crystals in some detail. These crystals, which are grown from aqueous solution, have received a lot of attention in solution growth [28]. The Hartman-Perdok theory predicts that the pyramidal {101} and the prismatic {100} faces are the most stable [29] and these are indeed the faces found when such crystals are grown. Figure 7a shows the habit of KDP. The exact shape turns out to depend strongly on the purity of the solution. For very clean growth conditions, all faces are approximately of the same size. For solutions containing small traces of trivalent metal ion impurities like Fe^{3+} and Cr^{3+}, the growth of the {100} faces is hampered, while the {101} faces are largely unaffected. This leads to elongated crystals with large {100} facets. We will show next how the microscopic origin of this macroscopic behaviour can be understood by determining the interface structure of the two faces.

The pyramidal {101} faces can terminate in either a K^+ or in a PO_4^- layer (fig. 7b). The
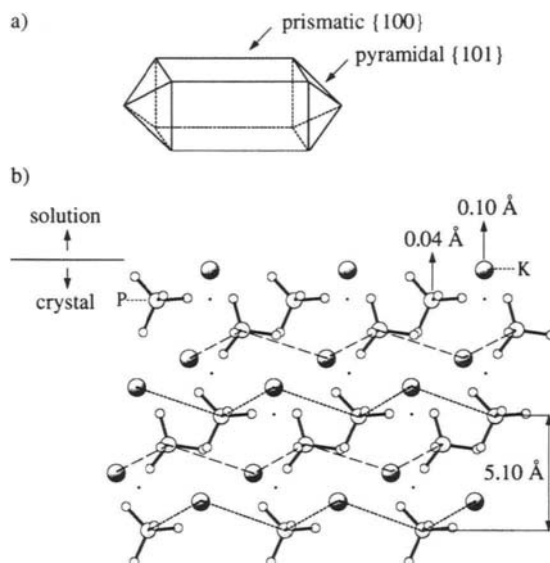
Fig. 7. (a) Growth habit of a KDP crystal with the prismatic and pyramidal faces indicated. In solutions containing metal impurities the growth of the prismatic faces is hindered, but the pyramidal faces are hardly affected. (b) Schematic side view of the pyramidal face, KDP{101}. From the bulk crystallographic structure one expects this face to either terminate in a potassium layer (dotted curve) or a phosphate layer (dashed curve). X-ray diffraction data show that the first possibility occurs. The outward relaxation of the top layers is shown as well. From ref. [30].

possibility of a mixed termination was ruled out on the basis of AFM measurements [2]. Which of the two terminations occurs can be determined using X-ray diffraction. This requires a special growth chamber in which the crystal is immersed in a saturated solution at a constant temperature [24].

The X-ray measurements require a flat and clean surface. In vacuum, this is typically achieved by sputter and annealing cycles. For the KDP crystals, the equivalent procedure is etching in an undersaturated solution. For KDP this leads to surfaces that are sufficiently flat to measure the crystal-truncation rod intensity over the entire range between bulk peaks. The results of the measurements are shown in figure 8, where the (10)-rod data are shown as open circles. The dashed and dotted curves are the calculated rod profiles for bulk PO_4^- and K^+ termination, respectively. The K^+ termination clearly describes the data much better.
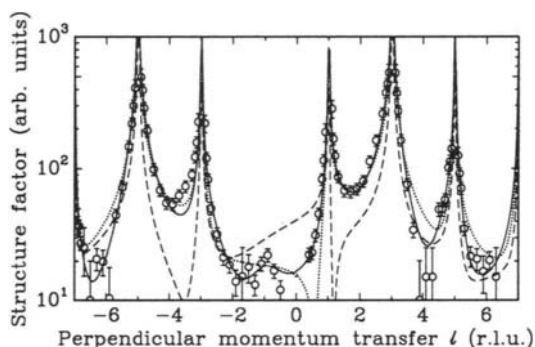
358

Fig. 8. Structure factor amplitudes along the (10) crystal truncation rod for KDP{101} as a function of the diffraction index l. The dotted line is a calculation for a bulk K^+-terminated surface, the dashed curve for a PO_4^--terminated one. The solid line is the best fit, starting from a K^+-terminated layer and allowing for perpendicular relaxation [30].

In this case, the data is of sufficient quality to include also relaxation of the two outermost layers in the fitting procedure. It is found that the K^+ ions relax outwards by an amount of 0.10 ± 0.05 Å and the PO_4^- groups by 0.04 ± 0.05 Å. A similar analysis was done for the {100} surface, where one termination was expected and found. This surface consists of alternating K^+ and PO_4^- groups.

Having established the atomic structure of both faces, we can now understand the effect of the positive metal ion impurities. The pyramidal face has only K^+ ions on the surface of the crystal, and will thus repel the positive metal impurities. The growth of this face is therefore unaffected by these impurities. The prismatic face, on the other hand, has both positive and negative ions at the surface, so that the impurities can adsorb easily and block the growth. This leads to the observed elongated crystal shape in impure solutions.

For KDP crystals we found a termination in a positive layer and a small amount of relaxation in the two topmost layers. We recently performed SXRD experiments on the {101} surface of the isomorphous ADP (ammonium dihydrogen phosphate, $NH_4H_2PO_4$) crystals [31]. The ADP crystal was found to terminate in a NH_4^+ layer, thus again in a positive layer. In contrast to KDP, however, the top layer shows a large *inward*
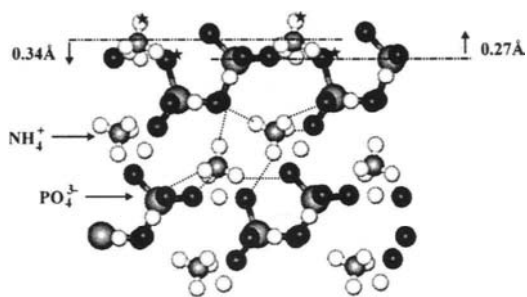
Fig. 9. A side view of the ADP{101} surface with the measured relaxation in the two top layers indicated. The stars in the topmost ammonium and phosphate groups indicate the positions where hydrogen bonds are broken due to the crystal termination. Owing to the large relaxation, new hydrogen bonds can form.

relaxation (fig. 9). The spacing between the topmost layers was found to be reduced by approximately 50%! This clearly shows that the assumption that the surface atoms are at bulk positions is not true for the case of ADP{101}. The difference with KDP is most likely caused by the hydrogen bonding of the NH_4 groups in ADP.

As discussed earlier, for growth from a solution also the structure of the liquid near the interface is of importance. The liquid may influence the growth and the resulting crystal shape in various ways. The interaction with the crystal may influence the relative stability of the various growth faces, the growth speed can change [32], concentration gradients may occur, impurities may be present or pre-ordering of the growth units may occur.

Using X-ray diffraction the ordering in the liquid can be measured without disturbing it. Solid-liquid interfaces are also of fundamental interest, and a number of theoretical studies have shown that a liquid in contact with a hard wall should show layering in the direction perpendicular to the surface [33]. Such layering over a few layers has been observed for a number of systems [34], but much remains to be learned. Recently some very nice electron microscopy studies were reported [35].

Layering is the ordering component in the direction perpendicular to the surface, but there should also be lateral components that are expected
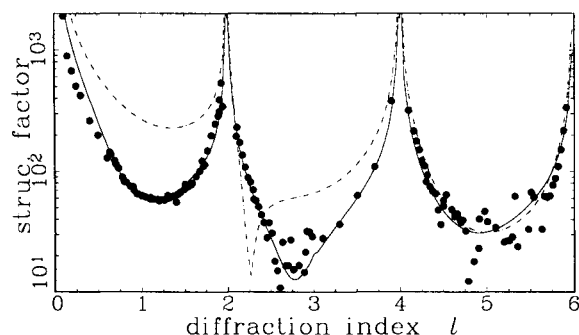
Fig. 10. The structure factor amplitude of the specular rod for KDP{101}. Circles are the experimental data points, the dashed and solid curves are model calculations in which the layering effects in the growth solution are ignored and included, respectively.

to be weaker. The liquid will therefore contribute weakly to the substrate crystal truncation rods. By measuring these rods with high accuracy, the full ordering characteristics of the liquid near the crystalline interface can be derived.

The ordering effects are strongest in the perpendicular direction and they will therefore have their largest influence in the specular or (0,0)-rod. Figure 10 shows the measured specular rod for KDP{101}, together with a calculation (dashed curve) based on the model derived from non-specular data and shown in figure 7. The fact that this calculated curve does not agree with the data immediately demonstrates the importance of the liquid structure. Including layering in the liquid at the interface leads to a proper fit (solid curve) [36].

The effects in non-specular rods are very weak. The signal-to-background ratio therefore needs to be optimised by using liquid films that are as thin as possible. The most detailed information of this type has therefore been obtained on a simple model system: a quasi-liquid monolayer of Pb on Ge(111) [37]. Only very recently the first results are emerging on thicker liquid films in contact with KDP [36].

5. CONCLUSIONS

Because of their penetrating power, X-rays allow the surfaces of growing crystals to be investigated *in situ* in a large variety of environments. Using X-ray diffraction, the surface roughness during MBE and CVD growth can be monitored and the development of stacking faults can be quantified. In the area of solution growth, X-ray diffraction is providing the

first detailed surface structure determinations. It appears that many surfaces deviate from their bulk extrapolated structure. The solution in contact with the crystal surface shows clear ordering.

REFERENCES

[1] P. Hartman and W.G. Perdok, Acta Cryst. 8 (1955) 49; R.F.P. Grimbergen, H. Meekes, P. Bennema, C.S. Strom and L.J.P. Vogels, Acta Cryst. A54 (1998) 491.

[2] J.J. De Yoreo, T.A. Land and B. Dair, Phys. Rev. Lett. 73 (1994) 838.

[3] M. Plomp, W.J.P. van Enckevort and E. Vlieg, J. Cryst. Growth 216 (2000) 413.

[4] R. Feidenhans'l, Surf. Sci. Rep. 10 (1989) 105; I.K. Robinson and D.J. Tweet, Rep. Prog. Phys. 55 (1992) 599; I.K. Robinson, in *Handbook on synchrotron radiation*, edited by G.S. Brown and D.E. Moncton (North-Holland, Amsterdam, 1991), Vol. 3, p. 221.

[5] S.R. Andrews and R.A. Cowley, Journal of Physics C: Solid State Physics 18 (1985) 6427.

[6] I.K. Robinson, Phys. Rev. B 33 (1986) 3830.

[7] E. Vlieg, J. Appl. Cryst. 30 (1997) 532.

[8] B.E. Warren, *X-ray diffraction* (Dover, New York, 1990).

[9] E. Vlieg, J.F. van der Veen, S.J. Gurman, C. Norris and J.E. Macdonald, Surf. Sci. 210 (1989) 301.

[10] E. Vlieg, J. Appl. Cryst. 33 (2000) 401.

[11] C.S. Lent and P.I. Cohen, Surf. Sci. 139 (1984) 121; S.K. Sinha, E.B. Sirota, S. Garoff and H.B. Stanley, Phys. Rev. B 38 (1988) 2297.

[12] H.A. van der Vegt, H.M. van Pinxteren, M. Lohmeier, E. Vlieg and J.M.C. Thornton, Phys. Rev. Lett. 68 (1992) 3335.

360

[13] E. Vlieg, A.W. Denier van der Gon, J.F. van der Veen, J.E. Macdonald and C. Norris, Phys. Rev. Lett. 61 (1988) 2241.

[14] J.H. Neave, B.A. Joyce, P.J. Dobson and N. Norton, Appl. Phys. A 31 (1983) 1; J.M. van Hove, C.S. Lent, P.R. Pukite and P.I. Cohen, J. Vac. Sci. Technol., B 1 (1983) 741.

[15] P.H. Fuoss, D.W. Kisker, G. Renaud, K.L. Tokuda, S. Brennan and J.L. Kahn, Phys. Rev. Lett. 63 (1989) 2389.

[16] G.B. Stephenson, J.A. Eastman, C. Thompson, O. Auciello, L.J. Thompson, A. Munkholm, P. Fini, S.P. DenBaars and J.S. Speck, Appl. Phys. Lett. 74 (1999) 3326.

[17] J. Vrijmoeth, H.A. van der Vegt, J.A. Meyer, E. Vlieg and R.J. Behm, Phys. Rev. Lett. 72 (1994) 3843.

[18] P. Bailey, T.C.Q. Noakes and D.P. Woodruff, Surf. Sci. 426 (1999) 358; S.A. de Vries, W.J. Huisman, P. Goedtkindt, M.J. Zwanenburg, S.L. Bennett, I.K. Robinson and E. Vlieg, Surf. Sci. 414 (1998) 159.

[19] S.A. de Vries, W.J. Huisman, P. Goedtkindt, M.J. Zwanenburg, S.L. Bennett and E. Vlieg, Phys. Rev. Lett. 81 (1998) 381.

[20] H.A. van der Vegt, J. Alvarez, X. Torrelles, S. Ferrer and E. Vlieg, Phys. Rev. B 52 (1995) 17443.

[21] P.H. Fuoss, D.W. Kisker, F.J. Lamelas, G.B. Stephenson, P. Imperatori and S. Brennan, Phys. Rev. Lett. 69 (1992) 2791.

[22] H.A. van der Vegt, W.J. Huisman, P.B. Howes and E. Vlieg, Surf. Sci. 330 (1995) 101.

[23] H. You, C.A. Melendres, Z. Nagy, V.A. Maroni, W. Yun and R.M. Yonco, Phys. Rev. B 45 (1992) 11288.

[24] S.A. de Vries, P. Goedtkindt, W.J. Huisman, M.J. Zwanenburg, R. Feidenhans'l, S.L. Bennett, D.-M. Smilgies, A. Stierle, J.J. De Yoreo, W.J.P. van Enckevort, P. Bennema and E. Vlieg, J. Cryst. Growth 205 (1999) 202.

[25] R.P. Chiarello and N.C. Sturchio, Geochim. Cosmochim. Acta 59 (1995) 4557.

[26] P. Fenter, P. Geissbühler, E. DiMasi, G. Srajer, L.B. Sorensen and N.C. Sturchio, Geochim. Cosmochim. Acta 64 (2000) 1221.

[27] D. Gidalevitz, R. Feidenhans'l, S. Matlis, D.-M. Smilgies, M.J. Christensen and L. Leiserowitz, Angew. Chem., Int. Ed. Engl. 36 (1997) 955.

[28] L.N. Rashkovich, KDP-family single crystals (Adam Hilger, Bristol, 1991).

[29] P. Hartman, Acta Cryst. 9 (1956) 721; B. Dam, P. Bennema and W.J.P. van Enckevort, J. Cryst. Growth 74 (1986) 118.

[30] S.A. de Vries, P. Goedtkindt, S.L. Bennett, W.J. Huisman, M.J. Zwanenburg, D.-M. Smilgies, J.J. De Yoreo, W.J.P. van Enckevort, P. Bennema and E. Vlieg, Phys. Rev. Lett. 80 (1998) 2229.

[31] J. Arsic, M.F. Reedijk, A.J.R. Sweegers, Y.S. Wang and E. Vlieg, to be published (2001).

[32] X.Y. Liu, E.S. Boek, W.J. Briels and P. Bennema, Nature 374 (1995) 342.

[33] W.A. Curtin, Phys. Rev. Lett. 59 (1987) 1228.

[34] M.F. Toney, J.N. Howard, J. Richer, G.L. Borges, J.G. Gordon, O.R. Melroy, D.G. Wiesler, D. Yee and L.B. Sorensen, Nature 368 (1994) 444; B.M. Ocko, Phys. Rev. Lett. 64 (1990) 2160; W.J. Huisman, J.F. Peters, M.J. Zwanenburg, S.A. de Vries, T.E. Derry, D.L. Abernathy and J.F. van der Veen, Nature 390 (1997) 379.

[35] S. Arai, S. Tsukimoto, S. Muto and H. Saka, Microsc. Microanal. 6 (2000) 358.

[36] M.F. Reedijk, J. Arsic, F.F.A. Hollander, S.A. de Vries and E. Vlieg, to be published (2001).

[37] F. Grey, R. Feidenhans'l, J.S. Pedersen, M. Nielsen and R.L. Johnson, Phys. Rev. B 41 (1990) 9519; S.A. de Vries, P. Goedtkindt, P. Steadman and E. Vlieg, Phys. Rev. B 59 (1999) 13301.

Using atomic force microscopy to investigate solution crystal growth

J.J. De Yoreo, C.A. Orme, and T.A. Land

Chemistry and Materials Science Directorate, Lawrence Livermore National Laboratory, P.O. Box 808, Livermore, CA 94551, USA

Over the past ten years, the atomic force microscope (AFM) has become a common tool for investigating the growth of crystal surfaces from solutions. In this chapter, we describe the basic operation of the AFM and present examples of its application to crystal growth science, we discuss the use of *ex situ* force microscopy to investigate surface morphology and coarsening in air, and describe the experimental arrangement used to investigate growth *in situ*. We use examples from a number of crystal systems to illustrate the use of *in situ* imaging to investigate dislocation source activity, 2D and 3D nucleation, critical step length, step kinetics, step roughness, and impurity-step interactions. Finally, we point out future directions for AFM investigations of solution crystal growth.

1. INTRODUCTION

Prior to the invention of the atomic force microscope, optical interferometry was the high-resolution tool of choice for investigating crystal surfaces in solutions. Although the development of phase shift interferometry [1]gave sufficient vertical resolution to detect the passage of single steps, the lateral resolution was still limited to about ten microns, far in excess of the length scale of molecules, steps, terraces, or islands. The atomic force microscope[2] has removed this limitation. With a lateral resolution of better than 10nm and the ability to resolve the lattice of even some simple inorganic crystals, one can [3] now address unanswered questions about the evolution of crystal surfaces.

Prior to the invention of the AFM, the scanning tunneling microscope[4] (STM), had been used for a number of years for atomic-scale investigations of metal and semiconductor thin films grown by vapor phase epitaxy. But due to the need for a substrate that could support a tunneling current, the STM is not applicable to the study of insulating crystals in ionic solutions. Although the AFM has a resolution that is about an order of magnitude lower than the STM, it can be operated with insulating crystals and in fluids. Consequently, it has provided the first technique with molecular-scale resolution that can be used to image the growth of crystal surfaces from solutions *in situ*. In addition, unlike the STM, the feature sizes that are accessible with the AFM range from about 5nm to over 100 microns, thus spanning length-scales from single molecules to entire crystals. This large dynamic range is illustrated in **Fig. 1** for rhombohedral crystals of the protein canavalin[5-8].

Gratz and Hansma[9] published the first *in situ* AFM results on crystal surfaces in solution, using crystalline SiO_2 as the subject of their investigation. In the decade since this landmark study, a wide range of crystal systems has been explored with the AFM. The nature of these systems has ranged from simple ionic solids such as quartz[9], calcite[10-19], and $KH_2 PO_4$[20-22], to hydrogen bonded solids such as diketopiperazine derivatives, to van der Waals bonded macromolecules including many protein[7, 8, 23-25] and virus systems[7, 23, 25-27].

The purpose of this chapter is to introduce the reader to the application of the atomic force microscope to investigations of solution crystal growth. No prior knowledge of scanned probe microscopy is assumed. However, this chapter is not intended to be a tutorial on instrumentation. Rather our intent is to give the reader enough familiarity with the instrument to feel comfortable planning a crystal growth research project involving the AFM, and sufficient exposure to investigations carried out to date to allow the reader to design a set of experiments that will produce useful results.
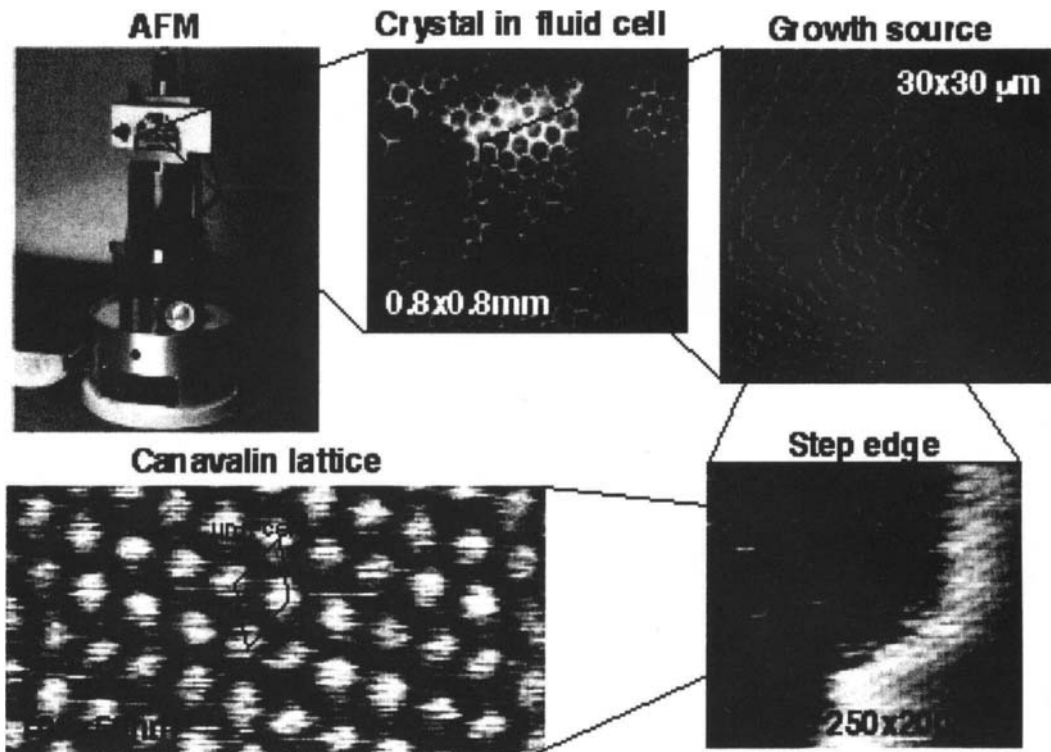
Figure 1. The dynamic range of the AFM makes it possible to access features with sizes from tens of nanometers to tens of microns.

Although crystal growth fundamentals are a crucial part of AFM data analysis, providing a primer in crystal growth physics is beyond the scope of this chapter. Rather, we give examples of crystal growth studies that have been or can be done and leave it to the reader to consult the references for a description of the theoretical underpinnings used to interpret the data. Most examples come form our own work, but numerous high quality AFM investigations have been performed by groups around the world. The bibliography at the end of this chapter provides a partial list of these studies.

The second section of this chapter covers the basic design and operation of the AFM. Many readers will find that they can skip this section. However, the novice will find particularly useful information about what the AFM is actually sensing and what information it provides. In the third section, we give examples of *ex situ* studies. The fourth section addresses *in situ* imaging of crystal surfaces and will include both a description of the experimental set-up and examples of research problems that have been addressed by either

ourselves or others. The fifth and final section contains some concluding remarks.

2. DESIGN AND OPERATION OF THE AFM

A schematic of the critical components of an atomic force microscope is shown in **Fig. 2**. The sample sits on top of a triple-axis piezo actuator tube and interacts with a sharp tip at the end of a cantilever, which is fixed to the stationary body of the AFM. A laser beam reflects off of the cantilever and onto a position-sensitive photo diode. As the piezo tube moves the sample laterally, the variations in the interaction force between the tip and the surface produces a deflection of the laser spot on the photo diode. The resulting voltage signal is used as the control signal for a feedback loop to the vertical axis of the piezo tube. The control software varies the vertical position of the sample to null the signal from the photo diode. In the most common mode of operation, the actuator tube is driven rapidly back and forth along one in-plane (x) axis and slowly back and forth along the other in-plane (y) axis, while the sample height is adjusted to keep the deflection of
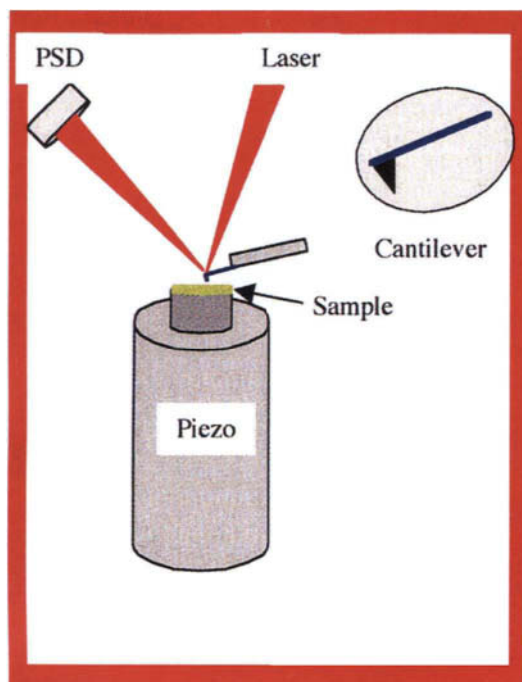
Figure 2. Schematic of AFM showing laser beam, piezo tube, cantilever and tip, sample and position sensitive detector (PSD).

the cantilever constant. When the voltage applied to the vertical piezo tube (z-axis) is plotted as a function of piezo deflection along x and y, the result is a map of topography. However in another common variation of this scheme, the slow scan axis is disabled, producing a time profile of a single line along the sample surface where the fast scan direction records position and the disabled axis records the time.

2.1 Surface-tip interactions

The nature of the interaction between the tip and surface depends on a number of operational choices. Because virtually all data presented in this chapter was collected by monitoring the vertical deflection of the cantilever, no discussion of lateral force, magnetic force, or other modes of detection is given. In contact mode, the AFM provides two primary pieces of information about crystal surfaces: surface morphology — or topography — and step motion. However, the nature of the tip-surface interaction that provides that information depends on what portion of the force vs. distance curve is being used for detection as illustrated in **Fig. 3**. As the figure indicates, surface detection can

either occur along the repulsive, attractive, or both regions of the curve, corresponding to what is commonly referred to as operating in contact, non-contact, and intermittent contact mode respectively.
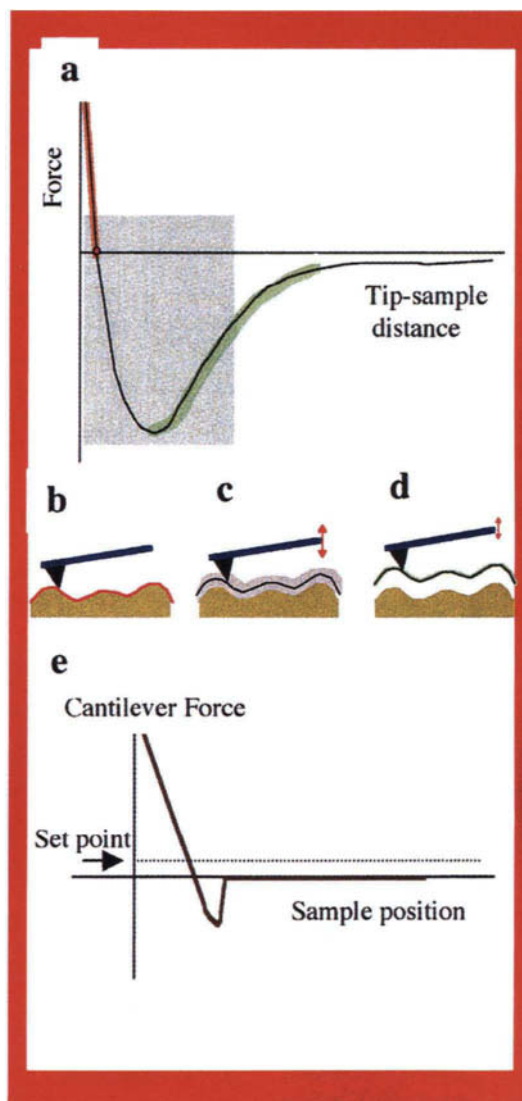
Figure 3. (a) Schematic of potential experienced by tip on an infinitely stiff cantilever approaching surface. (b-d) Schematic of tip-surface geometry during (b) contact, (c) intermittent contact, and (d) non-contact operation. Note tip trajectory is highlighted in color corresponding to portion of potential curve being probed. (e) Schematic of force curve experienced by real tip approaching surface showing relationship of set point voltage to that obtained at zero deflection. The region of the curve that dips below the later is corresponds to a "jump to surface" prior to establishing contact.

Figure 3e shows a schematic of the force vs. distance curve for the tip approaching the surface.

When imaging in air or vacuum, the attractive portion of the curve results in a phenomenon known as "jump to contact", which is due to capillary forces caused by water condensation at the tip-surface interface as illustrated in **Fig. 4**. When imaging in fluid, these capillary forces are eliminated or dramatically reduced. Consequently, the force curve exhibits a nearly constant (zero) force region when the tip is off the surface, followed by a monotonically increasing repulsive force as the tip contacts the surface (see **Fig. 4**). In order to obtain a control signal from the tip during scanning, some minimum amount of force must be established, even if only intermittently. Operationally, one minimizes this force as well as its capacity to influence the measurement by setting the voltage control point (the set-point voltage) of the instrument above the output voltage of the photo-diode when the tip is not in contact with the surface, but as close to it as possible while still maintaining the image. (See **Fig. 3**.) The total applied force can then be calculated by multiplying the difference between these two voltages times the detector sensitivity in V/nm times the spring constant of the cantilever in N/nm.
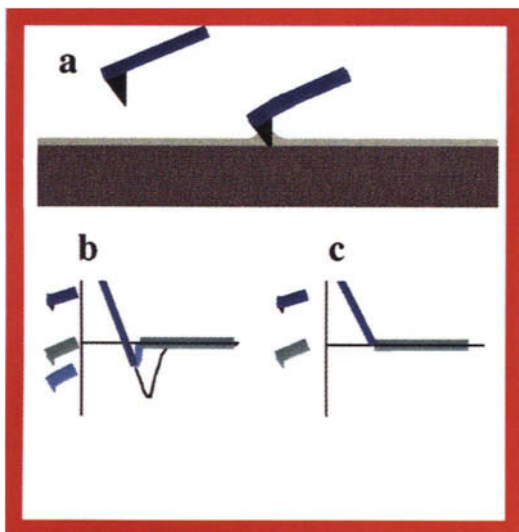
Figure 4. (a) Schematic of tip approaching surface with adsorbed water layer. (b-c) Force-displacement curve in (b) presence and (c) absence of capillary forces. (b) shows jump to contact on approach and adhesion on withdrawal.

The force required to image without damaging the surface depends on the strength of the intermolecular bonding in the solid, but is on the order of one nN. With such small forces, angstrom level variations in surface topography can only be detected if the cantilever is "soft". But soft cantilevers are more prone to capillary forces, which are typically of order 10-100nN. Consequently, imaging surfaces in fluids, where the capillary forces are minimal, nearly eliminates one of the obstacles to high-quality imaging of soft materials. Details of force calculations, cantilever deflection, piezo design and performance, and other instrument-related issues, can be found in the references given in the bibliography.

2.2 Image resolution

A number of factors influence image resolution. Not surprisingly, the sharpness of the AFM tip has a significant effect. As **Fig. 5** shows, standard –manufactured AFM tips can be modified to produce sharper tips by etching, oxide sharpening, or by attaching carbon nanotubes[28-30]. But in our experience, unidentified aspects of the sample-tip interaction have a significant impact on resolution. For no apparent reasons, some crystal systems yield molecular resolution data — even when comprised of angstrom-size inorganic molecules —– while in other systems, image quality is limited to tens of nanometers. In other cases, such as with crystals of acetaminophen, the resolution is poor because the adhesion force is still large, even in solution.

2.3 *Ex situ* vs. *in situ* imaging

Operation of the AFM is possible in many different environments including air, vacuum, vapor (including H_2O), and solution. Eggelston and coworkers[31, 32] have even constructed a hydrothermal cell for operation at temperatures of up to 150°C and pressures of 150psi and applied it to the growth of barite. The choice of environment clearly depends upon the objectives of the experiment. The majority of AFM studies on crystal surfaces have been performed either *ex situ* in air at room temperature or *in situ* in solution in the vicinity of room temperature. Despite the many reasons why *in situ* imaging is more desirable than *ex situ* imaging for investigating solution crystal growth physics, a simple analysis shows that many crystal systems simply grow too rapidly to allow for *in situ* operation. Typical scan rates for the fast scan direction are between 1 and 10Hz, (although for

some systems where our goal was only to measure step speeds, we have used scan rates as high as 110Hz). If we collect 256 lines of data per image, then at 10Hz, each image takes about 25s to collect. For most crystals that we have examined, images that allow us to distinguish steps for analysis are rarely in excess of 20x20 μm. This implies that, if we require that a step be visible in two successive frames, then the step must be moving at no more than about 1 μm /s. In practice, we have found that this can be pushed to about 10 μm /s by increasing the scan rate and the scan size.

these cases, we have found that there is much to be learned by imaging the surface *ex situ*.

3. *EX SITU* EXPERIMENTS

3.1 Surface preservation

The most critical factor in performing *ex situ* imaging of crystal surfaces is preservation of the growth surface. De Yoreo et al.[21] and Land et al.[22]successfully preserved surfaces of KDP (KH_2PO_4) crystals by pulling mm size crystals directly from solution through a stream of hexane at

Figure 5. Various kinds of AFM tips. (a) Standard SiN_4, (b) Oxide sharpened SiN_4, (c) Tip following carbon deposition., (d) Tip following attachment of carbon nanotube. Images courtesy A. Noy and Digital Instruments/Veeco, www.di.com, Santa Barbara, Ca

Now compare this to expected step speeds. To estimate the step speed in an arbitrary system we write it as $v = \Omega\beta(C-C_e)$ where Ω is the volume per molecule in the solid, β is the kinetic coefficient, and C and C_e are the actual and equilibrium concentrations respectively. Typical kinetic coefficients for inorganic systems are about 0.1cm/s[5]. Molecular volumes are on the order of 10^{-22} cm^3. Because growth mechanism, morphology and step speed are all dependent on supersaturation, σ, the ability to vary it over wide range is highly desirable. At a minimum , one would like to cover at least 10% of supersaturation. Writing $\sigma \approx (C-C_e)/C_e$, we require $C \geq 1.1C_e$. Solubilities vary over many orders of magnitude. For highly soluble salts it is not unusual, for C_e to be in excess of 10^{21} molecules/cm^3. Then, for $\sigma = 10\%$, $C-C_e > 10^{20}$ molecules/cm^3 and $v > 10^{-3}$cm/s or 10 μm /s, which is at the limit of the instrument. In a later section on *in situ* imaging we will describe an alternative method for determining step velocity involving disabling of the slow scan axis that allows this limit to be violated. But if one is interested in monitoring step and surface morphology, *in situ* imaging simply can not be used on these fast growing surfaces. In

the same temperature as the solution. These crystals were attached to the bars of rotating crystallizers to ensure that the surfaces were generated in a kinetic regime. Plomp et al.[33] used a similar scheme to examine the surface of $BaNO_3$ crystals. Ester et al.[34, 35] achieved exceptionally high quality crystal surfaces on KAP (potassium hydrogen pthalate) crystals using a system that directed a jet of Ar gas at the crystals as they were removed from solution. In each of these examples, the authors were confident that the overall morphology and average step spacing was representative of the growth surface, but that the details of the observed step edge and island morphologies could not be trusted. Examples of surfaces produced by these methods are shown in **Fig. 6**.

3.2 Growth hillock structure

The primary use of *ex situ* data has been to examine the structure of dislocation hillocks and determine the dependence of terrace width on supersaturation. Such studies have of course been performed *in situ* using interferometry[1] where the growth rate is not an issue until much higher supersaturations. Besides providing an indisputable record of the growth surface, interferometry also
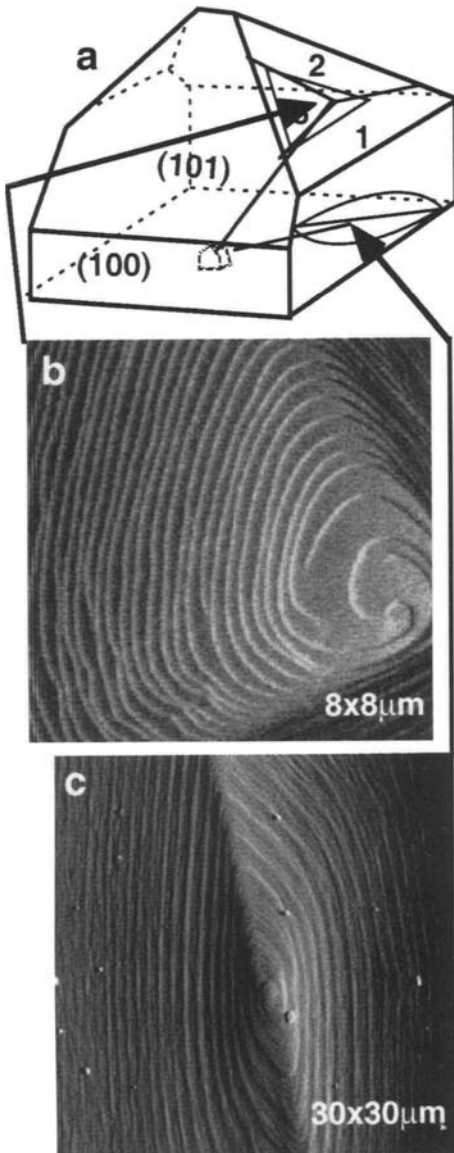
Figure 6. Example of KDP crystal surfaces preserved by pulling through hexane. (a) shows schematic of crystal structure. (b and c) Growth hillocks on the (b) {101} and (c) {100} face generated by dislocations emanating from the seed crystal interface.

averages over many terraces, giving little scatter in individual data points. But *ex situ* AFM has one distinct advantage: it provides direct information about the structure of the dislocation source. The

consequence of this advantage is best exemplified by the studies on KDP[20, 21]. As **Fig. 7** shows, hillocks on KDP {101} possess hollow cores at the dislocation sources. Even for dislocation sources with Burgers vectors of one unit step, these cores are larger than the critical radius even at relatively low supersaturations (~ 5%). Consequently, they are the primary factor determining the terrace width, which, as **Fig. 8** shows, is nearly independent of Burgers vector and weakly dependent on supersaturation. Although existing interferometric data on KDP and ADP contained the same information[36-38], understanding the source of this unexpected behavior only came about because of the AFM images.

3.3 Surface coarsening

Ex situ AFM can also be used to investigate the post-growth evolution of crystal surfaces. **Figure 9** shows the evolution of a KDP {101} surface in air at room temperature over the course of a week. The as-grown surface consists of a set of steps and islands with irregular boundaries. Over time, this surface coarsens until, after a week, the islands have disappeared and the steps have smoothed considerably. The reason for this phenomenon is, of course, surface diffusion of KDP — perhaps within a adsorbed film of H_2O — from regions of high to low curvature. Other results also show that vacancy islands fill-in through this process. The authors found that analysis of the island sizes gave a $t^{1/3}$ dependence of island radius on time. This dependence is expected for a surface diffusion limited process of coarsening and provides an estimate of the surface diffusivity of about $10^{-9} cm^2/s$.

4. *IN SITU* EXPERIMENTS

4.1 Experimental design

In situ AFM investigations of crystal growth present a number of experimental challenges. **Table I** gives a list of the variables that must be controlled in order to obtain quantitative information from an experiment. This control is achieved by using an experimental set-up like that illustrated in **Fig. 10**. The growth solution is held at fixed composition within a well mixed, temperature controlled reservoir. The flow rate of the solution is controlled by a peristaltic pump that lies between the reservoir and the AFM. (Alternatively, a syringe pump can be used in place of the peristaltic pump and solution reservoir in cases where active control
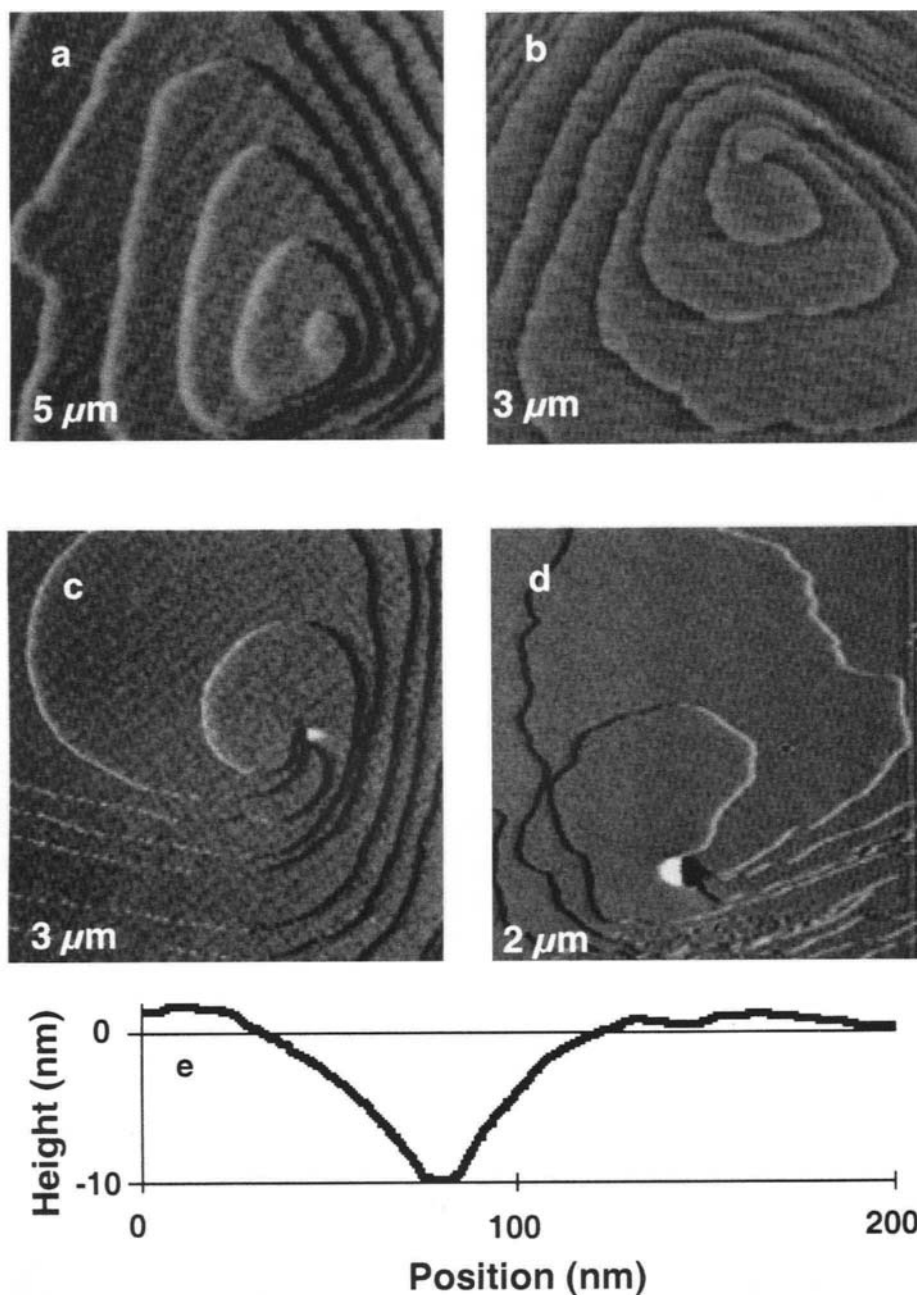
Figure 7. AFM images of typical growth hillocks on: (a-d) the {101} face of KDP showing hollow cores at dislocation sources. (e) Cross section of typical hollow core.

over pH or CO_2 content is not needed and the solution can simply be maintained at room temperature.) A riser in the flow line containing an air pocket serves to damp out any vibrations from the pump, and an in-line filter (not shown) removes any particles introduced during transfer of the solution to the reservoir. The sample sits on a glass coverr-slip, fixed to a metal puck which is turn fixed
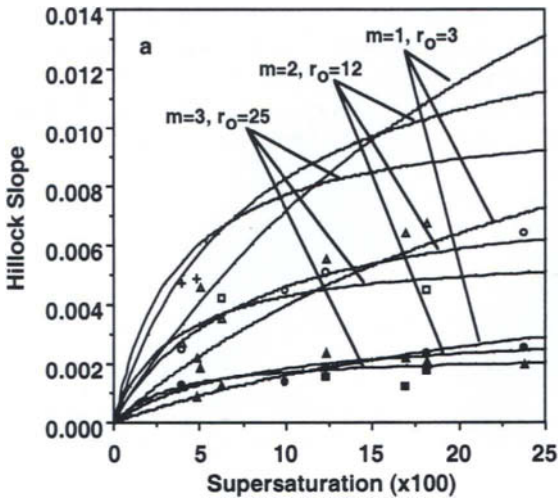
Figure 8. Dependence of hillock slope on supersaturation for hillocks on the {101} face of KDP. Solid lines are theoretical fits to data assuming hollow cores of radius r_o and burgers vectors of m unit steps. The three sets of curves correspond to the three sectors of the growth hillock (see Fig. 6). The data points are the measured slopes from the AFM images. Squares, triangles and circles correspond to m = 1, 2 and 3 respectively. The crosses correspond to m=2.

Figure 10. Schematic of system used for in-situ imaging of crystal surfaces in solution. (a) shows flow controls. (b) shows temperature controls.
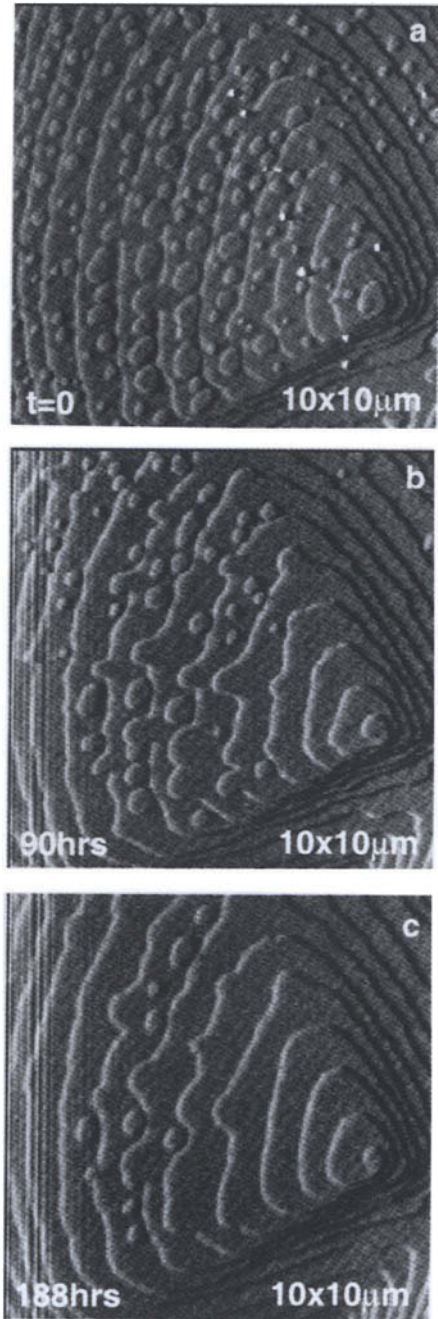
Figure 9. Coarsening of islands and steps on KDP {101}. With time, the material from the islands diffuses to the steps which in turn exhibit smoothing.

to a fluid cell with an inlet and outlet port makes a seal with the glass cover-slip via a silicone O-ring. Before passing into the inlet port, the solution flows

through the meander of a heat exchanger. The temperature of the heat exchanger is controlled using a thermoelectric element and the heat is removed from or added to the back side of the Peltier device by flowing chilled or heated water

control, which can be remedied using the set-up described above with a well tuned temperature controller. Finally, the pH of the solution exerts a strong control over step speeds and therefore must be held constant as other variables are changed if a

Table I
Experimental control parameters

Parameter	Primary effect
Solute concentration	Supersaturation, net flux to surface
Temperature	Activation rate for attachment and detachment, thermal roughness, equilibrium phase
Flow rate	Balance between bulk diffusion and surface kinetics
pH	Ratio of solution species, surface charge, activation barriers
Ionic strength	Screening between charges, activation barriers
Ratio of species	Relative fluxes of species, rate limiting adsorption barrier
Purity	Growth inhibition, growth morphology, surface energies

through a copper block. The control thermocouple is inserted into the fluid stream within the heat exchanger and a waterproofed measurement thermocouple is inserted into the outlet port of the fluid cell. The resulting system has temperature stability of about 0.1°C at the cell.

Flow rates are adjusted upward until the step speed on the crystal surface is no longer a function of flow rate. Typical flow rates are about 2ml/min. However, for very slowly growing crystals, it is not necessary to flow solution. Whether or not this is the case for any given system can be easily checked by monitoring the dependence of step speed on time. If flow is required, the step speed will immediately begin to drop when the pump is turned off.

Control over supersaturation is achieved in one of two ways: by holding the temperature fixed and changing the solute content of the reservoir, or by varying the temperature for fixed solute content, the latter being applicable only to systems in which the solubility is moderately temperature dependent.

There are a number of annoying but avoidable experimental problems common to many systems. The first is air bubbles, which are typically caused by leaky joints or a poor seal at the O-ring/cover-slip interface. Applying vacuum grease around the outside of the O-ring can alleviate this problem. A second obstacle is poor temperature

direct comparison of the datasets is to be performed. **Figure 11** shows the dependence of step speed on pH for rhombohedral canavalin[6].
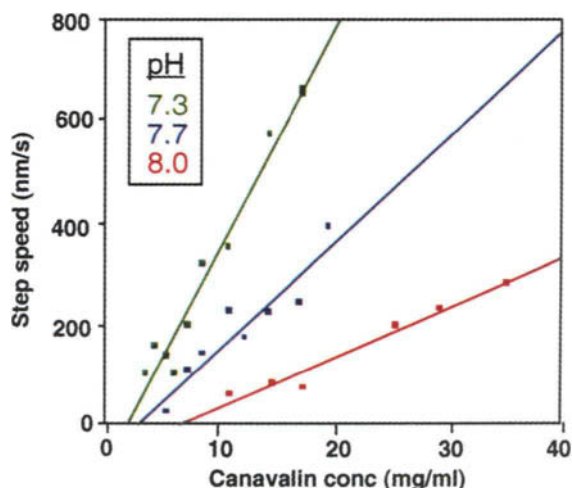
Figure 11. Dependence of step speed on concentration for rhombohedral canavalin at three different values of the solution pH.

Most AFM studies of solution crystal growth have utilized the contact mode of operation. However, in both our work and that of others, non-contact or intermittent contact has also been used, primarily when imaging biological materials. In our

370

experience, in nearly all cases, contact mode provides higher resolution. But on protein and virus crystals, the details of step structure and small adsorbed features can be altered in contact mode. In short, decisions on whether to use contact or tapping mode must be made on a case by case basis after initial experimentation.

The final aspect of the experimental set-up that we will discuss is sample preparation. We have found that the most successful approach varies from one system to another. For example, nucleation of crystals in small droplets of solution on transmission electron microscopy (TEM) grids glued to cover-slips works very well for protein and virus crystals. Alternatively, wires or fibers glued to the cover-slips can be used as clamps. Following nucleation of these crystals directly on the cover-slips, the crystals can be immobilized by manipulating them under the clamp. For most inorganic or organic crystals, dry, pre-nucleated or cleaved crystals are used as seeds for subsequent growth within the fluid-cell. These seeds are fixed to the cover-slip by gluing them. Some experimentation with glues is in order due to the variability of compatibility with various solutions. In general, urethane glues work well with aqueous solutions. Finally, when large, high-quality crystals that cleave well are available, the cleaved crystal surface can be used as the substrate in place of the cover-slip. The fluid-cell O-ring makes a seal directly on the surface of the cleaved crystal.

4.2 Dislocation source activity

In situ imaging can be used to observe the effect of supersaturation on dislocation source activity[22] and critical step length[15]. **Figure 12** shows the effect of supersaturation on terrace width for a dislocation hillock on brushite and on the competition between nearby dislocation sources on rhombohedral canavalin. Of course both of these effects are a result of the Gibbs-Thomson effect on critical step length. The AFM enables one to measure critical step length directly. By collecting sequential images of the emergence of steps from a dislocation source, one can bracket the critical length: The last image in which the emerging step is not yet moving gives a sub-critical lengths and the first image in which it is moving gives a super-critical length. **Figure 13** illustrates the determination of critical length and shows its dependence on supersaturation for the two unique step directions on calcite {104} surfaces. From the slopes of these curves, the step edge free energy associated with creation of the new step can be calculated.

Clearly, the ability to accurately determine
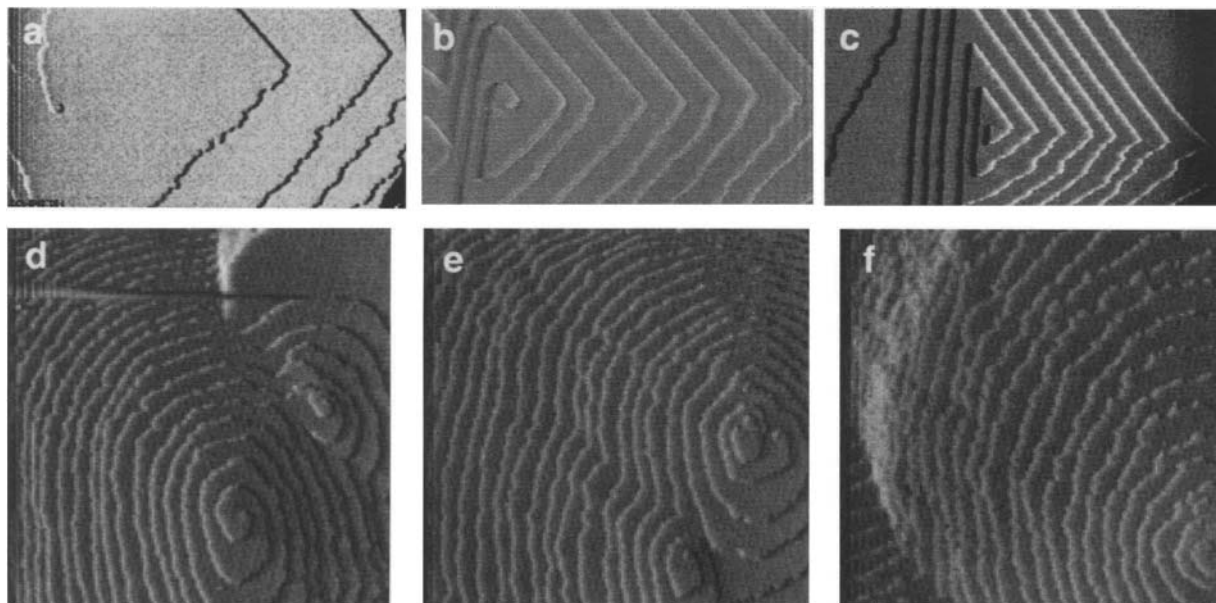
Figure 12. (a-c) 6 x 3μm images showing dependence of terrace width on supersaturation at dislocation hillocks on brushite. Supersaturations are (a) 0.044, (b) 0.065, (c) 0.277. (d-f) 25 x 25μm images showing change in activity of adjacent sources on canavalin as supersaturation decreases. The source in the upper right is composed of four closely spaced dislocations and the source in the lower left has just one dislocation. All emit one single unit height step. (d) Simple source dominates. (e) Sources are of comparable activity. (f) Complex source dominates and simple source is just a dislocation on the side of the main hillock.
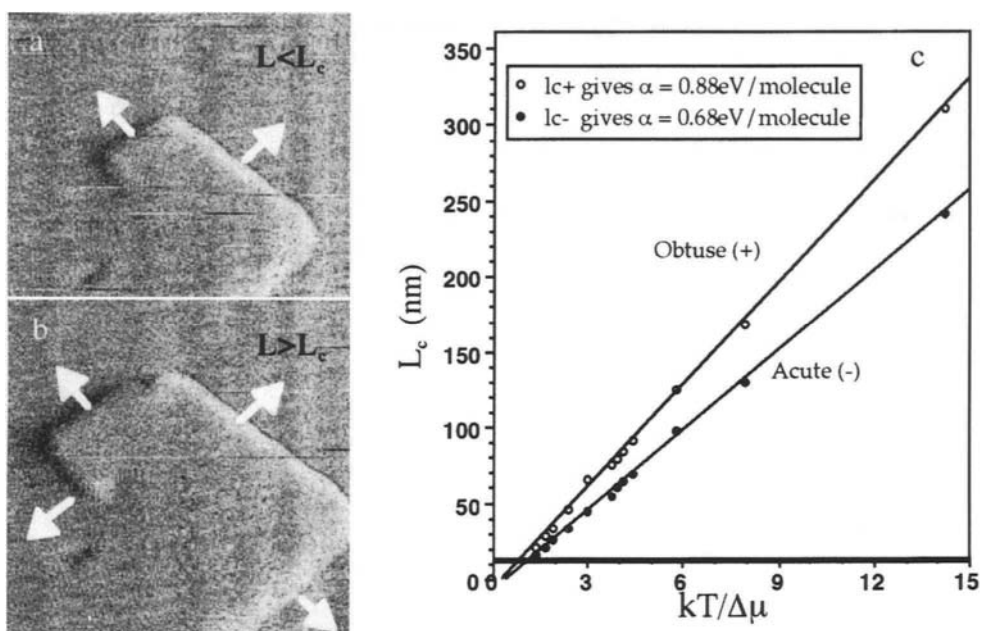
Figure 13. (a) and (b) 1x1µm images of a dislocation source on calcite illustrating measurement of critical step length. (c) Measured dependence of critical length on supersaturation for the two types of steps (obtuse (+) and acute (-)) on calcite. Step edge energy, α, is calculated from the slope.

critical length depends to a large degree on how rapidly the steps are moving relative to the scan rate. Thus high scan rates are useful for this purpose. However, in some cases, fast scan rates may still be insufficient to determine the critical length for certain step directions simply because the critical length is small and the adjacent step speed is rapid. Brushite is a case in point. For one step direction, these two factors make it nearly impossible to "catch" the step in a subcritical state. In these cases, only an upper limit can be determined.

Care must be taken in interpreting the critical length. It simply is not the case that the critical length for a given step direction gives the step edge energy for that step direction. The free energy that gives rise to the critical length is due to the creation of new step edge. But that step edge is not created along the step being measured. Rather it is created at the corners and adjacent sides of the step. For a more complete discussion we refer the reader to Teng et al[15].

4.3 Two-dimensional and three-dimensional nucleation

By increasing the supersaturation, one can unambiguously monitor the transition to 2D nucleation using the AFM. Malkin et al.[39] showed that through careful measurements, the rate of nucleation as a function of supersaturation, and therefore the interfacial energy associated with nucleation could be determined. In addition, a number of authors have used the size of critical islands to calculate the step edge energy[6, 8, 27, 40]. **Figure 14** shows examples of 2D nucleation on calcite, canavalin, and satellite mosaic tobacco virus (STMV).

The procedure for determining critical island size is somewhat different than that used for determining critical step length at dislocation sources. Numerous images of nucleating islands are captured. Some islands eventually shrink in size and disappear while some grow in size. By taking the average of the maximum island size obtained before shrinking, one can estimate the critical size. A warning about calculating step edge energies from critical island sizes is also in order here. The step edge energy that one calculates is an average for the entire island. But unless the step edge energy is isotropic (an unlikely situation), portioning it up

around the island requires quite a bit more work. One must make a polar plot of equilibrium island shape, use the Wulff construction to generate a polar plot of relative step edge energy, and then set the path integral of step edge energy per unit length times the displacement around the island to derive the magnitude of the step edge energy for any step orientation. While this has never been done for AFM studies of solution crystal growth, we refer the reader to the STM study of Icking-Konert et al.[41] for more detailed discussion, as well as the work of Teng et al.[15] and Orme at al.[42] where many of the individual pieces of this calculation are worked out.

A few authors have reported the occurrence of 3D nucleation on protein and virus crystals. An example is shown in **Fig. 15.** These nuclei consist of molecular clusters that land on the surface of the host crystal. But rather than become incorporated in a misoriented state, the molecules rearrange within the cluster to form multi-layer stacks in epitaxial registry with the underlying lattice. In some cases, such as on canavalin, these 3D nuclei are created bit ejection of material from the crystal surface by the tip as it strikes the surface with high force or moves across a crystal edge. But in other cases, such as on STMV[23] or apoferritin, the authors reported the occurrence of this phenomenon in the absence of these effects. Yau and Vekilov[40] used this phenomenon on glass substrates to investigate pathways to nucleation of apoferritin.

4.4 Measuring step speeds

Step speed is one of the simplest measures of attachment/detachment kinetics at crystal surfaces. Unlike other methods of monitoring crystal growth, there is no ambiguity about the step heights. Thus the speed of both elementary steps and macrosteps can be individually determined. There are three basic techniques for measuring step speed. In the first, one uses a dislocation as a fixed point on the surface and monitors the motion of steps away from or towards that point as illustrated in **Fig. 16 a** and **b**. In the second method, a sequence of upward and downward scans is collected. From the change in the angle of the steps between the upward and downward scans, one can determine both the true step direction and the step speed, v, as illustrated in **Fig. 16 c** and **d** using the relationship:
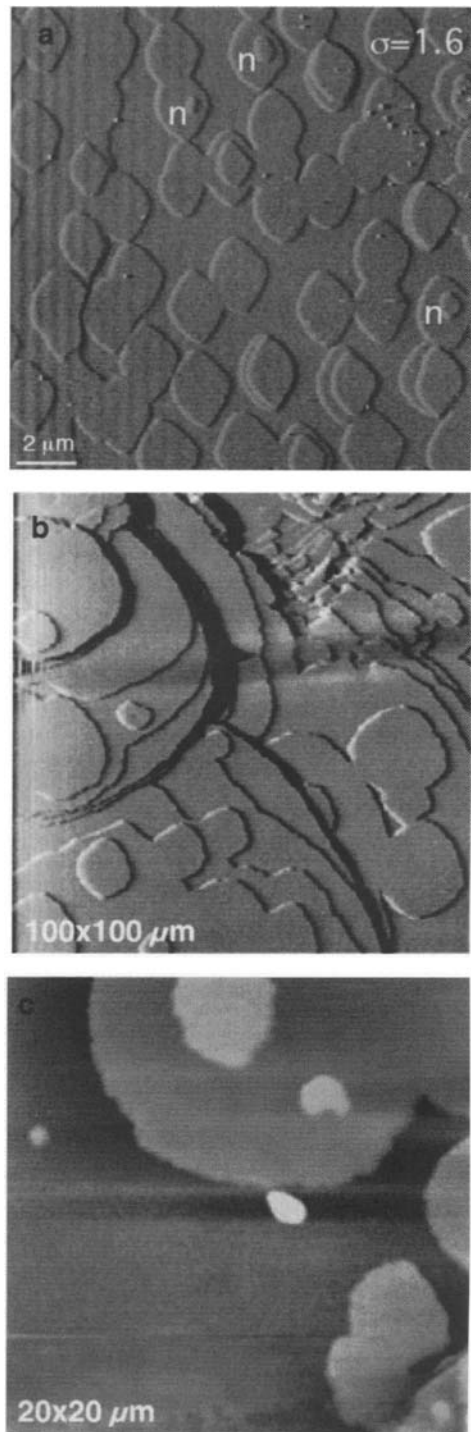
Figure 14. Examples of 2D nucleation on (a) calcite, (b) canavalin, and (c) STMV.
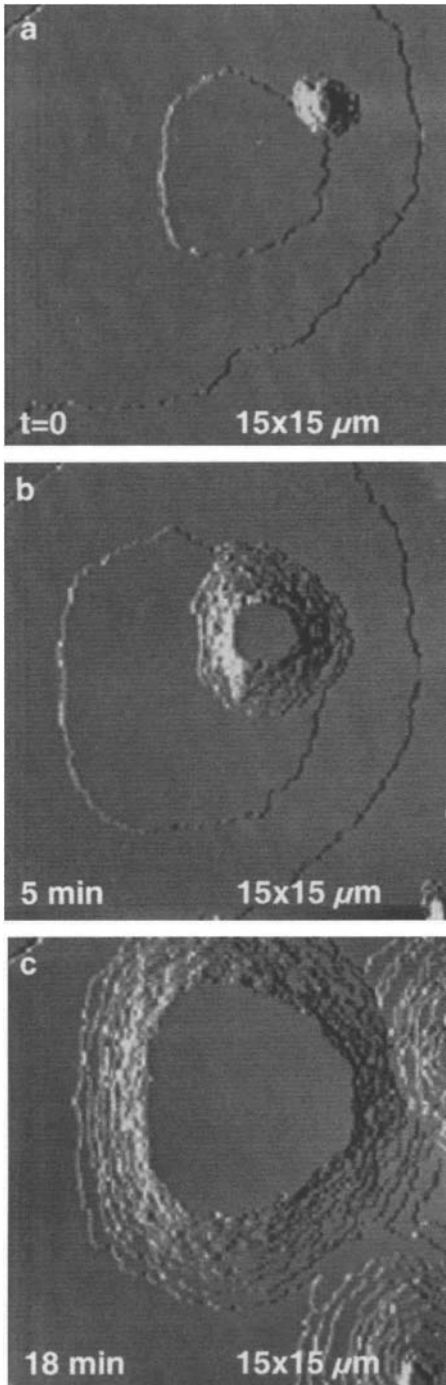
Figure 15. Example of deposition and growth of 3D nuclei on {111} face of STMV. After appearance, nuclei rearrange to grow in registry with underlying crystal. (Courtesy of Alex Malkin.)

$$v = v_{tip}[\cot\theta_D - \cot\theta_U] \sin\theta_T$$

$$2\cot(\theta_T) = \cot\theta_D + \cot\theta_U$$

$$v_{tip} = RL/2N$$

where θ_D is the angle during the down scan, θ_U is the angle during the up scan, θ_T is the true angle of the step train, R is the scan rate, L is the scan size, and N is the number of lines per scan. Finally, if one arranges the angle of the fast scan axis to be perpendicular to the true step direction, then disabling the slow scan axis provides a time profile of step position as illustrated in **Fig. 16 e** and **f**. **Figure 17** shows the dependence of step speed on solute concentration for brushite and calcite[16]. (The data for calcite show significant non-linear behavior that the authors attributed to impurity effects.)
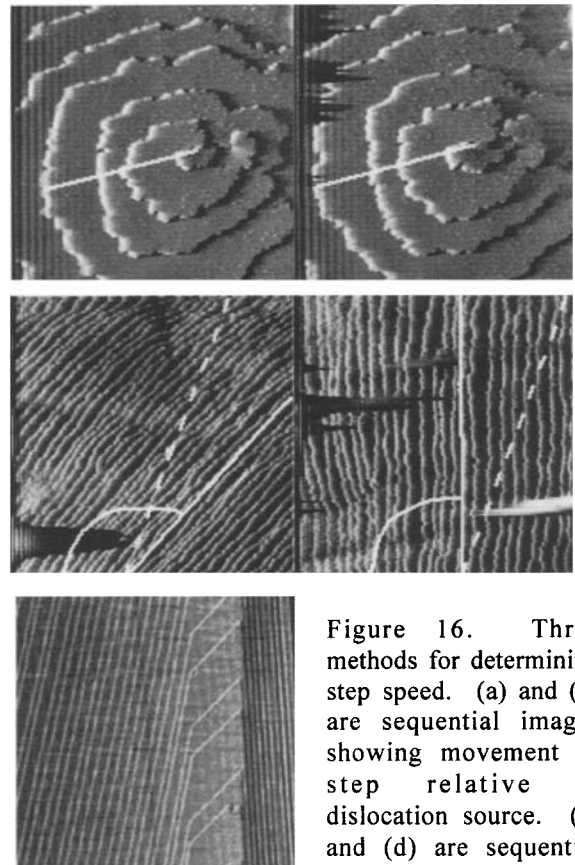
Figure 16. Three methods for determining step speed. (a) and (b) are sequential images showing movement of step relative to dislocation source. (c) and (d) are sequential images showing the change in apparent angle of step train from a downward to upward scan. (e) shows image of steps near top of hillock with the slow scan direction (vertical) disabled. The horizontal position gives distance and the vertical position gives time.
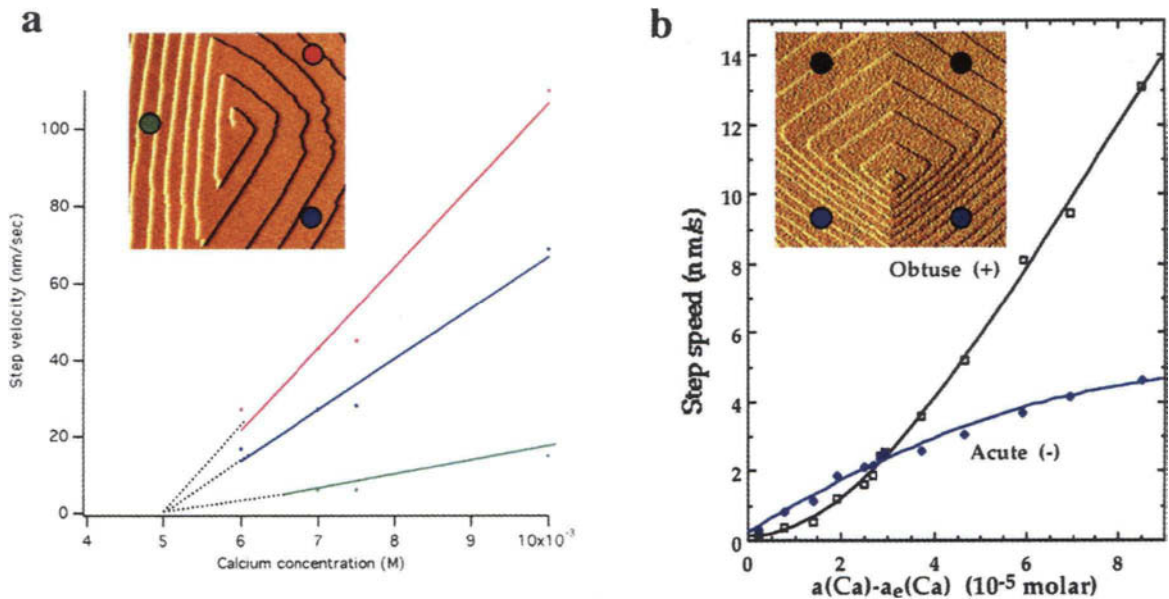
Figure 17.   Dependence of step speed on (a) Ca concentration for brushite (courtesy A. Villacampa) and (b) Ca^{2+} activity for calcite.

Due to the Gibbs-Thomson effect, when the step length is near that of the critical length, the step speed should depend on length as well as supersaturation.   The AFM is the only instrument capable of probing this phenomenon.   For a system in which step speeds are small and critical lengths are large, this measurement is readily performed. However, the same caveats about measuring critical length in "uncooperative " systems discussed above apply here.

Teng et al.[15] and, later, Higgins et al.[43], measured the dependence of step speed on length and found a sharp discrepancy between the experimental results and that expected from standard crystal growth theory.   These measurements were the first to provide evidence that the rough-step model of growth is incorrect for many systems.   Rather, a one-dimensional nucleation model may be needed to understand the growth data.   **Figure 18** shows the dependence of step speed on length for calcite.

4.5 Step roughness

The roughness of the step edge, temporal and spatial fluctuations of step position, and fluctuations in terrace width can be related to attachment/detachment kinetics and step stiffness. A number of STM investigations of these phenomena
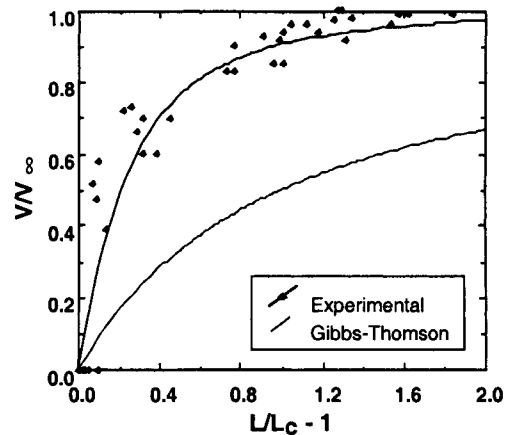
Figure 18.   Dependence of step speed on step length for calcite {104} surface.   Data points were collected at multiple supersaturations.   Red curve shows the Gibbs-Thomson prediction.   v = actual step speed, v_∞ = limiting step speed, L = step length, L_c = critical step length.

have been done on metal and semiconductor surfaces. But until very recently, no such studies had been performed on crystals during growth from solution.   Some of the systems that have been examined or are currently under investigation include silver[44], apoferritin[45], KDP[3], lysozyme[46], brushite, and calcite. What has been

published to date shows variable behavior and is not yet sufficient to allow one to draw a general picture of step edge dynamics in solution.

Figure 19 shows some examples of step edge roughness for a number of systems. As the figure shows, roughness varies dramatically from one system to another, with kink densities ranging from less than one in a hundred to nearly unity. Disabling the slow scan axis and monitoring the position of a point along the step, allow one to extract the time dependence of the height-height correlation function for that position. From the slope of this curve, one can extract information about the pathway by which solute molecules make their way to the step edge. Similar information can be obtained from the distribution of terrace widths. **Figure 20** shows the terrace width distribution for brushite along with the distribution expected from a simple model that considers only strong repulsive interactions between the steps.

4.6 Impurity effects

The final example that we present in this chapter, is the use of AFM to examine the effect of impurities on growth and dissolution. Three types of impurities have been examined quantitatively: those that pin step motion, impurities that get incorporated and change the solubility but have little effect on step kinetics, and those that act like surfactants to modify the surface energies, but are neither incorporated nor cause step pinning. As an example of the first type, Land et al.[47], looked at the effect of trivalent cations of Fe, Cr and Al on the growth of KDP. This study revealed behavior that could not have been detected or understood without the resolution that the AFM provides. The authors showed that, after doping the growth solution with Fe^{3+} of sufficient concentration to drive the surface into the dead zone, as the supersaturation was

250x200nm   3x3μm   610x610nm   225x225nm

Figure 19.   Examples of variability in step roughness or solution grown crystals. (a) canavalin, (b) two step directions on brushite, (c) calcite and (d) lysozyme. (d is courtesy of L.N. Rashkovich.)

Figure 20. Terrace width distributions for trains of steps on brushite P(l) dl gives the expected number of terraces with width between l and l+dl. Here, l is the normalized width: l=L/L$_{av}$, where L is the width measured from the AFM image. [Since the images were not processed for the scan speed, etc., L is not the true width of the terrace; thus the advantage of normalizing by L$_{av}$.] There are 320 data points (i.e., distinct pairs of steps) in the set. The standard deviation is 0.28 and the skewness is +0.04. The area under the curves or the data bars is unity, i.e., P(l) is normalized. The solid line is a Gaussian fit (it does not capture the skewness). The dashed line is the fit using a generalized Wigner distribution, $P(l)=a_r l^r exp(-b_r l^2)$, with $b_r=6$ corresponding to strongly repulsive interactions between the steps (it captures the skewness but does not recover the peak very well). Courtesy M. Bartelt and A. Villacampa.

increased, the emergence of the crystal surface from the dead zone was led by the motion of macrosteps. Furthermore, at supersaturations where the elementary steps were moving, the macrosteps moved at higher velocity. A sample of these results is shown in **Fig. 21**.

As an example of an incorporating impurity species, Davis et al.[10] looked at the effect of Mg on calcite growth. The authors found that the well documented inhibiting effect of Mg on calcite growth was not due to step pinning, but rather to a shift in the equilibrium solubility of the crystal with increasing Mg content. The dependence of step speed on Ca concentration for increasing Mg level is shown in **Fig. 22**.

Finally, the effect of acidic amino acids on calcite growth was examined by Orme et al.[42] in an attempt to understand some of the controls on crystallization exerted by living organisms during the formation of biominerals. Related experiments were carried out by Walters et al.[18] and Thompson et al.[48] **Figure 23** shows the effect of aspartic acid on calcite growth and dissolution. Individual acid enantiomers induce chirality in the growth hillock and etchpit geometries. Orme et al.[42] used these experiments to deduce binding relationships between the impurities and the crystal surface.

5 CONCLUSIONS

The application of atomic force microscopy to solution crystal growth has had and continues to have a major impact on our understanding of solution crystal growth physics. In certain areas, such as dislocation source dynamics, there appears

Figure 21. Effect of Fe^{3+} on morphology of growing KDP surface. The AFM images in **a** (upward scan) and **b** (downward scan) show that just above the dead zone, macro steps (heavy white step) are mobile while elementary steps remain stationary. Elementary steps are stripped off of the macrostep in a piecewise fashion (see black arrows) leaving a finger-like pattern of steps. The images in **c** (downward scan) and **d** (upward scan) show that, well above the dead zone, the macrosteps still move faster than the elementary steps. (The change in angle of the steps between images is larger for the macrosteps than the elementary steps.) (From [47])

to be little left to learn. However in the area of step kinetics, there are clear discrepancies between standard models and system behavior. We do not have a clear picture of mass transport process near

Figure 22. Measured dependence of step velocity on Ca^{2+} activity, a_{Ca2+} as a function of Mg^{2+} concentration. The solubility shifts to higher activities while the kinetic coefficient remains roughly constant. The solution activity of magnesium is expressed in a shorthand form where 320xMg = 5(64xMg) = 1.59×10^{-4} (from [10]).
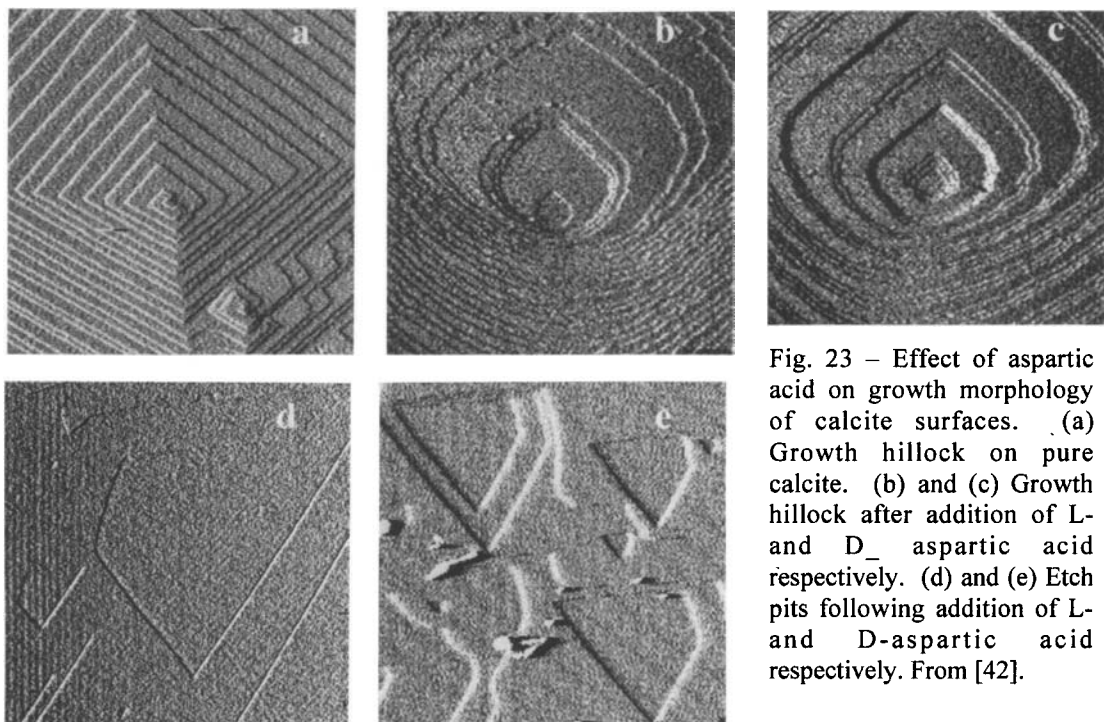
Fig. 23 – Effect of aspartic acid on growth morphology of calcite surfaces. (a) Growth hillock on pure calcite. (b) and (c) Growth hillock after addition of L- and D_ aspartic acid respectively. (d) and (e) Etch pits following addition of L- and D-aspartic acid respectively. From [42].

the surface, nor do we understand the extent to which creation and destruction of new units (often referred to as 1D nuclei) along step edges controls the kinetics of growth. In short, a detailed picture of step edge dynamics has yet to emerge. Thus one of the most important results on the horizon is the precise determination of the supersaturation dependence of kink density and step edge roughness. The development of mass produced nanotube probes, ultra-small cantilevers to reduce thermal noise, and the ability to scan at faster rates is currently underway. These advances will help enormously in sorting out the details of molecular attachment and detachment at step edges.

A second area of fertile ground is that of impurity-surface interactions. Only a handful of quantitative studies has been done in this area, yet impurities are ubiquitous in virtually all natural and synthetic solution growth systems. In biological systems in particular, impurities are intentionally added to control the location and timing of nucleation, crystallographic orientation of the nuclei, and the growth morphology of the resulting crystal. The effort to understand the physics of these processes is in its infancy and the AFM promises to play a prominent role in developing that understanding.

REFERENCES

1. K. Onuma, T. Kameyama and K. Tsukamoto, *J Cryst Growth* **137**, (1994) 610.
2. G. Binnig, *Ultramicroscopy* **42**, (1992) 7.
3. L. N. Rashkovich, O. A. Shustin and T. G. Chernevich, *Phys Solid State* **42**, (2000) 1921.
4. G. Binnig and H. Rohrer, *Ibm J Res Develop* **44**, (2000) 279.
5. T. A. Land and J. J. De Yoreo, *J Cryst Growth* **208**, (2000) 623.
6. T. A. Land, J. J. DeYoreo and J. D. Lee, *Surface Sci* **384**, (1997) 136.
7. T. A. Land, A. J. Malkin, Y. G. Kutznesov, A. McPherson and J. J. Deyoreo, *J Cryst Growth* **166**, (1996) 893.
8. T. A. Land, A. J. Malkin, Y. G. Kuznetsov, A. McPherson and J. J. Deyoreo, *Phys Rev Lett* **75**, (1995) 2774.

378

9. A. J. Gratz, S. Manne and P. K. Hansma, *Science* **251**, (1991) 1343.

10. K. J. Davis, P. M. Dove and J. J. De Yoreo, *Science* **290**, (2000) 1134.

11. A. J. Gratz, P. E. Hillner and P. K. Hansma, *Geochim Cosmochim Acta* **57**, (1993) 491.

12. P. E. Hillner, A. J. Gratz, S. Manne and P. K. Hansma, *Geology* **20**, (1992) 359.

13. P. E. Hillner, S. Manne, A. J. Gratz and P. K. Hansma, *Ultramicroscopy* **42**, (1992) 1387.

14. G. T. Paloczi, B. L. Smith, P. K. Hansma, D. A. Walters and M. A. Wendman, *Appl Phys Lett* **73**, (1998) 1658.

15. H. H. Teng, P. M. Dove, C. A. Orme and J. J. DeYoreo, *Science* **282**, (1998) 724.

16. H. H. Teng, P. M. Dove and J. J. DeYoreo, *Geochim Cosmochim Acta* **63**, (1999) 2507.

17. H. H. Teng, P. M. Dove and J. J. De Yoreo, *Geochim Cosmochim Acta* **64**, (2000) 2255.

18. D. A. Walters, et al., *Biophys J* **72**, (1997) 1425.

19. A. Wierzbicki, C. S. Sikes, J. D. Madura and B. Drake, *Calcified Tissue Int* **54**, (1994) 133.

20. J. J. DeYoreo, et al., *J Cryst Growth* **182**, (1997) 442.

21. J. J. DeYoreo, T. A. Land and J. D. Lee, *Phys Rev Lett* **78**, (1997) 4462.

22. T. A. Land, J. J. De Yoreo, T. L. Martin and G. T. Palmore, *Crystallogr Rep* **44**, (1999) 655.

23. A. J. Malkin, Y. G. Kuznetsov, T. A. Land, J. J. Deyoreo and A. McPherson, *Nature Struct Biology* **2**, (1995) 956.

24. R. E. Hirsch, et al., *Protein-Struct Funct Genet* **42**, (2001) 99.

25. Y. G. Kuznetsov, A. J. Malkin, W. Glantz and A. McPherson, *J Cryst Growth* **168**, (1996) 63.

26. A. J. Malkin, Y. G. Kuznetsov, R. W. Lucas and A. McPherson, *J Struct Biol* **127**, (1999) 35.

27. A. J. Malkin, T. A. Land, Y. G. Kuznetsov, A. McPherson and J. J. Deyoreo, *Phys Rev Lett* **75**, (1995) 2778.

28. S. Akita, H. Nishijima, Y. Nakayama, F. Tokumasu and K. Takeyasu, *J Phys-D-Appl Phys* **32**, (1999) 1044.

29. S. S. Wong, J. D. Harper, P. T. Lansbury and C. M. Lieber, *J Amer Chem Soc* **120**, (1998) 603.

30. J. H. Hafner, C. L. Cheung and C. M. Lieber, *J Amer Chem Soc* **121**, (1999) 9750.

31. S. R. Higgins, C. M. Eggleston, K. G. Knauss and C. O. Boro, *Rev Sci Instr* **69**, (1998) 2994.

32. S. R. Higgins, G. Jordan, C. M. Eggleston and K. G. Knauss, *Langmuir* **14**, (1998) 4967.

33. M. Plomp, K. Maiwa and W. J. P. van Enckevort, *J Cryst Growth* **199**, (1999) 246.

34. G. R. Ester, R. Price and P. J. Halfpenny, *J Cryst Growth* **182**, (1997) 95.

35. G. R. Ester and P. J. Halfpenny, *J Cryst Growth* **187**, (1998) 111.

36. P. G. Vekilov, Y. G. Kuznetsov and A. A. Chernov, *J Cryst Growth* **121**, (1992) 44.

37. P. G. Vekilov, Y. G. Kuznetsov and A. A. Chernov, *J Cryst Growth* **121**, (1992) 643.

38. L. N. Rashkovich, *KDP-family single crystals*, The Adam Hilger series on optics and optoelectronics (A. Hilger, Bristol, Eng. ; Philadelphia, 1991).

39. A. J. Malkin, Y. G. Kuznetsov, W. Glantz and A. McPherson, *J Phys Chem* **100**, (1996) 11736.

40. S. T. Yau and P. G. Vekilov, *Nature* **406**, (2000) 494.

41. G. S. Icking-Konert, M. Giesen and H. Ibach, *Phys Rev Lett* **83**, (1999) 3880.

42. C. A. Orme, et al., Nature, (2001) accepted.

43. S. R. Higgins, D. Bosbach, C. M. Eggleston and K. G. Knauss, *J Phys Chem B* **104**, (2000) 6978.

44. M. Giesen, M. Dieterle, D. Stapel, H. Ibach and D. M. Kolb, *Surface Sci* **384**, (1997) 168.

45. S. T. Yau, B. R. Thomas and P. G. Vekilov, *Phys Rev Lett* **85**, (2000) 353.

46. L. N. Rashkovich, N. V. Gvozdev and I. V. Yaminsky, *Crystallogr Rep* **43**, (1998) 696.

47. T. A. Land, T. L. Martin, S. Potapenko, G. T. Palmore and J. J. De Yoreo, *Nature* **399**, (1999) 442.

48. J. B. Thompson, et al., *Biophys J* **79**, (2000) 3307.

BIBLIOGRAPHY

1. S. Akita, H. Nishijima, Y. Nakayama, F. Tokumasu and K. Takeyasu, *J Phys-D-Appl Phys* **32**, (1999) 1044.

2. G. Binnig, *Ultramicroscopy* **42**, (1992) 7.

3. G. Binnig and H. Rohrer, *Ibm J Res Develop* **44**, (2000) 279.

4. D. Bosbach, A. Putnis, U. Bismayer and B. Guttler, *J Phys-Condens Matter* **9**, (1997) 8397.

5. D. Bosbach, C. Hall and A. Putnis, *Chem Geol* **151**, (1998) 143.

6. P. A. Campbell, G. R. Ester and P. J. Halfpenny, *J Vac Sci Technol B* **14**, (1996) 1373.

7. A. A. Chernov, L. N. Rashkovich, I. V. Yaminski and N. V. Gvozdev, *J Phys-Condens Matter* **11**, (1999) 9969.
8. K. J. Davis, P. M. Dove and J. J. De Yoreo, *Science* **290**, (2000) 1134.
9. S. A. de Vries, et al., *J Cryst Growth* **205**, (1999) 202.
10. S. A. deVries, et al., *Phys Rev Lett* **80**, (1998) 2229.
11. J. J. Deyoreo, T. A. Land and B. Dair, *Phys Rev Lett* **73**, (1994) 838.
12. J. J. DeYoreo, et al., *J Cryst Growth* **182**, (1997) 442.
13. J. J. DeYoreo, T. A. Land and J. D. Lee, *Phys Rev Lett* **78**, (1997) 4462.
14. S. D. Durbin and W. E. Carlson, *J Cryst Growth* **122**, (1992) 71.
15. G. R. Ester, R. Price and P. J. Halfpenny, *J Cryst Growth* **182**, (1997) 95.
16. G. R. Ester and P. J. Halfpenny, *J Cryst Growth* **187**, (1998) 111.
17. G. R. Ester, R. Price and P. J. Halfpenny, *J Phys-D-Appl Phys* **32**, (1999) A128.
18. G. R. Ester and P. J. Halfpenny, *Phil Mag a* **79**, (1999) 593.
19. M. Giesen, M. Dietterle, D. Stapel, H. Ibach and D. M. Kolb, *Surface Sci* **384**, (1997) 168.
20. A. J. Gratz, S. Manne and P. K. Hansma, *Science* **251**, (1991) 1343.
21. A. J. Gratz, P. E. Hillner and P. K. Hansma, *Geochim Cosmochim Acta* **57**, (1993) 491.
22. N. Gvozdev, L. Rashkovich and I. Yaminsky, *Macromol Symposia* **160**, (2000) 49.
23. J. H. Hafner, C. L. Cheung and C. M. Lieber, *J Amer Chem Soc* **121**, (1999) 9750.
24. S. R. Higgins, C. M. Eggleston, K. G. Knauss and C. O. Boro, *Rev Sci Instr* **69**, (1998) 2994.
25. S. R. Higgins, G. Jordan, C. M. Eggleston and K. G. Knauss, *Langmuir* **14**, (1998) 4967.
26. S. R. Higgins, D. Bosbach, C. M. Eggleston and K. G. Knauss, *J Phys Chem B* **104**, (2000) 6978.
27. A. C. Hillier and M. D. Ward, *Science* **263**, (1994) 1261.
28. P. E. Hillner, A. J. Gratz, S. Manne and P. K. Hansma, *Geology* **20**, (1992) 359.
29. P. E. Hillner, S. Manne, A. J. Gratz and P. K. Hansma, *Ultramicroscopy* **42**, (1992) 1387.
30. P. E. Hillner, S. Manne, P. K. Hansma and A. J. Gratz, *Faraday Discuss* , (1993) 191.
31. R. E. Hirsch, et al., *Protein-Struct Funct Genet* **42**, (2001) 99.
32. G. S. Icking-Konert, M. Giesen and H. Ibach, *Phys Rev Lett* **83**, (1999) 3880.
33. Y. G. Kuznetsov, A. J. Malkin, W. Glantz and A. McPherson, *J Cryst Growth* **168**, (1996) 63.
34. Y. G. Kuznetsov, et al., *Biophys J* **72**, (1997) 2357.
35. Y. G. Kuznetsov, A. J. Malkin and A. McPherson, *J Struct Biol* **120**, (1997) 180.
36. Y. G. Kuznetsov, A. J. Malkin and A. McPherson, *Phys Rev B-Condensed Matter* **58**, (1998) 6097.
37. Y. G. Kuznetsov, J. Konnert, A. J. Malkin and A. McPherson, *Surface Sci* **440**, (1999) 69.
38. T. A. Land, A. J. Malkin, Y. G. Kuznetsov, A. McPherson and J. J. Deyoreo, *Phys Rev Lett* **75**, (1995) 2774.
39. T. A. Land, A. J. Malkin, Y. G. Kutznesov, A. McPherson and J. J. Deyoreo, *J Cryst Growth* **166**, (1996) 893.
40. T. A. Land, J. J. DeYoreo and J. D. Lee, *Surface Sci* **384**, (1997) 136.
41. T. A. Land, J. J. De Yoreo, T. L. Martin and G. T. Palmore, *Crystallogr Rep* **44**, (1999) 655.
42. T. A. Land, T. L. Martin, S. Potapenko, G. T. Palmore and J. J. De Yoreo, *Nature* **399**, (1999) 442.
43. T. A. Land and J. J. De Yoreo, *J Cryst Growth* **208**, (2000) 623.
44. A. J. Malkin, Y. G. Kuznetsov, T. A. Land, J. J. Deyoreo and A. McPherson, *Nature Struct Biology* **2**, (1995) 956.
45. A. J. Malkin, T. A. Land, Y. G. Kuznetsov, A. McPherson and J. J. Deyoreo, *Phys Rev Lett* **75**, (1995) 2778.
46. A. J. Malkin, Y. G. Kuznetsov, W. Glantz and A. McPherson, *J Phys Chem* **100**, (1996) 11736.
47. A. J. Malkin, Y. G. Kuznetsov, R. W. Lucas and A. McPherson, *J Struct Biol* **127**, (1999) 35.
48. A. J. Malkin, Y. G. Kuznetsov and A. McPherson, *J Cryst Growth* **196**, (1999) 471.
49. G. Z. Mao, L. Lobo, R. Scaringe and M. D. Ward, *Chem Mater* **9**, (1997) 773.
50. A. McPherson, A. J. Malkin and Y. G. Kuznetsov, *Annu Rev Biophys Biomol Struc* **29**, (2000) 361.
51. J. D. Ng, et al., *Nucl Acid Res* **25**, (1997) 2582.
52. F. Ohnesorge and G. Binnig, *Science* **260**, (1993) 1451.
53. K. Onuma, T. Kameyama and K. Tsukamoto, *J Cryst Growth* **137**, (1994) 610.
54. C. A. Orme, et al., Nature, (2001) accepted.
55. G. T. Paloczi, B. L. Smith, P. K. Hansma, D. A. Walters and M. A. Wendman, *Appl Phys Lett* **73**, (1998) 1658.
56. N. S. Park, M. W. Kim, S. C. Langford and J. T. Dickinson, *Langmuir* **12**, (1996) 4599.
57. D. N. Petsev, B. R. Thomas, S. T. Yau and P. G. Vekilov, *Biophys J* **78**, (2000) 2060.

380

58. C. M. Pina, U. Becker, P. Risthaus, D. Bosbach and A. Putnis, *Nature* **395**, (1998) 483.

60. R. Price, G. R. Ester and P. J. Halfpenny, *Proc Roy Soc London Ser a* **455**, (1999) 4117.

61. L. N. Rashkovich, *KDP-family single crystals*, The Adam Hilger series on optics and optoelectronics (A. Hilger, Bristol, Eng. ; Philadelphia, 1991).

62. L. N. Rashkovich, N. V. Gvozdev and I. V. Yaminsky, *Crystallogr Rep* **43**, (1998) 696.

63. L. N. Rashkovich, O. A. Shustin and T. G. Chernevich, *J Cryst Growth* **206**, (1999) 252.

64. L. N. Rashkovich, O. A. Shustin and T. G. Chernevich, *Phys Solid State* **42**, (2000) 1921.

65. L. Scudiero, S. C. Langford and J. T. Dickinson, *Tribol Lett* **6**, (1999) 41.

66. H. D. Sikes and D. K. Schwartz, *Science* **278**, (1997) 1604.

67. H. H. Teng, P. M. Dove, C. A. Orme and J. J. DeYoreo, *Science* **282**, (1998) 724.

68. H. H. Teng, P. M. Dove and J. J. DeYoreo, *Geochim Cosmochim Acta* **63**, (1999) 2507.

69. H. H. Teng, P. M. Dove and J. J. De Yoreo, *Geochim Cosmochim Acta* **64**, (2000) 2255.

70. J. B. Thompson, et al., *Biophys J* **79**, (2000) 3307.

71. P. G. Vekilov, Y. G. Kuznetsov and A. A. Chernov, *J Cryst Growth* **121**, (1992) 643.

72. P. G. Vekilov, Y. G. Kuznetsov and A. A. Chernov, *J Cryst Growth* **121**, (1992) 44.

59. M. Plomp, K. Maiwa and W. J. P. van Enckevort, *J Cryst Growth* **199**, (1999) 246.

73. D. A. Walters, et al., *Biophys J* **72**, (1997) 1425.

74. A. Wierzbicki, C. S. Sikes, J. D. Madura and B. Drake, *Calcified Tissue Int* **54**, (1994) 133.

75. A. Wierzbicki, J. D. Sallis, E. D. Stevens, M. Smith and C. S. Sikes, *Calcified Tissue Int* **61**, (1997) 216.

76. S. S. Wong, J. D. Harper, P. T. Lansbury and C. M. Lieber, *J Amer Chem Soc* **120**, (1998) 603.

77. S. T. Yau, D. N. Petsev, B. R. Thomas and P. G. Vekilov, *J Mol Biol* **303**, (2000) 667.

78. S. T. Yau and P. G. Vekilov, *Nature* **406**, (2000) 494.

79. S. T. Yau, B. R. Thomas and P. G. Vekilov, *Phys Rev Lett* **85**, (2000) 353.

80. S. T. Yau and P. G. Vekilov, *J Am Chem Soc* **123**, (2001) 1080.

81. C. M. Yip and M. D. Ward, *Biophys J* **71**, (1996) 1071.

82. C. M. Yip, M. R. DeFelippis, B. H. Frank, M. L. Brader and M. D. Ward, *Biophys J* **75**, (1998) 1172.

83. C. M. Yip, C. C. Yip and M. D. Ward, *Biochemistry-Usa* **37**, (1998) 5439.

84. C. M. Yip, M. L. Brader, M. R. DeFelippis and M. D. Ward, *Biophys J* **74**, (1998) 2199.

85. C. M. Yip, M. L. Brader, B. H. Frank, M. R. DeFelippis and M. D. Ward, *Biophys J* **78**, (2000) 466.

Crystal Morphology Control with Tailor-Made Additives; A Stereochemical Approach

Isabelle Weissbuch, Meir Lahav, Leslie Leiserowitz

Department of Materials and Interfaces,
The Weizmann Institute of Science, Rehovot 76100 Israel

Tailor-made molecules have been designed and used as additives, in crystallization processes from solution, to control the nucleation and growth of molecular crystals. These additives can be either inhibitors or promoters of crystal growth. Tailor-made inhibitors can be used for a variety of purposes that include crystal morphology engineering, kinetic resolution of racemates, reduction of crystal symmetry, assignment of absolute configuration of chiral molecules and polar crystals and control of crystal polymorphism. As for crystal growth promoters tailor-made additives, that self-assemble at the air-aqueous solution interface due to their intrinsic hydrophobic nature, have been used to induce oriented crystallization by means of a structural match.

1. INTRODUCTION

Crystals and the crystallization process from solutions play an important role in pure and applied sciences. However, the ability to design crystals with desired properties and the control over crystallization, based on structural understanding, are still limited when compared with the controlled crystal growth provided in biomineralization processes.

The ground rules for a quantitative determination of crystal morphology based on the internal crystal structure were laid by Hartman, Perdok and Bennema[1-5] in their landmark papers on the prediction of crystal morphology. Regarding the effect of additives and solvent on crystal growth and morphology, a variety of studies have been reported around the mid 20th century. The control on nucleation, growth, dissolution and morphology of molecular crystals, however, remained primarily a matter of 'mix and try'.

In the last 20 years, a stereochemical approach has been adopted involving growth and dissolution of molecular crystals in the presence of 'tailor-made' additives that interact with specific crystal faces and may be crystal growth inhibitors or nucleation promoters.[6-14] The surfaces of a growing crystal can be thought of as composed of 'active sites' that interact stereospecifically with molecules in solution, in a manner similar to

enzyme-substrate or antibody-antigen interactions. The repetitive arrangement at crystal surfaces and the knowledge of their structures offer simpler means to pinpoint intermolecular interactions by using the concept of molecular recognition at interfaces. With the assistance of tailor-made additives, it is possible to examine a variety of processes concerning open questions in the field of crystal nucleation and growth, interactions with the growth environment and crystal polymorphism.

2. CRYSTAL MORPHOLOGY ENGINEERING

The habit of a crystal is defined by the relative rates of its growth in different directions; the faster the growth in a given direction the smaller the face developed perpendicular thereto. Consequently, when growth is inhibited in a direction perpendicular to a given face, the area of this face is expected to increase relative to the areas of other faces of the same crystal. Differences in the relative surface areas of the various faces can therefore be directly correlated to the inhibition in different growth directions. Such a change in crystal morphology is shown in Scheme 1.

Dramatic changes in morphology observed during growth of molecular crystals in the presence of additives revealed a high degree of specificity in the interaction of the foreign material with the different structures of the growing surfaces.

Unaffected crystal growth
rate of growth B>A

Stereoselective additive adsorption
inhibits growth ⊥ B

Affected crystal growth
rate of growth A>B
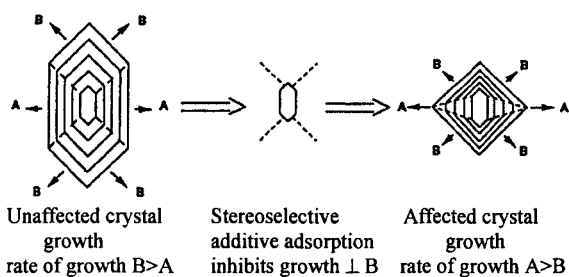
Scheme 1

Crystal morphology engineering became possible with the use of tailor-made additives of molecular structure similar to the structure of the corresponding host molecules. From a systematic study of a variety of organic compounds crystallized in the presence of additives, it was possible to deduce a stereochemical correlation between the structures of the affected crystal surfaces and the molecular structure of the additive. We could infer that the tailor-made additive is recognized only at certain surfaces on the growing crystals, in a manner similar to the host molecules, and then adsorbed at those faces with the part of the additive that differs from that of the host emerging from the crystal. Such selective adsorption of the tailor-made additive at crystal faces and subsequent inhibition of their growth are depicted in Scheme 2.
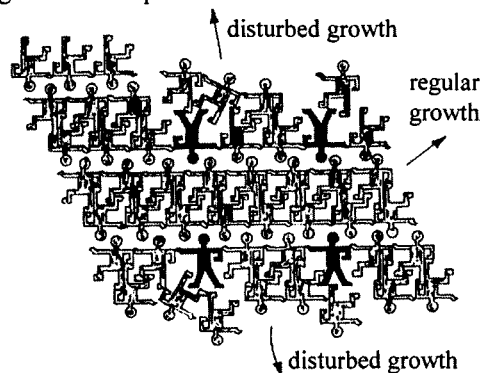
Scheme 2

Once this mechanism was established, it became possible to exploit it to systematically modify the morphology of crystals by tailoring additives that bind at preselected faces and thus inhibit growth in a predictable manner.

An example[6] where the host-tailor-made additive interactions at the growing solid-liquid interface can be pinpointed and correlated with the change in crystal morphology is benzamide, $C_6H_5CONH_2$. This compound crystallizes from

ethanol as plates, in the monoclinic space group $P2_1/c$. The packing arrangement in this crystal can be described in terms of hydrogen-bonded cyclic molecular dimers interlinked by hydrogen bonds to yield ribbons parallel to the b axis (Fig 1a). The ribbons are stacked along the a axis to form stable (001) layers.
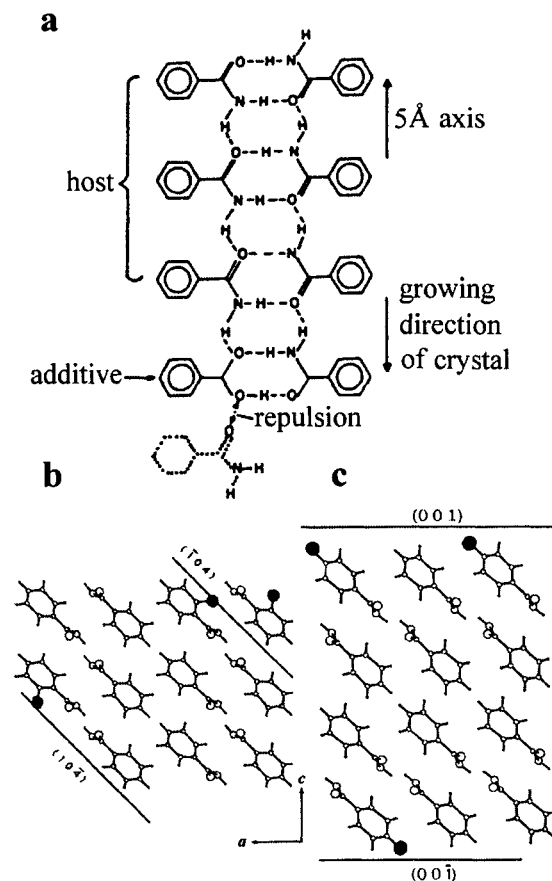
Fig. 1. (a) Schematic representation of the ribbon motif of hydrogen-bonded dimers of benzamide molecules interlinked along the 5 Å b-axis, showing the effect of benzoic acid additive.
(b, c) Packing arrangement of benzamide, viewed along the b axis, showing the effect of o-toluamide and p-toluamide, CH_3-$C_6H_4CONH_2$, respectively.

These tightly packed layers juxtapose along the c direction, interacting via weak van der Waals interactions between phenyl groups, thus accounting for the {001} plate-like shape of the crystals.

Benzoic acid C_6H_5COOH additive, adopting the stable O=C-OH syn planar conformation, can replace a molecule of benzamide at the end of a

ribbon (Fig. 1a); however, at the site of the additive, an attractive N-H$\cdots$O bond is replaced by repulsion between adjacent oxygen lone-pair electrons of the bound additive molecule and of the oncoming benzamide molecule. As predicted, benzoic acid as additive inhibits growth of the benzamide crystals along b, transforming the pure plates (Fig. 2a) into [100] needles (Fig. 2b). Inhibition of growth along a was accomplished with o-toluamide that yielded [010] bars (Fig. 2c). This additive can easily be adsorbed in the hydrogen-bonding chain, without disturbing growth along b, but with the methyl group, that emerges from the $(10\bar{4})$ side-face (Fig. 1b) interfering with growth in the a direction along which the dimers are stacked. Thinner plates, (Fig. 2d), are obtained by addition of increasing amounts of p-toluamide since its methyl group perturbs the already weak van der Waals interactions between the phenyl layers in the c direction (Fig. 1c).
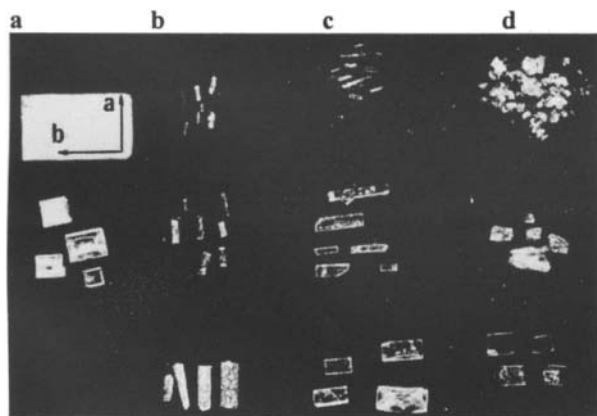
Fig. 2. Photographs of benzamide crystals: (a) pure; (b-d) grown in the presence of increasing amounts (from bottom to top) of (b) benzoic acid, (c) o-toluamide, (d) p-toluamide.

The two-step mechanism of stereospecific binding of tailor-made additives at the growing crystal surfaces followed by inhibition of growth perpendicular to these faces has been applied to a variety of molecular crystals including α-aminoacids, di-peptides, primary and secondary amides, carboxylic acids, sugars, steroids, etc., and is therefore general. It has been also successfully applied in the separation of enantiomers by kinetic resolution of conglomerates.

3. TAILOR-MADE INHIBITORS FOR RESOLUTION OF RACEMATES[7,8]

Application of the above two-step mechanism to racemic mixtures in which each enantiomer R- or S- crystallizes in enantiomorphous crystals (Scheme 3left) suggested the possibility of performing by these means a manual Pasteur-type separation, as well as a kinetic resolution. Enantiomorphous crystals behave differently when grown in the presence of enantiomeric tailor-made additives since diastereomeric interactions occur at the surfaces of the growing crystals. It follows directly from the mechanism illustrated in Scheme 3right that an enantiomeric additive S' consisting of a appropriately modified S molecule will in general be adsorbed only at the S crystal surface and not at the R crystal. As a consequence, the S crystal undergoes changes in morphology which allow easy visual identification and separation of the two enantiomorphs. This principle is illustrated with the following example.

| unmodified morphology | modified morphology |

Scheme 3
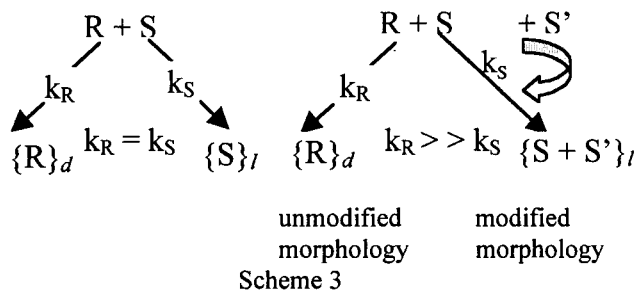
Racemic (R,S)-glutamic acid.HCl (Glu.HCl), $HOOC(CH_2)_2C^*H(NH_3^+)COOH$ crystallizes as a conglomerate of (R)- and (S)- enantiomorphs. The packing arrangement and morphology of (S)-Glu.HCl crystal are shown in Fig. 3. The molecular side chains are directed along the c axis and form hydrogen bonds in this direction. Therefore, various other α-aminoacids can be regarded as tailor-made additives since these molecules bear the residue $^+H_3NC^*HCOOH$ of the α-aminoacid group, whereas the side chain is different. When growing (S)-Glu.HCl in the presence of such S'-α-aminoacids, the additive will enantiospecifically adsorb at the two {001} crystal faces and inhibit crystal growth perpendicular to them since the hydrogen bonding along c will be disrupted. In contrast, R'-α-aminoacid additives cannot be adsorbed at the same surfaces because the molecular structure of the

substrate and additive are of opposite handedness. By symmetry, (R')-α-aminoacid additive can be adsorbed only at the {001} faces of (R)-Glu.HCl crystals. Indeed, crystals grown from racemic R,S-Glu.HCl in the presence of S-lysine additive (1-3% wt/wt of racemate) appear as plates of (R)-Glu.HCl covered by a crystalline powder of the (S)-Glu.HCl (Fig. 4). The pure plate-like crystals of (S)-Glu.HCl (Fig. 5a) grow thinner and become a powder with increasing S-lysine concentration (Fig. 5). Kinetic resolution of the racemic R,S-Glu·HCl can be achieved with the assistance of larger (up to 10% wt/wt) amounts of S-lysine additive that prevents the growth of (S)-Glu.HCl yielding preferred crystallization of (R)-Glu.HCl. This principle has been applied to other racemic systems, including (R,S)-asparagine monohydrate, (R,S)-threonine, (R,S)-*p*-hydoxy-phenylglycine.*p*-toluene-sulfonate.

The efficiency of the tailor-made additives could be significantly increased by anchoring the α-aminoacid moieties to a polymeric backbone. Addition of only 0.1-1% poly-(N$^{\varepsilon}$-metacryloyl-R-lysine), **1**, to supersaturated solutions of R,S-Glu.HCl could bring about the preferred crystallization of (R)-Glu.HCl.
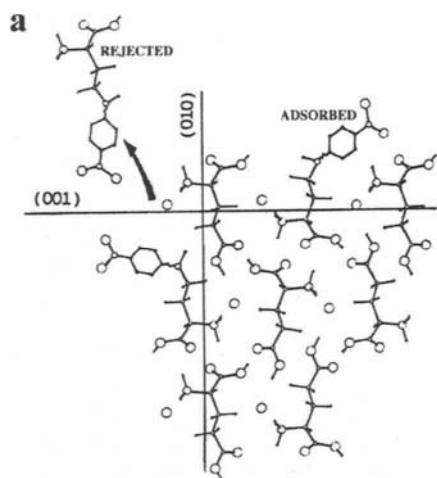
Fig. 3. (a) Packing arrangement of (S)-Glu.HCl viewed along the *a* axis and showing the effect of S-N$^{\gamma}$(*p*-nitro-phenyl)-α,γ-diamino-butyric acid, O_2N-C_6H_4-NH-$(CH_2)_2$CH(NH_3^+)COO$^-$, additive. (b) Morphology of pure crystals.
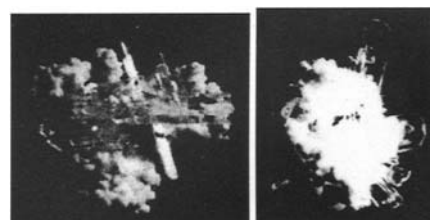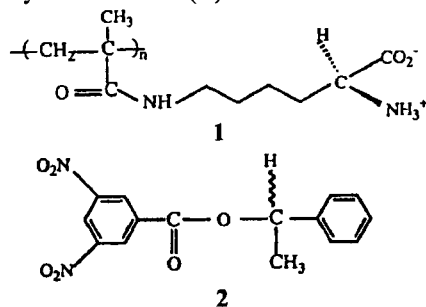
Fig. 4. Racemic Glu.HCl grown in the presence of S-lysine; the plates are the (R)-Glu.HCl, whereas the powder is the (S)-Glu.HCl.
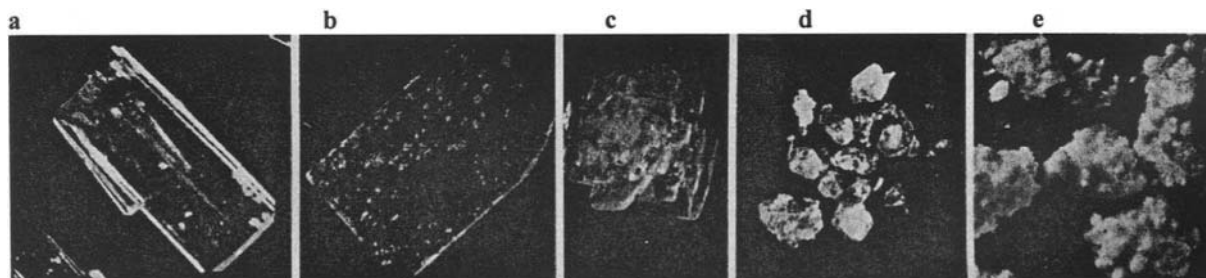
Fig. 5. Photographs of (S)-Glu.HCl crystals: (a) pure; (b-e) Grown in the presence of increasing amounts of S-lysine additive: (b) 2 mg/ml; (c) 5 mg/ml; (d) 10 mg/ml; (e) 50 mg/ml.

An industrially feasible kinetic process for the resolution of methionine.HCl has been reported [9]. The method is also applicable to molecular crystals lacking ionic and hydrogen bonds as in the kinetic resolution of racemic (R,S)-*sec*-phenethyl-3,5-dinitro-benzoate, **2**, with poly-(N-acryloyl-(*p*-aminobenzoyl)-R-*sec*-phenethyl-amide), **3**, additive.

The kinetic resolution may encompass preferred precipitation, from a racemate solution, of one enantiomorph at the expense of the more stable racemic crystal phase. Naturally, the energy gap between the racemic crystal and the mixture of resolved enantiomorphs must be relatively small. For example, racemic R,S-histidine.HCl, $C_3H_3N_2$-CH_2-$C^*H(NH_3^+)COO^-$, precipitates at 25°C as a stable racemic dihydrate crystal (R,S)-His.HCl.2H$_2$O. Above 45°C it precipitates as a metastable conglomerate of enantiomorphous monohydrate (R)- and (S)-His.HCl.H$_2$O crystals. Addition of as little as 1% of the enantiomeric poly-(*p*-acrylamido-phenylalanine), **4**, tailor-made inhibitor induced complete resolution of the desired enantiomer.[8] The polymeric inhibitor was designed by considering the packing arrangements of the two crystalline forms of histidine (Fig. 6). Following our adsorption/inhibition approach, we anticipated that an enantiomeric α-aminoacid bearing aromatic side groups grafted onto a polymeric backbone (to yield a water soluble polymer with free α-aminoacid head groups) should be enantiospecifically adsorbed onto the {011} faces of the racemic (R,S)-His.HCl.2H$_2$O form and inhibit its growth along the *b* direction (Fig. 6a).
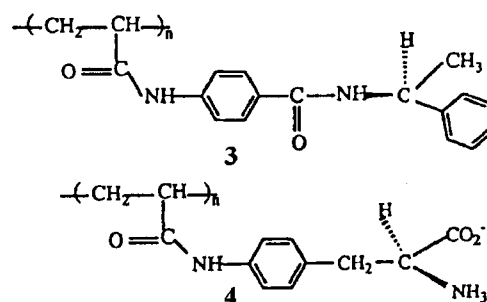
Analogously, the same polymer of say R-configuration should be nucleation and growth inhibitor of (R)-His.HCl.H$_2$O (Fig. 6b). Consequently, resolution of the racemate could be accomplished.

The need for a fine geometric match between the molecular structure of the α-aminoacid group grafted onto the polymer and that of the substrate crystal was further demonstrated by the fact that poly-(N$^\epsilon$-methacryloyl)lysine, **1**, could not prevent crystallization of the racemate even when present in high concentrations (~10% wt/wt).
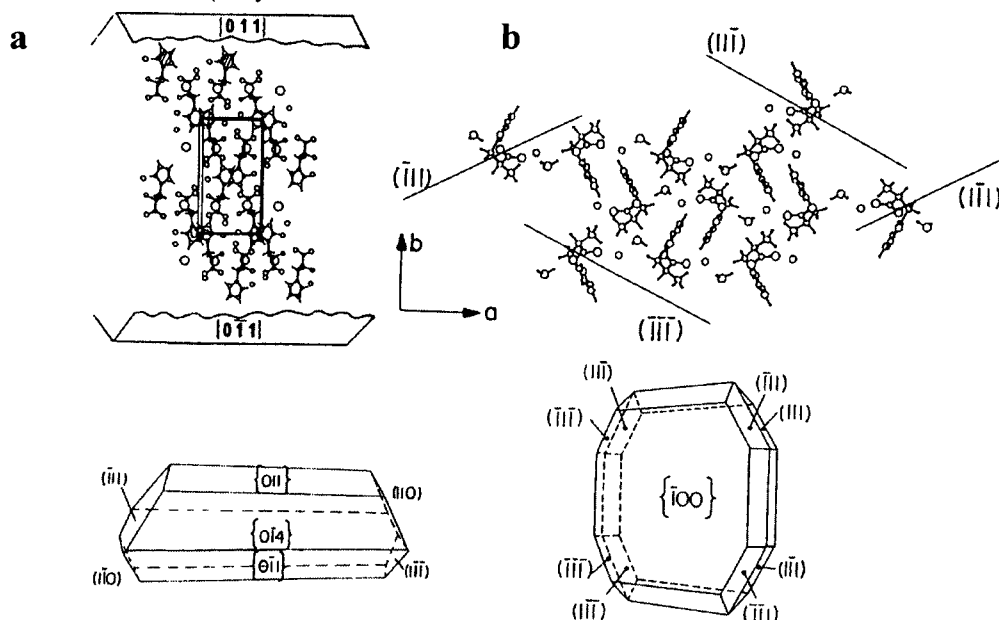
Fig. 6. Packing arrangement and computer-drawn morphology of: (a) racemic (R,S)-His.HCl.2H$_2$O; (b) (R)- or (S)-His.HCl.H$_2$O.

4. EFFECT OF TAILOR-MADE ADDITIVES ON GROWTH OF POLAR CRYSTALS

The effect of tailor-made molecular additives on crystal morphology and structure could be taken advantage of to demonstrate the interplay between molecular chirality and various macroscopic phenomena. The method has been applied to assign the absolute structures of polar crystals whose constituent molecules are either chiral, racemic mixtures or nonchiral.

4.1. Polar crystals composed of enantiomerically pure molecules[10]

We first focus on polar crystals composed of enantiomerically pure molecules. The method involves the principle of fixing the orientation of the constituent molecules *vis-a-vis* the polar axis, or axes, of the crystal and subsequently determining the absolute configuration of the chiral molecules. This principle is depicted in Scheme 4 that shows two enantiomorphous sets of hands arranged in a lattice; right hands forming an (R)-crystal and left hands forming an (S)-crystal. The fingers of the hands are exposed at the (010) face of the right-handed "molecules" and, by symmetry, at the (0$\bar{1}$0) face for the left-handed "molecules". Thus, by determining at which face of the crystal specimen the fingers or wrists are exposed, the handedness of the constituent molecules may be assigned.

Scheme 4
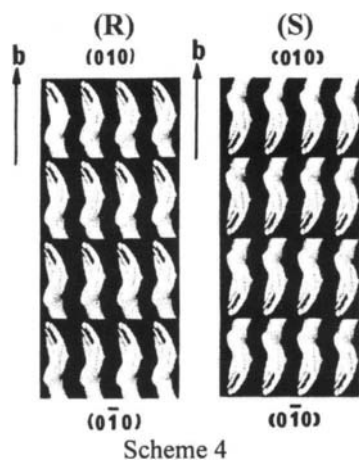
By applying the two-step adsorption-inhibition mechanism with appropriate tailor-made additives,

such assignment can be performed. We illustrate this approach with several examples.

N-(E-cinnamoyl)-S-alanine[7,10]

The molecules of N-(E-cinnamoyl)-S-alanine, C_6H_5-CH=CH-CONH-C*H(CH_3)-COOH, crystallize in space group $P2_1$. The packing arrangement (Fig. 7) and the morphology of the pure crystal grown from methanol solution (Fig. 8a) show that molecules are arranged with the COOH groups emerging at the two {1$\bar{1}$1} faces and with the C*(chiral)-H bonds directed along the +b axis. We found, as expected, that the methyl ester, C_6H_5-CH=CH-CONH-C*H(CH_3)-COOCH$_3$, of the same absolute configuration as the host molecule induced the appearance of large {1$\bar{1}$1} faces (Fig. 8b) since the O-CH$_3$ group replaces an O-H group emerging from such faces. Therefore the absolute direction of the O-H bond with respect to the polar b axis is fixed and so the absolute configuration of the substrate molecules. Tailor-made additive N-(E-cinnamoyl)-R-alanine of configuration opposite to that of the host induced, as expected, formation of an (010) face (Fig. 8c) because of an interchange of CH$_3$ and H at the guest asymmetric carbon (C*) so that the C*-CH$_3$ guest moiety replaces the emerging host C*-H bond.
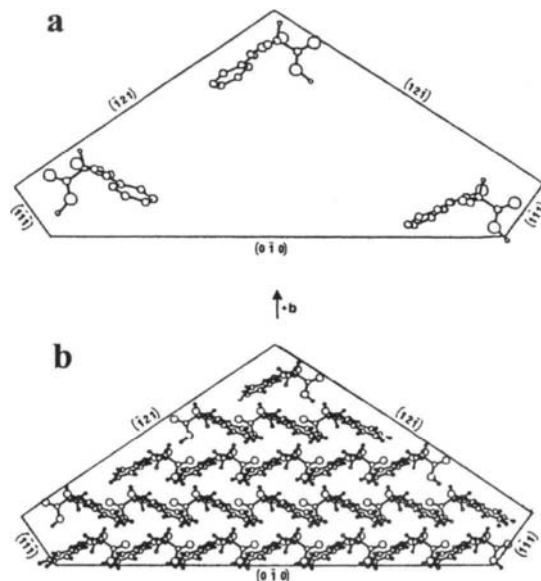
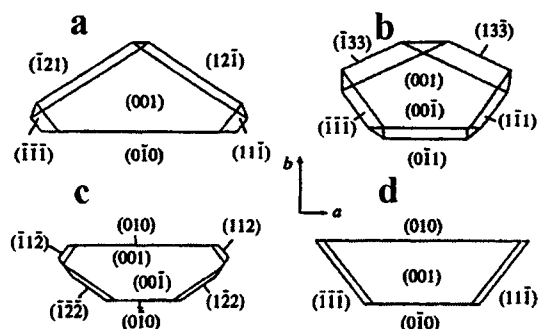Fig. 7. Packing arrangement of N-(E-cinnamoyl-(S)-alanine delineated by the faces of the pure crystal.

Fig. 8. Computer-drawn morphology of measured crystals of N-(E-cinnamoyl)-(S)-alanine.
(a-c) Grown from methanol: (a) pure, (b, c) grown in the presence of the methyl ester and N-(E-cinnamoyl)-(R)-alanine, respectively;
(d) Grown from glacial acetic acid.

Solvents may act in a manner similar to the tailor-made additives as illustrated by growing the crystals of N-(E-cinnamoyl)-S-alanine from acetic acid instead of methanol. Acetic acid is a solvent which can selectively bind at the exposed carboxylic acid groups of the two {1$\bar{1}$1} faces forming a hydrogen bonded dimer **5a**. Acetic acid can also bind to the C*HC=O moiety of N-(E-cinnamoyl)-S-alanine via a cyclic dimer **5b** on the (010) face. The morphology of the crystals grown from glacial acetic acid (Fig. 8d) is in keeping with expectation.

5a   **5b**
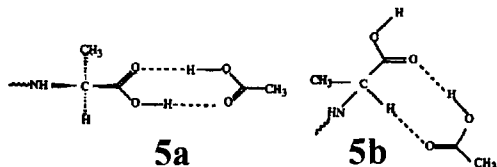
(R)- or (S)-lysine.HCl.2H₂O[7]

The dihydrate crystals of (R)- or (S)-lysine.HCl, ^+H_3N-$(CH_2)_4$-C*H(NH$_3^+$)COO$^-$, crystallized from water, appear in the monoclinic space group $P2_1$. The packing arrangement is shown in Fig. 9 and the morphology of the pure form in Fig 10a. Lysine molecules are aligned parallel to the *b* axis, with the ^+H_3N-C*H-COO$^-$ moiety emerging from the +*b* end of the crystal. The ε-NH$_3^+$ points toward the −*b* and is hydrogen bonded to neighboring molecules in the *ac* plane. Tailor-made α-aminoacid additives with a modified carboxyl or α-amino group, such as lysine methyl ester, inhibited growth in the +*b* direction, inducing development of the (010) face (Fig. 10b). Conversely, additives that bear a modified side

group, such as norleucine or norvaline, inhibited growth along the −*b* direction with a concomitant pronounced increase in the areas of the {1$\bar{1}$0} faces (Fig. 10c).
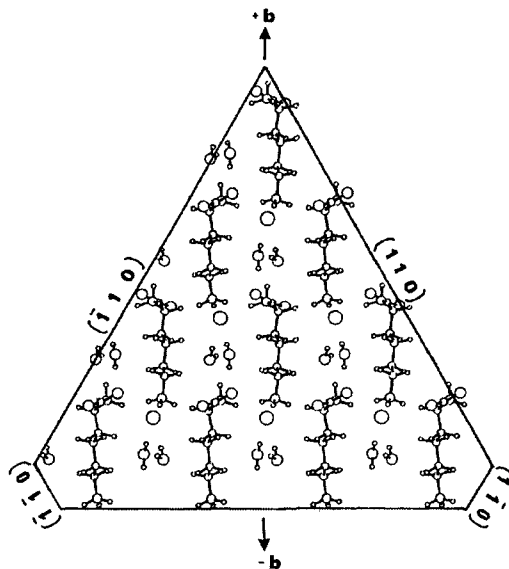
Fig. 9 Packing arrangement of (S)-lys.H₂O viewed along *c* axis.
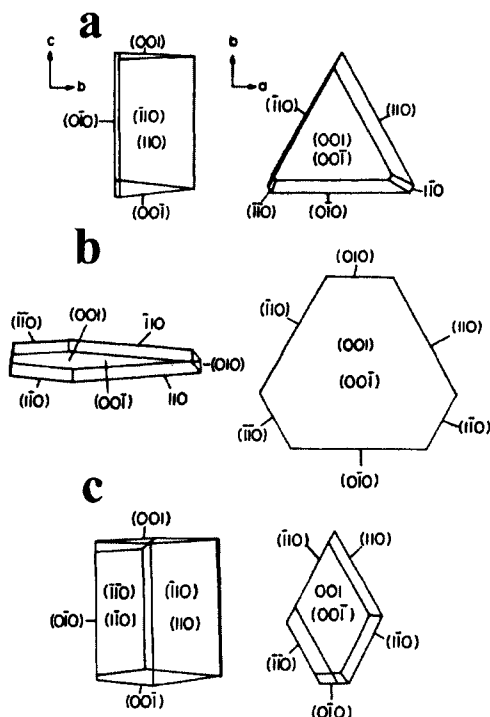
Fig. 10. (a-c) Computer-drawn morphology of measured (S)-lys.H₂O crystals: (a) pure, (b) grown with methyl ester (c) grown with S-norleucine.

The selective adsorption of the tailor-made additive molecules at one end of the polar axis implies that the additive must be occluded only in that part of the crystal that had exposed the adsorbing face to solution during crystal growth (because the additive can be occluded only through adsorbing faces). Therefore, analysis of the occluded additive at the crystal extremities along the polar axis must reveal an anisotropic distribution, as confirmed experimentally by high performance liquid chromatography (HPLC) analysis (Fig. 11) showing the preferential occlusion of the S-norleucine additive at the $-b$ end of the crystal.
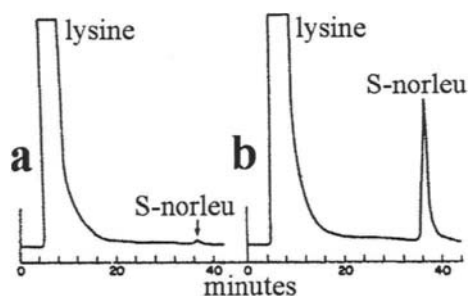
Fig. 11. HPLC analysis of the occluded S-norleucine in crystals as in Fig. 10c: (a) $+b$ side of the crystal; (b) $-b$ side of the crystal.

4.2. Polar crystals composed of racemic molecules[10]

(R,S)-alanine

The absolute orientation of the racemic (R,S)-alanine, $H_3C-C^*H(NH_3^+)COO^-$, polar crystals, of space group $Pna2_1$, could also be assigned from growth experiments in the presence of additives designed to affect either of the two polar ends of the crystal. The packing arrangement delineated by the crystal morphology is shown in Fig. 12a. In the crystal, which is needle-like along the c axis, the R- and S-molecules are oriented with respect to the polar needle c axis so that the carboxylate COO^- groups are exposed at one end of the polar axis and the amino NH_3^+ groups at the opposite end. The crystals exhibit different (hemihedral) faces at opposite ends of the c axis since they display one end face that is flat and the opposite end capped (Fig. 12b). Conventional X-ray crystallography does not allow one to assign the absolute molecular orientation with respect to the polar axis, and so at

which ends are the COO^- and the NH_3^+ groups exposed.
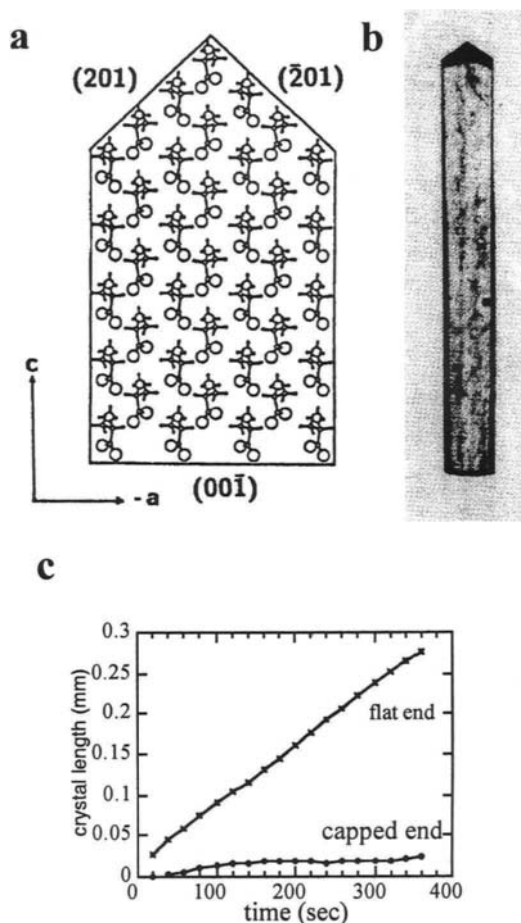
Fig. 12. (a) Packing arrangement of (R,S)-alanine delineated by crystal faces as viewed along the b axis; (b) Photograph of the pure crystal; (c) Graph of the relative growth at the opposite ends of the polar axis.

Monitoring crystal growth of (R,S)-alanine from aqueous solutions led to the conclusion that the needle-like crystals display a fast-growing flat end and a slow-growing capped end (Fig. 12c). Growth experiments in the presence of methyl-alaninate additive yielded needles much shorter than the pure crystals fixing the absolute orientation of the molecules with the COO^- groups exposed at the fast-growing flat end of the crystal. N-methyl-alanine additive, which is expected to affect the slow-growing capped end of the crystal, yielded needles as long as from the pure solution, fixing the orientation of the NH_3^+ in the crystal.

4.3. Probing early stages of nucleation via twinning of (R,S)-alanine crystals[10]

The arrangements of the R- and S-molecules *vis-à-vis* the four symmetry-related {210} side faces of the (R,S)-alanine crystals are shown in Fig. 13 in the two possible orientations of the crystal structure with respect to the polar axis. Naturally, all four symmetry-related molecules are exposed at each of the four symmetry-related {210} faces, but only one of the four molecules is oriented such that its C*-CH3 group emerges from a particular {210} crystal face. Once these orientations in a specimen crystal are determined, the absolute molecular arrangement *vis-a-vis* the polar axis is assigned.
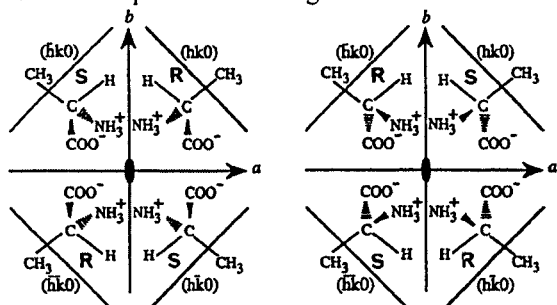
Fig. 13. Schematic (*mm*2 point group) representation of the orientation of the four molecules of (R,S)-alanine vis-à-vis the symmetry related {210} faces, when viewed along the polar *c* axis.

Crystallization in the presence of enantiomerically pure α-aminoacid additives, such as R-threonine, induced a dramatic change in the morphology of the (R,S)-alanine crystals, yielding propeller-shape crystals (Fig. 14) with {$\overline{2}$1-Δl} (Δl = 0.05 to 0.1) affected side faces. Growth in the presence of the opposite enantiomeric S-threonine additive yielded propellers of the mirror image with {21-Δl} affected side faces. The chiral nature of the propeller arises macroscopically from the fact that the affected pair of side faces, say (21-Δl) and ($\overline{2}$1-Δl), and the unaffected pair, (2$\overline{1}$0) and ($\overline{2}$10), have different areas and slopes. The assignment of the affected faces was explained in terms of a twinning model where each end of the crystal grows along the –*c* end (i.e. where COO⁻ groups are exposed) and so the propeller is twinned about the central *ab* plane. This model is in keeping with the observation that the two crystal halves appear to be stitched across this central *ab* plane.
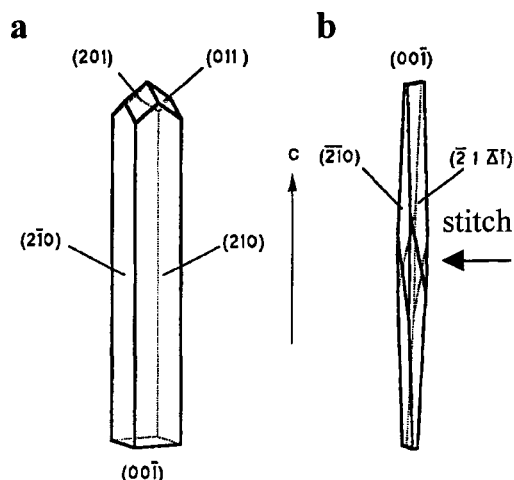
Fig. 14. Computer-drawn morphology of measured crystals of (R,S)-alanine: (a) pure; (b) twin crystal when grown in the presence of 1% R-threonine or R-phenylalanine additive.

However, the two surfaces across such a twinning plane would each expose NH_3^+ groups, which must be energetically unfavorable. One explanation of the observed effects is the inhibition of crystal nucleation of both racemic (R,S)- and chiral, say (R)-, forms of alanine with the same handedness as the additive (R') present in solution, leading to nucleation of the opposite (S)-enantiomorph through absence of steric interaction with the additive. Once formed, the chiral (S)-"nuclei" serve as templates for the epitaxial nucleation and growth of (R,S)-alanine at their two opposite {001} faces yielding the twinned crystals by virtue of the particular similarity between the racemic and chiral crystal forms.

The proposed arrangement at the epitaxial interface is shown in Fig. 15. Following this mechanism, twinning should be avoided when (R,S)-alanine is grown in the presence of racemic additives and chiral seeds should act as templates for the epitaxial growth of (R,S)-alanine, as was indeed experimentally observed. X-ray diffraction measurements of a twinned crystal revealed a misalignment between the two halves and only from the crystal center the presence of reflections that are symmetry forbidden for the space group $Pna2_1$ and so could arise from an enantiomeric alanine "nucleus" of $P2_12_12_1$ symmetry.
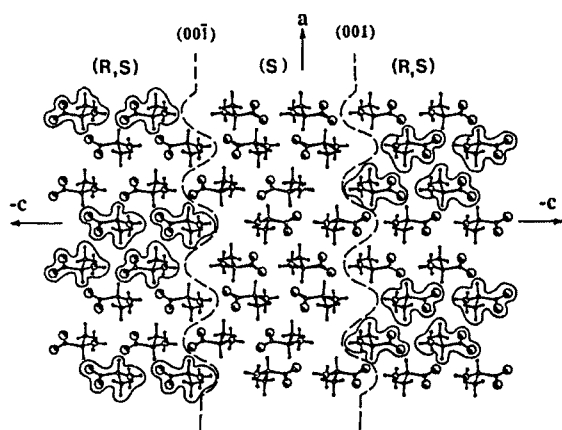
Fig. 15. Proposed molecular arrangement of the "nucleus" composed of a central (S)-"nucleus" onto whose (001) and (00$\bar{1}$) "faces" (R,S)-alanine have been epitaxially grown.

5. EFFECT OF TAILOR-MADE ADDITIVES ON GROWTH OF CENTROSYMMETRIC CRYSTALS

There is a basic difference between the effect of enantiomeric tailor-made additives on polar and centrosymmetric crystals. The morphological changes induced by enantiomeric tailor-made additives on one set of enantiotopic faces of centrosymmetric crystals with appropriate packing features can also be used for the direct assignment of the absolute configuration of the additives. A prerequisite for application of this method is that within the centrosymmetric racemic crystal a specific functional group attached to an R-molecule points toward the face f1 but not toward the opposite face f$\bar{1}$ (Scheme 5).
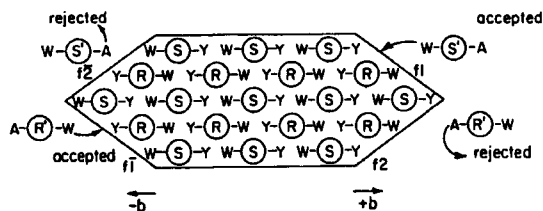
Scheme 5

By symmetry, the same functional group attached to an S molecule emerges at the enantiotopic face f$\bar{1}$, but not f1.

Crystallization of a centrosymmetric crystal in the presence of a R'-additive, designed so that it will fit in the site of a R-molecule on the growing crystal faces f$\bar{1}$ or f$\bar{2}$ but not on the enantiotopic faces f1 or f2, will hinder growth only in the -b direction (Scheme 5). By virtue of symmetry, the enantiomeric S'-additive will inhibit growth perpendicular to faces f1 and f2 only in the +b direction, while racemic R',S'-additives will inhibit growth in both directions, +b and -b. This approach is not limited to racemic crystals, but the constituent molecules should at least be prochiral for ease of interpretation.

The interaction of tailor-made additives with the different crystal faces has also bearing on the general question regarding the distribution and arrangement of such molecules within the affected crystals and on the symmetry and structure of the formed crystalline solid solutions. We illustrate these aspects by examining the effect of a tailor-made additive on various centrosymmetric crystals.

5.1. Assignment of the tailor-made additive molecular chirality[10]

Threonine / (R,S)-serine[7,10]

Racemic (R,S)-serine, HO-CH$_2$-C*H(NH$_3$+)COO$^-$, crystallizes in space group $P2_1/a$. From the packing arrangement (Fig. 16) one can see that the C$_\beta$-H$_{Si}$ bond vector of the rigid methylene group of R-serine molecule has a major component along +b and, by symmetry, the C$_\beta$-H$_{Re}$ bond vector of the S-serine has a major component along -b. Thus, H$_{Re}$ and H$_{Si}$ replacement by a methyl group, as in threonine molecule, will inhibit growth in the b direction. That is, an R-threonine HO-C*H(CH$_3$)-C*H(NH$_3$+)COO$^-$, molecule, with a side-chain β-carbon of chirality S, will inhibit growth along +b and S-threonine will inhibit growth along -b.

As expected, whereas (R,S)-serine forms tabular crystals with point symmetry 2/m (Fig. 17a), the crystals affected by either R- or S-threonine exhibit reduced morphological point symmetry 2 (the mirror plane is lost) and are enantiomorphous (Fig. 17b,c).

When R,S-threonine is used as additive, the morphological point symmetry 2/m is left unchanged because the effects induced by each additive separately combine.
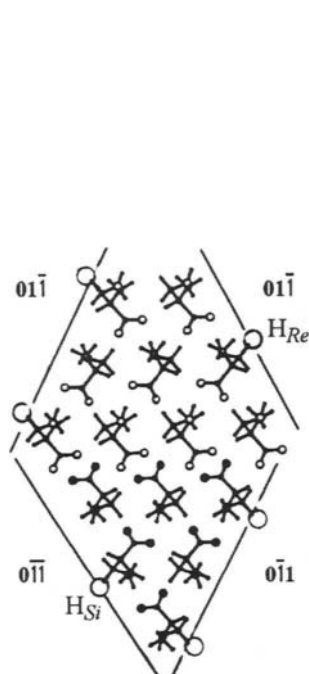
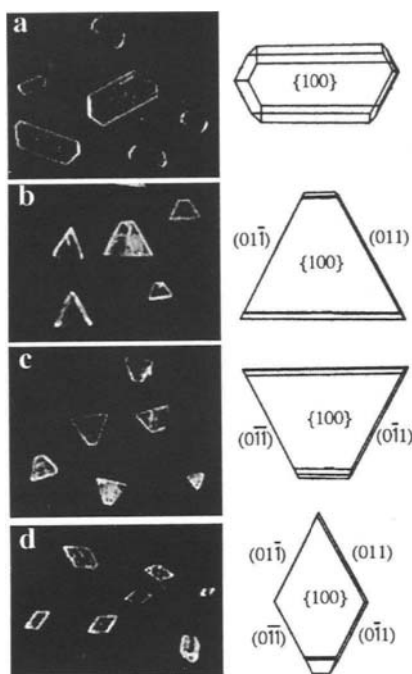Fig. 16. Packing arrangement of (R,S)-serine viewed along the *a* axis.

Fig. 17. **Photographs and computer-drawn morphology of (R,S)-serine crystals: (a) pure. (b-d) grown in the presence of R-, S- and R,S-threonine.**
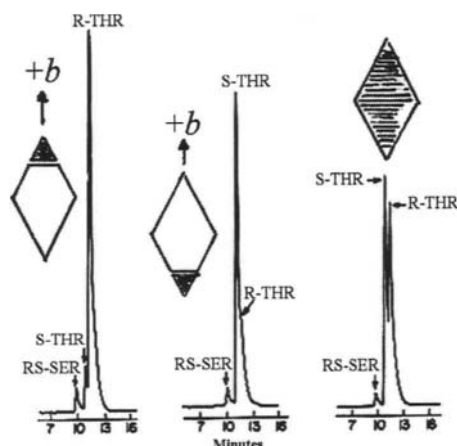
Fig. 18. Enantiomeric analysis of threonine occluded in the (R,S)-serine crystals. Material taken from: tips of the crystal at +b and −b end and the whole crystal, respectively.

Such crystals appear as rhombs (Fig. 17d) with a clear increase in the area of the {011} side faces relative to those of the pure form. The observed changes in crystal morphology, complemented by surface binding energy computations[11], can be interpreted only in terms of enantioselective adsorption of the R-enantiomer at faces (011) and (01$\bar{1}$), and of S-enantiomer at (0$\bar{1}$1) and (0$\bar{1}\bar{1}$). These results fix the absolute chirality of the chiral resolved threonine additive molecules.

The observed changes in crystal morphology and their interpretation in terms of an adsorption-inhibition-occlusion mechanism, imply that during crystal growth of (R,S)-serine in the presence of R,S-threonine, the occluded additive must be distributed anisotropically along the *b* axis. R-threonine, being occluded through the (011) and (01$\bar{1}$) faces, will predominate at the +b half of the crystal, whereas S-threonine will predominate at the −b half, being occluded through the (0$\bar{1}$1) and (0$\bar{1}\bar{1}$) faces. Thus the threonine molecules, racemic in solution, must, on occlusion into (R,S)-serine

crystal, be enantiomerically segregated into domains of opposite chirality along the *b* axis, as was indeed experimentally shown by HPLC enantiomeric analysis (Fig. 18).

5.2. Enantiomeric distribution of selectively occluded tailor-made additives[10]

The interaction of tailor-made additives with the different crystal faces has bearing on the general question regarding the distribution and arrangement of such molecules within the affected crystals and on the symmetry and structure of the formed crystalline solid solutions. We illustrate this aspect by examining the effect of a tailor-made additive on a schematic arrangement, of space symmetry $P2_1/c$, delineated by four symmetry-related, diagonal faces, labeled {011}, similar to (R,S)-serine (Fig. 19a). One can easily see that an additive molecule, bearing an appropriately modified moiety, can be recognized and substitute a substrate molecule at any diagonal face, say (011), but only at one of the four different surface sites where the modified group

does not disturb the regular pattern of interactions at the crystal surface.
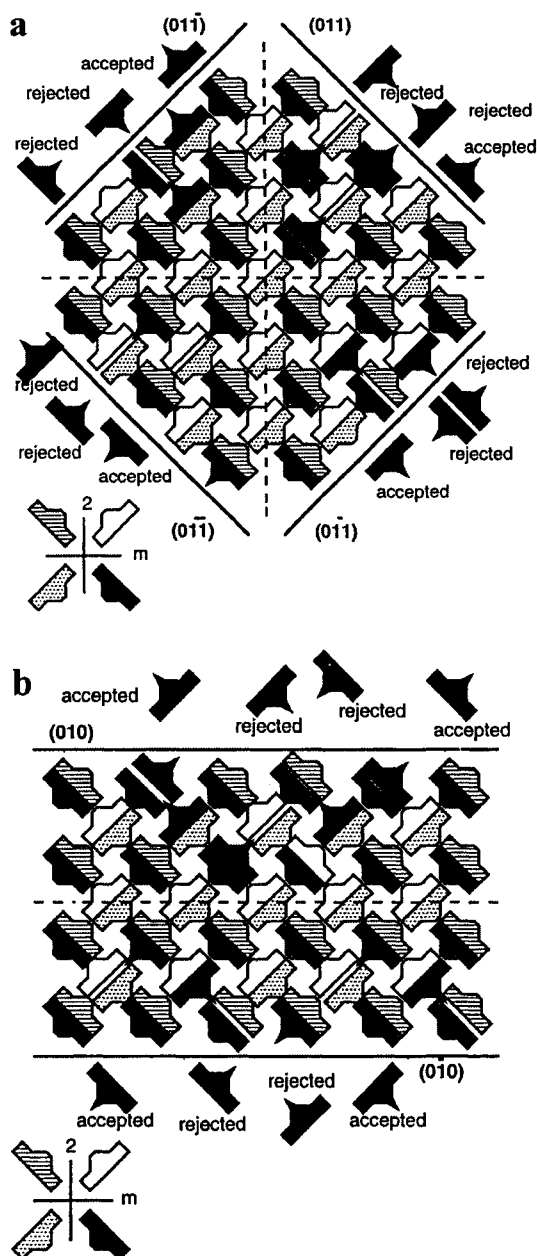

Fig. 19. Schematic representations of adsorption and occlusion of tailor-made additive molecules (in black) through: (a) the slanted {011} faces and (b) the top and bottom {010} faces in a centrosymmetric crystal of space group $P2_1/c$, containing four symmetry-related molecules (point symmetry $2/m$).

Occlusion of such a molecule through the (011) face will lead to a crystal sector in which the occluded molecules therein are related by translation symmetry only. Occlusion of additive through all four diagonal faces {011} (Fig. 19a) will lead to a crystal composed of four sectors. Consequently, in general, an additive molecule will be anisotropically distributed within the grown crystal, preferentially occluded through different surface sites on the various faces, leading to a mixed crystal composed of sectors coherently intertwined, although the occluded additive may occupy only a small fraction of the bulk sites. The reduced symmetry of each sector was predicted in several host-additive systems and experimentally demonstrated directly, by X-ray and neutron diffraction, and by a variety of methods including changes in crystal morphology, HPLC, nonlinear optical properties and solid state asymmetric photodimerization.

Symmetry lowering of somewhat similar nature has been observed in a centrosymmetric crystal when the additive molecules are occluded through the {010} crystal faces leading to a reduction in symmetry from $P2_1/c$ to $P2_1$ (Fig. 19b).

Some examples are given to illustrate these arguments and principles.

α-Aminoacids / α-form of glycine[7,10]

Glycine, $^+H_3NCH_2COO^-$, is a prochiral molecule containing two enantiotopic hydrogen atoms at the central carbon atom; namely, replacing one of these hydrogens by a different group yields a chiral molecule. Glycine crystals display polymorphism. The α-form of glycine, that appears when grown from aqueous solutions, has space group $P2_1/n$ and contains four symmetry related molecules, labeled *1*, *2*, *3*, and *4* in Fig. 20. Molecules *1* and *2*, related by twofold screw symmetry, have their C-H_{re} bonds pointing in the +*b* direction and emerging from the (010) face. By symmetry, molecules *3* and *4*, related to *1* and *2* by a center of inversion, have their C-H_{si} bonds pointing toward -*b* and emerging from the (0$\bar{1}$0) face.

Only the R-α-aminoacid additives can substitute a glycine molecule at the *1* and *2* sites, and then only on face (010), while only S-α-aminoacids can be adsorbed at sites *3* and *4* on face (0$\bar{1}$0). This constraint arises from the steric requirement that the additive molecule be recognized on the {010} surfaces as a "host" molecule so that the

α–aminoacid side chain can replace the H atom of glycine only at the C-H bond emerging from the crystal surface. Otherwise, repulsion would occur between the additive and the surrounding molecules on the crystal surface.
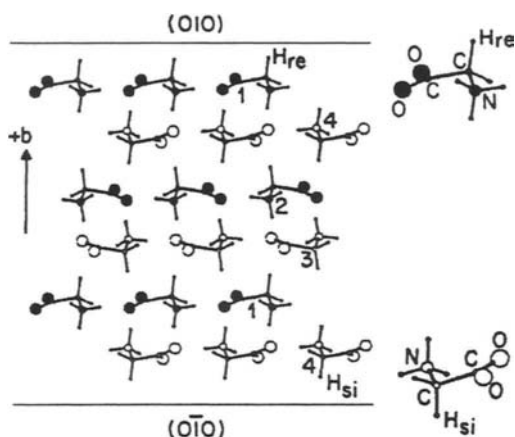
enantiomorphous morphology (Fig. 21c). Racemic α-aminoacid additives cause the formation of {010} plates (Fig. 21d) because growth at both the +*b* and -*b* sides of the crystal is inhibited. As expected, the *R*-enantiomers populate exclusively the +*b*, and the *S*-enantiomers the -*b* half of the crystal, according to HPLC enantiomeric analysis (Fig. 21).

Fig. 20. Packing arrangement of the α–form of glycine viewed along the *a* axis.

α-Glycine crystallizes from water as bipyramids (Fig. 21a). Crystallization experiments in the presence of R-α-aminoacid additives leads to the formation of pyramids with a (010) basal plane, because growth in the +*b* direction is inhibited (Fig. 21b). S-α-aminoacid additives induce the
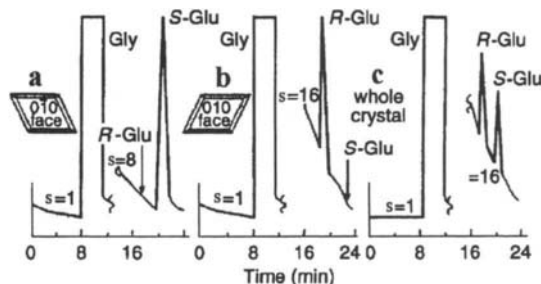
Fig. 22. HPLC enantiomeric distribution of R,S-glutamic acid occluded in {010} platelike crystals of α-glycine: (a, b) material shaved from the (0$\bar{1}$0) and (010) faces; (c) whole crystal.
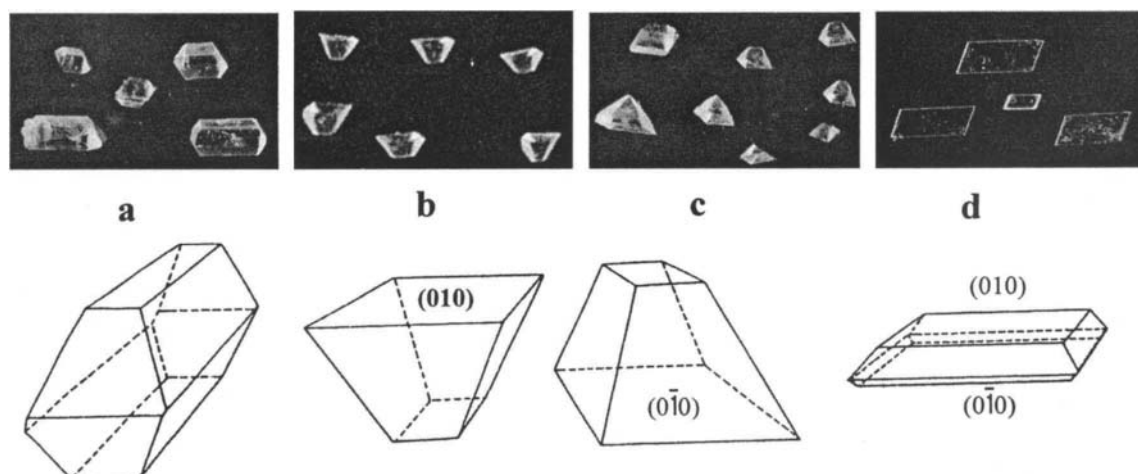

Fig. 21. Photographs and computer-drawn morphology of the α-glycine crystals: (a) pure, (b-d) grown in the presence of R-, S-, R,S-additives.

Glycyl-leucine / glycyl-glycine[7]

The centrosymmetric single crystal of glycyl-glycine $+H_3NCH_2-CONH-CH_2COO^-$ (gly-gly) was found suitable for probing the conformation and mode of occlusion of several chiral dipeptides by changes in crystal morphology, distribution of occluded additive and conformational analysis. Crystals of pure gly-gly, of space group $P2_1/c$, (Fig. 23) are diamond-shaped {100} plates (Fig. 24a). The enantioselective changes in the morphology of the gly-gly crystals grown in the presence of tailor-made additives proved that gly-R-leu induces the development of the $(12\bar{1})$ and $(\bar{1}21)$ faces whereas the gly-S-leu enantiomer adsorbs and consequently occluds through the enantiotopic $(1\bar{2}\bar{1})$ and $(\bar{1}2\bar{1})$ faces (Fig. 24b, c).
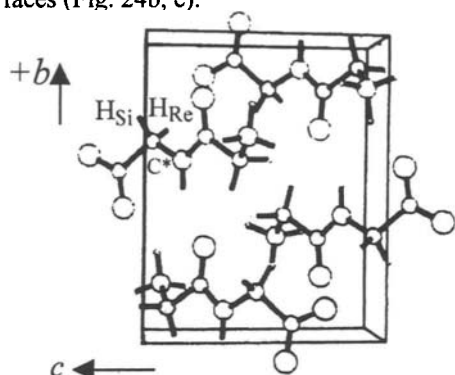
Fig. 23. Packing arrangement of the gly-gly centrosymmetric crystal viewed along the *a* axis.

When the racemic additive was used, all four faces were expressed (Fig. 24d). As expected, the occluded additive is segregated into enantiomers along the *b* axis of the host crystal, as determined by the HPLC enantiomeric analysis (Fig. 25).

Different morphologies were obtained when gly-gly crystals were grown in the presence of gly-R,S-ala and ala-R,S-gly additives, (Fig. 26), proving the effect of preferred conformation of the additive molecule during crystal growth. Atom-atom potential energy calculations[11] involving the preferred conformations of the guest molecules and their enantiospecific binding at the affected crystalline faces were in agreement with the experiments.

Fig. 25. HPLC enantiomeric distribution of R,S-gly-leu inside the crystals of gly-gly.
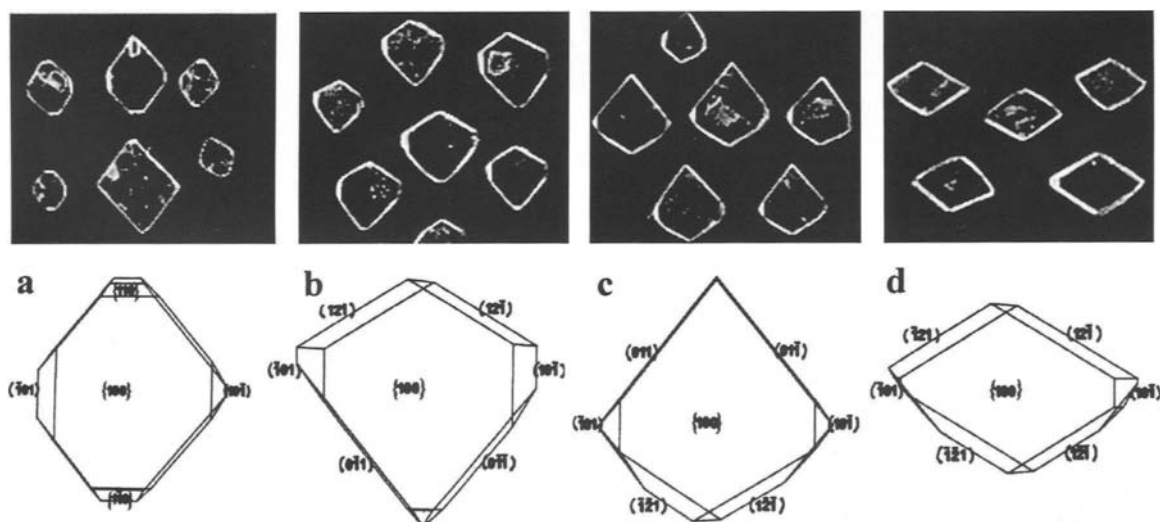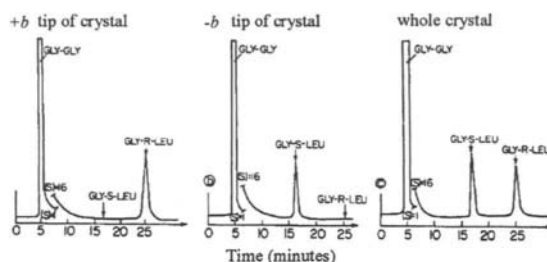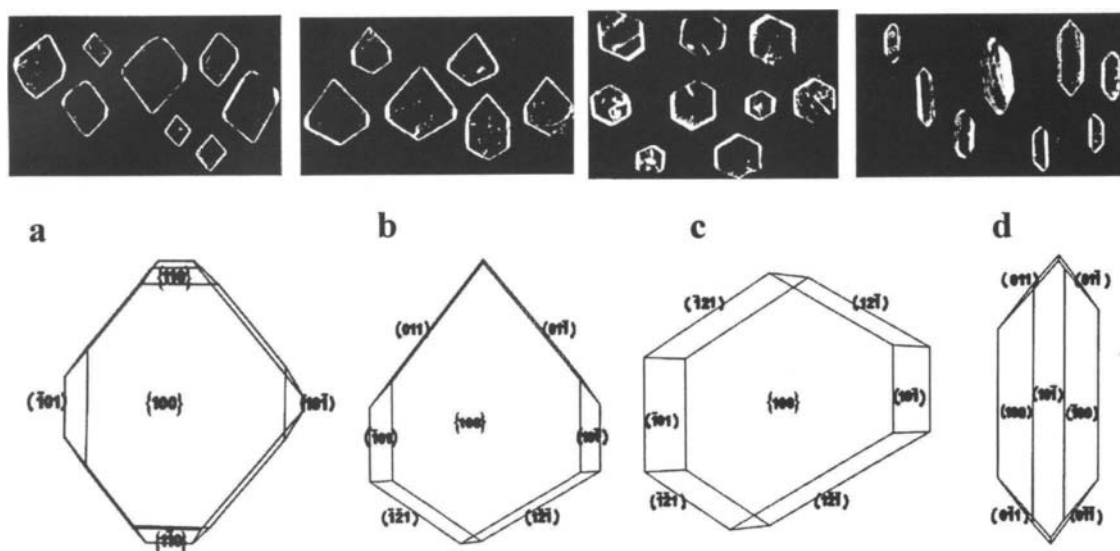

Fig. 24. Photographs and computer-drawn morphologies of the gly-gly crystals: (a) pure (b-d) grown in the presence of R-, S- and R,S- glycyl-leucine, $^+H_3NCH_2-CONH-C^*H(C_4H_9)COO^-$, tailor-made additive.

Fig. 26. Photographs and computer-drawn morphologies of the gly-gly crystals: (a) pure (b-d) grown in the presence of S-, R,S-glycyl-alanine, $^+$H$_3$NCH$_2$-CONH-C*H(CH$_3$)COO$^-$, and R,S-alanyl-glycine, $^+$H$_3$NC*H(CH$_3$)-CONH-CH$_2$COO$^-$, tailor made additives.

5.3. Probing intermolecular forces through reduced crystal symmetry[10]

We focus now on the minimal modification of a tailor-made additive with respect to the host molecule that would still be recognized and discriminated by the growing crystal surface and lead to reduction in symmetry. Indeed, solid solutions composed of host and additive molecules of similar structure and shape have been generally believed to exhibit the same symmetries as those of the host crystals. We consider solid solutions of carboxylic acids (X-COOH) in primary amides (X-CONH$_2$), where a NH$_2$ group of the host is substituted by an OH moiety of the guest (Fig. 1a). A strong inhibition of growth develops along the direction of the O=C-N-H···O=C hydrogen bond. Inhibition arises when an adsorbed carboxylic acid additive molecule causes repulsion between its oxygen lone-pair electrons and those of the oncoming host-amide molecule. The carboxylic acid additive molecule will thus avoid surface sites that require its OH groups to be oriented toward the surface and be preferentially adsorbed at sites where the OH group emerges from the surface. This principle is illustrated here with several examples.

Aspartic acid / (S)-Asparagine

Asparagine H$_2$NOCCH$_2$C*H(NH$_3$$^+$)COO$^-$ crystallizes from water as a monohydrate with a tight three-dimensional net of hydrogen bonds in the $P2_12_12_1$ space group (Fig. 27a). The morphology is prismatic, with 18 developed faces (Fig. 27b). Crystallization of (S)-asparagine.H$_2$O in the presence of S-aspartic acid HOOCCH$_2$-CH(NH$_3$$^+$)COO$^-$ yields {010} plates (Fig. 27c). In terms of the arguments already given for amide-acid systems, the additive aspartic acid molecule should be more easily adsorbed at sites 1 and 3, on the growing (010) surface, than at sites 2 and 4 (Fig. 27d); the reverse situation holds for the opposite (0$\bar{1}$0) face. If, on growth of the mixed crystal, the (0$\bar{1}$0) face is blocked so that the amide and acid molecules would be occluded only through the (010) face, the symmetry of the mixed crystal should be reduced to $P12_11$. This effect was demonstrated by a low temperature (18 K) neutron diffraction study of a protonated asparagine crystal grown with deuterated S-aspartic acid additive.
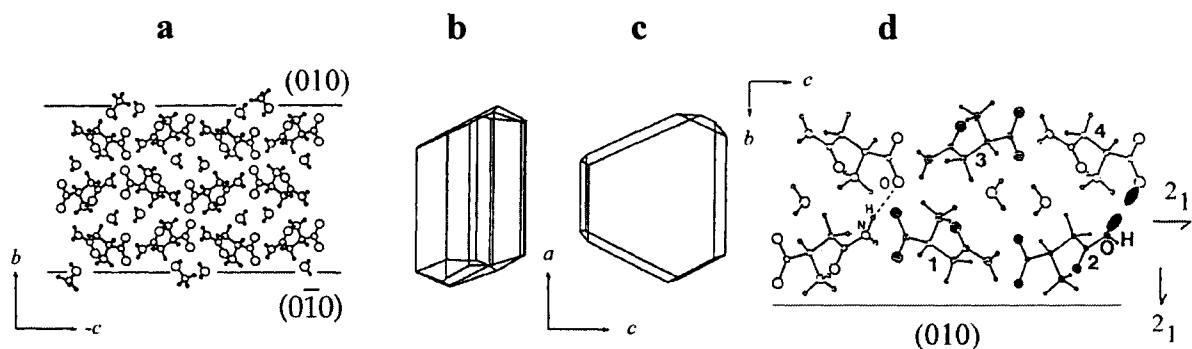
Fig. 27. (a) Packing arrangement of (S)-asparagine.H_2O viewed along the a axis. (b, c) Computer-drawn morphology of measured crystals: (b) pure (c) grown in the presence of S-aspartic acid. (d) Preferential adsorption of aspartic acid on the (010) surface of the crystal.

E-cinnamic acid / E-cinnamamide; asymmetric synthesis in centrosymmetric host crystals[10,12]

The site selective adsorption and occlusion of E-cinnamic acid, C_6H_5-CH=CH-COOH, additive into the crystal of E-cinnamamide C_6H_5-CH=CH-$CONH_2$ is based on the same principle of repulsion between the lone-pair electrons as discussed above. The pure crystal (Fig. 28b) of E-cinnamamide appears in a centrosymmetric monoclinic arrangement, space group $P2_1/c$. The crystal structure (Fig. 28a) is composed of hydrogen-bonded dimers interlinked by N-H···O bonds to form a ribbon-like motif that runs parallel to the b axis. During crystallization in the presence of E-cinnamic acid, the additive is preferentially

occluded through half of the {011} surface sites of the crystal at the opposite ends of the ribbons (Fig. 28a). This effect leads to a modified mixed crystal (Fig. 28c) with two enantiomorphous halves of at most $P2_1$ symmetry. This reduced symmetry was proven photochemically. Ultraviolet irradiation of E-cinnamamide yields centrosymmetric photodimers, by virtue of a cyclobutane ring formation involving pairs of close packed >C=C< bonds across centers of inversion (Fig. 29). Replacement of one of such a pair by E-cinnamic acid results in the formation of asymmetric cinnamamide-cinnamic acid photodimers of opposite chirality at the two enantiomorphous halves of the mixed crystal, with an enantiomeric ratio of 60:40 at each opposite half (Fig. 30).
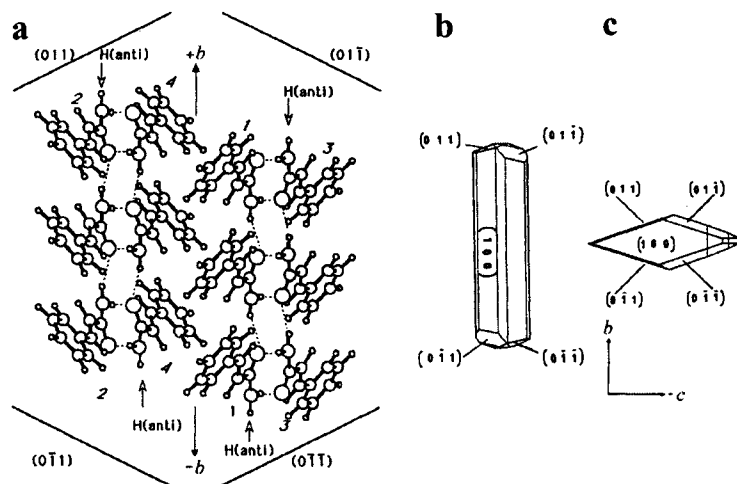

Fig. 28. (a) Packing arrangement of E-cinnamamide viewed along the a axis. (b, c) Crystal morphology: (b) pure; (c) grown in the presence of E-cinnamic acid.
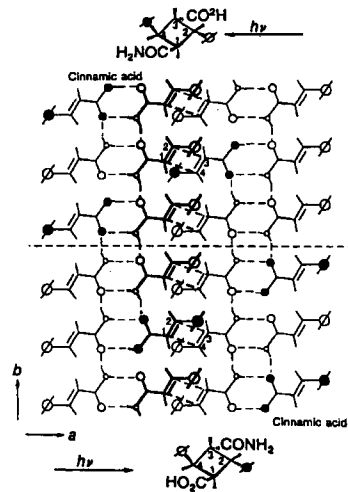
Fig. 29. Packing arrangement of E-cinna-mamide viewed along the c axis showing the photodimerization reaction.
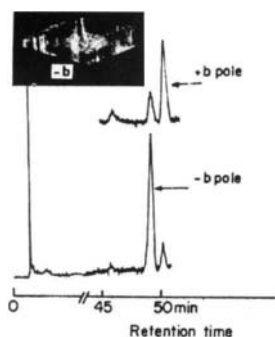
Fig. 30. GC enantiomeric analysis of photodimers isolated from the +b and -b poles of a crystal of E-cinnamamide:E-cinnamic acid.
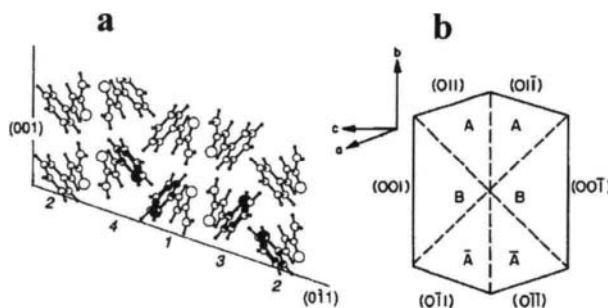
Fig. 31. (a) Packing arrangement of E-cinnamamide showing the β-thienyl rings of the additive that replace the phenyl rings of the host molecules. (b) morphology of the mixed crystal. (c) GC enantiomeric analysis of the photodimers generated in the A, Ā and B sectors of an E-cinnamamide: E-2-(β-thienyl)acrylamide mixed crystal.

2-E-thienylacrylamide / E-cinnamamide[12,13]

When crystals of E-cinnamamide were grown in the presence of E-2-(β-thienyl)acrylamide additive, the four {011} faces were more developed implying that the thienyl moieties were occluded primarily through these faces at which the host molecules display herring-bone contacts between aromatic C-H groups and π electron clouds of neighboring phenyl rings related by two-fold screw symmetry. These contacts were replaced in the host-additive mixed crystal by unfavorable contacts between sulfur lone-pair electrons and the π electron system. The additive can easily be adsorbed through each of the four {011} faces at only one of the four sites (Fig. 31a) at which the sulfur emerges from the face, e.g. at site 1 on the (01̄1) face, reducing crystal symmetry to $P1$. In addition, E-2-(β-thienyl)acrylamide is occluded also through the {001} faces which exhibit pm plane symmetry and such occlusion should reduce the crystal symmetry to Pc. This mixed crystal should be thus divided into six sectors of reduced symmetry (Fig. 31b). The sectors A and Ā are related to one another by the $2/m$ point symmetry of the host crystal. Low temperature X-ray and neutron diffraction studies on A, Ā and B type sectors, cut from a crystal specimen proved the reduction in crystal symmetry. The chiral photodimers generated from sectors A and Ā are of opposite handedness, as determined by gas chromatography enantiomeric analysis (Fig. 31c), circular dichroism and NMR measurements.

6. TAILOR MADE ADDITIVES AS ENANTIOSELECTIVE NUCLEATION PROMOTERS[13]

Tailor-made additives may act as inhibitors of crystal nucleration and growth as well as promoters of crystal nucleation. The latter effect is illustrated here for the system of α-glycine / α-aminoacid additives. As described above, α–glycine pure bipyramids become pyramids with either a (010) or (01̄0) basal plane when grown in the presence of R- or S-α-aminoacid additives, respectively (Fig. 21). Racemic additives cause the formation of {010} plates that contain 0.02-0.2% racemic additive molecules occluded into the crystal bulk, with the two enantiomers segregated in the two sectors at the +b and –b crystal halves (Fig. 22).

Systematic studies demonstrated that the presence of the hydrophobic α-aminoacids, such as valine, leucine, norleucine, phenylalanine, as co-solutes to aqueous solutions of glycine, induce crystallization of the glycine crystals floating at the air-solution interface and oriented with one of the {010} faces exposed towards the air. In the presence of hydrophobic R- or S-additives (1%wt/wt of

glycine) all the pyramidal crystals float with their either (010) or (0$\bar{1}$0) basal plane oriented towards the air, respectively. Crystallization in the presence of hydrophobic R,S-additives induces the formation of an equal number of plate-like crystals oriented with either (010) or (0$\bar{1}$0) exposed to air (Fig. 32a). HPLC analysis of the additives occluded into such floating single crystals demonstrated that the (010)-oriented plates contain only S-additives whereas the (0$\bar{1}$0)-oriented plates contain only the R-additives (Fig. 32b), in agreement with the occlusion mechanism through the enantiotopic face exposed to the solution. These results were explained in terms of the formation, at the air-solution interface, of ordered two-dimensional clusters of the hydrophobic, say, R-α-aminoacid additive molecules oriented with their side chains exposed to air and their α-aminoacid moiety exposed to solution. Glycine solute molecules can form hydrogen-bonded centrosymmetric dimers with such exposed moieties and mimic the 010 layer of the

glycine crystal that thus can serve as a template for the nucleation of (010)-oriented glycine crystals. By symmetry, hydrophobic S-additives should mimic the 0$\bar{1}$0 layer of glycine crystals and induce nucleation of floating (0$\bar{1}$0)-oriented glycine crystals, in keeping with the experimental results.

Based on the above mechanism, the induced resolution of α-aminoacids via occlusion inside glycine crystals was performed by oriented crystallization of glycine in the presence of, say, S-leucine and mixtures of hydrophilic R,S-α-aminoacids such as p-hydroxy-phenylglycine, glutamic acid and methionine. All the floating glycine plates were formed with their (0$\bar{1}$0) faces oriented towards the air (Fig. 32c) and occluded exclusively the R-additives through the (010) face exposed towards solution, according to HPLC enantiomeric analysis (Fig. 32d). When polycrystalline crusts of glycine were formed, the enantiomeric excess of the occluded additives was higher than 70%.
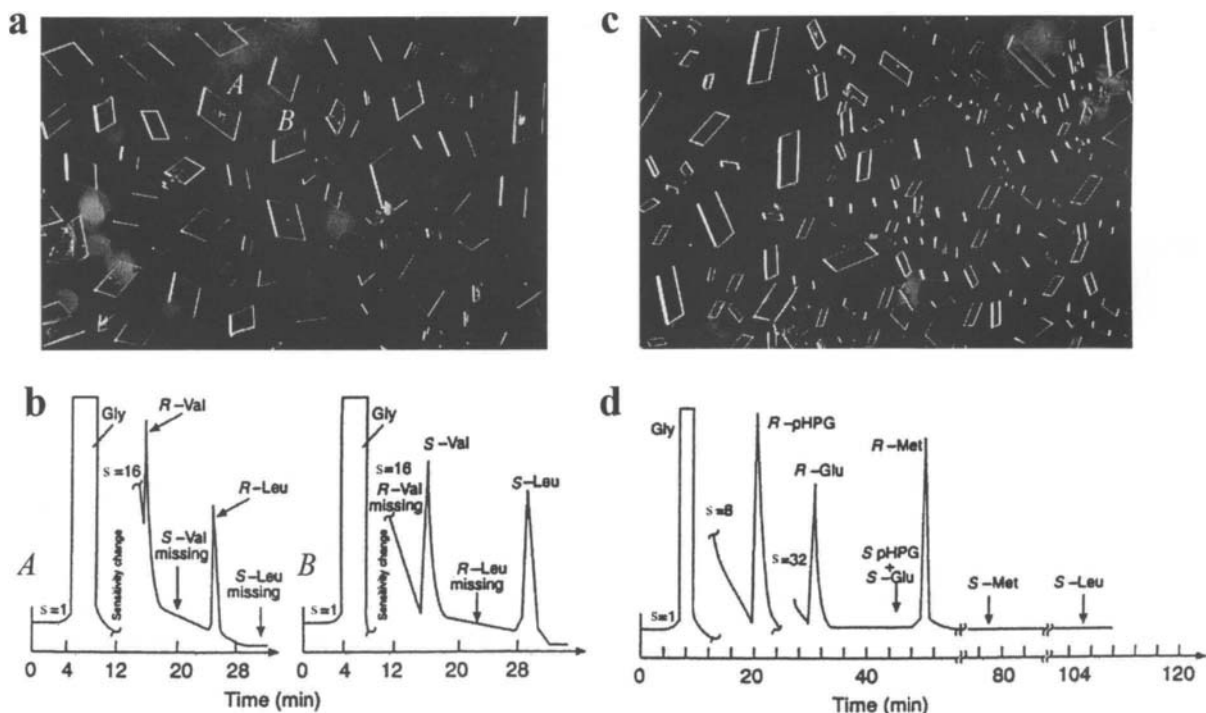

Fig. 32. (a, b) Plate-like crystals of α-glycine floating at the solution surface as grown in the presence of R,S-leucine and R,S-valine and the corresponding HPLC enantiomeric analysis of occluded additives in single crystals of type A and B; (c, d) Plate-like crystals of α-glycine floating at the solution surface as grown in the presence of S-leucine and racemic p-hydroxyphenylglycine, glutamic acid and methionine and the corresponding HPLC enantiomeric analysis of occluded additives inside the crystals.

1. CONTROL OF CRYSTALLINE PHASE FORMATION[14]

We now consider preferred precipitation of a crystal containing a polar axis at the expense of the dimorph which is centrosymmetric and so nonpolar. As described above, in crystals with a polar axis, all molecules are parallel vis-a-vis the polar axis (Scheme 4) whereas in centrosymmetric crystals neighboring molecules are arranged in antiparallel manner (Scheme 5). An appropriate tailor-made additive will inhibit growth of the centrosymmetric form at the opposite ends of the crystal and so prevent its appearance, but will inhibit growth of the polar form only at one end of the crystal. This concept has been demonstrated for the system of N-(2-acetamido-4-nitrophenyl)-pyrrolidine (PAN) **6a**, whose metastable polar form is important for its nonlinear optical properties. As little as 0.03% of the polymer **6b** inhibited precipitation of the pseudo-centrosymmetric stable form of PAN allowing crystallization of the metastable form only.

6a   **6b**
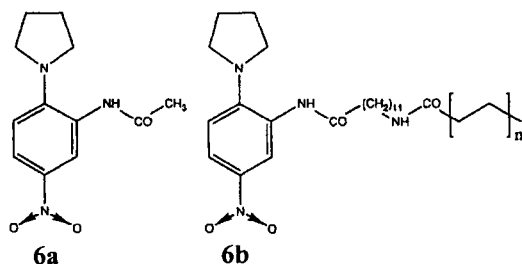
γ-glycine polymorph

As already mentioned, glycine usually crystallizes from aqueous solutions in the centrosymmetric α-form (Fig. 33a) but, when grown from acetic acid or ammonia solutions, glycine can precipitate as a mixture with the polar γ-form (Fig. 33d). Preferential crystallization from aqueous solution of the γ--form of glycine was achieved in the presence of racemic hexafluorovaline $(CF_3)_2CHC^*H(NH_3^+)COO^-$ and chloranil as tailor-made additive (Fig. 33e, f).

The crystal of γ-glycine appears in space group $P3_1$ (or $P3_2$), arranged in a way similar to (R,S)-alanine crystal, so that the molecules expose COO^- groups at one end of the polar trigonal axis and NH_3^+ groups at the opposite end (Fig. 34a).

As previously described, common α-aminoacid additives bind only at the {010} faces of α-glycine

crystal and so inhibit its growth only along b, inducing formation of the {010} plates (Fig. 33b). In contrast, racemic hexafluorovaline is a more effective inhibitor of growth of α-glycine since it binds at the four {011} faces (Fig. 34b) and so blocks growth along both b and the fast-growing c directions (Fig. 33c). The abnormal behaviour of hexaflurovaline can be understood in terms of its molecular structure and conformation, as found in its native crystal. Steric repulsion imposed by the hexafluro-isopropyl moiety prevents adsorption onto the {010} face, but not at the {011} faces to which it is strongly bound through charge-transfer (CT) interactions. Furthermore, hexafluorovaline does not inhibit nucleation and growth of the γ-glycine form because the additive is bound thereto primarily at the NH_3+ end of its polar axis and the crystal is growing unidirectionally along its polar c axis at the COO-end.

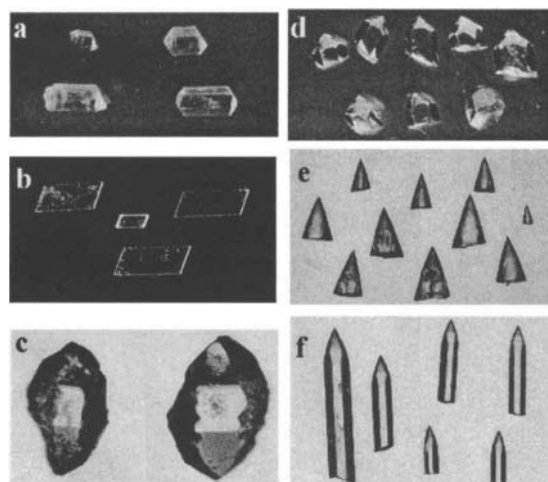Fig. 33. (a-c) Crystals of α-glycine grown from water: (a) pure, (b) in the presence of common R,S-α-aminoacids and (c) R,S hexaflurovaline. (d) Crystals of γ-glycine grown from aqueous acetic acid solutions. (e, f) Crystals of γ-glycine grown from water in the presence of: (e) R,S-hexafluorovaline, (f) chloranil (tetrachlorobenzo-quinone).

However, the hexafluorovaline molecules can be also adsorbed on the newly developed {¡51} side faces of the trigonal pyramidal crystals (Fig. 33e). The strong inhibition of α-glycine via charge-transfer interactions was proved by using as additive
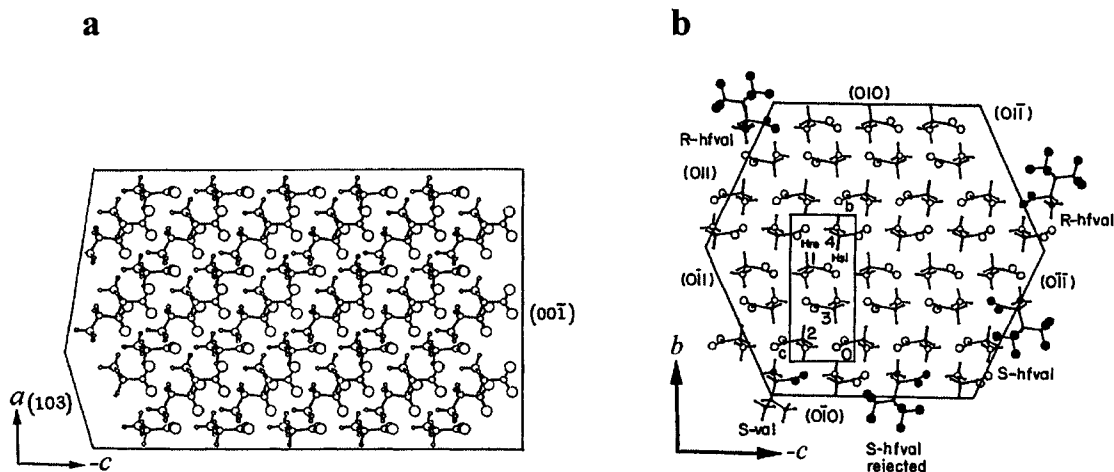
400

Fig. 34. (a) Packing arrangement of the γ-form of glycine viewed along the *b* axis; (b) Packing arrangement of the α-form of glycine viewed along the *a* axis and showing the effect of R,S-hexafluorovaline additive.

chloranil, a molecule known to form a CT complex with glycine as an electron donor. This CT complex was anticipated to bind at the various faces of α-glycine clusters that expose NH_3^+ groups such as {011}, {110} and {010}, and so inhibit nucleation of this phase. The CT complex can also bind at the slow-growing capped end of γ-glycine but not at the fast-growing end (fig. 33f). The conclusion of these experiments is that the inhibitor, to be effective, should bind at the fast-growing directions.

8. CONCLUSIONS

A variety of crystal properties can be controlled and modified with the use of tailor-made additives, including growth and dissolution, morphology, crystal texture and polymorphism. Such studies allows one to help understand basic scientific questions such as pinpointing fine molecular interactions at interfaces and the correlation between molecular chirality and macroscopic phenomena in crystals. Kinetic resolution of racemates and the control in the very early stages of crystal formation to yield the polymorph with desired properties have large applications in the pharmaceutical industry in particular and in applied sciences in general.

REFERENCES

1. P. Hartman, W.G. Perdok, Acta Crystallogr. 8, (1955), 49; Ibid, (1955), 525.
2. P. Hartman, in: P. Hartman (ed.), Crystal Growth An Introduction, North-Holland, Amsterdam, p. 367, 1973.
3. P. Bennema, G. Gilmer, in: P. Hartman (ed.), Crystal Growth, An Introduction, North-Holland, Amsterdam, p. 272, 1973.
4. P. Hartman, P. Benema, J. Crystal Growth 49, (1980), 145.
5. P. Bennema, J. Crystal Growth 122, (1992), 110.
6. L. Addadi, Z. Berkovitch-Yellin, I. Weissbuch, M. Lahav, L. Leiserowitz, in: E.L. Eliel, S.H. Willen and N.L. Allinger (eds.), Topics in Stereochemistry; Wiley& Sons Ltd., New York, 16, 1-85, 1986.
7. L. Addadi, Z. Berkovitch-Yellin, I. Weissbuch, J. van Mil, L.J.W. Shimon, M. Lahav, L. Leiserowitz, Angew. Chem., Int. Ed. Engl., 24, (1985), 466.
8. I. Weissbuch, L. Leiserowitz, M. Lahav, in: A. Mersmann (ed.), Handbook of Crystallization, M. Decker, New York, p. 401, 199
9. D. Zbaida, M. Lahav, K. Drauz, G. Knaup, M. Kottenhahn, Tetrahedron, 56, (2000), 6645.
10. I. Weissbuch, R. Popovitz-Biro, M. Lahav, L. Leiserowitz, Acta Crystallogr. A, LEAD ARTICLE B51, (1995), 115.
11. Z. Berkovitch-Yellin, J. Am. Chem. Soc., 107, (1985), 3375.
12. M. Vaida, L.J.W Shimon, Y. Weisinger-Lewin, F. Frolow, M. Lahav, L. Leiserowitz, R.K. McMullan, Science, 241, (1988), 1475.
13. I. Weissbuch, R. Popovitz-Biro, L. Leiserowitz, M. Lahav, in: J.P.Behr (ed.), 100 Years of the The-Lock-and-Key Priciple, Wiley&Sons Ltd., New York, Ch. 6, p. 173, 1994.
14. I. Weissbuch, L. Leiserowitz, M. Lahav, Advanced Materials, 6, (1994), 952.

Crystal engineering of biological soft materials

Kiyotaka Sato

Faculty of Applied Biological Science, Hiroshima University,
Higashi-Hiroshima, 739-8528, Japan

Crystallization processes of biological soft materials are discussed on the basis of results of experiments using amino acids and lipids. Particular interest is paid to the effects of templates on control of the rate of nucleation, crystal structure (polymorphism) and crystal morphology through specific crystal-additive interactions. Solvent crystallization of long-chain lipid molecules was found to be better controlled, in terms of heterogeneous nucleation rate, polymorphism and molecular orientation through specific molecular interactions occurring at the template-solution interfaces, by template films prepared by physical vapor deposition. As for the crystallization in emulsion and microemulsion systems, it was shown that heterogeneous nucleation of lipids and amino acid was initiated at the emulsion interface through van der Waals interactions. As a result, the nucleation rate and crystal morphology were largely modified from those occurring in bulk crystallization systems.

1. INTRODUCTION

Biological soft materials are categorized as organic crystals that are present in biological tissues and are synthesized for applications in various industries such as pharmaceutical, nutraceutical, food and cosmetics industries. Bones, teeth and other biological minerals, as counterparts, may be categorized as biological hard materials. Typical biological soft materials include amino acids, n-alkanes (paraffins), lipids, peptides, carbohydrates, vitamins, pharmaceuticals (Figure 1). For example, triacylglycerols, fatty acid esters of glycerine, are used in food, cosmetics and pharmaceuticals [1]. Amino acids are the main constituents of protein molecules and used as food ingredients and as nutritional and pharmcological additives [2]. Biological soft materials can be considered as key target substances to be focused on in life science and technology, together with proteins which are biological macromolecules (not discussed here).
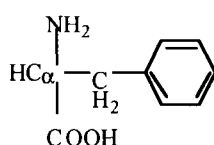
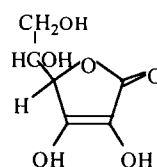Much interest has been shown in crystal engineering of biological soft materials from
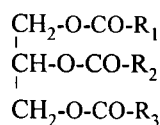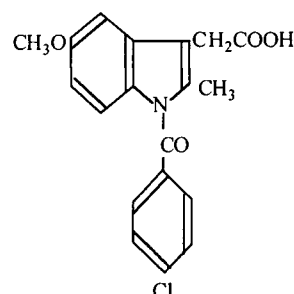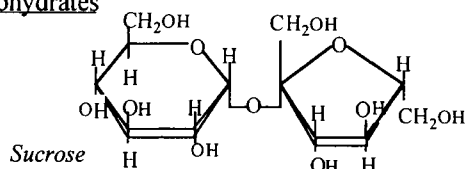
Figure 1. Typical biological soft materials.

both practical and scientific view points. The word *SOFT* was chosen not simply because those materials are mechanically soft and fragile compared to materials such as bones and teeth but more importantly because biological soft materials exhibit diversified polymorphism (crystal structure) and morphology (crystal shape) in various crystallization systems; e.g., bulk and emulsions (see below). This property may be characterized from both as a soft system.

In this chapter, the key concepts underlying crystal engineering processes of biological soft materials are discussed, and recent works on heterogeneous nucleation processes of amino acids and lipids examined in bulk and emulsion states are reviewed.

2. KEY CONCEPTS IN CRYSTAL ENGINEERING

2.1 Crystals and functionality

In actual applications, biological soft materials are often synthesized in a crystalline state.

It should be noted that the major properties in the crystals of such materials are different from, for example, semiconductor and optical materials that are grown in large bulky single crystals as starting materials for advanced fabrications. Biological soft materials, however, are eventually crystallized in polycrystalline particles, in which the crystal structure (polymorphism) and crystal shape may be two major properties to be focused on. This is because these two properties are closely related to the behavior of melting and solidification, dissolution, aggregation and dispersion, crystal network formation, suspension of foreign substances, stabi-lity of crystal-containing systems, which are funct-ional properties that the biological soft mater-ials reveal in the applications.

Another difference between the crystallization properties of biological soft materials and those of electronic and optical materials may appear in growth media or systems: biological soft materials are crystallized not only in bulk liquid or solution but also in confined spaces called emulsions [3] (Figure 2). Emulsions are meta-stable dispersion phases of one liquid in a second immiscible liquid. Adsorption of emulsifiers (surfactants) which comprise hydrophobic (nonpolar) and hydrophilic (polar) moieties, causes a decrease in oil/ water interfacial energy and, thereby, makes it easy for small emulsion droplets with diameters ranging from 10^{-5} to 10^{-7} m to be dispersed. Crystallization of biological soft materials even occurs at air/water or air/oil interfaces (foams), as is often observed in whipped cream in which lipid crystals surround the air bubbles for stabilization. More interestingly, crystallization occurring at the foam interfaces is applied to synthesize better-controlled crystal aggregation [4].
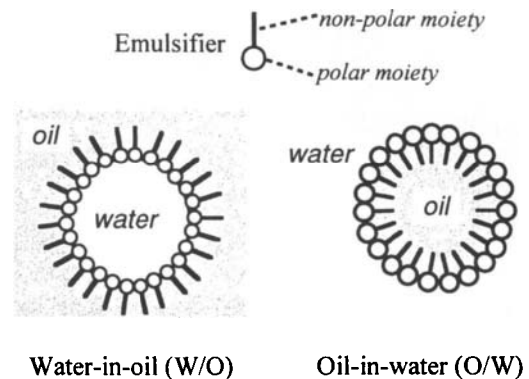

Figure 2. Two types of emulsions.

Figure 3 shows the principal properties of lipid crystals that are formed from neat liquid in physical states of bulk, oil-in-water (O/W) emulsion and water-in-oil (W/O) emulsion. Lipid crystals are used, in confections in the bulk state, margarine in the O/W state, and cream in the W/O state. Among the major functional properties of the lipid crystals, melting and solidification are affected by the size and particularly by the polymorphic forms of the crystals. The appearance of the end product depends on the morphology and aggregation of the lipid crystals. Viscosity, consistency and spread-ability are determined by the aggregation and net-work formation of lipid crystals, which are influenced by the size distribution and morphology of crystals. The stability of emulsion is often affected by crystallization of the lipid phase [5]. Blooming is basically a result of uncontrolled crystal growth of a more stable polymorph at the expense of a meta-stable form that is a polymorphic state under normal conditions [6]. Therefore, the crystallization of
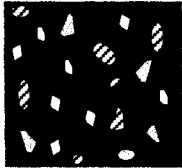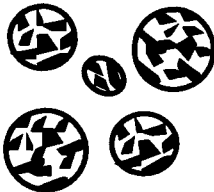
Physical states	Functionality
Bulk	

melting
solidification
appearance
viscosity
blooming

Oil-in-water emulsion

melting
solidification
stability
whipping
viscosity
de-emulsification

Water-in-oil emulsion

melting
solidification
stability
plasticity
consistency
spreadability
blooming

Figure 3. Crystal-function relationships of lipids in bulk and emulsions.

lipids is processed with particular purpose of controlling the nucleation and transformation of specific polymorphic forms, optimal shape, and size and network of the crystals [1].

Crystals of pharmaceutical molecules are grown from a solution. Optimal crystal shape and size distribution are required to achieve efficient filtering during industrial crystallization, and a better suspension is formed when crystalline particles are mixed as ingredients in solid matrices. The rate of dissolution is an essential factor to attain suitable bioavailability of drugs, and it is mainly governed by solubility. When a pharmaceutical substance reveals polymorphism, it follows that less stable polymorphs have higher solubility, and *vice versa* [7]. Therefore, rapid dissolution requires crystal polymorphs with lesser stability, and slow release requires crystals

of highly stable polymorphic modification. For example, indomethacin, a generic name for 1-(*p*-chlorobenzoyl)-5-methoxy-2-methylindolylacetic acid (see Figure 1), an anti-infla-mmatory drug, has at least three and probably nine polymorphs involving differently solvated states [8, 9]. As for not-solvated crystal forms, solubility of the most stable γ form is lower than that of the meta-stable α form by at least 30 %. Therefore, bioavail-ability of indomethacin is remarkably different bet-ween the two polymorphs.

The crystal shape also influences the rate of dissolution; needle and dendrite shapes are more soluble than are bulky shapes. Thus, there is par-ticular interest in polymorphism and crystal shape in the case of crystallization of pharmaceutical mole-cules. A quite similar situation occurs for amino acid crystallization.

When polymorphism of biological soft materials is complicated and thermodynamic stability and meta-stability region of every polymorph is far different from each other, the effects of polymorphic structure and transformation in crystallization pro-cesses are critical [10]. Figure 4 summarizes the major crystallization processes relevant to biological soft materials discussed above.

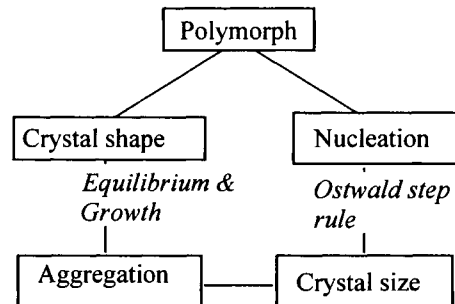Figure 4. major crystallization properties of biological soft materials.

2.2 Modifying nucleation and crystal growth

In the case of crystal engineering of biological soft materials performed in a batch system, crystallization is induced by cooling, evaporation, chemical reactions at various temperatures and/or pressures. Spontaneous nucleation (homogeneous or heterogeneous), secondary nucleation and

subsequent growth are the major crystallization processes [11].

In addition, several attempts are made to modify the crystallization feature of biological soft materials as briefly summarized in the following.

1) Tempering

Tempering process involves cyclic variation of temperature during crystallization that is applied to obtain optimal polymorphic forms or crystal size distribution in the most efficient way. A typical example is tempering in cocoa butter crystallization. Cocoa butter is a chocolate fat, revealing six polymorphs (Form I through Form VI), among which the second stable Form V is most favored [12].

In chocolate tempering, temperature is varied in the manner shown in Figure 5 that involves nucleation of metastable forms (stage A), transformation from the metastable forms to Form V (stage B) and crystal growth of Form V (stage C) [6].

Microscopically, the tempering process involves nucleation of less stable polymorphs in accordance with *Ostwald step rule* [13], and melting of more stable forms at the expense of the less stable forms (melt-mediated transformation) [14]. The stable form is obtained by the tempering procedure shown in Fig.15 more efficiently than that of direct crystallization caused by simple cooling of neat liquid from high temperature, because of successive transformation from liquid to more stable forms through metastable forms [15].
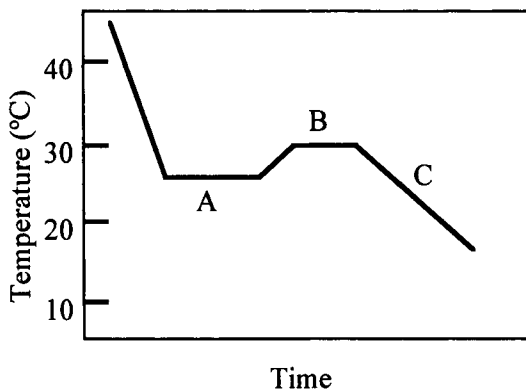

Figure 5. Temperature variation in chocolate tempering.

2) Ultrasonication

Ultrasound at high power and low frequencies can affect various chemical reactions. As for the effects on crystallization, it has been reported that the high power ultrasound remarkably influences both nucleation and crystal growth, by creating fresh and more nucleation sites in the crystallization medium.

The following main mechanisms have been presented to account for the effects of ultrasonication [16-20], (a) violent collapse of the cavitation bubbles may form active sites of nucleation centers, (b) enhanced agitation may effect profound mobility of crystallizing molecules, (c) cooling caused by evaporation from the surface of the cavity during the growth of a cavitation bubble may increase supercooling, and (d) local pressure may increase melting point in the vicinity of a collapsing cavity which turns to be increased supercooling.

3) Shear effect

Crystallization from neat liquid is influenced by shear force. As for lipid crystallization, remarkable effects were observed as increase in the rate of crystallization [21] and in the accelerated occurrence of more stable forms [22]. Few microscopic interpretations have been given to clarify the shear effects.

4) Magnetic field effect

Recent studies have unveiled that strong magnetic field affects the crystallization of organic materials that are diamagnetic. For example, orientation of the crystals was varied by magnetic field [23], rate of crystal growth and dissolution was retarded [24], and crystallinity of the as-grown crystals was improved [25]. The latter two effects were observed for protein crystals. It is thought that the magnetic filed effects may not be operative at a molecular level but at cluster or even post-nucleation level, and that kinetic properties of aqueous solution such as diffusion are influenced by magnetic field.

5) Effects of growth medium

As mentioned before, one of the major characteristics of crystallization of biological soft materials is revealed in growth media or systems; namely, crystallization is performed not only in

bulk system but also in emulsion systems. In the emul-sion systems, nucleation occurs in a different manner from that in the bulk system. In particular, rate of nucleation is reduced and effects of the interface between the oil and water phases may play key roles in the nucleation process in the emulsions. The latter effects are categorized into template effects, which will be further discussed in the next section.

It is widely considered that heterogeneous nucleation is predominantly occurs in the emulsion, al-though some workers assume homogenous nucleation [26]. In the emulsion system, the dispersed phase is divided into a number of droplets that vary in size (poly-disperse). Catalytic impurities are hence dis-tributed unequally through some of the droplets. This will lead to isolation of impurities that catalyze heterogeneous nucleation in the bulk system, and thereby the value of supercooling is enlarged. For nucleation to be initiated, a critical radius (size) of nuclei must be formed. When a pure material is emulsified into many droplets, each droplet has a minute size of several micrometers, then it become difficult to assume that the nucleation process can be homogeneous, since the scattered inner phase mass has no ability to spontaneously assemble into an ordered domain greater than the critical nucleus size [27]. Therefore, nucleation in emulsion is assumed to be heterogeneous by reducing surface energy of droplets.
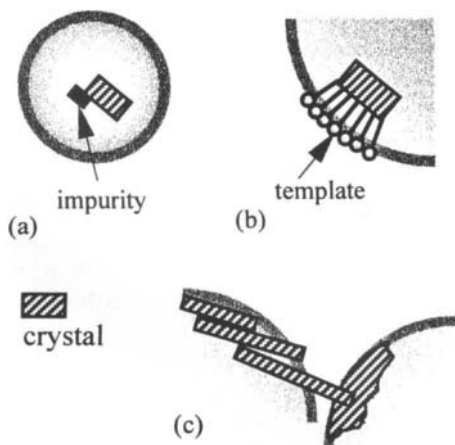

Figure 6. Three types of heterogeneous nuclea-tion occurring in emulsion droplets.
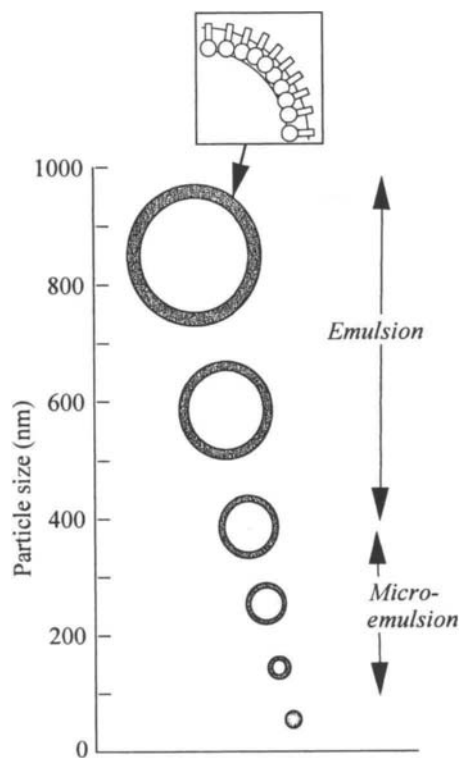
Figure 7. Particle size of water-in-oil emulsion and microemulsion.

Heterogeneous nucleation in the emulsions is subdivided into three types, depending on the location where catalytic reactions resulting hetero-geneous nucleation occur (Figure 6):(a) Volume heterogeneous nucleation initiated by catalytic im-purities scattered randomly in the droplets [28], (b) surface heterogeneous nuclea-tion that is induced by additive molecules performing templates at the interface [29], and (c) inter-droplets nucleation caused by interaction between solid and liquid droplets [30].

The emulsions are metastable, whereas micro-emulsions having nanometers-sized droplets are homogeneous and thermodynamically stable [30] (Figure 7). Two types of microemulsions are formed, like emulsions; water phase dispersed in oil (W/O microemulsion), and O/W micro-mulsion having an opposite phase relation. Crystallization in the microemulsions is also of particular interest, showing unique properties.

W/O microemulsions are formed only in specific ranges of temperature, pressure, and composition, and are therefore best described

with phase diagrams [32, 33]. The presence of supramolecular aggre-gates has made W/O microemulsions solubilize the materials which are otherwise sparingly soluble in aqueous phase. Therefore, O/W emulsions have been applied for specific reactions as "micro-reactors" or "nano-reators" [34, 35].

In contrast, O/W microemulsions containing lipids have recently investigated as novel drug carrier systems [36]. O/W microemulsions are formed by very strong shear force with the size dimensions smaller than 150 nm. Crystallization and poly-morphic transformation of lipids in the O/W micro-emulsions are largely different from those in the bulk phase.

6) Effects of additives

As an advanced crystal engineering strategy, many researchers have attempted to clarify the effects of specific additives that may modify the rate of nucleation and growth of crystals under a given set of thermodynamic conditions in different growth systems (bulk or emulsion).
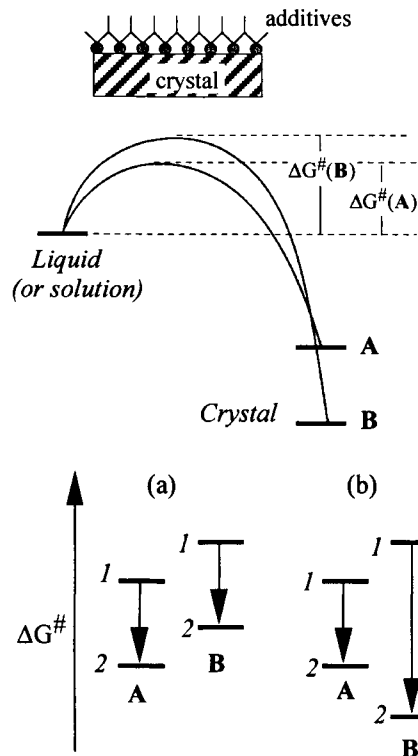

The additives may play roles in modification of the activation free energy values ($\Delta G^\#$) for crystallization in the manner depicted in Figure 8 [37]. A and B mean different crystal structures or different crystal faces of a given crystalline material, and states 1 and 2 indicate conditions without and with the additive, respectively. Since $\Delta G^\#$ of A is larger than B without the additive, the rate of nucleation or crystal growth of A is higher than that of B. In Fig.8(a), it can be seen that the additive has no effect on any cases, since the $\Delta G^\#$ values of both A case and B case are syn-chronously varied by the presence of the additive in the growth condition. This means that there is no effect from the additive.

On the other hand, in Fig.8(b), it can be seen that the $\Delta G^\#$ values of A and B are reversed by the additive, because of the specific molecular interactions acting at the crystal-additive interface. In general, $\Delta G^\#$ can be a free energy barrier for nucleation of a crystal or for crystal growth of a specific crystal face. Additives playing the former role have effects on selective nucleation of crystal structures, and the crystal shape is modified by additives that play the latter role. Taylor-made additives employed for crystal shape modification are of the latter type, as is discussed in another chapter.

The additives that control the rate of nucleation of particular polymorphic structures and the rate of crystal growth of particular crystal faces are classified into three groups: impurity, crystal seedings and templates. Since the effects of impurity [38] and seeding [39] are discussed in recent reviews, here considered is the effect of template.

7) Effects of template

A template is prepared specifically to promote nucleation of selected materials in growth systems. The template is arranged in a well-defined manner in or around the growth media as depicted in Figure 9.

Crystallization occurring in a template system is closely related to elementary processes of biological mineralization and to crystal engineering processing for the production of advanced materials. It has been shown that one of the properties most specific to crystallization processes in complex systems is revealed in the

Figure 8. Effects of additive on activation free energy ($\Delta G^\#$) of crystallization.

nucleation process: active sites present at certain positions play key roles as templates in the nucleation processes. A template promotes nucleation of selected materials through specific interactions between the active sites and crystallizing materials. This process may be categorized as interfacial heterogeneous nucleation. In Fig. 9(a), the template is an amphiphilic film, the polar head groups of which catalyze the heterogeneous nucleation of polar crystals in an aqueous solution. In Fig. 9(b), the amphiphilic template film constructs W/O emulsions in which polar crystals are nucleated in the aqueous solution encapsulated in the W/O emulsion. The nucleation processes of nonpolar crystals at a planar template and an emulsion interface are shown in Fig. 9(c) and (d), respectively. In the former case, the non-polar head groups of the template film catalyze the heterogeneous nucleation of non-polar crystals in an organic solution. In Fig. 9(d), the amphiphilic template molecules construct O/W emulsions in which non-polar crystals are grown in interior phases of droplets, and the template may be prepared in the hydrophobic compartment of the oil/water interface.
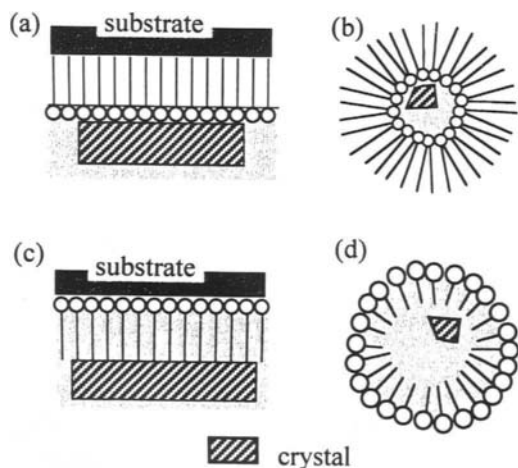

Figure 9. Template-assisted crystallization systems.

Many studies on crystallization of various advanced materials have been carried out using template growth systems such as those illustrated in Fig.9. Interestingly, it has been revealed that well-defined molecular interactions occur at heterogeneously nucleating interfaces through specific molecular interactions having the nature of electrostatic interactions [40] and hydrogen bonding [41].

Template-accelerated crystallization of the type shown in Fig. 9(a) has been well documented [42, 43]. Therefore, in this chapter, recent works on template-accelerated crystallization in the same systems as those of Fig. 9 (b), (c) and (d) are discussed.

3. CRYSTALLIZATION OF AMINO ACID IN W/O MICROEMULSION

W/O micro-emulsions have the ability to solubilize polar crystals and therefore have been used to synthesize micro- and nano-particles of metal oxides and semiconductors. Compared to those inorganic substances, crystallization of biological soft materials, such as aspartame [44], an artificial sweetener, in W/O microemulsions has only been attempted quite recently. This is mainly because the high molecular weights of organic materials result in lower solubility, and the stabilization of W/O emulsions is often disrupted by the pH control of an aqueous solution that is necessary to increase the solubility of polar organic substances.

As for amino acids, polymorphic crystallization has been investigated in the bulk solution [45, 46], whereas crystallization of amino acids in a W/O microemulsion has been initiated recently. Hydrophobic amino acid, such as leucine, can cause template nucleation of glycine crystals at the air-water interface [47]. This phenomenon was examined in W/O microemulsions [48, 49].

Yano et al. investigated the solubilization of glycine and L-phenylalanine (Fig.1) in a water-in-isooctane microemulsion stabilized by an emulsifier, sodium di-2-ethylhexyl sulfosuccinate (AOT) [50, 51]. The maximum soluble amount of amino acid was determined, and the effects of the addition of amino acids in aqueous solution on the size and shape of the microemulsion droplets and on their thermal properties were determined by small-angle X-ray scattering (SAXS) and differential scanning calorimeter (DSC) measurements. It was found that the extent of solubilization strongly depended on the

hydrophobicity of the amino acid, which also determines the location of guest molecules within the microemulsion.

The solubilization properties of glycine, *L*-phenylalanine and hystidine are summarized as follows. [50]:

1) The maximum amount of amino acid that could be solubilized was determined by the solid-liquid extraction method, and the effects of the addition of amino acids on the size and shape of the microemulsion droplets and on their thermal properties were determined using SAXS and DSC measurements, respectively.

2) The solubilization of glycine molecules, which primarily dissolve in a water pool, was slightly lower than their solubility in pure water, decreasing with increasing concentration of AOT and increasing with increasing water content in the microemulsion. In contrast, the solubilization of *L*-phenylalanine, which is primarily located at the water/oil interface, was several times greater than its solubility in water, the solubilized amount increasing with increasing AOT and/or water concentrations. Histidine had characteristics intermediate between these two extremes.

3) Solubilization of those molecules effected an increase in droplet size. The results of thermal analysis showed that loading of the microemulsion droplets with glycine has a much stronger effect on the thermal behavior of the emulsified water than does loading with *L*-phenylalanine. The low solubilization of glycine compared to its solubility in pure water can be explained by the state of water within the microemulsion droplets, i.e., part of it is present as free water and part as water bound to the AOT head groups. The loading of *L*-phenylalanine changed the shape of microemulsion droplets from spherical to ellipsoidal, and the [phenylalanine]/[AOT] molar ratio at the interface increased as the droplet size increased.

To explain the results in more detail, Figure 10 shows that solubilization of glycine in microemulsions was slightly lower than that in bulk water, whereas the solubilization of *L*-phenylalanine was several times greater than that in bulk

water. This difference may be due to the difference in hydrophobicity of the total molecules, e.g., *L*-phenylalanine is more hydrophobic than glycine, and *L*-phenylalanine molecules may therefore be entrapped at the hydrophobic moieties of the AOT molecules, forming microemulsion droplets.
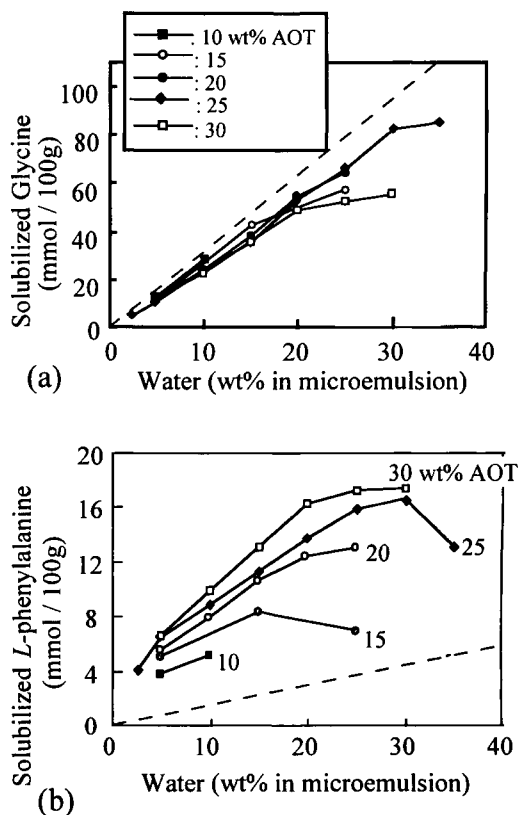
Figure 10. Solubilization of amino acid in W/O microemulsions, (a) glycine and (b) *L*-phenylalanine.

On the basis of the solubilization properties, it was found that the crystallization of glycine and *L*-phenylalanine from the W/O microemulsions was strongly affected by the localization of solubilized molecules within the microemulsion droplets [51]. In the case of glycine, a significant reduction in crystal size was observed; e.g., the crystals formed in bulk aqueous solution usually grew to a size in the order of mm, whereas the glycine crystals grown from microemulsions were of the order of sub-μm to μm in size. In addition, polymorphic occurrence also varied: the γ-form

Table 1
Polymorphism and morphology of amino acid crystals grown from bulk solution and water-in-oil microemulsion.

	Polymorph	Morphology	
		Bulk	Microemulsion
Glycine	α (P2$_1$/n)	(010) plate	(010) plate
	γ (P3$_1$ or P3$_2$)	Trigonal pyamid	n.d.*
	β (P2$_1$)	Needle	n.d.
L-phenylalanine	α (n.d.)	Needle	n.d.
	β (C2)	(001) plate	Pyramid

* not determined

dominantly crystallized from microemulsion, whereas the α-form was predominant from the bulk solution.

In the case of L-phenylalanine, the morphology, polymorphism and size of crystals grown from the bulk solution and microemulsion were different. L-phenylalanine crystallized in two polymorphs from the bulk solution, i.e., needle-like α-form and plate-like β-form, whereas, only the β-form occurred in the crystallization from microemulsions. It would be reasonable to assume that the β–form, having two dimers in a hydrophobic-layered structure, is preferentially nucleated at the hydrophobic moieties of the AOT molecules. Furthermore, the plate shape of the β-form became a triangular shape with dominant crystal faces, quite different to the bulk-grown crystal (Figure 11, Table 1).

In the above study, it was found that interaction of microemulsion droplets is involved in the crystallization of rather hydrophilic glycine and hydrophobic L-phenylalanine from a W/O microemulsion. In the case of glycine, the microemulsion serves as a medium for com-partmentalization without contribution of the W/O interface to the crystallization process, which was initiated by heterogeneous nuclei and proceeded due to the addition of solute molecules supplied through dynamic droplet interactions. In the case of L-phenylalanine, however, the W/O interface is involved in the crystallization process. The crystallization was initiated by hetero-geneous nuclei, consisting of molecules that were preferentially oriented at the W/O interface, and

proceeded through contact with interfaces of other droplets. Under these conditions, both the crystal structure and morphology are controlled by the crystallizing environment in which microemulsion droplets and the molecules involved in them play dynamic roles.
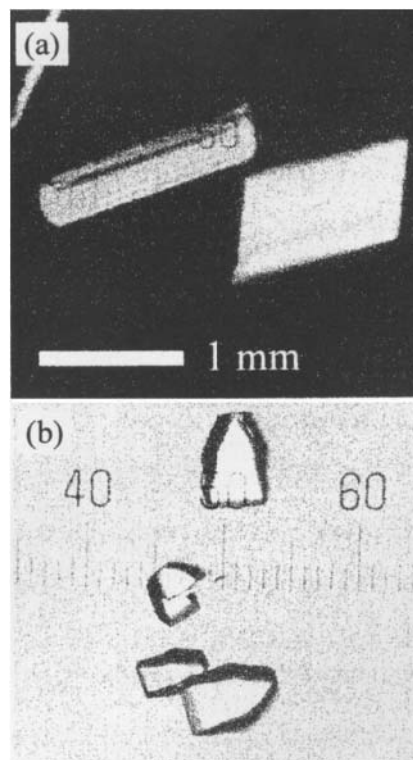

Figure 11. Optical micrographs of L-phenylalanine b crystals grown from (a) bulk solution and (b) microemulsion.

The results of present work on the crystallization of amino acids from W/O microemulsions indicates that there is great potential for application of this technique to other biological soft materials, such as pharmaceuticals and neutraceuticals.

4. TEMPLATE-ACCELERATED SOLVENT CRYSTALLIZATION

It appears that biological soft crystals interacting with templates through nonpolar forces, as depicted in Fig.9 (c) and (d), may be less specific and much weaker compared to interactions of a polar nature (electrostatic and hydrogen bonding). However, it has been observed that acceleration of heterogeneous nucleation occurs by the template films through van der Waals interactions. In this section, template-accelerated crystallization of the type shown in Fig. 9 (c) is discussed.

Morphological and kinetic observations of heterogeneous nucleation of long-chain n-alcohol crystals (Cn-OH, referred to here as guest crystals, n being the number of carbon atoms) from solution, which was accelerated by the presence of vapor-deposited thin films of long chain mole-

cules (fatty acid, Cn-COOH, and mono-acylglycerol, Cn-MG, both referred to as hosts), were carried out [52-56] . Cn-OH crystals were chosen as the model material of lipids. The host templates were formed by physical vapor deposition and placed in slightly supersaturated solution (decane solvent). No crystallization occurred over a period of several hours without the template films, whereas the presence of template films shortened the induction times to several minutes. In addition to the rate of nucleation, the host-guest relationships in terms of polymorphic matching, molecular orientation and chain length limitation were examined.

Figure 12 (a) and (b) show the influences of molecular orientation of the host films (C18-MG) on growth patterns of the guest crystals of C22-OH [56]. Figure 12 (c) shows the model of template-assisted nucleation, which heterogeneously occurs at the growth steps of the template crystalline films. The template film was vapor-deposited on cleaved mica [52]. In Fig.12a, the basal (001) plane of the γ form of the C22-OH crystals is almost parallel to the template films, e.g., the long-chain axes of the C22-OH molecules in the crystal are normal to the substrate. In contrast, the guest crystals accele-
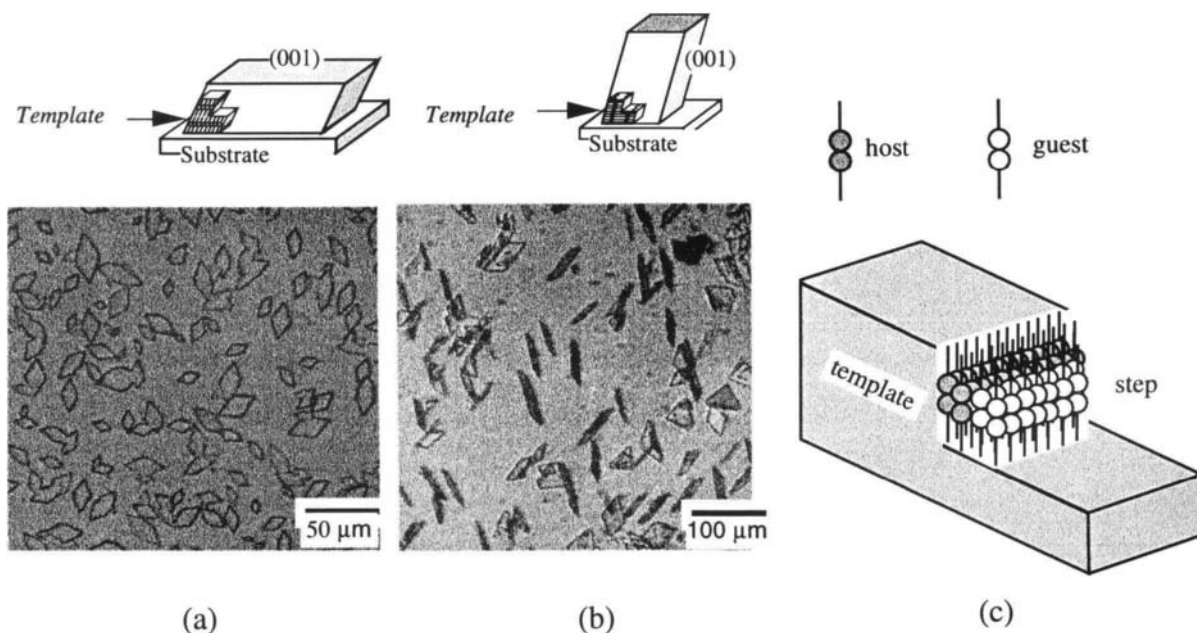

Figure 12. Optical micrographs of solvent-crystallized n-alcohol from (a) normally-oriented and (b) laterally oriented template films. (c) A model of heterogeneous nucleation.

Table 2
Structural properties of template and guest crystals

Host template			Guest $(C_n\text{-OH})$	
$C_n\text{-MG}$ β' form	$C_n\text{-COOH}$		γ form	β form
	E form	C form		
monoclinic	monoclinic	monoclinic	monoclinic	monoclinic
$O_\perp$ *	$O_\perp$	$O_\perp$	$O_\perp$	$O_\perp$
a = 0.88	a = 0.88	a = 0.936	a = 0.889~0.903	a = 0.503~0.507 **
b = 0.51	b = 0.51	b = 0.495	b = 40.97~0.498	b = 0.738~0.743
$\alpha = \gamma = 90°$	$\alpha = \gamma = 90°$	$\alpha = \gamma = 90°$	$\alpha = \gamma = 90°$	$\alpha = \gamma = 90°$
$\beta = 124°$	$\beta = 124°$	$\beta = 128°$	$\beta = 121.7~122.5°$	$\beta = 89.7~92.28°$

* subcell : orthorhombic perpendicular   ** Unit, nm.

rated by the template films in which the long-chain axes are arranged parallel to the substrate, show the same molecular orientation with respect to the substrate as shown in Fig.12b. This difference is due to the geometric relations of the host and guest crystals illustrated in the inserted figures. As for the polymorphic matching between the fatty acid templates and Cn-OH crystals, the occurrence of the γ form was predominant when the host template was the C

form of Cn-COOH and β' of Cn-MG, yet the β form of Cn-OH was dominantly nucleated with the template film of the E form of Cn-COOH. This polymorphic matching was explained by taking into account the unit cell structures of the host templates and guest crystals. Table 2 shows the crystal structure data of polymorphic forms of the crystals commonly occurring in Cn-MG, Cn-COOH and Cn-OH [56].

The relative occurrences of crystallization of

Table 3
Extent of crystallization of guest n-alcohol crystals accelerated by host template thin films

host template films	guest crystals			
	$C_{18}\text{-OH}$	$C_{20}\text{-OH}$	$C_{22}\text{-OH}$	$C_{24}\text{-OH}$
Monoacylglycerols (Cn-MG)				
$C_{16}\text{-MG}$	+	+++	++	+
$C_{18}\text{-MG}$	+	++	+++	+
$C_{22}\text{-MG}$	+	++	++	++
Fatty acids (Cn-COOH)				
$C_{18}\text{-COOH}$	+++	-	-	-
$C_{20}\text{-COOH}$	++	+++	-	-
$C_{22}\text{-COOH}$	-	++	+++	-
$C_{24}\text{-COOH}$	-	-	++	+++

+++ : most accelerated   ++ : very accelerated   + : fairly accelerated   - : no acceleration

three *n*-alcohol crystals (Cn-OH) grown under the condition of template-guest combinations were compared, and the results are summarized in Table 3. Special attention was paid to the effects of chain length matching, as expressed in n of the guest and host template molecules, where two types of template were tested: fatty acid and monoacylglycerol. The Cn-MG templates accelerated the crystallization of Cn-OH, even when the n values of the guest crystals were longer than those of the host films by 8. However, the most enhanced acceleration was observed in such a combination that the n value of the template material was smaller than that of the guest material by 4. This is in clear contrast to the template-guest relations between the combination of Cn-COOH/ Cn-OH, in which acceleration was only observed when the n value of the template films was the same as or longer by 2 than that of the guest crystals.
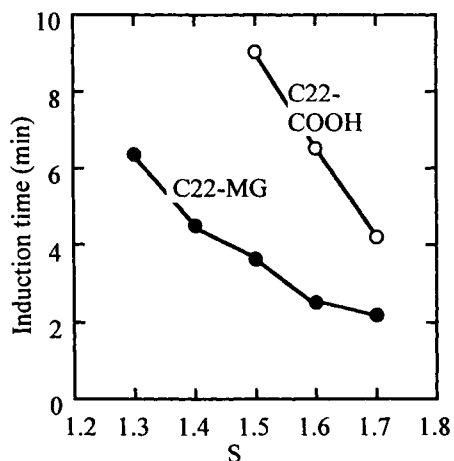

Figure 13. Induction times of solvent crystallization of C22-OH with host template films of C22-COOH and C22-MG.

The induction time measurements for overall crystallization showed two common properties: (a) the induction time increased as supersaturation ratio (S) value decreased in crystallization experiments of the all the host-guest combinations examined, and (b) when a particular template film was set, preferred acceleration was observed for the guest crystal that was specific in regard to the difference between chain lengths of the host template and guest solute molecules, as

revealed by shorter induction times. As an example, Figure 13 shows the induction times of crystallization of the guest material of C22-OH with template films of C22-COOH and C18-MG. The induction time increased as the S value decreased, yet a remarkable difference was seen between the template films of fatty acids and monoacylglycerols, as evidenced in the two sets of experiments. As shown in Fig. 13, the induction time value in the case of the C22-COOH film was 9 minutes, about 3-times longer than that in the case of C18-MG film. As already mentioned, several hours are needed for the nucleation of guest crystals without template films in this range of S values. This means that the acceleration effects of the template films became more enhanced as the value of supersaturation decreased and that the influence of the template-guest molecular interactions was also enhanced as the S value decreased.

In the bulk phase, binary mixture phases of Cn-COOH and Cn-OH are eutectic, yet molecular compound formation was observed in the mixture of Cn-OH and Cn-MG due to intimate hydrogen bonding of the two molecules. This difference in the mixing behavior may account for the relative occurrences shown in Table 3 and induction times shown in Fig.13. Most of the above results were explained by considering a basic model of the template-assisted nucleation illustrated in Fig.12c. This model assumes that heterogeneous nucleation occurs from the steps of the template crystalline films exposed to the solution. The van der Waals molecular interactions between the hydrophobic hydrocarbon chains and hydrophilic polar groups of glycerol groups (Cn-MG), carboxyl groups (Cn- acid) and carbonyl groups (Cn-OH) are operative at the steps exposed at the outer surfaces of the template films. Preferential adsorption of the guest molecules, followed by condensation around the steps, may result in the heterogeneous nucleation of guest crystals. It is assumed that the attractive forces between the mono-acylglycerol and *n*-alcohol molecules are much stronger than those between the fatty acid and alcohol molecules, due to the presence of two OH-groups in the monoacylglycerol template films. This is exhibited in the binary mixture phase behavior forming the molecular compound crystals. It is also thought that hydrogen bonding
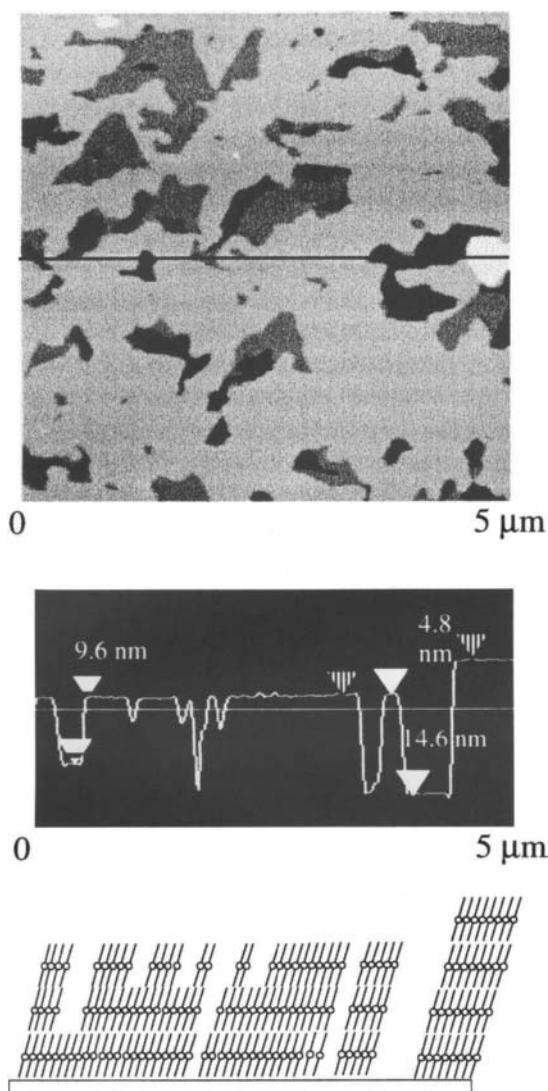
from the solution (supersaturation of 60%) in contact with a C22-COOH template film. The crystals were counted *in situ* under an optical microscope every 30 seconds, soon after the template film had been placed in the solution. Nucleation occurred after an induction time of 150 seconds. The number of crystals rapidly increased within 100 seconds and then stopped increasing. To evaluate the data shown in Fig. 15, a kinetic model describing the three dimensional heterogeneous nucleation and subsequent growth on active centers of heterogeneous nucleation was resolved numerically, taking depletion of the active centers on the template films into account [57-59]. The interfacial energy was chosen to get the best agreement between the calculated and experimental data of Fig.15, giving 0.0029 J/m^2 for C20-OH.
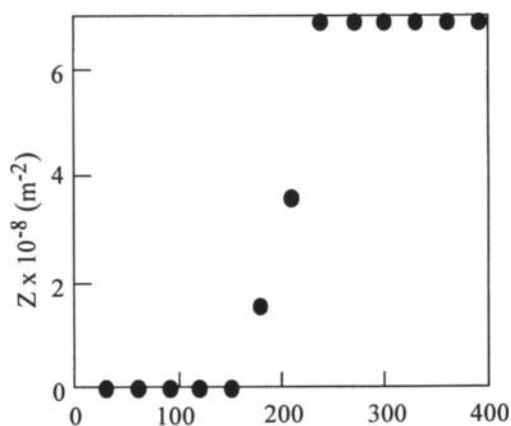

Figure 14. Atomic force microscopic (AFM) image of template film of C18-MG.

Figure 15. Number of crystals (Z) of C20-OH nucleated from template film of C22-COOH.

between the mono-acylglycerol and alcohol molecules occurs at the steps. Figure 14 shows atomic force microscopic feature of the template films of C18-MG deposit-ed on mica surface. The molecules in the film are arranged normal to the substrate, and may step patterns are shown, as illustrated at the bottom of Fig.14.

Numerical calculation of experimental results of the kinetics of heterogeneous nucleation using the host (fatty acid) and guest (alcohol) combinations was performed. Figure 15 shows time variation in the number of crystals of C20-OH nucleated

By applying the template-accelerated heterogeneous nucleation shown above, it is possible to crystallize the guest crystals with specific molecular orientations and polymorphic forms at a reduced rate of nucleation, even when the major molecular interactions are of a nature of van der Waals forces.

5. TEMPLATE-ACCELERATED CRYSTALLIZATION IN O/W EMULSION

The lipid crystallization in the O/W emulsion droplets influences stability, rheology and appear-

414

ance of the O/W emulsions. For example, controlling the lipid crystallization in the O/W emulsion phase is essential in de-emulsifying process of creams, freezing of creams, and coagulation of the O/W emulsions at chilled states [5]. Production, quality and stability of lipid products in emulsion states are greatly influenced by crystallization of the oil phase, and many of recent studies have therefore been aimed at investigation of the lipid crystallization in the O/W emulsion.

The complexity of the lipid crystallization phenomena in an O/W emulsion is revealed in many factor influencing the crystallization processes, such as the rate and extent of crystallization, influences of emulsion droplet size, effects of emulsifier, droplet-droplet interaction, polymorphism, effects of cooling rate and subsequent temperature history. Since a pioneering work made by Skoda and Tempel [60], it has been obvious that complexity in the lipid crystallization phenomena in the O/W emulsion are revealed in the rate and extent of crystallization, influences of emulsion droplet sizes, effects of emulsifiers, droplet-droplet interactions, polymorphism, effects of cooling rate and sub-sequent temperature history, etc.[61-64]. To clarify the crystallization mechanisms of an O/W emulsion, the authors have employed ultrasonic velocity measurement, DSC and synchrotron radiation X-ray diffraction (SR-XRD) [65-67]. In particular, the ultrasonic velocity and SR-XRD measurements enabled *in situ* and nondestructive monitoring of lipid crystallization in the emulsion.

The kinetic properties of the crystallization processes of lipids in the O/W emulsions, which are remarkably modified by highly hydrophobic emulsifier additives having high melting points, are discussed here. *n*-Hexadecane (melting point, 18 °C) was chosen as the lipid. Tween 20 was put in a water phase to make the emulsions (20 wt.% oil and 80 wt.% water) having a mean droplet size of 0.8 μm in diameter.

As additives, sucrose fatty acid oligoesters, polyglycerine fatty acid esters and diacylglycerols, put in *n*-hexadecane before emulsification, were examined. Figure 16 shows the ultrasonic velocity (V) values of *n*-hexadecane-in-water emulsions measured by cooling-heating in the tem-

perature range of 60 °C to - 6 °C. The results revealed the following features [65]:

1) On cooling from 60 to 3 °C, the V value first increased and decreased with decrease in temperature. The increase in V value was mainly due to the contribution of the *n*-hexadecane phase, and the subsequent decrease was due to the contribution of the water phase.

2) On further cooling, an abrupt increase in the V value was observed around 2 °C, corresponding to the crystallization of the *n*-hexadecane phases in the emulsion. Therefore, 2 °C is defined as the crystallization temperature (T_c). With further cooling below 0 °C, V value decreased again due to the contribution of the water phase.

3) On heating from -6 °C, the ultrasonic velocity increased up to about 12 °C and then rapidly decreased above that temperature up to 18 °C, where the V values were stabilized. This means that the melting of *n*-hexadecane in the O/W emulsion started at 12 °C and ceased at 18 °C.

4) The results of the DSC heating study showed melting peaks of *n*-hexadecane at onset and offset temperatures of 16.1 °C and 19.0 °C, respectively (data not shown). These results indicate that the melting property of *n*-hexadecane in the emulsion was not greatly different from that in the bulk, although reduced melting might occur in the case of crystals having small sizes in small emulsion droplets.
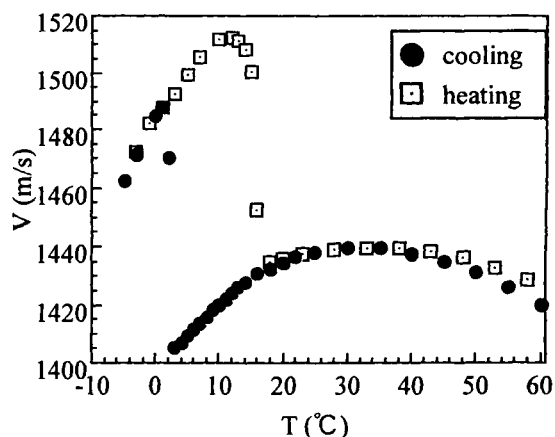

Figure 16. Variation of ultrasonic velocity values (V) of *n*-hexadecane-in-water emulsion with temperature.

5) The final V value, which was reached after the thermal cycle of 60 →-6 →60 °C, was larger than the initial value at 60 °C. This may be due to an increase in the oil-liquid droplet size in the emulsion due to coalescence, which might occur during a thermal process lasting for over 24 hours.

From these results, it was evident that the crystallization of n-hexadecane in an emulsion was diminished with a supercooling value of about 16 °C, whereas no change was observed in the melting behavior. This meant that the crystallization kinetics was solely influenced by the emulsification. The next results show how the reduced crystallization was enhanced by the addition of hydrophobic emulsifier additives.
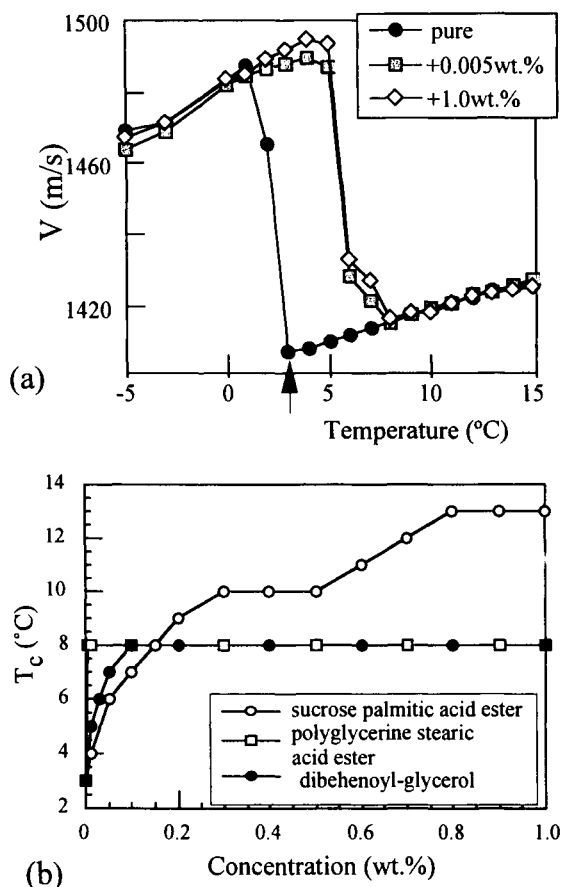

(a)

(b)

Figure 17. (a) Ultrasonic velocity value (V) of n-hexadecane-in-water emulsion with an additive of polyglycerine stearic acid ester, and (b) crystallization temperature (T_c) vs concentrations of three additives.

Figure 17 (a) shows the V values of n-hexadecane-in-water emulsions measured by cooling the emulsion with the additive of polyglycerine stearic acid ester (DAS-750) of different concentrations [66]. Fig. 17 (b) shows the variation in the crystallization temperature of the emulsion (defined below) with increasing concentrations of three additives. The following features are noted from Fig. 17:

1) Without the additive, the emulsification decreased T_c compared to that in the case of the bulk systems (the same as the result in Fig.16).

2) Crystallization was accelerated with an increasing amount of the additive, as revealed by the increases in T_c values with increasing additive concentrations.

3) It was shown that the higher the melting points of the additives are, the higher is the extent of the acceleration of crystallization (data not shown here).

4) A separate experiment showed that the rates of crystal growth of n-hexadecane were retarded by the all additives examined (data not shown here).

These results clearly showed that the additives accelerated the nucleation but retarded the crystal growth. This feature was not observed in crystallization in the bulk phase.

It is assumed that the acceleration of crystallization is due to heterogeneous nucleation processes caused by the additives [68]. Therefore, simultaneous DSC and SR-XRD measurements were carried out in n-hexadecane-water emulsion with the additive (1 wt.%) of polyglycerine stearic acid ester (DAS-750) as shown in Figure 18 (unpublished data).

In the case of pure emulsion without an additive, when the emulsion was cooled from 20 °C, an exothermic DSC peak appeared at 2 °C (onset temperature) and corresponding wide-angle SR-XRD spectra were detectable at the same temperature as shown in Fig.16a. This is due to the crystallization of n-hexadecane. However, during the same cooling process, an exothermic DSC peak first appeared at 7.7 °C, which is not due to the crystallization of n-hexadecane, because the corresponding SR-XRD wide-angle spectra are not of n-hexadecane (arrows in Fig.19b). The n-hexadecane crystallized at 5.5 °C,
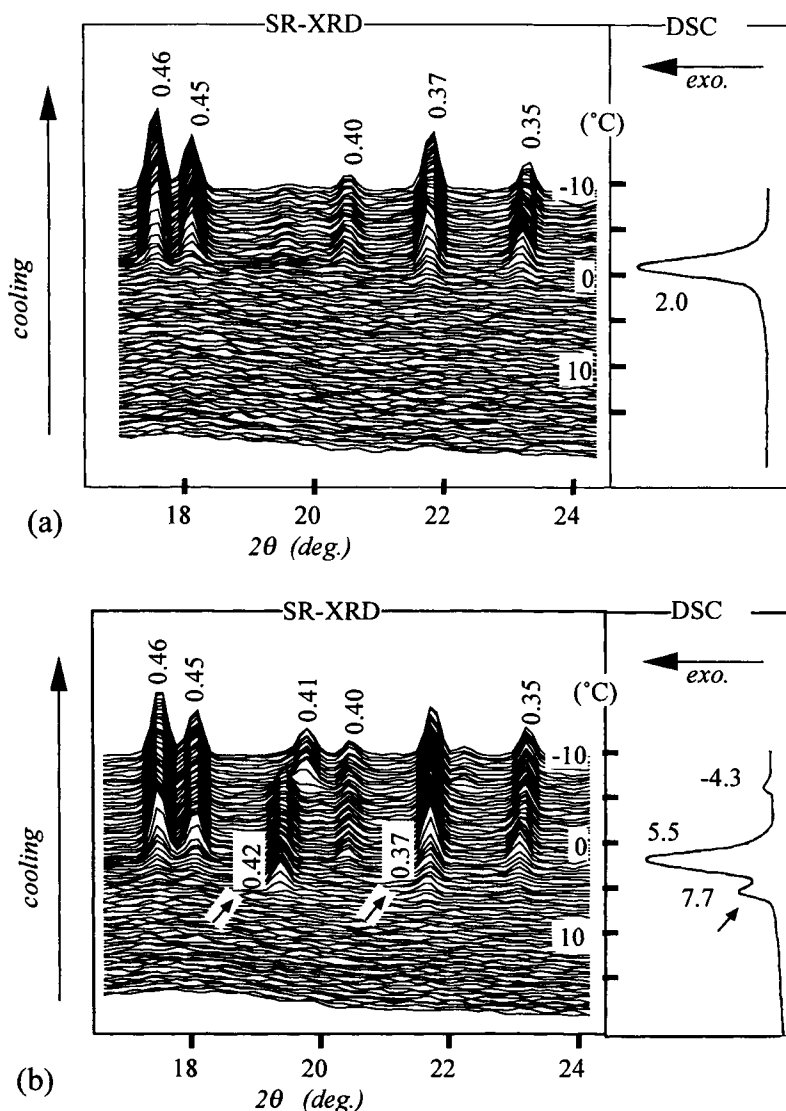
416

Figure 18. Simultaneous Synchrotron-X-ray diffraction (SR-XRD) and DSC measurement of crystallization of n-hexadecane-water emulsion. (a) without additive and (b) with additive (DAS-750). Unit, nm.

since the corresponding SR-XRD spectra appeared at this temperature. On further cooling, subtle changes were detectable around – 4 °C for the SR-XRD spectra which appeared at 7.7 °C, and correspondingly an DSC exothermic peak appeared at the same temperature. This means that the DSC exothermic peaks at 7.7 °C and – 4.3 °C are due to phase transformations of the molecular aggregate that was formed by the addition of polyglycerine stearic acid ester. It is thought that this aggregate acts as a template for the nucleation of n-hexadecane.

Figure 19 illustrates the template-accelerated heterogeneous nucleation model that explains the above experimental results usin ultrasonic velocity and simultaneous SR-XRD and DSC techniques [67]. The additive molecules put in liquidus n-hexadecane are preferentially adsorbed at the oil-water interface after the emulsification, making a molecular aggregate together with Tween 20, water and probably n-hexadecane itself. This occurs easily since the additives examined have high melting points and are therefore less soluble in n-hexadecane. On
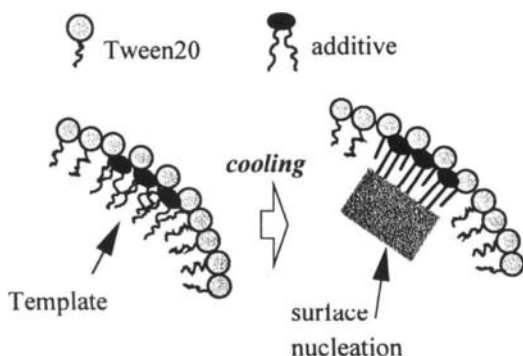
Figure 19. Template-assisted nucleation model of oil-in-water emulsion.

cooling, the liquidus phase of the molecular aggregate transforms to a solid phase, giving rise to the exothermic DSC peak observed at 7.7 °C (Fig. 19b). The frozen aliphatic tail chains of the molecular aggregate serve as the template for the heterogeneous nucleation for n-hexadecane, increasing the T_c of n-hexadecane in proportion to the concentration of the additive molecules.

The acceleration of crystallization in the emulsion by the additives may be indicative of crystal engineering of functional materials in encapsulated systems, such as an emulsion. This is because the nucleation rate is reduced in an emulsion and, thereby, the overall extent of the crystallization is minimized. The use of an additive may raise the total extent of the crystallization due to the accelerated nucleation rate and eventually raise the yield of the crystallizing materials, even if the rate of crystal growth is retarded by the additive. This was confirmed in the case of palm oil crystallization in the O/W emulsions [69].

6. CONCLUSION

The crystallization processes of biological soft materials are closely related to life science and technology, as mentioned in the introduction. Details, however, of crystallization of these materials have not thoroughly be elaborated. Possible reasons for this are (a) multiple polymorphism must complicate the crystallization phenomena, (b) crystal growth systems range

from bulk to emulsions and (c) no adequate technique has been applied to the monitoring of crystallization of biological soft molecules in, particularly those in an emulsion system. The use of SR-XRD techniques combined with DSC or even infra-red (IR) absorption techniques might resolve many of the above difficulty. Time-resolved measurements of crystallization and subsequent transformation are possible, with detailed information of molecular and thermal properties.

The roles of template additives have been discussed in this chapter. In addition to this, sonication, shearing and seeding are effective external factors affecting the crystallization behavior of the biological soft materials, as mentioned at the beginning. The use of these advanced techniques should further elucidate the crystal engineering of biological soft materials, thus making their crystallization processes more understandable and controllable.

REFERENCES

1. F. D. Gunstone and F. B. Padley (eds.), Lipid Technoogy and Applications, Marcel Dekker, New York, 1997.
2. G. C. Barrett (ed.) Chemistry and Biochemistry of the Amino Acids, Chapman and Hall, London, 1985.
3. P.Becher (ed.), Encyclopedia of Emulsion Technology, Vol.1, Marcel Dekker, New York, 1983.
4. Bao-Dong Chen, J. J. Cillers, R. J. Davey, J. Garside and E. T. Woodburn, J. Am. Chem. Soc., 120 (1998) 1625, and references therein.
5. K. Boode, C. Bisperink and P. Walstra, Colloid Surf., 61 (1991) 55.
6. K. Sato and T.Koyano, in Crystallization Processes of Fats and Lipid ystems, N.Garti and K.Sato (eds.) Marcel Dekker, New York, 2001, pp.429-456.
7. J. O. Henck, U. J. Griesser and A. Bueger, Pharm. Ind., 59 (1997) 65.
8. L. Borka, Acta Pharm. Suecica, 11 (1974) 295.
9. N. Kaneiwa, M.Otsuka and T.Hayashi, Chem. Pharm. Bull., 33 (1985) 3447.
10. A. Gavezzotti and G. Filippini, J. Am. Chem. Soc., 117 (1995) 12299.
11. A. S. Myerson, (ed.) Handbook of Indistrial Crystallization, Butterworth-Heinemann, Boston, 1993.

418

12. W. Wille, R.L. and Lutton, J.Am.Oil Chem.Soc. 43 (1966) 491.
13. J. W. Mullin, Crystallization, Third edition, Butterworth-Heinemann 1993, pp.172-201.
14. K. Sato, J.Phys. D., Appl.Phys. B77 (1993) 26.
15. D. Kashchiev and K.Sato, J. Chem. Phys., 109 (1998) 8530.
16. T. J. Mason, L. Paniwnyk, and J. P. Lorimer, Ultrasonics. Sonochem. 3 (1996) S253.
17. L. H. Staney, Ultrasonics (1967) 202.
18. N. Enomoto, T. H. Sung, Z. E. Nakagawa, and S. C. Lee, J. Materials. Sci. 27 (1992) 5239.
19. G. I. Eskin, Ultrasonics. Sonochem. 1 (1994) S59.
20. K. Ohsaka and E. H. Trinh, Appl. Phys. Lett,. 73 (1998) 129.
21. Z. Ziegleder, Intern. Z.Lebensm.Techn.Verf. 36 (1985) 412.
22. S. D. MacMillan, K. J. Roberts, A. Rossi, M. Wells, M. Polgreen and I. Smith, in: The Proceedings of World Congress on Particle Technology, Brighton, 1998 pp.96-103.
23. M. Fujiwara, M. Fukui and Y. Tanimoto, J.Phys.Chem. B, 103 (1999) 2627.
24. S. Yanagiya, G. Sazaki, S.D. Durbin, S. Miyashita, K. Nakajima, H. Komatsu, K. Watanabe and M. Motokawa, J. Cryst. Growth, 208 (2000) 645.
25. T. Sato, Y. Yamada, S. Saijo, T. Nakaura, T. Hori, R. Hirose, N. Tanaka, G. Sazaki, K. Nakajima, N. Igarashi, M. Tanaka and Y. Matsuura, Acta Cryst., D56 1079 (2000) 1079.
26. W. Kloek, P. Walstra and T. V. Vliet, J.Am.Oil Chem.Soc., 77 (2000) 643.
27. M. J. W. Povey, in Crystallization Processes of Fats and Lipid Systems, ed N. Garti and K.Sato, Marcel Dekker, New York, 2001, pp.
28. D.Turnbull and R. 1. Cormia, J.Chem.Phys., 34 (1961) 820.
29. J. P. Cordiez, G. Grange and B. Mutaftschiev, J. Colloid Interface Sci., 85 (1982) 431.
30. S. Hindle, M. J. W. Povey and K. Smith, J.Colloid Interface Sci., 232 (2000) 370.
31. M. Bourrel and R. S. Schechter Microemulsions and related systems, Marcel Dekker, New York, 1988, pp.127-205.
32. W. F. C. Sager, Langmuir, 14 (1998) 6385.
33. W. F. C. Sager, Curr. Opin. Colloid Interfcae Sci. 3 (1998) 276.
34. M. P.Pileni, J.Phys.Chem., 97 (1993) 6961.
35. M.P.Pileni, I.Lisiecki, L.Motte, C.Petit, J.Tanori and N.Moumen, in Micelles, Microemulsions, and Monolayers, Science and Technology, ed.D.O.Shah, Marcel Dekker, Inc., New York, (1998) pp.289-304.
36. H. Bunjes, M. H. J. Koch and K. Westesen, Langmuir, 16 (2000) 5234, and references therein.
37. S.Mann and G.A.Ozin, Nature 382 (1996) 313.
38. K. Sangwal, Prog.Crystal Growth & Charact. 32 (1996) 3.
39. W. Beckmann, Org. Process Res. Dev., 4 (2000) 372.
40. M.Antonietti and C.Goltner, Angew. Chem. Int. Engl. 36 (1997) 910.
41. J. Majewski, L. Margulis,L, I. Weissbuch, R. Popovitz-Biro, T. Arad, Y. Talmon, M. Lahav and L. Leiserowitz, Adv. Mater., 7 (1997) 26.
42. S. Mann, Biomimetic Materials Chemistry; VCH, Germany : New York, 1996.
43. M. Sarikaya and I. A. Aksay (eds.) Biomimetics, Design and Processing of Materials, ed., AIP Press, Woodbury, 1995.
44. H. Fudredi-Milhofer, N. Garti and A. Kamyshny, J.Cryst.Growth 198/199 (1999) 1365.
45. M. Kitamura and T. Ishizu, J. Cryst. Growth 209 (2000) 138
46. M. Kitamura and K. Onuma, J. Colloid & Interface Sci.,224 (2000) 311.
47. E. Landau, M. Levanon, L. Leiserowitz, M. Lahav and J. Sagiv, Nature 318 (1985) 353.
48. T. Hirai, S. Horiguchi, I. Komasawa and R. J. Davey. Langmuir 13 (1997) 6650.
49. K. Allen and R. J. Davey, Industrial Crystallization, ICHemE, (1999) 1.
50. J. Yano, H. Furedi-Milhofer, E. Wachtel and Nissim Garti, Langmuir, 16 (2000) 9996.
51. J. Yano, H. Furedi-Milhofer, E. Wachtel and Nissim Garti, Langmuir, 16 (2000) 10005.
52. H. Takiguchi, M.I zawa, K. Yase, S. Ueno, M. Yoshimura, T. Yao, K.S ato, J. Crystal Growth, 146 (1995) 645.
53. K. Sato, H. Takiguchi, S. Ueno, J. Yano and K. Yase, .in Advances in the Understanding

of Crystal Growth Mechanisms, T. Nishinaga et al., (eds.) North-Holland, Amsterdam, 1997, pp.349-363.

54. H. Takiguchi, K. Iida, S. Ueno, J. Yano and K. Sato, J. Cryst. Growth, 193 (1998) 641.

55. H. Takiguchi, J. Yano, T. Nakada, S. Miyashita, H. Komatsu and K.Sato, J. Cryst. Growth, 205 (1999) 575.

56. K. Fujiwara, S. Nagahisa, J. Yano, S. Ueno and K. Sato, J. Phys. Chem.B, 104 (2000) 8116.

57. Z. Kozisek, P. Demo and M. Nesladek, J. Chem. Phys., 108, (1998) 9835.

58. Z. Kozisek, P. Demo and K. Sato, J. Cryst. Growth 209 (2000) 198.

59. Z. Kozisek, T. Koga, K.Sato and P. Demo, J. Chem. Phys, to be published.

60. W. Skoda and M. Tempel, J.Colloid Sci., 18 (1963) 568.

61. E. Dickinson, D. J. McClements and M. W. Povey, J. Colloid Interface Sci., 142 (1991) 103.

62. J. Coupland, E. Dickinson, D. J. McClements, M. W. Povey and C. R. Mimmer, in: E.

Dickinson, and P. Walstra (eds.), Food Colloids and Polymers: Stability and Mechanical Properties, Roy. Soc. Chem., Cambridge, 1993, 243-248.

63. D. J. McClements, E. Dickinson, S. R. Dungan, J. E. Kinsela, J. G. Ma and M. W. Povey, J. Colloid Interface Sci., 160 (1993) 293.

64. D. Kashchiev, N. Kaneko and K.Sato, J. Colloid Interface Sci., 208 (1998) 167.

65. 64. N. Kaneko, T. Horie, S. Ueno, J. Yano, T. Katsuragi and K.Sato, J. Cryst. Growth, 197 (1999) 263.

66. T. Aoyama, S. Ueno and K. Sato, J. Jon.Oil Chem. Soc., 49 (2000) 809.

67. T. Katsuragi, N. Kaneko and K.Sato, Coll.& Surf.B. Biointerfaces, 20 (2001) 229.

68. D. Kashchiev, Nucleation, Basic Theory with Applications, Butterworth-Heinemann, Oxford, 2000 pp. 30-44, 293-299.

69. Y. Hodate, S. Ueno, J. Yano, T. Katsuragi, Y. Tezuka, Y. Tagawa, N. Yochimoto and K. Sato, Coll. Surf., 128 (1997) 217.

This Page Intentionally Left Blank