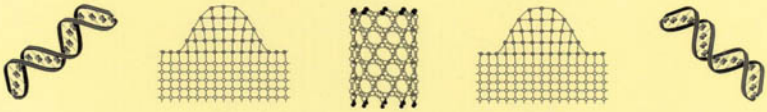
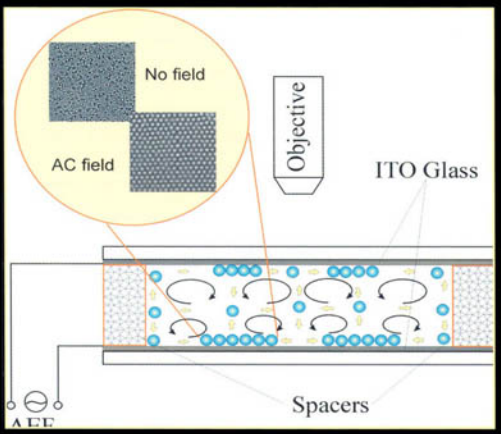


NANOSTRUCTURE SCIENCE AND TECHNOLOGY
Series Editor: David J. Lockwood

Nanoscale Structure and Assembly at Solid-Fluid Interfaces

Volume I: Interfacial Structures versus Dynamics



Edited by Xiang Yang Liu and James J. De Yoreo

NANOSCALE STRUCTURE AND ASSEMBLY AT SOLID–FLUID INTERFACES

VOLUME I

Interfacial Structures versus Dynamics

Nanostructure Science and Technology

Series Editor: David J. Lockwood, FRSC
National Research Council of Canada
Ottawa, Ontario, Canada

Current volumes in this series:

Alternative Lithography: Unleashing the Potentials of Nanotechnology
Edited by Olivia M. Sotomayor Torres

Introduction to Nanoscale Science and Technology
Edited by Massimiliano Di Ventra, Stephane Evoy and James R. Heflin

Nanoparticles: Building Blocks for Nanotechnology
Edited by Vincent Rotello

Nanoscale Structure and Assembly at Solid–Fluid Interfaces (two-volume set)
Vol I: Interfacial Structures versus Dynamics
Vol II: Assembly in Hybrid and Biological Systems
Edited by Xiang Yang Liu and James J. De Yoreo

Nanostructured Catalysts
Edited by Susannah L. Scott, Cathleen M. Crudden, and Christopher W. Jones

Nanotechnology in Catalysis (two-volume set)
Edited by Bing Zhou, Sophie Hermans and Gabor A. Somorjai

Polyoxometalate Chemistry for Nano-Composite Design
Edited by Toshihiro Yamase and Michael T. Pope

Self-Assembled Nanostructures
Jin Z. Zhang, Zhong-lin Wang, Jun Liu, Shaowei Chen and Gang-yu Liu

Semiconductor Nanocrystals: From Basic Principles to Applications
Edited by Alexander L. Efros, David J. Lockwood and Leonid Tsybeskov

---

NANOSCALE STRUCTURE AND ASSEMBLY AT SOLID–FLUID INTERFACES

VOLUME I

Interfacial Structures versus Dynamics

edited by

Xiang Yan Liu
Department of Physics
National University of Singapore
Singapore

James J. De Yoreo
Department of Chemistry and Materials Science
Lawrence Livermore National Laboratory
USA

KLUWER ACADEMIC PUBLISHERS
NEW YORK, BOSTON, DORDRECHT, LONDON, MOSCOW

eBook ISBN: 1-4020-7810-0
Print ISBN: 1-4020-7792-0

Created in the United States of America

Visit Springer's eBookstore at: http://www.ebooks.kluweronline.com
and the Springer Global Website Online at: http://www.springeronline.com

Table of Contents

Volume I
Interfacial Structures versus Dynamics

This page intentionally left blank

Nanoscale structure and assembly at solid–fluid interfaces

Series of Two Volumes

Editors:

Xiang Yang Liu

Department of Physics, National University of Singapore, Singapore

&

James J. De Yoreo

Department of Chemistry and Materials Science, Lawrence Livermore National Laboratory, USA

Foreword

An old joke states that physicists apply good techniques to poorly characterized substances, while chemists work with good substances making use of poor techniques. Nowadays, biologists and biochemists also work with materials poorly characterized because of their complexity. Nevertheless, emerging now and reflected in this book is an era in which high-level techniques are applied to chemically well-controlled substances. The authors provide an excellent collection of fascinating stories on what happens at interfaces between crystalline materials and solutions. Scanning atomic force, tunneling, and vertical interferometric microscopies, modern optics, grazing X-ray diffraction, and molecular dynamic simulations in ionic systems are the major tools that the authors have applied and that have revealed the processes and interactions that people only dreamed of tackling half a century ago. The authors pay attention to the essential "know-how" of all these techniques. Certainly, such a book would not have been possible 20 or even 10 years ago.

This collection of articles reflects the breakthroughs that are happening in understanding surface pheonomena in chemically complex systems—crystals and other solid surfaces—in contact with mainly aqueous solutions containing ions of major components and additives to be adsorbed—from Mg^{2+} to protein ions, to viruses and antibodies. The last couple of decades brought about significant progress in understanding "tailor-made impurities" on the stereo-chemical level. This book reflects further steps towards more detailed quantitative physico-chemical insights. Inorganic crystal growth, dissolution, adsorption, nucleation in connection with specific crystal surface structures are the major subjects of the articles collected in Volume I. Visualization of structural details of viruses, protein crystals and formation of the latter, properties of "soft" biologically related objects and surfaces, like membranes, protein molecules immobilized for analyses, probing their structure and properties, and processes in formation of emulsion droplets provide a second group of objects and phenomena. This group is covered in Volume II. Here, too, crystal growth and adsorption are the key physico-chemical issues.

The demands to monitor environment, to create new nanoelectronics, and to bring more rationale to the endless ocean of exciting but still spotty findings in biology and medicine provide long-term and strong driving forces to move from art to science in all these areas. Yes, the general framework of crystal growth and surface science is now reaching its maturity. Unfortunately, existence of this framework does not yet mean quantitative predictive power, especially for the chemically complex systems, that are in our hands. We still have a long way to go to achieve the aimed spirit through precise quantitative knowledge from physics to chemistry and—which is a much more difficult task—to biology. This book provides a timely step in that direction.

On a more specific note, this book reflects the revival of an interest in crystal growth from solutions. It was boosted about two decades ago by needs to obtain crystals for non-

linear optics, in particular, for laser induced nuclear fusion (an effort in which one of the editors—J.J. De Yoreo—was successfully involved). Another strong incentive was, and is, growth of crystals either built of biological macromolecules or through biomineralization. These are extreme examples of crystallization in systems that are very complex chemically. Hydrothermal crystallization of zeolites and similar purely inorganic systems already exhibit significant complexity. The much larger diameter of biomacromolecules (from several to hundreds of nm) as compared both to conventional ions and to the range of molecular forces, adds additional degree of freedom to the system. Even a moderate change in conventional ionic solution composition, pH and temperature noticeably alters the surface of a biomacromolecule controlling crystallization of these species. Consequently, such a change means, in fact, that other species are to be crystallized. The several polymorphic modifications into which the majority of protein molecules crystallize under different conditions represents a convincing illustration. Thus biomacromolecular crystallization still depends on screening and serendipity. Bringing scientific rationale to this big and hot area is a challenge.

The reader may easily identify the papers dealing with each of the problems listed above so that the only thing that remains to be mentioned is that the authors of the papers are excellent experts in the field. I hope that the editors, X.Y. Liu and J.J. De Yoreo, active and highly recognized contributors to the effort aimed at merging physics, chemistry and biology on a rationale basis, will earn a big credit from our scientific community for having compiled this book.

A.A. Chernov
June 2003, Huntsville, AL

Preface

The initiative and motivation for this book arose from the constructive discussion of one of the editors with Dr. David Lockwood, who visited Singapore in 2001. One of the key questions raised, which ultimately led to this book, was how the structure of the solid–fluid interface could exert a direct impact on the formation of nano structures/phases. Of course, as more sparkling ideas were adopted from our fellow colleagues, who constructively contributed to the various chapters, the scope has been spun off far beyond this point. Nevertheless, this remains as the theme of this book.

All of us have read about the vast potential inherent in nanotechnology and the exciting impact it has in changing our lifestyle in the 21st century. One of the basic issues confronting us is how to fabricate devices or materials on the nano scale. What is the basic physics governing the formation of nano phases? How can biological systems inspire us to formulate nano scale architectures, in the way nature has always done and continues to do? These are two main areas of focus in this book.

The aim of this book is to take you to the root of these issues: the solid–fluid interfacial structures and the basic interactions between nano particles determine the kinetics of nano particles and formation of structures, and consequently the resulting structures and functionalities of the nano particle systems and devices. By taking a fresh look at the novel assembly and surface probing techologies from the global viewpoint of fundamental principles, the two volumes of this book allow us to zoom our focus from the macroscopic phase to the nano structures within a broad spectrum of phenomena. Featuring contributions from a number of international experts in the related fields, this book offers a comprehensive and synergistic look into these challenging issues in tackling topics like theoretical modeling, computer simulations, advanced surface probing and fabrication and interface characterizations. The book also provides a link to the nanostructure engineering of some novel materials playing an important role in advancing technologies in this field.

Volume I focuses on both the theoretical and experimental aspects of crystal surfaces and solid–fluid interfacial structures, the effect of ordering and disordering of the interfacial molecules on the kinetics of crystalline matter, and in particular on the nano crystalline phase and the formation of self-organised nanofiber networks. Expertly written chapters expand the existing knowledge on surface/interfacial structures associated with bulk phases to the micro/nano structural domain, presenting examples of the characterization and manipulation of self-organized nano structures. The subject matter covers the principles and technologies concerning surface/interface structure formation and characterisation, nanophase fabrication and biomineralisation.

Volume II is devoted to the assembly of bio-related nano structures and hybrid nano materials and systems. The part on bio-related nanostructures focuses on the aggregation and pattern formation of protein molecules, or biomaterials in the bulk and on the surface of

designated substrates. The way to control the process of biomineralization, biological matrices, hierachical structures and other bio-related materials by relying on robust methods, such as templating, will also be treated in detail. Finally, innovative techniques to engineer complex nano structures and similar systems such as molecular design and molecular self assembly are highlighted.

The book covers a broad range of subjects relevant to nano/life sciences and technologies. Due to its fundamental nature, this book functions as a valuable resource for scientists and engineers working in the related areas, as well as an important reference manual for both scientists and graduate students in nano/life sciences, majoring in the design of nano systems and related technologies.

X.Y. Liu
on behalf of the Editors
November 2003, Singapore

List of contributors

Jason B. Benedict, University of Washington, USA

P. Bennema, University of Nijmegen, The Netherlands

Melissa A. Hines, Cornell University, USA

Sei-Hum Jang, University of Washington, USA

Bart Kahr, University of Washington, USA

Werner Kaminsky, University of Washington, USA

Miki Kurimoto, University of Washington, USA

Xiang Yang Liu, National University of Singapore, Republic of Singapore

Andreas Lüttge, Rice University, USA

H. Meekes, University of Nijmegen, The Netherlands

Aleksandr Noy, Lawrence Livermore National Laboratory, USA

Michael R. Philpott, National University of Singapore, Republic of Singapore

Elias Vlieg, University of Nijmegen, The Netherlands

This page intentionally left blank

1

Atomic scale modeling of the solid–liquid interface

"Man masters Nature not by force but by understanding. That's why science has succeeded where magic failed; because it has looked for no spell to cast on Nature".

Jacob Bronowski, "Science & Human Values", 1956.

Michael R. Philpott*
Department of Materials Science, Building SIA, Room 02-10, Science Drive 4, National University of Singapore, Lower Kent Ridge Road, Singapore 117543, Republic of Singapore

1.1. INTRODUCTION

The surface of Earth is seventy percent covered by water. The subject area of solid–liquid interfaces is one with profound consequences from the weathering of the planet surface to understanding the science that underpin important technologies like batteries, metal plating and the crystallization of pharmaceuticals. In this review we will survey a number of areas where atomic scale modeling of interface structure and dynamics continues to provide insight into mechanisms and processes that are scientifically significant, technologically important, naturally occurring on a large scale, or just plain interesting.

This article is arranged as follows. We will first briefly survey the scope of atomic scale modeling methodologies in general. Then we will move on to survey briefly some of the science and technology areas where an understanding of the solid–liquid interface is important. Then, drawing unabashedly on the authors' research collaborations during the 1990s, this article illustrates a number of specific areas of solid–liquid interface science. The examples chosen include: intermediate neglect of differential overlap (INDO) calculations of reaction pathways for metal ion dissolution, modified limited Car–Parinello (CP) type calculations of metal dissolution, classical molecular dynamics (MD) simulations of electric double layers at charged metal–aqueous electrolyte interfaces, and finally simple modeling mineral surfaces in connection with speculations about the pre-biotic chemical origins of primitive life.

The three main regimes of modeling materials and their interfaces can be categorized as: atomic scale modeling, continuum modeling, and kinetic method modeling. Figure 1.1

*E-mail address: chmmp@nus.edu.sg

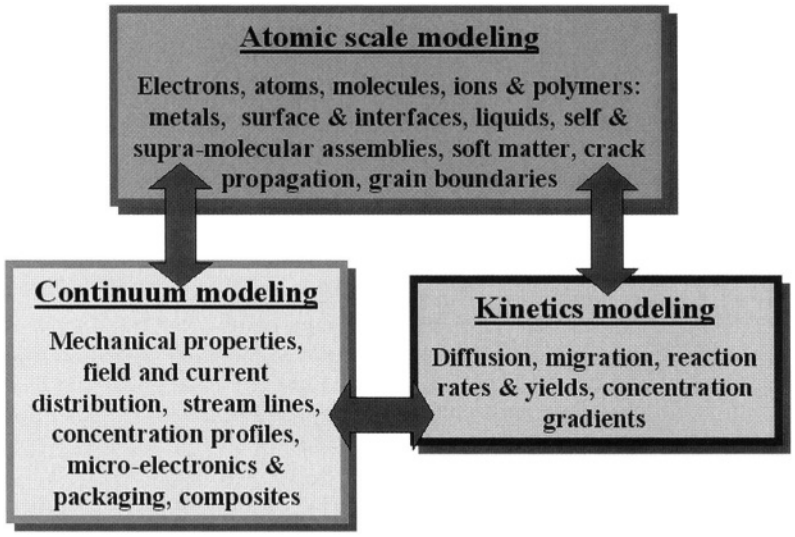


FIGURE 1.1. The three regimes of modeling used in the study of materials and their interfaces: atomic scale, continuum scale and kinetic modeling.

depicts schematically the regimes. Atomic scale modeling is the subject of this review. It includes all aspects of the electronic structure, chemical bonding and atomic motion. It includes the recent arrival of graphic user interfaces (GUI) with sophisticated computer codes that permit the researcher to build graphic pictures, like stick and ball models, to visualize complex interfaces and apply chemical knowledge and intuition to either understand or even predict chemical and physical phenomena. In continuum modeling all the atomic details have been averaged out (e.g., when molecular electric polarization effects are expressed as a dielectric constant) to simpler quantities at length and time scales that are much larger than the size of individual atoms and molecules. In continuum systems the emphasis is usually on modeling complex macroscopic geometry using finite grid methods (FGM). The third basic regime involves the use of kinetics methods to follow time changes in the composition arising from competing reactions or variations in temperature and pressure. As already stated the main interest in this article is not with modeling in general, but with a discussion of atomic scale modeling with as it turns out some emphasis on developments in molecular dynamics (MD). Space does not permit mention of all the work of the outstanding group of investigators active in the fields covered by this article. This is especially true of the electrochemical field where this author has worked most recently. For these omissions the author offers a sincere apology.

Computing with personal computers (PCs) has become cheap. Moore's "law" still operates in CPU power and disk drive storage density (chip speed 2 GHz, information storage density $50\ \mathrm{Gb\,in^{-2}}$). Though some fundamental physical limits are in sight (e.g., chip feature sizes around 20 nm, and the old bogy of the supra-paramagnetic limit in magnetic information storage) the electronics industry still seems to find orders of magnitude with which to assail each others products. These terrific advances in power and density, the concomitant reduction in price of PCs and development of a lot of user-friendly software, has made the old fashion kind of 1980s supercomputing affordable. Indeed the combina-

tion of price, hardware advance and user-friendly software has put advanced atomic scale modeling capabilities firmly in the hands of the masses in office and small laboratory environments.

One comment in closing. It is sobering thought to remember that in our simulations, sample sizes (with some notable exceptions) are still much closer to the zepto-mole (ca. 602 molecules) than atto-mole (ca. 602,486 molecules).

1.2. ATOMIC SCALE MODELING METHODOLOGY

There are several categories of tools for performing atomic scale modeling. The first are the readily available computational chemistry and physics codes for calculating electronic structure, for example the chemical bonding in molecules, covalent solids and metals. There are excellent references and source materials available [1]. The second set consists of MD codes for following the dynamic movement of atoms and molecules. There is also a set of Monte Carlo (MC) codes for calculating molecular distributions for systems in thermodynamic equilibrium. There is also an excellent book describing MD and MC simulations by Frankel and Smit [2]. For a comprehensive coverage all the methods needed for materials science calculations see the excellent review in Ohno *et al.* [3].

Since this review is about computer based modeling we will say next to nothing about the classic statistical mechanics methods dating from the time of Boltzmann, and in more modern times associated with the names of Kirkwood, Onsager, Rice and their students. These tools use analytical potentials and the powerful methodology of statistical mechanics methods to calculate experimental properties directly. These methods have been widely used for many years in the analysis of specific problems. In solid–liquid interfaces they include a large body of work connected to the structure of electrolyte solutions and electric double layers. These statistical mechanics methods, though an important part of modern modeling methods will not be discussed any further here. For the reader interested in electrolyte systems we refer the reader to publications by Henderson and co-workers [4–8].

1.2.1. Calculations of electronic structure and chemical bonding

In recent years many electrochemists have taken advantage of the impressive range of electronic structure codes available on workstations and PCs. The software for doing quantum electronic structure (chemical bonding) and geometry optimizations has evolved with the advent of easy to learn and use graphic user interfaces (GUI). Molecule structure builders and optimizers are readily available with the user being able to choose the level of computational sophistication. You do not have to know any Fortran, C^{++} or use a slide rule. Chemical energy accuracy is available for small clusters. Methods like complete neglect of differential overlap (CNDO), which are fast, have always been good for ground state geometries of organic molecules [1]. Gaussian is the method of choice for molecules [9]. Density functional theory methods are now available as a component of almost all the commercial packages and can solve problems for larger systems because of their analytic treatment of key aspects of electronic correlation [3,9,10]. Many university groups have now packaged their research codes making them easier for the novice. Some like Wein 2K and *Crystal* are very good for calculating the electronic properties of metals.

1.2.1.1. INDO calculations *Ab initio* methods can be time-consuming in the study of the electronic and spatial structure of a system containing hundreds of atoms. In order to be able to study the metal dissolution and deposition in aqueous solution, in conjunction with accurate *ab initio* methods, there is a need to do complementary research using, for example, a semi-empirical quantum chemistry method. In some of the studies described later in this chapter, extensive use was made of the INDO (intermediate neglect of differential overlap) [1]. The INDO method, which is a simplified semi-empirical version of the Hartree–Fock formalism, provides a rapid method for exploring PE surfaces and is applicable when the part of the PE surface of most interest is outside the solid (metal). This method has been parameterized for several different systems including oxide, semi-conductor and ferroelectric materials by Shluger *et al.* [11–13]. Though the INDO method is parameter dependent, it has a parameter set which is more or less transferable for a group of chemical constituents. In recent years the INDO method has been used extensively to study defects in bulk crystals and on surfaces of many oxide materials and some semiconductor materials [11–15].

1.2.2. Molecular dynamics and Monte Carlo methods

For those interested in reading a practical account for getting into modeling and simulation there are many excellent texts available. The book by Frenkel and Smit [1] includes a good set of references and includes some history. We continue with a brief comment about the interaction of greatest importance in electrochemistry of solutions, namely electrostatics. In three dimensions sums of interactions between ions $q_1 q_2 / r_{12}$ are shape dependent. Accordingly great care must be exercised when evaluating interactions between point charges and dipoles. The shape of the volume being summed over is dictated by the physics of the system. Several methods in common use have implied shapes that may be at variance with the actual geometry of the system. The most obvious is a sum over a finite radius sphere inside a thicker slab as when the electrolyte system is confined by two parallel planar electrodes. This latter method will give the wrong answer including results that ions of like charge can attract. There several mathematically acceptable methods including: Ewald summation [16]; reaction field method [17]; fast multipole methods [18,19]; particle mesh [20]. Ohno *et al.* [21] have conveniently explained and summarized almost all the methods in use.

1.2.2.1. Classical MD and MC In a MD simulation all the equations of motion for the degrees of freedom of the atoms (or molecules) are solved subject to the boundary conditions on the entire system. Information about the trajectories is stored. For systems in thermodynamic equilibrium the Monte Carlo method can be used because only the potential energy function of the complete system is used to calculate time independent properties like the molecular distribution functions. In practice MC is often combined in a hybrid scheme with MD to improve the efficiency of the random moves of the atoms in the MC part. Typically in MD the simulations are continued for one or two nanoseconds.

Quantum MD calculations are much more time consuming and have mostly been performed on systems containing He atoms at low temperature (4 K) in the liquid state or adsorbed on a surface [22]. Though this is a hot topic in low temperature physics, it has not yet charged up many electrochemists. Accordingly we will not consider this important area of physics any further here. In the case of a few isolated electronically excited molecules in

a medium that is in the ground electronic state (examples are electronically excited solute, or dye molecules adsorbed on a semiconductor electrode) can be modeled without resorting to a quantum treatment of all species present. This is a much simpler problem than liquid He.

In classical MD it is assumed that the interaction potentials acting amongst the atoms are known (usually an analytical formula) and starting from a set of initial coordinates and velocities the entire system is allowed to evolve. The evolution is governed by a set of equations of motion and a specification of a thermostat. For example using Newton's equations simulations of electrochemical phenomena have substantial increased our appreciation for how solvent and ions are distributed and how their motion is correlated amongst themselves and near objects large and small including infinite planar surfaces (electrodes). The main drawbacks are the short time length over which a simulation can be performed and the low quality of the interatomic potentials. Pair potentials and the absence of electric polarization potentials limit the applicability. Improving on pair potentials by including many body corrections is a difficult proposition unless there is some guiding physics (as in metals). Electric polarization in molecular systems has some systematic methodology to follow [23–25]. For example one can add dynamic charge variables to the Lagrangian or introduce atomic polarizability terms and iterate the local electric filed to self-consistency. Interfaces between condensed matter phases are important regions where induced fields are important. It is important to use what every physical measurement is available. For systems with an interface second harmonic generation provides an experimental test because the signal is vanishingly small in regions with a center of symmetry.

Some of the earliest simulations were performed on liquid argon and liquid water. The empirical potentials developed for the rare gases are generally acknowledged as the best available. John Barker who was first to develop accurate potentials for the rare gases also showed that the surface tension of argon could be calculated to within about 10% of the experimental value when the Axilrod–Teller triple dipole many body interaction was included [26]. Calculations of liquid water surface tension over a wide e range of temperatures have been reported by Tildesley *et al.* [27,28]. The results obtained seem remarkably good considering the simplicity of the model and the effort required to calculate a value for the surface tension that is in reasonable agreement with experiment for liquids Ne, A, Kr and Xe.

For water there are about one dozen different empirical models used to simulate the properties of the bulk liquid at room temperature and pressure. These models are used for simulations all the way from crystal hydrates and clathrates to electric double layers surrounding bio-macro-molecules. Ambitious attempts to improve on empirical potentials by fitting to the results of electric fields obtained from *ab initio* quantum chemistry calculations have often not resulted in a patently superior model for the condensed phase properties. Part of the problem is our inability to systematically devise good many body potentials for molecules that are not spherical. The pair potential models optimized for bulk properties like a given dielectric constant value or for diffusion do not include enough flexibility to predict good values of the surface tension except as noted above.

In general in classical MD the potentials used do not allow chemical bond breaking. The electrons are not considered explicitly unless there is a dynamic charge in the Lagrangian or an explicit atomic or molecular electronic polarization potential term. There

are a couple of interesting exceptions like the Stillinger–David potential for ionizable water [29] and the improvements made by Halley and Rustad [30,31]. The classical mechanics method has been applied to studying the statistical properties of numerous molecular systems. None of this work would be possible without the use of computers. Nowadays MD and MC simulations are part of laboratory packages and should be viewed as another tool to help the investigator gain understanding and insight and solve the problem at hand. For each set of potentials there is a hypothetical materials world in which all the nuclei follow the force law based on the input potential. How good the result is depends almost solely on the accuracy of the assumed analytic potential to reproduce natural phenomena.

1.2.2.2. Bond order potentials and "reactive" MD In the 1980s a new set of potentials was devised. For convenience we will call the dynamics using this type of potential "bond order MD". The bond order potentials were originally devised by Tersoff to describe covalently bonded solids like silicon, germanium and diamond carbon [32]. Later a set of potentials specialized to C and H containing materials was developed by Brenner, which included the two-atom bond stretching, three-atom angle bending and four-atom chain torsions [33]. Notable work on another set of potentials has been developed by the Oxford Group of Pettifor [34]. These potentials are superior to Brenner and can account for significant organic chemistry due to a more fundamental approach. Consequently it is able to accommodate aspects due to unpaired electrons (radicals) and resonance stabilization of π bonds.

The basic form of the bond order potential is:

$$E_{\text{bonds}} = \frac{1}{2}\sum E_i = \sum\sum \left[V_{\text{R}}(r_{ij}) - B_{ij}\, V_{\text{A}}(r_{ij})\right],$$

where the labels i and j run over all atoms in the system. The components V_{R} and V_{A} are repulsive and attractive Morse-like potentials multiplied by a cut-off function $f_{ij}(r_{ij})$. These latter functions are chosen to be differentiable and to restrict the potential nearest neighbor atom pairs. This combination of two Morse functions mimics what happens as the bond is stretched to the homolytic breaking point. The bond order function B_{ij} is a nonlinear function of atomic positions. It carries the dependence on well-understood molecular properties like such as: the angles between chemical bonds and the torsional angles along chains of bonds.

Notable recent developments include extending the potentials other atoms (rare gas atoms, fluorine, nitrogen, oxygen, silicon) and adding terms to the potential that extend its range to account for non bonded interactions like van der Waals interactions. The potentials are analytic functions and so the bond order MD describes an imaginary world of hydrocarbon chemistry that through careful parameterization to the physical properties of real molecules reproduces most if not all vibration and bonding characteristics (closed shell all electrons paired, but not radicals or unbonded lone pairs) molecules with sp^3, sp^2 and sp^1 bonding. An extended set of bond order potentials for the atoms C, H, N, O, S and P would open up the possibility of doing MD simulations including reactions in organic and biochemical materials without the need to explicitly include any of the electrons involved. All the reactions would have to occur on the ground electronic state surface. Because the bond order potentials contain no explicit reference to the electronic degrees of freedom the computational cost of performing bond order MD

is much smaller than the cost of explicitly including the electrons involved in the chemistry.

1.2.2.3. Car–Parrinello MD Most of the difficulties experienced in performing MD simulations are due to the use of empirical potentials (especially pair potentials) that only poorly represent the complex many body forces acting amongst atoms. In the condensed phase these forces are intrinsically many bodied because they are dependent on the local environment around each atom. The use of empirical potentials is completely avoided in the third quite possibly the most important molecular dynamics method yet devised. In 1985 a method was proposed by Car and Parrinello that combined DFT and MD [35]. This method introduced a new paradigm. In the Car–Parrinello MD scheme the effective potentials determining the forces acting on individual atomic nuclei are found simultaneously using DFT calculations for the electron density and classical mechanics for the atomic nuclei. Space does not permit any but a very brief description. There is a good account by in the article by Remler and Madden [35] and Ohno *et al.* [3]. The hang up prior to 1985 was the Born–Oppenheimer (BO) viewpoint [16]. The BO scheme was developed for finite molecules with widely separated electronic potential energy surfaces. In the BO method the electronic state wave function is first calculated with the atomic nuclei in fixed positions. So the electronic wave function depends parametrically on nuclear coordinates. The potential energy surface on which the nuclei move is calculated from a sparse linear combination of electronic state functions. For a material with a highly degenerate set of electronic ground states like liquid or solid metals and semiconductors the BO approach cannot work because the electronic spatial structure changes as the atomic nuclei move. The brilliance of the CP method was the simple methodology (electron functions as dynamical parameters), by passing the issues provoked by the BO separation.

The Car–Parrinello MD method has been developed extensively and the code has become widely available. Applications to all types of materials and their combinations have been reported. For electrochemistry these include all the important problems: bulk water, ionic solutions, water on oxide surfaces, and water between metal electrodes [37–41]. The main draw back is the limitation to small system size and short simulation times. The use of the massively expensive hardware referred to earlier alleviates this problem for some groups. Another strategy is to extend the time scale by placing some constraints on the nuclear or electronic motion in order to reduce the computational burden.

Recently our group in Singapore developed the notion of *chemistry in a box MD,* a modification of the Car–Parrinello MD scheme useful in systems where all the electrons involved in the chemistry are localized. All the molecules of the surroundings not involved in the chemistry are treated by classical MD using empirical potentials [42,43]. Though this is a departure from the spirit of the Car–Parrinello MD it does have a pragmatic twist in that simulation times of several tens of picoseconds are reachable on a single processor workstation.

1.2.2.4. Improvements in MD methods Some of the problems besetting MD methods have been mentioned in passing. Larger system size can be achieved by parallel computer systems. In longer time scales can be achieved in some dynamical systems. For example Voter [44,45] has described two schemes that applicable to diffusion of atoms on surfaces that give an order of magnitude extension in time range. In summary the issues are electrostatics, time scales, balanced sets of potentials, and the development of time dependent methods for incorporating localized excited electronic states.

1.3. MECHANISMS AT THE SOLID–LIQUID INTERFACE

1.3.1. Introduction

In this section we conduct a limited survey of some of the basic or fundamental problems in solid–liquid interfaces with a focus on electrochemistry and comments on how theory, modeling and simulation can lead to improved insight and understanding. We start by asking the question: "What is fundamental?" The fundamental structures or mechanisms in any area of science are here defined to be those that impact large conceptual areas. On this basis a prioritized list of basic problems for an area like electrochemistry would read as follows: water > water and ions > electrons > protons > ions at interfaces > catalysis > … etc. This is a prioritized list. Since a lot of interfaces support electrochemistry this list should have a broad relevance. Of course this list is not necessarily one of what is currently most exciting. We often work in areas where progress can be measured and excitement or interest from peer groups is substantial. Water is a case in point. Understanding the properties and dynamics of water is an important endeavor but it currently lacks the attention and excitement found in other fields like for example nanotubes. However we just do not know enough about water or ionic solutions. In calculations the engineers get around this problem by using "steam tables", numerical tabulation of aqueous electrolyte properties. The models used in statistical mechanics calculations are designed for specific scenarios. The hydrogen bond in water and biological systems is poorly understood. The bond is weak and there have been very few experiments able to probe electron distribution and directionality [46]. In the prioritized list, electrons come after water and solutions. Electron transfer processes pervade much of electrochemistry and the description by Marcus theory [47,48] is very widely applied to interpreting experiments. Detailed time dependent understanding requires a quantum theory. For protons there is debate over whether it is ever sufficient to treat their dynamics by classical mechanics provided the potential surface is determined by a quantum electronic calculation.

Processes involving ionic adsorption or dissolution are in the fundamental class because metal dissolution–deposition and oxide formation are basic to major technologies and understanding them better would perhaps resolve some of the complex issues in plating and planarization in computer chip manufacturing. Catalysis is a basic issue to cost efficient chemical production, chemical product range and fuel cell development.

Having made broad comments on the basic or fundamental phenomenological areas of electrochemistry we turn now to a slightly more detailed look at areas in terms of physical, chemical and device related aspects. There are many problems that do not specifically involve electron or proton transfer. These physical problems are ones where considerable understanding already exists as a result of clever experimentation, or analytic model theory or fairly simple computation and simulation. The best known of these are the venerated Debye–Hückel theory of electrolyte solutions [49] and the older Gouy–Chapman theory of electric double layers [50].

1.3.2. Debye–Hückel theory of ion solvation and electric charge screening

Theory and computer simulation have revealed details of the inner structure of the first solvation shell. In particular distortion of the solvent molecule geometry and polarization, especially for high valence, in simple closed shell ions like Al^{3+} and transition metal ions

involved in outer or inner sphere electron transfer reaction as in the Fe^{2+}–Fe^{3+} redox reaction [51]. The Car–Parrinello MD method has been applied to several highly charged cations and details of water geometric distortions examined [37–40]. Also exchange of molecules in solvation shells is an important experimentally measurable quantity using modern NMR. The continuation from this inner zone aspect of hydration into the longer ranged effects that give rise to shielding of the electric field by the dielectric constant of the solvent is poorly understood. In some geometrically well-defined systems like a pair of charged impurities in a crystal this problem has been solved formally by Mahan and Mazo [52]. Over many years the analytical work of Henderson and coworkers [4–8] on solutions is of great important in providing a foundation from which to build more intricate models with properties mentioned above.

1.3.3. Theory of electric double layers and the structure of the inner layer

The Gouy–Chapman theory of electric double layer screening of charged electrodes preceded Debye's historically later work on spherical ions. There are seminal experimental and theoretical papers that provide a wealth of detail (see for example Grahame [53] and Bockris [54]). The main difficulty is the extreme complexity of including metal, ions and water molecules in the same calculation. There are no good empirical potentials that measure up the range of forces involved and no statistical theories or methodologies that can handle the large range of time scales. Another area of difficulty is in whether to use constant surface charge in calculations or a constant electrode potential. Simulations that address this issue are described in a later section. Most theoretical calculations use surfaces with constant charge. However this is not what experimentalists do in the laboratory. They almost always work at a chosen value of the external potential. Constant potential implies fluctuating electronic charge on the metal. This problem can be solved and computed in classical statistical mechanics. However in a quantum mechanical treatment it means that the number of metal electrons is not a constant of the motion. This problem has not been rigorously solved for applications useful to electrochemical simulations. Currently the most sophisticated treatment is in the study of electron currents flowing through nano-wires. There the formulation is in terms of scattering theory methods.

Structure of the inner part of the double layer (Stern layer). There have been very beautiful experimental studies of adsorption of ions and neutrals on metal and semi-conductor electrode surfaces. The application of *in situ* STM and AFM has revolutionized the experimental fields and created whole new subfields. The challenges for the theorist in the face of all this new and very exciting experimental work is immense even some what over whelming: i) metal and semiconductor substrates, their adions and reconstruction, ii) charged and neutral adsorbates, electro-adsorption valence, iii) contact inner layer adsorbed solvent structure, and iv) contact adsorbed hydrated ions.

Classical MD has studied some of these problems. For example an intact double layer consisting of Stern layer and outer layer has been simulated. The system was sufficiently large that the bulk electrolyte region was clearly identifiable [55–57]. Simulations at medium salt concentration (0.1 M) where the entire double layer extends 10 nm or more are challenging because of system size.

1.3.4. Deposition and dissolution of noble metals

The deposition and dissolution of metal on substrates is a subject with a very long history. It is a basic process of a host of electro-chemical phenomena that lie at the heart of many current and emerging technologies. For example it is an important step in microelectronics materials processing, electroplating, electro-crystallization, battery action, corrosion, and the electrochemical restructuring of surfaces using scanning probes [58,59]. This is an area of immense scientific and technological importance. From corrosion protection plating of chrome, to cosmetic plating of jewelry to the low cost electroless plating of micro-electronics circuitry, a better understanding of the basic chemical mechanisms involved in metal deposition and dissolution is an immensely important subject. The continuing work of Kolb [58,59] on the underpotential deposition and metal deposition from probes exemplifies some of the fine experimental work in progress. The recent book by Budevski *et al.* [60] has summarizes the traditional viewpoint quite well. Movement of adions from kink sites onto terraces and then into solution is a computational grand challenge by almost any measure. The solvent stabilizes the cation, but it does not decompose as with alkali and alkaline earth metals for example. However the composition of the supporting electrolyte influences the reaction path as shown in the beautiful experiments by Magusson and Behm in which STM experiments of copper metal dissolution in halide and sulphate media were followed [61].

The dissolution appears to involve the following steps:

- Metals dissolve in presence of halide in the aqueous sub-phase.
- When the electrode is under potentio-static control electrode the surface charge fluctuates as ions and water molecules move to and from the electrode.
- Metal atoms that leave the surface are ad-atoms.
- Ad-atoms leave the surface when they are positively charged.
- Ad-atoms leave the surface when a solvent cage exists ready to accept them.

Accordingly we can think of the process of metal dissolution or deposition as involving three steps: diffusion and aggregation processes on the metal surface; act of metal atom transitioning on or off the surface; dynamical processes occurring in the aqueous electrolyte phase close to metal surface.

The long time interest is in the chemistry of technologically useful metals like: Cu, Ag, Au, Pt, Ni… etc. Especially we are interested in conditions where the potentio-statically controlled deposition of Cu on single crystal Cu metal, or ordered monolayers and bilayers of Cu or Ag deposited at under potentially deposited condition on Au metal [58,59].

The following effects play a role in the metal dissolution/deposition process: i) diffusion of atoms from kink to step edge to terrace site, ii) terrace to bridge to top site, iii) electric charge fluctuation on the surface including that induced by anions approaching close to or even contact adsorbing on the surface, iv) solvent fluctuation at the adatom site, and v) chemical complexation of metal ions by halide or pseudo-halide ions [61].

The primary source of adatoms are terrace edge kink sites [60]. First the atom moves to an edge site, then leaves the edge for the terrace as a diffusing adatom. Finally it “hops” into solution. Little has been discussed about how the electric charge varies as the metal atoms move about on the surface or as they leave (or arrive) the surface from the solution phase. Dissolution occurs faster at larger positive potentials due to lowering of the Fermi level. It is known that over potentials exist so that for metals like Cu on Au the first and

second monolayers deposit at different potentials to the bulk. It is known that dissolution of transition metals requires the presence halide ions. This effect could be due to anion-induced surface or to the formation of specific chemical complexes [61].

1.3.5. Battery

In the last few years there has been considerable activity in the use of quantum calculations of the electronic structure of Li ion battery materials and other rocking chair battery systems. Using modern DFT codes that predict the band structure of metals it has been able to predict the theoretical voltage of battery electrode materials. This gives rise to the possibility of performing survey calculations of new electrode materials in order to speed the search for higher voltage materials [62,65]. There has also been exploratory work on mechanisms of phase transitions as either Li ions are depleted of stored. Previously there has been interest in how and where Li ions are stored in graphite. Clearly then in the area of Li ion battery theory, modeling and simulation has a lot to offer. The theoretical battery potential can be calculated from the computed solid state quantum electronic ground state energy difference with another intercalation system. This could be Li/graphite. In principle theory provides an alternative to laboratory screening of new materials free of sample size, shape and purity issues. The intercalation of metal and other species into graphite has been studied using quantum chemistry cluster models. Issues concerning the electronic structure and possible phase transitions in the cathode materials like Li_xCoO_4, provide unprecedented insights and reinforce the conceptual foundation for materials development.

The motion of Li ions in and out of both electrodes presumably affects the kinetics and ability of the battery system to deliver power. Currently ideas about these processes are somewhat primitive and certainly are not at the same level of sophistication found in electric double layer studies. The existence of a thick heterogeneous layer of products that accumulate on the graphite and Li_xCoO_4 grains during battery assemble and filling or during the first few cycles of operation are also important. There have been several intriguing calculations of the dynamics and structure of Li ions in polymer electrolyte systems [66].

1.3.6. Mineral surfaces and pre-biotic chemistry

There are a multitude of scenarios for how life may have got started about four giga years ago (Gya). These scenarios have involved ocean, atmosphere, ocean and atmosphere, multiple origins in time, a start in two-dimensions on or inside various mineral materials, and even seeding from an extraterrestrial sources. The hypotheses involving the oceans have been the strongest favorites. Now there are several new serious contenders that postulate that life got a jumpstart on mineral surfaces. Several groups: Wächtershäuser [67–69], Corliss [70], Russell, Cairns-Smith and Hall [71–73], have elaborated on how life started on mineral surfaces. The hypothesis that is best developed in terms internal connectivity from surface adsorbates all the way to metabolizing systems is described in the series of papers published by Wächtershäuser [67–69]. As is to be expected when new and audacious ideas are proposed, there has been plenty of criticism from the supporters of the ocean-soup theories. The interest here from the perspective of this review, is that 4 Gya the mineral surface had a pivotal role as reactant reservoir, general substrate, enthalpy source, catalyst and product repository. It did not matter if some processes were poorly catalysed, because at the time this was the only show in town.

The hypothesis of Wächtershäuser is as follows. Early life was chemautotrophic not heterotrophic. Organic molecules were built by a series of reactions. The beginning could be a reaction involving CO_2, H_2S and FeS to simple organics on surfaces with the mineral iron pyrite FeS_2 being one of reaction products. Enthalpic drivers like pyrite formation lead to more complex organics that organised into surface metabolic systems drawing feed stock from a environment streaming past the surface localized metabolizer. From here it could be a 'short' step through a set of anabolic pathways to a cycle with all the attributes of a reverse Krebs cycle. This new system would synthesis complex biochemical molecules like membrane molecules and coenzymes that would work with the catalytic centres on the mineral. Eventually catalysis on surfaces would be superseded by catalysis on proteins surfaces. Fine tuning the peptide bond would necessitate the development of precursors to RNA. A more complete description can be found in the papers by Wächtershäuser [67–69]. There are however numerous significant problems and challenges. The proposed enthalpy source has to be qualified and alternative delineated, reaction partners identified and reaction determined. Finally there has to be experimental verification. Some of the current problems are for example: whether suitable ionic surfaces are stable on pyrite FeS_2, the possibly marginally low value of the mineral formation enthalpy in relation to the free energies of formation of simple organics [74–77]. In addition conditions on the early Earthly were extreme, knowledge of the effect on the reactions at the solid–fluid interface of high flow rate, high temperature and high pressure on all the proposed processes has to be understood. We refer the reader to more comments and discussion in a later section. In closing we point out that this is a fabulous multi-disciplinary solid–liquid interface area for new experimental and theoretical work.

1.4. TOPICS IN SOLID–LIQUID INTERFACE MODELING

In this section we discuss the results of a variety of methods of exploring interfaces. The example all have a strong electrochemical flavor. They involve semi-empirical quantum chemistry cluster calculations, classical MD, modified Car–Parrinello MD and 'simple' stick and ball modeling. All the examples considered are taken from most published and little unpublished work of the author and his collaborators.

1.4.1. Metal dissolution

The semi-empirical quantum chemistry cluster method INDO described in a previous section has been used to study how an atom on a metal surface converts into an ion in a solvent cage close to the surface. In these calculations the metal is represented as a finite cluster with a cage of fixed solvent molecules above the (001) surface. An atom is moved from the surface into the cage and the total electronic potential energy of the system is calculated with the atom position as a parameter.

The following represents an approximate summary of what is relevant to the subject as well as comments on what should be studied theoretically.

1. Metal surfaces are covered with a weakly chemisorbed water layer that may be H-bonded in an ice like structure involving a second layer of water. Geometric compatibility of these layers with a low index surface which also has the role of template

as in epitaxial grown systems is likely to be the key effect. Surface planes like (111) may be better than (001) surfaces in this regard provided the surface cell lattice parameters are in the right size range. Therefore it is important to understand not only the surface bonding of one molecule but also that of small water clusters. The first calculation is well posed because all quantum chemistry calculations show that the top site has the strongest binding. These calculations also indicate that water is tilted so that the angle between the surface normal and molecular plane is about 70 degrees. In practice we need to determine the well depth and geometry for at least one top site adsorbed water molecule. In principle it is also important to examine whole monolayers of adsorbed water because there is a possibility of a collective interaction involving partial charge transfer between water layer and metal surface that could change the surface electronic density of states.

2. The binding curve for removal of one atom either from a surface plane site or from an adatom site is important for determining the relative stability of these two very different types of site.
3. The binding curve for removal of one atom in the presence of a water cage is important in order to determine the effect of solvent in stabilizing the charged ion over the neutral atom. The size, shape and number of water molecules in the cluster comprising the cage have importance. Some possible water cages of different shape and size are: tetrahedron, four square planar pyramid, and square bi-pyramid. Typically one water molecule can provide up to 0.5 eV of stabilization to an ion over the corresponding neutral atom.
4. Important additional stabilization is provided by a halide ion located in the vicinity of the departing ion. This can be studied in terms of the binding curve for removal of one atom as an ion in a water cage and a polarizing ion pair with anion above the separated atom is important to determine the effect of halide in localizing positive surface around the separating atom and assist its transition to a cation.

In a separate study large clusters with 100 atoms or more were examined in order to build a more consistent model of dissolution/deposition. The largest clusters with 101 atoms had features corresponding to the following sites: kink, step edge, hollow, bridge and top site. For simplicity the metal atoms were modeled using the pseudo-potential for Li, which is a soft metal with small work function. This is a great convenience in order to perform fast calculations on large clusters. The interest was not in quantitative accuracy or in the chemical reactivity between this metal and water. The focus was on qualitative features of a model that define or control the path followed by ions arriving or departing the surface. The cluster consisted of a given number of body centered unit cells. All the atoms in the cluster except one were kept fixed at lattice positions corresponding to the 3D crystal. To trace out the PE surface for dissolution, calculations were performed in which a metal atom (ion) was moved along a cluster symmetry axis perpendicular to a low index plane like (100). Fitting them to the results of ab initio calculations and available experimental data optimized INDO parameters for Li, O, and H. Explicit information can be found in several research publications [78–80].

1.4.1.1. Atom from the surface (100) plane: tetrahedral solvent cage We performed comparative calculations for the binding energy of Li atom to 35 atom Li cluster located in the middle of the Li (100) surface. In order to estimate Li atom binding energies numerically we first pulled out middle Li atom from (100) surface in [100] direction. INDO calcula-

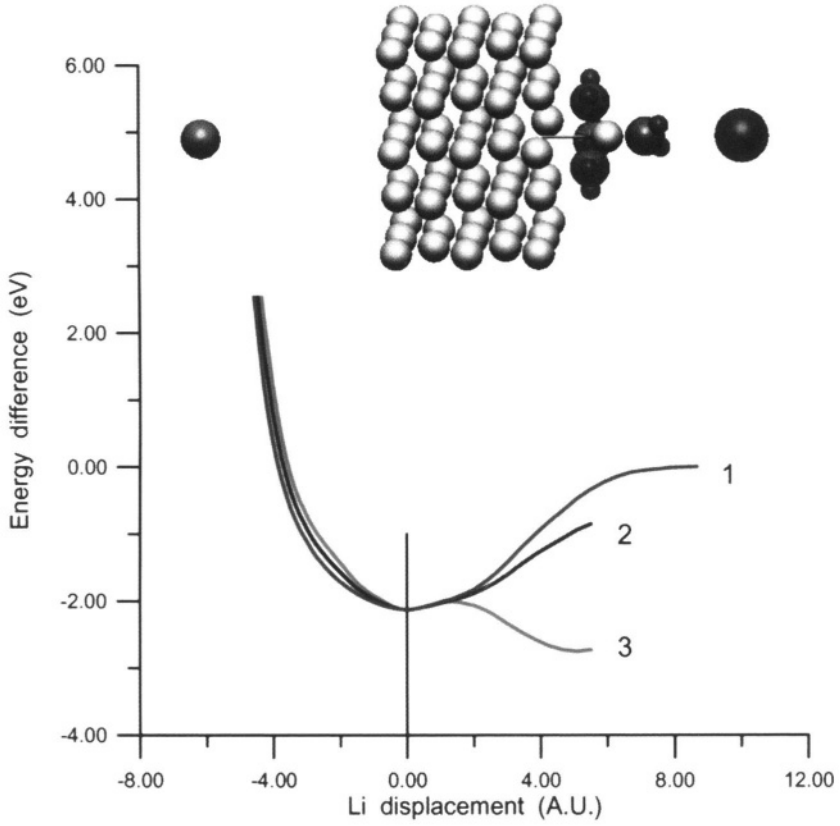


FIGURE 1.2. The electronic potential energy curves calculated using the INDO method for a cluster M_{35} of thirty five metal atoms. In each case a metal atom from the central top surface plane is moved along the z axis normal. Curve 1 is for cluster in vacuum when an atom is removed. Curve 2 is for the case of a tetrahedral cage of four water molecules located above the surface plane. The difference from curve 1 is due to water molecule stabilization. Curve 3 is the potential energy curve when a chloride ion Cl^- is located on the z axis above the central atom site at a distance of 15 au and a Na^+ ion is located below the cluster on the corresponding negative z axis at –5 au.

tions reveal Li atom binding energy for this case to be approximately 2.13 eV. According the calculations, the binding energy for corner Li atom pulled out in [111] direction from the same 35 Li atom cluster into vacuum is approximately 1.17 eV. These results are consistent with according our results Li surface atom are more strongly bound to the cluster in [100] direction as compared to the Li 35 atom cluster corner atom in [111] direction.

Figure 1.2 shows the PE curves for M35 cluster. For each case the central M atom is removed along z. Curve 1 is for cluster in vacuum where an atom is removed. Curve 2 is for case where a tetrahedral water cage is located above the lattice atom site, and the shift is due to water stabilization. Curve 3 is the PE curve when a chloride ion Cl^- is located on the z-axis above the central atom site at a distance of 0.7 nm (15.4 au). Simultaneously a Na^+ ion is located below the cluster on the negative portion of the z-axis. Next a cage of four water molecules at the corners of a tetrahedron was placed face down (three waters parallel

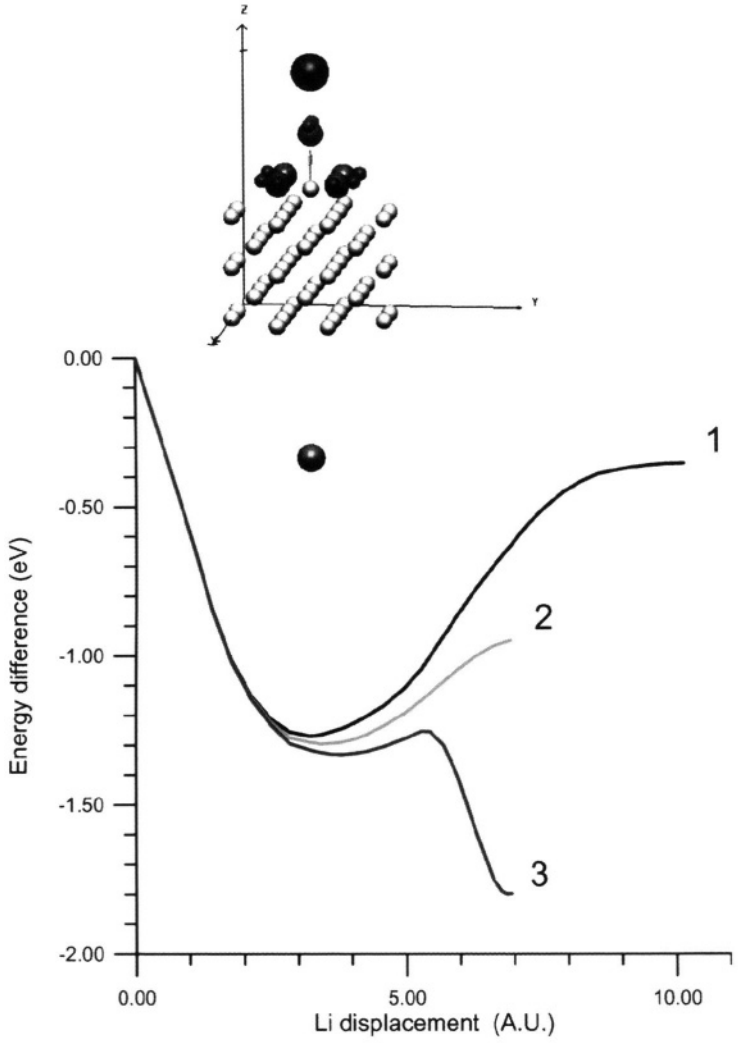


FIGURE 1.3. The electronic potential energy curves calculated using the INDO method for a cluster M_{63} of sixty three metal atoms. In this case the surface atom starts at an adatom 4H site on the cluster (001) surface. The solvent cage consists of five water molecules located at the apices of a square pyramid. The three curves are for the same three cases described in Figure 1.2.

to (001)) and one apical water on the axis furthest away from the surface. The PE curve was calculated for the same range of distance forms the surface plane. This is a good test of how water molecules can stabilize a departing species as an ion instead of an atom. Finally a chloride ion is placed above the surface and a sodium ion placed behind the cluster, both ions on the symmetry axis. The chloride ion simulates the presence of halide ion near the surface and the sodium ion insures charge neutrality in the entire system.

Figure 1.3 shows three PE curves and the arrangement of atoms metal cluster, water cage and sodium and chloride ions on the four-fold axis on either side of the water metal-

cluster. In reality the sodium ion is at z_{Na^+} at about –0.7 nm (actually –15.40 au) outside the frame of the figure. It was moved closer for convenience of making a smaller graphics picture. The chloride ion was located at about $z_{Cl^-} = 0.7$ nm (actually 15.40 au) above the (100) surface of Li cluster. INDO results reveal binding energy of Li atom to the 63-atom cluster in the vacuum equal to 2.13 eV. The 1.3 eV high-energy barrier hinders Li cluster dissolution in water, while our calculations predict the possibility to dissolve Li cluster in water in the presence of electrolyte Na^+ and Cl^- ions. In addition we also calculated the transfer and distribution of charge on Li and water molecules of the cage. These results showed how polarization charge flowed within the solvated ion as it moves away from the surface. There was some indication of charge transfer to the water molecules of the cage.

1.4.2. Metal clusters in aqueous solution

In the condensed phase interatomic forces are intrinsically many bodied because they depend on the environment around each atom. The use of empirical potentials is avoided completely in a brilliantly straightforward approach that that combines density functional theory (DFT) with MD in the Car–Parrinello approach [35]. The effective potentials determining the forces acting on individual atomic nuclei are found simultaneously using DFT for the electron density and classical mechanics for the atomic nuclei. Applications to all types of materials and their combinations have been reported. However there is a considerable cost in computer time in this methodology. Typically only a few picoseconds to a few tens of picoseconds of MD simulation time can be followed. Fortunately a lot of essential science can begin and end in this time frame. In electrochemistry applications these include: bulk water, ionic solutions, water on oxide surfaces, and water between metal electrodes. Because the electronic part of the simulation is very compute time intensive the calculations have been performed only on systems with small numbers of electrons for short times of the order of a pico-second.

Our strategy to extend the time scale in Car–Parrinello MD places some constraints on the nuclear or electronic motion to reduce the computational time. We call this *chemistry in a box MD,* a modification of the Car–Parrinello MD scheme useful when all the electrons involved in the chemistry are localized, and when there are molecules in the surroundings not involved in the chemistry that can be treated using empirical potentials [42,43]. This is a departure from the spirit of CP MD, but it does have the practical consequence that simulation times of tens of pico seconds are accessible on a single processor workstation.

In this paper a cube shaped cluster of thirty-five one-electron atoms was surrounded by aqueous electrolyte (NaCl) in the SPC/E model [81]. For computational simplicity the atoms were taken to be Li atoms (sp-bonded metal). The dynamics of one cluster surface ion was considered with all other cluster ions at fixed sites. Non-local norm conserving Bachelet–Hamann–Schluter (BHS) pseudo-potential approximations were used [82]. The water molecules and ions comprising the electrolyte that surrounds the metal, are treated as objects obeying the laws of classical mechanics. As it was demonstrated on this model the developed simulation approach enables to extent simulation times to large time scale keeping fairly good accuracy in representation of metal electronic structure. With this approach it is already possible to collect reliable thermodynamic data, and study such relatively large time scale events as formation of hydration shell, etc. Not all issues of dissolution and deposition can be addressed by the present scheme.

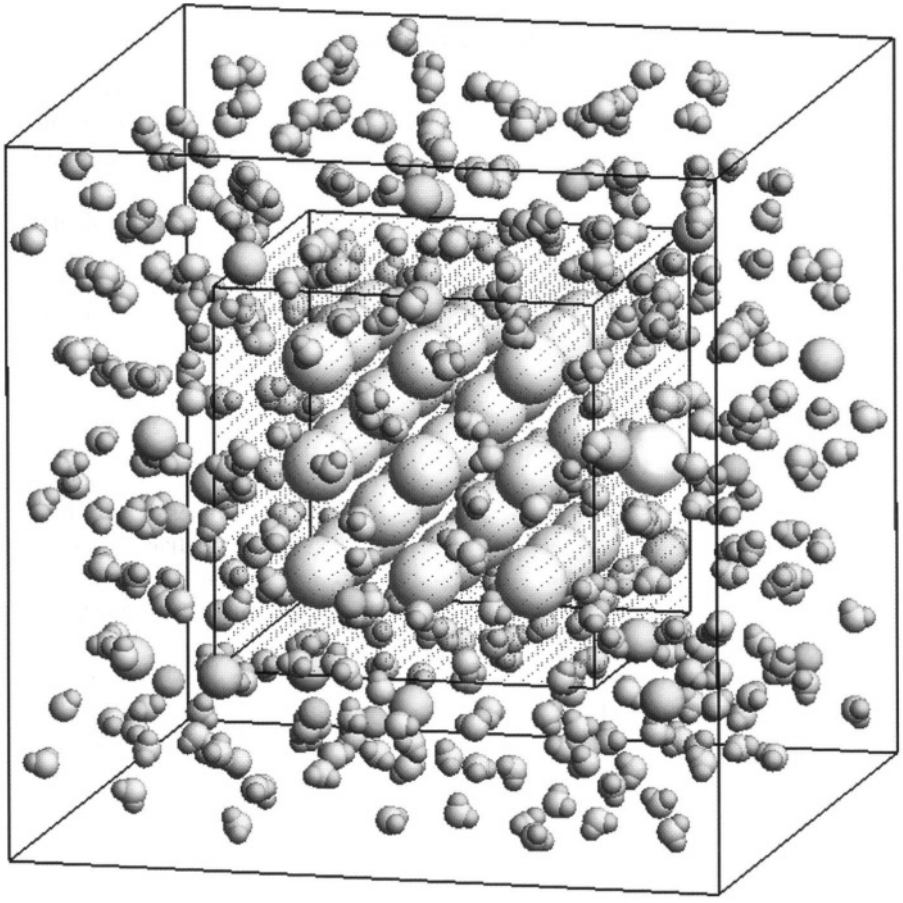


FIGURE 1.4. Snap shot configuration of a modified Car–Parrinello molecular dynamics simulation of a small metal cluster surrounded by aqueous electrolyte. The small wire frame box depicts the electronic box containing a 35 atom metal cluster. The large wire frame defines the molecular dynamics box containing aqueous electrolyte surrounding the cluster. In the simulation the electronic structure of the cluster is computed as the dynamics simulation of the entire system evolves.

Figure 1.4 shows the dynamical system. There are two "boxes". All the electrons of the metal cluster are confined to a cubical box of edge length L_{el} ("electronic box"). The electronic wave functions set zero on the box and outside of this box by expanding its into plane wave basis which satisfies zero boundary conditions on the box walls. Introducing of zero boundary conditions is reasonable because the charge density goes to zero outside the cluster sharply. There is no periodic boundary condition imposed on the system. In general in our calculation scheme to achieve good accuracy a size of a box can be chosen to only ensure that walls fall in the region where valence charge density is close to zero. The entire system was enclosed in a larger cubical box with edge length L_{MD} with repulsive walls (MD box). We followed Car–Parinello scheme in LD and BHS pseudo-potential approximations to simulate dynamics of electronic structure of cluster. To extend the MD

simulation to larger time scales the Car–Parinello dynamics for valence electronic functions was performed in the basis of lowest energy Kohn–Sham orbitals with some specific predetermined number in the set. These Kohn–Sham orbitals were calculated for a reference geometry chosen to be the one of an ideal cluster in vacuum in the special plane wave basis.

The equations of motion of the system were integrated in multi-scale MD fashion. The equations for motion of the electrolyte were integrated with a time step around 1.5 fs, while the equation for charge density and metal ion motion integrated with a time step 0.5 fs. The constant temperature 298 K was maintained using an Anderson thermostat [83,84].

We were able to calculate the potential of mean force (PMF) curve for the metal, which governs the dissolution event. The other part of our simulation was concerned with the structure of solvation of the metal ion during the "dissolution" event. The development of the solvation shell structure was studied in this part of simulation. The formation of solvation shell naturally correlates with shape of the PMF curve. The solvation shell forms as a result of the interaction of the water with both ionic core of the cluster and the charge density of the cluster. These calculations required the collection of statistical data for periods of several tens ps. This was the clear advantage of our technique.

Figure 1.4 shows a metal cluster immersed in a 2 M salt solution. The snap shot of the cluster and molecules illustrates two features of the method. A small frame box around the cluster confines the DFT electronic calculation and the larger frame box the confines the MD calculation.

Figure 1.5 shows on the plot of the radial distribution function $g(r)$ of water around the ion and the solvation number function $n(r)$ with a plateau at $n(r)$ equal to four. The radial distribution function $g(r)$ and the hydration number $n(r)$, the integral of $g(r)$, were calculated for the ion fixed at d = 0.0, 4.0, and 8.0 au. These separations correspond respectively to the cases of the ion in the equilibrium position in the cluster, in the transition state to physically absorbed state, and almost in the bulk solute. These radial distribution functions have a large error bar because of a limited time of data collection, the geometry of the system, and the presence of freely moving solute ions. However, the first peak of the radial distribution function we were mainly interested usually formed in first 5 ps of simulation time. When ion was in the equilibrium position in the cluster first hydration shell was formed from one molecule. This water molecule spent most of the time on top of the ion. The first change in the slope of free energy curve occurred at 2.0 au and this corresponds to the Lennard–Jones interaction between the Li^+ ion and the water molecule on top. The transition to a fully hydrated ion occurs when there are four water molecules in the first hydration shell. The average coordination number for Li^+ ion in bulk SPC/E water is 4.1 [85]. Our preliminary simulations showed that the first hydration shell becomes actually completed (running integration number changed from three to four molecules) in the vicinity of the transition state. The first solvation shell progressively added water molecule from one at d = 0.0 au to 4–5 at d = 8 au. The metal ion then appeared fully hydrated because the second hydration shell had a well-defined second peak at 7.5 au in the radial distribution function. When the Li^+ ion was in the bulk solvent region (d = 8 au) the first solvation shell comprised of five, occasionally of four molecules. In Figure 1.5 (bottom panel at 8.0 au) showed a second peak in the radial distribution function also consistent with bulk solution environment.

The proposed approach to simulate a changing in the electronic structure of metals has some promise. It could be easily extended to the periodic systems like infinite metal slabs in

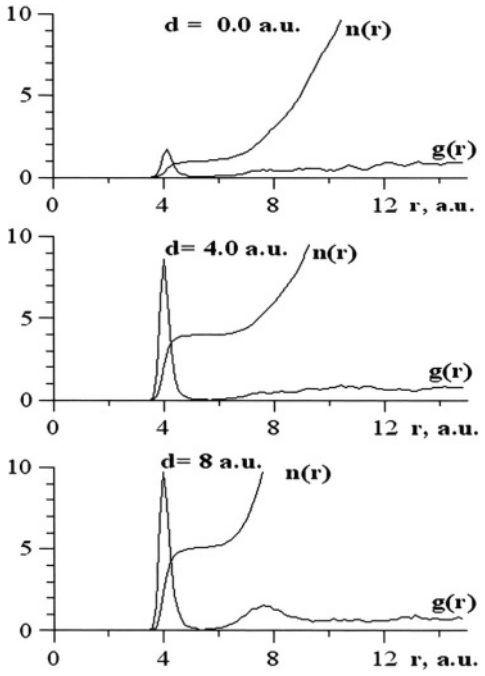


FIGURE 1.5. Evolution of the hydration shell of the metal ion with distance d from surface. Radial distribution function $g(r)$ and corresponding running integration number $n(r)$ for Li^+ ion separated by a distance d from the cluster surface plane.

contact with an electrolyte. We see another possible area of applications of developed numerical technique that goes beyond metal–electrolyte interface. It could be applied to study dissociation of single molecules like alkali halides in the water solute when a changing in the electronic structure of system before and after formation of ionic pair is not crucial and seems to be well approximated by proposed above formalism. In addition we also have performed the MD simulation for adatom placed on the surface of cluster of cubic shape formed from 91 atoms. The adatom is more weakly bound than the surface atom and since it protrudes into the electrolyte has a higher degree of solvation.

1.4.3. Molecular dynamics simulations of electric double layers

For systems where the independent units (atoms or molecules) only interact weakly it is possible to represent the potential as a sum of atomic pair potentials, for example the Lennard–Jones 12-6 potential. When the form of the potential energy function is known, the problem reduces to solving the set of Newton equations. There is one equation for each degree of freedom, so in a typical simulation with thousands N of atoms there will be $3N$ equations of motion to solve. This huge set of coupled equations is solved by numerical integration using very small time steps, typically 0.5 to 2 fs ($1\ \text{fs} = 10^{-15}$ seconds). The force acting on any atom is found by differentiating the potential energy function with respect to the coordinate of the atom.

1.4.3.1. Aqueous electrolyte film between metal plates: constant field vs. constant charge
There has been a lot of interest in studying the structure and dynamics of the electric double layers in their thermodynamically stable regime. Pioneers in the MD simulation field have been Heinzinger and Spohr [86–91], Guidelli [92], Halley [93], Sprik [94,95], Berowitz [96] and Benjamin [97]. Glosli and Philpott [55–57] have studied electrochemical systems at constant N, V, T and either constant electric potential or constant surface charge (you cannot have both) in order to model electric double layers of 1:1 electrolytes between charged metal plates. These simulations were carried out with velocity scaling to keep the temperature constant. The integration time step was 2 fs. All electric fields were calculated without cut-offs by the fast multipole method of Greengard and Rokhlin [3,18,19]. These simulations show that many of the experimentally known features of the double layer can be modeled qualitatively. For example: compact layer, diffuse layer, surface-oriented water layers and in large enough systems at high concentration, an identifiable bulk zone. The electric field and potential across the slab deviate considerably from Gouy–Chapman predictions due to finite size of atoms, water layering and electrostatic quadrupole effects. In these calculations we used the SPC/E water model, the simulation cell was cubic with coordinate origin at the center of the cell, and the z axis perpendicular to the metal plates. Snapshot configurations were stored every few ps ($1\ \text{ps} = 10^{-12}$ seconds) for a total duration of about one ns ($1\ \text{ns} = 10^{-9}$ seconds). This data was used to calculate atom and charge density profiles across the film along the z axis. The electric charge density in the system $\rho(z)$ was found by binning the configurations in bins typically $L/400$ nm wide. The so-called 9-3 potential was used to represent the Pauli repulsion and attractive dispersive interaction of molecules and ions with the walls. The interaction of conduction electrons was modeled by a separate image potential for every charge in the system. Since the metal walls were close together there are image multiple reflection terms in the electrostatic potential, which were included by an extension of the Greengard finite multipole algorithm to systems between perfect conductors. Figure 1.6 shows a typical snap shot of a simulation where the metal walls have atomic corrugation. It shows water molecules and ions between two (001) metal surfaces. The snap shot configuration is shown without perspective. Note how water forms distinct layers near the metal. The inclusion of atomic corrugation is not discussed in this paper. In this note we focus on the variation in electric potential across the system. The potential used in this calculation was developed by Zhu and Philpott [98].

To calculate the electric field across the system the electric charge density $\rho(z)$ was integrated once. The electric potential was found by integrating a second time ($\times -1$). Figure 1.7 shows potential profiles across the liquid film. In these calculations we were interested primarily in the electric field and potential. Therefore the surface was modeled as flat and featureless unlike Figure 1.6. The 9-3 atom-surface wall parameters describing interaction with non-conduction electrons were chosen to be the same as that used by Lee *et al.* [99], $A = 17.447 \times 10^{-6}\ \text{kJ nm}^9\ \text{mol}^{-1}$ and $B = 76.144 \times 10^{-3}\ \text{kJ nm}^3\ \text{mol}^{-1}$ for the oxygen atom. The A and B parameters for hydrogen were set to 0. The potential corresponding to these parameters describes a graphite-like surface. The edge length $L = 2.495$ nm, with metal surfaces at ± 1.247 nm. We use the SPC/E water model: a fractional charge located at each atomic position, $q_H = 0.4238$ e, $\angle\text{HOH} = 109.5°$, OH bond length 0.1 nm, and a Lennard–Jones sphere (centered on oxygen) with radius $\sigma = 0.317$ nm and well depth $\epsilon = 0.650\ \text{kJ mol}^{-1}$. In the absence of water each plate was given a charge

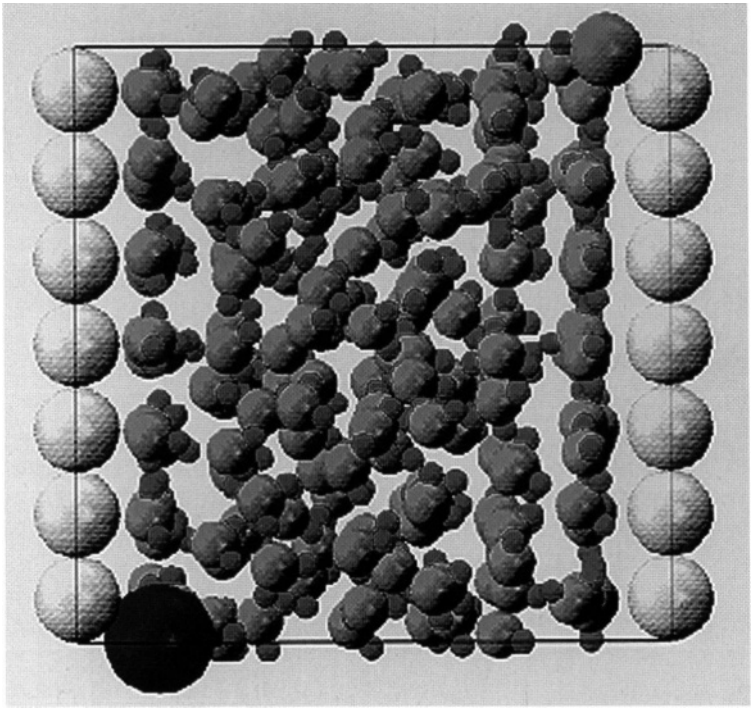


FIGURE 1.6. Configuration of a classical mechanics molecular dynamics simulation of water molecules, Li^+ and I^- ions between corrugated Pt (001) charged metal surfaces. The Li^+ and I^- are adsorbed on oppositely charged surfaces. The picture of the configuration is shown without perspective.

of 1.0 e per plate, corresponding to a charge density of 0.0257 $C\,m^{-2}$. In Figure 1.7 the electric potential is shown for 3 M NaCl solution (top) and pure water (bottom). The oscillations in potential near the surfaces were due to layering of water molecules similar to that visible at the electrode edges in Figure 1.6. Layering has been detected in grazing incidence X-ray measurements of silver electrodes in contact with NaF solution [100].

In the case of zero charge (top step) the drop of potential from edge to film center is due primarily to the static quadrupole moment of the water molecule. Wilson *et al.* [101, 102] have pointed out, the quadrupole moment makes a large contribution. For ionic solutions the charge on the walls is completely shielded by ions so that a electric double layer formed. In this case the staircase in the top of Figure 1.7 has flat steps. In contrast in water without ions there is a residual field across the water film, and consequently the stairs are slanted towards positive z in the lower part of the figure. The slope gives the electric field. Extrapolating the central linear part of the potential curve to the wall edges gives the surface excess potential. It is possible to estimate this quantity from experimental electric double layer differential capacitance measurements.

When both surfaces are metals, the simulation is performed under the condition of a constant electric potential difference between the plates. The charge on the metal surfaces is not constrained, but fluctuates about an equilibrium value determined by the ensem-

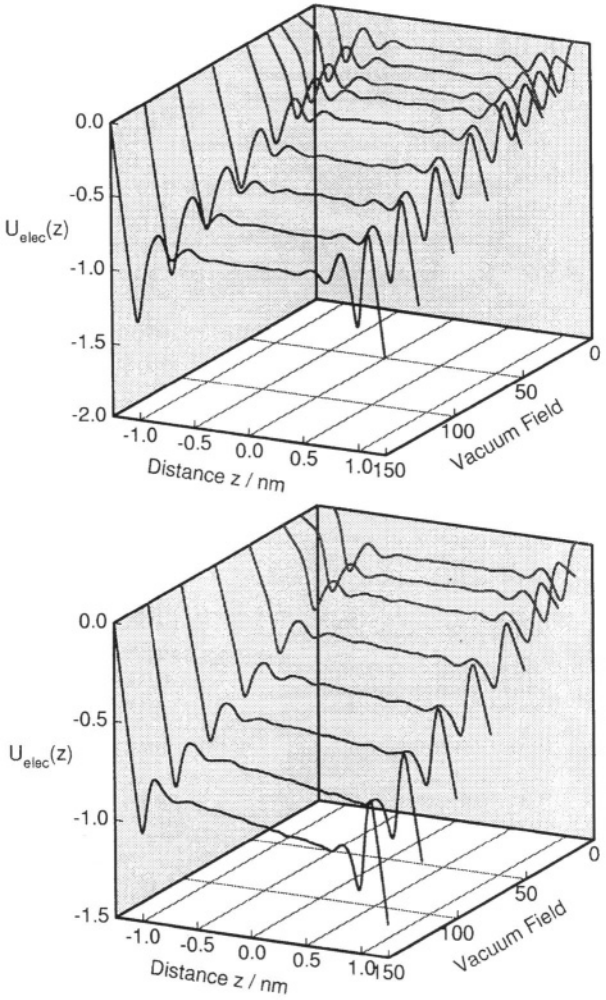


FIGURE 1.7. Three dimensional plot of the electric potential profiles across an electrolyte film between charged flat parallel conducting plates for different vacuum electric fields or vacuum surface electric charge density. Top picture 500 model SPC/E water molecules with 3 M NaCl. Bottom picture is water without any ions. Image charge planes at $z = -1.25$ nm and $z = 1.25$ nm. The quantity on the axis pointing toward the observer is the vacuum electric field which is linearly proportional to the surface charge density present on the metal electrodes in the absence of electrolyte.

ble average of the screening charge in the water layer. The charge and its fluctuation are most easily understood in terms of the electrostatic images of the water molecules near the metal. If the walls were two dielectrics the charge would be fixed and the electric potential difference across the system would fluctuate. As stated previously calculation of the molecular dynamics forces is performed without truncation of long range electrostatic interactions, and the complete microscopic charge distributions are used to obtain the time independent ensemble average electric field and potential profiles of the systems.

1.4.4. Models for mineral surfaces and pre-biotic chemistry

Microbial life on planet Earth started in aqueous electrolytes about 4.2 to 3.8 billion years ago, possibly at high temperature and pressure deep inside hydrothermal vents at spreading centers on the sea floor. Cairns-Smith, Hall and Russell [72] commented: "The gap between reactions on mineral surfaces and chemical processes occurring in the simplest living organisms is enormous." In the last twenty years there have been several proposals that the origins of terrestrial life were reactions on mineral surfaces and that the conditions were similar to that found in or associated with hydrothermal seabed vents. Conditions deep inside these vents can be extreme with temperatures up to 400°C in the fracture zone near the magma chamber and pressures up to 2 kbar.

Over ten years ago Wächterhäuser [67–69] in a detailed hypothesis called the iron–sulphur world, proposed that the origin of life on Earth began on the surface of iron pyrites FeS_2. In this section we focus on FeS_2 minerals and consider stick and ball models of the surface in order to look for possible catalytic sites on the bulk terminated (001) surface. This is part of a larger program in which *ab initio* quantum electronic structure and geometry calculations and molecular dynamics simulations are being used to explore structure-property-function relationships in extreme environments. In our approach they constitute a corner stone of a strategy for postulating pre-biotic era reaction partners and mechanisms and calculating rates of reaction in mineral systems with adsorbed organic molecules.

1.4.4.1. Mineral surface crystallography To this end we have constructed a family of stick and ball models to display increasingly more complicated features of the structure of low index crystal faces. Only sites on the (001) surface of bulk terminated iron pyrite FeS_2 crystal will be considered in any detail in what follows. We have constructed extended models of all the low index crystal faces. These models include flat perfect surfaces, simple terraced surfaces, and terraces with the simplest imaginable kink sites. The purpose of this effort is to look at a range of hypothetical sites where atomic, ionic and molecular adsorption might be preferred and reactions catalyzed. Recently Rosso, Becker and Hochella [77] reported STM images of UHV fractured (001) pyrite surfaces and noted evidence of step edges and kink sites. Simple models convey the possible range of adsorption and catalytic sites and pathways available on these mineral surfaces.

1.4.4.2. Models of (001) surfaces Pyrite cell constant is 0.54281 nm. Perfect bulk terminated crystals have an electrically neutral (001) face. Figure 1.8 shows a schematic of the (001) surface. An important feature of the (001) surface is the array of dangling bonds from each surface Fe^{2+} and S_2^{2-} groups. Half the S atoms on the (001) surface have the same bonds as in the bulk of the crystal and the other half protrude with a dangling bond directed outwards. In the surface each Fe atom is connected to five S_2^{2-} groups, four of which lie in the surface. Surface electronic state neutrality requires electron donation from Fe dangling bonds to S dangling bonds so that the Fe bonds become an empty conduction band and the S bonds become a filled surface valence electronic band. What is important on the surface is the possibility that the Fe sites can act cooperatively in a catalytic event. One Fe might be part of the same surface or part of a growth or defect feature. Here we comment only on the (001) terrace. There are two types of five atom rings. On one ring there is one dangling S atom bond and on the other there are two S dangling bonds. The opportunities for divergent catalysis (in a variety electrolytes) compared say to the active sites of the Fe-S containing proteins are obvious.

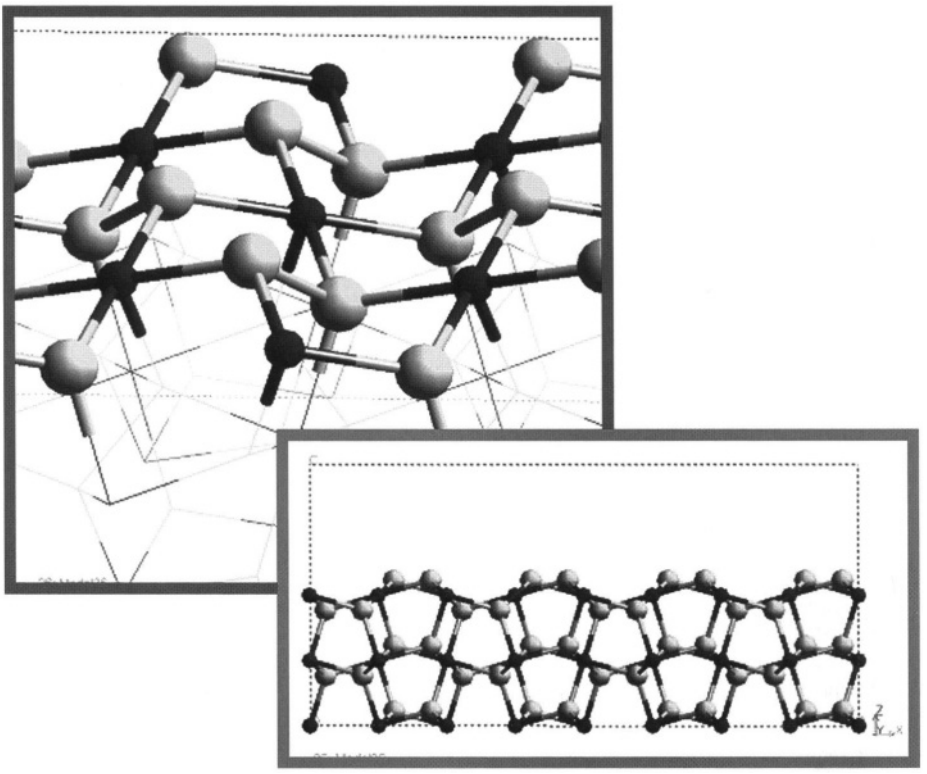


FIGURE 1.8. Diagram of the (001) surface of pyrite FeS_2 showing the two types of five atom ring structures on the surface. Brown Fe and yellow S atoms. One ring has two raised sulphurs each with a dangling bonds and the other only one raise sulphur with a dangling bond.

Figure 1.9 shows a larger fragment of the (001) surface with a "symbolic" representation of H_2O (H_2S) coordinated through O (or S) to the Fe atom. Bonds to surface and subsurface atoms not belonging to the fragment are again all shown symbolically as "H" atoms in order to display the directionality of the covalent bonding network of the bulk FeS_2 crystal. Note that a dangling bond on a next neighbor S points towards the Fe dangling bond. These next neighbor site pairs form a crenellation network across the (001) surface. Locally because of this surface corrugation, adsorption of small neutral or ionic ligands like H_2O, H_2S, OH^- or SH^- could be favored by H bonding to the second neighbor S atom as shown in the figure.

Figure 1.10 shows two types of single atom step edges and kink sites of the FeS_2 crystal in cross section parallel to the (010) plane. Only atoms composing the kink site are shown as balls. The rest of the crystal is shown as stick bonds. The corrugation of terrace surfaces is clearly evident. Each step edge atom has two dangling bonds. At the top step edge the tilt of the dangling bonds is "down" whereas at the second terrace edge the tilt is "up". Catalysis would be higher because of the higher density of dangling bond along the edge. These sites would be attractive to bidentated ligands like bicarbonate HCO_3^- and ionized carboxylic acids.

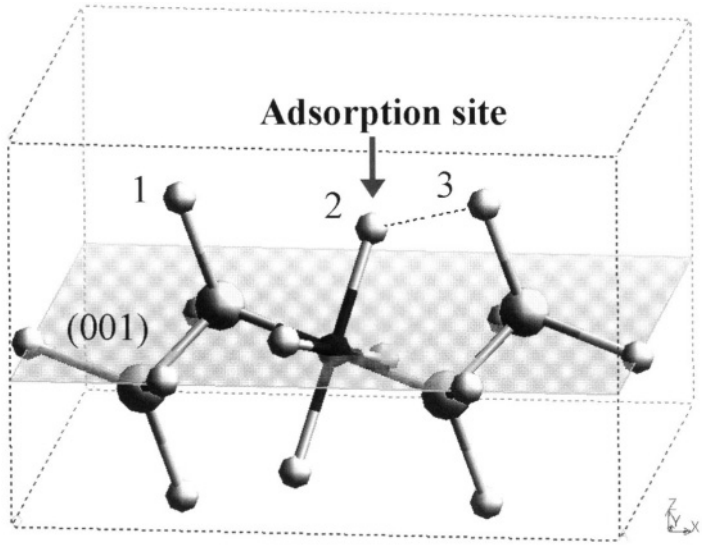


FIGURE 1.9. A fragment of the (001) surface with a symbolic representation of a triatomic water-like adsorbate (e.g., H_2O or H_2S) coordinated through O (or S) to the Fe atom. Bonds to surface and subsurface atoms not belonging to the fragment are shown symbolically as H atoms in order to display the directionality of the covalent bonding network of within the bulk FeS_2 crystal. Note that a dangling bond on a next neighbor S points towards the Fe dangling bond. These next neighbor pairs of sites form a crenellation network across the (001) surface.

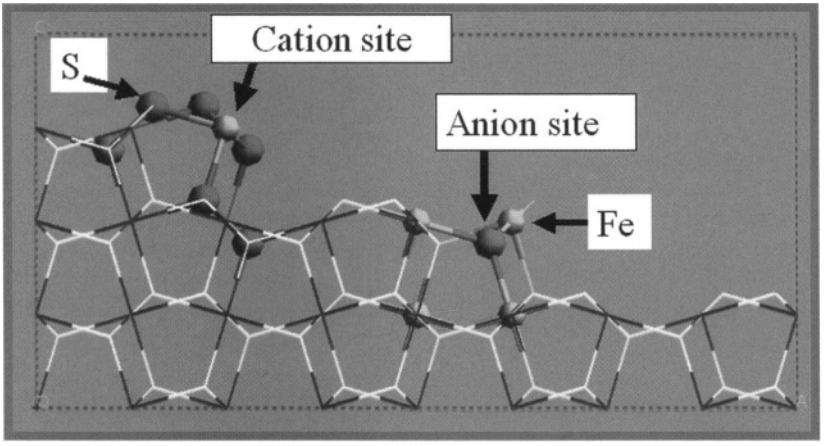


FIGURE 1.10. Two types of atomic step edge of the pyrite FeS_2 crystal displayed in cross section parallel to the (010) plane. The corrugation of the terrace surface separating the steps is clearly evident. At each step edge the Fe atoms have two dangling bonds. At the top step edge the tilt of the dangling bonds is "down" whereas at the second terrace edge the tilt is "up". The step sites would be attractive to charged bidentated ligands such as the bicarbonate ion HCO_3^-. Also shown in contrasting colors are two kink sites one on each of the steps. Only atoms comprising the steps, e kink sites and adatoms are shown. The rest of the crystal is drawn with stick bonds. Red atoms are S atoms belonging to S_2 pairs. The Fe atom sites are shown in blue. Kink site on left side is composed of three S_2 pairs and the blue atom shows where an iron atom could adsorb and bond through three dangling bonds to the surface. Kink site on right side is composed of three Fe atoms each with a dangling bond pointing towards the first red atom.

One kink site attracts Fe and is composed of three S_2 pairs and the atom shows where an iron atom could adsorb and bond through three dangling bonds to the surface. The second kink site on right side is composed of three Fe atoms each with a dangling bond pointing towards the first atom. This site would also attract tridentated ligands like phosphate and this could well be one site where inorganic PO_4^{3-} is sequester. The identification of catalytic sites on pyrite in terms of the number of accessible dangling bonds, groups of sites and their latent chirality's are important steps in understanding the accessible chemical pathways of the surface. Clearly even simple considerations such as the ones described here indicate the power of a mineral surface to catalyze a broad range of chemical processes. On the early Earth it would have been sufficient that a process could occur. There was no competition or fitness criterion to satisfy in the beginning other than time scales related to tectonic plate activity. The emergence of dissipative structures as in other branches of science can be deemed inevitable. However their ability to combine into more complex cooperating arrays does not seem at all obvious. This is one of the great challenges confronting the field.

ACKNOWLEDGEMENT

Over the years it has been my good fortune to work with a number of outstanding computational scientists. In particular I want to acknowledge the debt I owe to Paul S. Bagus, James N. Glosli, and Sergey Izvekov. Ting Ting Lin (IMRE) is thanked for competent assistance with some of the simulations and figures. Discussions with Igor Goliney and Richard Jaffe about catalysis on the pyrite surfaces contributed to the views described above.

REFERENCES

1. J.A. Pople and D.L. Beveridge, *Approximate Molecular Orbital Theory* (McGraw-Hill, New York, 1980).
2. D. Frankel and B. Smit, *Understanding Molecular Simulation. From Algorithms to Applications* (Academic Press, San Diego, 1996).
3. K. Ohno, K. Esfarjani and Y. Kawazoe, *Computational Materials Science: From Ab Initio to Monte Carlo Methods.* Springer Series in Solid-State Sciences (Springer-Verlag, Berlin 1999).
4. D. Henderson, F.F. Abraham and J.A. Barker, The Ornstein–Zernike equation for a fluid in contact with a surface, *Mol. Phys.* **31,** 1291–1295 (1976).
5. D. Henderson and M. Plischke, Some comments on the self-consistency of the Poisson–Boltzmann theory for inhomogeneous coulomb systems, *Mol. Phys.* **62,** 801–804 (1987).
6. D. Henderson, L. Blum and M. Lozada-Cassou, The statistical mechanics of the electric double layer, *J. Electroanal. Chem.* **150,** 291–3003 (1983).
7. D. Henderson, L. Blum and M. Lozada-Cassou, Fundamentals of Inhomogenous Fluids (Marcel Dekker, New York, 1992).
8. F. Vericat, L. Blum and D. Henderson, Nonprimative electrolyte near a charged wall: generalized mean spherical approximation, *J. Electroanal. Chem.* **150,** 315–324 (1983).
9. Gaussian94: 1994, Users Reference. Gaussian Inc., Carnegie Office Park, Bldg 6, Pittsburgh PA 15106, USA.
10. W. Kohn and L.J. Sham, Self-consistent equations including exchange and correlation effects, *Phys. Rev.* **140**, A1133–A1138 (1965).
11. A. Shluger, Application of molecular models to the electronic structure calculations of defects in oxide crystals. I. Parameterization of the modified INDO method, *Theoret. Chim. Acta* **66,** 355–363 (1985).
12. E. Stefanovich, E. Shidlovskaya, A. Shluger and M. Zakharov, Modification of the INDO calculation scheme and parameterization for ionic crystals, *Phys. Stat. Sol. (b)* **160,** 529–540 (1990).
13. A. Shluger and E. Stefanovich, Models of the self–trapped exciton and nearest-neighbor defect pair in SiO_2, *Phys. Rev.* **B42,** 9664–9673 (1990).

14. R.I. Eglitis, A.V. Postnikov and G. Borstel, Semiempirical Hartee–Fock calculations for pure and Li-doped $KTaO_3$, *Phys. Rev.* **B55,** 12976–12981 (1997).
15. R.I. Eglitis, N.E. Christensen, E.A. Kotomin, A.V. Postnikov and G. Borstel, First-principles and semi-empirical calculations for F centers in $KNbO_3$, *Phys. Rev.* **B56,** 8599–8605 (1997).
16. M. Born and K. Huang, *Dynamical Theory of Crystal Lattices* (The Clarendon Press, Oxford, England, 1954).
17. J.A. Barker, Reaction field, screening and long ranged interactions in simulations of ionic and dipolar systems, *IBM Research Division Preprint.*
18. L.F. Greengard, *The Rapid Evaluation of Potential Fields in Particle Systems* (The MIT Press, Cambridge, MA, 1987).
19. L. Greengard and V. Rokhlin, A fast algorithm for particle simulations, *J. Comp. Phys.* **73**, 325–348 (1987).
20. R.W. Hockney and J.W. Eastwood, *Computer Simulation Using Particles* (Adam Hilger, Bristol and New York, 1988).
21. K. Ohno, K. Esfarjani and Y. Kawazoe, *Computational Materials Science: From Ab Initio to Monte Carlo Methods.* Springer Series in Solid-State Sciences (Springer-Verlag, Berlin 1999). Chap. 6.
22. F.F. Abraham, J.Q. Broughton, P.W. Leung and V. Elser, Second layer solidification of 3He on graphite, *Europhys. Letters* **12,** 107–112 (1990).
23. M. Wilson and P.A. Madden, Polarization effects in ionic systems from first principles, *J. Phys. Condens. Matter* **5**, 2687–2706 (1993).
24. M. Sprik, Computer simulation of the dynamics of induced polarization fluctuations in water, *J. Phys. Chem.* **95**,2283–2291(1991).
25. J.E. Roberts, B.L. Woodman and J. Schnitker, The reaction field method in molecular dynamics simulations of point polarizable water models, *Mol. Phys.* **88**, 1089–1108 (1996).
26. J.A. Barker, Surface tension and atomic interactions in simple liquids: argon and xenon, *Mol. Phys.* **80**, 815–820(1993).
27. M. Allen and D.J. Tildesley, *Computer Simulation of Liquids* (Clarendon Press, Oxford, 1990).
28. J. Alejandre, D.J. Tildesley and G.A. Chapela, Molecular dynamics simulation of the orthobaric densities and surface tension of water, *J. Chem. Phys.* **102,** 4574–4583 (1995).
29. F.H. Stillinger and C.W. David, Polarization model for water and its ionic dissociation, *J. Chem. Phys.* **69,** 1473–1484 (1978).
30. J.W. Halley, J.R. Rustad and A. Rahman, A polarizable, dissociating molecular dynamics model for liquid water, *J. Chem. Phys.* **98**, 4110–4119 (1993).
31. J.R. Rustad, B.P. Hay and J.W. Halley, Molecular dynamics simulation of iron(III) and its hydrolysis products in aqueous solution, *J. Chem. Phys.* **102,** 427–131 (1995).
32. J. Tersoff, New empirical approach for the structure and energy of covalent systems, *Phys. Rev.* **B37**, 6991–7000(1988).
33. D. Brenner, Empirical potential for hydrocarbons for use in simulating the chemical vapor deposition of carbon films, *Phys. Rev.* **B42,** 9458–9471 (1990).
34. (a) P. Horsfield, P.D. Godwin, D.G. Pettifor and A.P. Sutton, Computational materials synthesis. I. A tight-binding scheme for hydrocarbons, *Phys. Rev.* **B54,** 15773–15775 (1996); (b) D.G. Pettifor and I.I. Oleinik, Bounded analytic bond order potentials for σ and π bonds, *Phys. Rev. Lett.* **84**, 4124–4127 (2000); (c) D.G. Pettifor, *Bonding and Structure of Molecules and Solids* (Clarendon press, Oxford, 1995).
35. R. Car and M. Parrinello, Unified approach for molecular dynamics and density functional theory, *Phys. Rev. Letts.* **55**, 2471 (1985).
36. D.K. Remler and P.A. Madden, Molecular dynamics without effective potentials via the Car–Parrinello approach, *Mol. Phys.* **70,** 921–966 (1990).
37. D. Marx, J. Hutter and M. Parrinello, Density functional study of small aqueous Be^{2+} clusters, *Chem. Phys. Lett.* **241**, 457–462(1995).
38. K. Laarsonen, M. Sprik, M. Parrinello and R. Car, *Ab initio* liquid water, *J. Chem. Phys.* **99,** 9081–9089 (1993).
39. M. Tuckerman, K. Laasonen, M. Sprik and M. Parrinello, Ab initio water molecular dynamics simulation of the solvation and transport of H_3O^+ and OH^-, *J. Phys. Chem.* **99,** 5749–5752 (1995).
40. L.M. Ramaniah, M. Bernasconi and M. Parrinello, Ab initio molecular-dynamics simulation of K^+ solvation in water, *J. Chem. Phys.* **111,** 1587–1591 (1999).

41. P. Raybaud, M. Digne, R. Iftimie, W. Wellens, P. Euzen and H. Toulhoat, Morphology and surface properties of boehmite (γ-AlOOH): a density functional theory study, *J. Catal.* **201**, 236–246 (2001).
42. S.V. Izvekov, M.R. Philpott and R.I. Eglitis, *Ab initio* simulation of metal cluster surrounded by electrolyte, *J. Electrochemical Soc.* **147**, 2273–2278 (2000).
43. S.V. Izvekov and M.R. Philpott, Ab initio molecular dynamics simulation of LiBr association in water, *J. Chem. Phys.* **113**, 10676–10684(2000).
44. M.R. Sorensen and A.F. Voter, Temperature accelerated dynamics for simulation of infrequent events, *J. Chem. Phys.* **112**, 9599–9606 (2000).
45. A.F. Voter, A method for accelerating the molecular dynamics of infrequent events, *J. Chem. Phys.* **106**, 4665–4677 (1997); Hyperdynamics: accelerated molecular dynamics of infrequent events, *Phys. Rev. Lett.* **78**, 3908–3911 (1997); Parallel replica method for dynamics of infrequent events, *Phys. Rev.* **B57**, R13985–R13988(1998).
46. E.D. Issacs, A. Shukla, P.M. Platzman, D. Hamann, B. Barbiellini and C.A. Tulk, Covalency of the hydrogen bond in ice: a direct X-ray measurement, *Phys. Rev. Lett.* **82**, 600–603 (1999).
47. R.A. Marcus, Electron transfer at electrodes and in solution: comparison of theory and experiment, *Electrochim. Acta* **13**, 995–1004 (1968).
48. See also J.M. Hale and M.E. Peover, in: *Reactions of Molecules at Electrodes,* Ed. N.S. Hush (Wiley-Interscience, New York, 1871). Chaps. 4 and 5.
49. N. Davidson, Solutions of electrolytes. The Debye–Huckel theory, in: *Statistical Mechanics* (McGraw-Hill, New York, 1962). Chap. 21, pp. 485–527.
50. A. Bard and L.R. Faulkner, *Electrochemical Methods: Fundamentals and Applications* (Wiley, New York, 1980). Chap. 12.
51. J.S. Bader and D. Chandler, Computer simulation study of the mean forces between ferrous and ferric ions in water, *J. Phys. Chem.* **96,** 6423–6427 (1992).
52. G.D. Mahan and R.M. Mazo, Coulomb interactions in an atomic dielectric, *Phys. Rev.* **175**, 1191–1200 (1968).
53. D.C. Grahame, The electrical double layer and the theory of electro-capillarity, *Chem. Rev.* **41**, 441–501 (1947).
54. J.O'M. Brockris, M.A.V. Devanathan and K. Mueller, On the structure of charged interfaces, *Proc. Roy. Soc. (London)* **A274**, 55–79, 1963.
55. M.R. Philpott and J.N. Glosli, Electric potential near a charged metal surface in contact with aqueous electrolyte, *J. Electroanal. Chem.* **409**, 65–72 (1996).
56. M.R. Philpott and J.N. Glosli, Screening of the charged electrodes in aqueous electrolytes, *J. Electrochem. Soc.* **142**, L25–L28(1995).
57. J.N. Glosli and M.R. Philpott, Molecular dynamics study of interfacial electric fields, *Electrochimica Acta* **41**, 2145–2158(1996).
58. D.M. Kolb, Physical and electrochemical properties of metal monolayers on metallic substrates, *Advances in Electrochemistry and Electrochemical Engineering* **11**, 125–270 (1978).
59. D.M. Kolb, R. Ullmann and J.C. Ziegler, Electrochemical nanostructuring, *Electrochimica Acta* **43**, 2751–2760(1998).
60. E. Budevski, G. Staikov and W.J. Lorentz, *Electrochemical Phase Formation and Growth* (VCH Publishers, NY, 1996).
61. O.M. Magnussen and R.J. Behm, Atomic-scale processes in Cu corrosion and corrosion inhibition, *Mat. Res. Soc. Bulletin,* July 1999, pp. 16–23.
62. C. Wolverton and A. Zunger, First principles prediction of vacancy order–disorder and intercalation battery voltages, *Phys. Rev. Lett.* **81**, 606–609 (1998).
63. M.K. Aydinol, A.F. Kohan and G. Ceder, Ab initio calculation of the intercalation voltage of lithium–transition-metal oxide electrodes for rechargeable batteries, *J. Power Sources* **68,** 664–668 (1996).
64. Van der Ven, M.K. Aydinol, G. Ceder, G. Kresse and J. Hafner, First principles investigation of the phase stability in Li_2CoO_2, *Phys. Rev.* **B58**, 2975–2987 (1998).
65. G. Ceder, Y.-M. Chiang, D.R. Sadoway, M.K. Aydinol, Y.-I. Jang and B. Huang, Identification of cathode materials for lithium batteries guided by first-principles calculations, *Nature* **392**, 694–696 (1998).
66. F. Mueller-Plathe and W.F. van Gunsteren, Computer simulation of a polymer electrolyte: lithium iodide in amorphous poly(ethylene oxide), *J. Chem. Phys.* **103**, 4745–1756 (1995).

67. G. Wächtershäuser, Pyrite formation, the first energy source for life: a hypothesis, *Syst. Appl. Microbiol.* **10**, 207–210(1988).
68. G. Wächtershäuser, Before enzymes and templates: theory of surface metabolism, *Microbiol. Reviews* **52**, 452–484 (1988).
69. G. Wächtershäuser, Ground works for an evolutionary biochemistry: the iron–sulphur world, *Prog. Biophys. Mol. Biol.* **58**, 85–201 (1992).
70. J.B. Corliss, J.A. Baross and S.E. Hoffman, An hypothesis concerning the relationship between submarine hot springs and the origin of life on earth, *Oceanologica Acta,* No. SP, 59–69 (1981). Proceedings of the 26th International Geological Congress, Geology of the Oceans Symposium, Paris, July 7–17, 1980.
71. M.J. Russell, R.M. Daniel and A.J. Hall, On the emergence of life via catalytic iron sulphide membranes, *Terra Nova* **5**, 343–347 (1988).
72. A.G. Cairns-Smith, A.J. Hall and M.J. Russell, Mineral theories of the origin of life and an iron sulphide example, *Origins of Life and Evol. Biosphere* **22**, 161–180 (1992).
73. M.J. Russell, R.M. Daniel and A.J. Hall, A hydrothermally precipitated catalytic iron sulphide membrane as a first step toward life, *J. Mol. Evol.* **39**, 231–243 (1994).
74. E.L. Shock, T. McCollom and M.D. Schulte, Geochemical constraints on chemolithoautotrophic reactions in hydrothermal systems, *Origin of Life* **25**, 141–159 (1995).
75. J. Bebié, M.A.A. Schoonen, M. Fuhrmann and D.R. Strongin, Surface charge development on transition metal sulphides: an electrokinetic study, *Geochim. Cosmochim. Ada* **62**, 633–642 (1998).
76. M.A.A. Schoonen, Y. Xu and J. Bebié, Energetics and kinetics of the prebiotic synthesis of simple organic acids and amino acids with the $FeS\text{-}H_2S/FeS_2$ redox couple, *Origins Life Evol. Biosphere* **29**, 5–32 (1999).
77. K.M. Rosso, U. Becker and M.F. Hochella, Jr., The interaction of pyrite {100} surfaces with O_2 and H_2O: fundamental oxidation mechanisms, *Amer. Mineralogist* **84**, 1549–1561 (1999).
78. R.I. Eglitis, S.V. Izvekov and M.R. Philpott, Metal dissolution in aqueous electrolyte. Semi-empirical Hartree–Fock and ab initio MD calculations, *Computational Mat. Sci.* **17**, 275–278 (2000).
79. R.I. Eglitis, S.V. Izvekov and M.R. Philpott, Modeling metal dissolution in aqueous electrolyte: Hartree–Fock and molecular dynamics calculations, in: *Solid–Liquid Interface Theory,* Ed. J. Woods Halley (ACS Symposium series 789, distributed by OUP, 2001). Chap. 5, pp. 51–65.
80. S.V. Izvekov, M.R. Philpott and R.I. Eglitis, *Ab initio* simulation of metal cluster surrounded by electrolyte, *J. Electrochemical Soc.* **147**, 2273–2278 (2000).
81. H.J.C. Berendsen, J.R. Grigera, T.P. Straatsma, The missing term in effective pair potentials, *J. Phys. Chem.* **91**, 6269–6271 (1987).
82. G.B. Bachelet, D.R. Hamann and M. Schluter, Psuedo potentials that work: from H to Pu, *Phys. Rev.* **B26**, 4199–4228(1982).
83. H.C. Andersen, Molecular dynamics simulations at constant pressure and/or temperature, *J. Chem. Phys.* **72**, 2384–2393(1983).
84. H.C. Andersen, Rattle: a "velocity" version of the shake algorithm for molecular dynamics calculations, *J. Comp. Phys.* **52**, 24–34 (1983).
85. S.H. Lee and J.C. Rasaiah, Molecular dynamics simulation of ion mobility. 2. Alkali metal and halide ions using the SPC/E model for water at 25°C, *J. Phys. Chem.* **100**, 1420–1425 (1996).
86. E. Spohr and K. Heinzinger, A molecular dynamics study of an aqueous LiI solution between Lennard–Jones walls, *J. Chem. Phys.* **84**, 2304–2309 (1986).
87. E. Spohr and K. Heinzinger, Computer simulations of water and aqueous electrolyte solutions at interfaces, *Electrochimica Acta* **33**, 1211–1222 (1988).
88. E. Spohr and K. Heinzinger, A molecular dynamics study on water/metal interface potential, *Ber. Bunsenges. Phys. Chem.* **92**, 1358–1363 (1988).
89. J. Boecker, Z. Gurskii and K. Heinzinger, Structure and dynamics at the liquid mercury–water interface, *J. Phys. Chem.* **36**, 14969–14977 (1996).
90. E. Spohr, Computer simulation of the structure of the electrochemical double layer, *J. Electroanal. Chem.* **450**, 327–334(1998).
91. E. Spohr, Molecular simulation of the electrochemical double layer, *Electrochimica Acta* **44**, 1697–1705 (1999).
92. G. Aloisi, M.L. Foresti and R. Guidelli, A Monte Carlo simulation of water molecules near a charged wall, *J. Chem. Phys.* **91**, 5592–5596 (1989).

93. J. Hautman, J.W. Halley and Y.-J. Rhee, Molecular dynamics simulation of water between two ideal classical metal walls, *J. Chem. Phys.* **91**, 467–72 (1989).
94. J.I. Siepmann and M. Sprik, Influence of surface topology and electrostatic potential on water/electrode systems, *J. Chem. Phys.* **102**, 511–524 (1995).
95. J.I. Siepmann and M. Sprik, Ordering of fractional monolayers of H_2O on Ni(110), *Surf. Sci. Letters* **279**, L185–L190 (1992).
96. M. Berkowitz and W. Wan, The limiting conductivity of Na^+ and Cl^- ions in aqueous solutions: molecular dynamics simulation, *J. Chem. Phys.* **86**, 376–382 (1986).
97. D.A. Rose and I. Benjamin, Solvation of Na^+ and Cl^- at the water–platinum (001) interface, *J. Chem. Phys.* **95**, 6856–6865 (1991).
98. S.-B. Zhu and M.R. Philpott, Interaction of water with metal surfaces, *J. Chem. Phys.* **100**, 6961–6968 (1994).
99. C.Y. Lee, J.A. McCammon and P.J. Rossky, The structure of liquid water at an extended hydrophobic interface, *J. Chem. Phys.* **80**, 4448–455 (1984).
100. 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, Distribution of water molecules at the silver (1 1 1)–electrolyte interface, *Nature* **368**, 444–446 (1994).
101. M.A. Wilson, A. Pohorille and L.R. Pratt, Surface potential of the water liquid–vapour interface, *J. Chem. Phys.* **88**, 3281–3285(1988).
102. M.A. Wilson, A. Pohorillc and L.R. Pratt, Comment on the "Study on the liquid–vapour interface of water. Simulation results of the thermodynamic properties and orientutional structure", *J. Chem. Phys.* **90**, 5211–5213(1989).

2

Characterization of solid–liquid interfaces using X-ray diffraction

Elias Vlieg*
NSRIM Dept. of Solid State Chemistry, University of Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands

Abstract: X-ray diffraction has found in recent years an increasing use as a tool to study the atomic-scale structure of solid–liquid interfaces. By using an appropriate scattering geometry, suitable positions in reciprocal space and the current generation of synchrotron radiation sources, information on the few monolayers that make up both sides of a solid–liquid interface can be obtained. The major applications have so far been in the areas of crystal growth and electrochemistry. From these studies it becomes clear that the structure at both the solid and the liquid side of the interface can deviate strongly from that in the bulk. Such structural knowledge is important for a full understanding of the processes occurring at solid–liquid interfaces.

2.1. INTRODUCTION

In all the areas where one encounters solid–liquid interfaces, e.g. crystal growth, lubrication and electrochemistry, a proper understanding of the physical and chemical phenomena requires structural knowledge of the interface at an atomic scale. For a long time suitable experimental techniques were not available and the main ideas were derived from theoretical models [1]. Such studies made it for example clear that the first few liquid layers have special properties. Since about ten years, a number of techniques have become available that draw also an experimental picture of the solid–liquid interface. One of these is scanning-probe microscopy (SPM), a technique that has revolutionized the study of almost all interfaces. SPM is particularly useful to observe the morphology of the solid part, but less suitable for deriving its atomic-scale structure or for observing the structure in the liquid [2]. A second technique is transmission-electron microscopy (TEM). Though this technique requires an elaborate sample preparation and is therefore not suitable for many types of systems, very beautiful observations of the solid–liquid interface have been made [3]. This chapter deals with a third technique, X-ray diffraction. X-ray diffraction is

*E-mail address: vlieg@sci.kun.nl

of course much older than the other techniques (the first experiment dates from 1912), but its use to study surfaces and interfaces started only about twenty years ago with the advent of powerful synchrotron radiation sources. The crystallographic capabilities of X-rays are fully known and developed in bulk crystallography and have more recently been employed for the study of surfaces as well [4]. The penetrating power of X-rays make them suitable as a probe for *in situ* investigations of buried interfaces. Using X-ray diffraction, an interface structure can in principle be determined with sub-Ångström resolution, but this is always done by averaging over a large sample area. It is therefore advantageous to combine diffraction with an imaging technique, such as SPM.

A solid–liquid interface has of course two sides, and both are expected to show a structure that deviates from that in the corresponding bulk phase. While the solid side has received more attention than the liquid side, X-ray diffraction is increasingly utilized to reveal the liquid structure as well. The most important types of solid–liquid interfaces that have been studied using X-ray diffraction are:

1. Crystal growth. When crystals are grown from a solution or melt, the actual growth takes place at a solid–liquid interface. The structure at this interface is one of the properties that determines the morphology of the crystal, and this morphology is known to depend e.g. on the solvent used [5,6].
2. Electrochemistry. In electrochemistry, chemical reactions take place at the interface between an electrode and an electrolyte. X-rays allow the nature of the so-called electrical double layer to be probed, as well as the structural changes at the interface as a function of the potential [7].
3. 2D melting transitions. Whereas melting is a discontinuous transition in the bulk, the melting of 2-dimensional layers is typically a continuous process [8]. X-ray diffraction can observe the changes in long-range order occurring during such melting transitions.
4. Surface melting. This is related to 2D melting, but more specifically deals with the onset of the melting of a crystal at the surface [9]. Though a number of X-ray diffraction experiments have been performed, few details of the solid–liquid interface have been revealed.

Solid–liquid interfaces are also important in lubrication [10]. The fact that in this case the liquid is confined between two solid interfaces poses difficult boundary conditions on the experimental geometry. No atomic-scale X-ray diffraction experiments have therefore been performed; the only report to date concerns a colloidal fluid [11]. We may expect more results in the future, but in this chapter this subject will not be discussed any further.

The remainder of this chapter starts with a description of the method, after which we will discuss the various topics in which X-ray diffraction has been employed.

2.2. METHOD

2.2.1. Surface sensitivity

The technique of surface X-ray diffraction (SXRD) has established itself as a valuable tool for surface crystallography [4]. The main reason why X-rays, given their dominant role in bulk crystallography, were not used in surface crystallography before the 1980's, was the

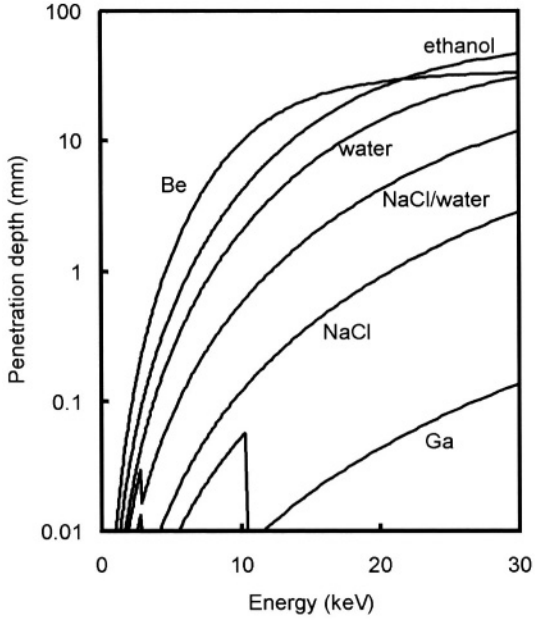


FIGURE 2.1. The penetration depth of X-rays for various materials as a function of the X-ray energy.

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. These properties make X-ray diffraction an ideal tool to study solid–liquid interfaces. Figure 2.1 shows the penetration depth for a number of compounds as a function of the X-ray energy. Earlier synchrotron radiation sources provided X-rays typically in the energy range around 10 keV. In that case the penetration depth is only several millimeters in solvents like water and ethanol. However, the latest generation of synchrotron radiation sources, such as the ESRF in Grenoble shown in Figure 2.2, provides large fluxes of X-ray at energies easily up to 30 keV, leading to penetration depths of tens of millimeters in these solvents, and thus facilitating the design of experimental environments. But even at such high energies, the penetration depth in heavier elements remains very small, see e.g. Ga that has a penetration depth of 0.1 mm at 30 keV. As will be discussed later, the design of sample environments should therefore typically be such that only a limited amount of material is penetrated. As a note of caution, the figure also shows that the penetration depth of a saturated NaCl solution is much less than that of pure water.

While the large penetration depth of X-rays is a necessity for *in situ* studies of solid–liquid interfaces, we need to discuss how this is compatible with surface sensitivity. How can the signal from the few layers that make up the interface be distinguished from that from the thousands of illuminated layers in the bulk? The main trick here is to look at the

FIGURE 2.2. An aerial photograph of the European Synchrotron Radiation Facility (ESRF) in Grenoble, France. A synchrotron source consists of a ring-shaped vacuum pipe in which high-energy electrons are circulating. Radiation is produced at each location where the electron path is curved. The ESRF is one of the three hard X-ray synchrotron radiation sources of the third generation that are currently in operation. The other two are located in the USA (APS, in Argonne) and in Japan (SPring-8, in Nishi Harima). (Source: ESRF).

appropriate places in reciprocal space. The situation is illustrated in Figure 2.3. Consider a crystal with lattice vectors $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ and $\boldsymbol{a}_3$. The reciprocal lattice vectors $\boldsymbol{b}_j$ are defined by

$$\boldsymbol{a}_i \cdot \boldsymbol{b}_j = 2\pi\delta_{ij}. \tag{1}$$

The momentum transfer vector $\boldsymbol{Q}$, which is the difference between the wavevectors of the incident and scattered X-rays, is denoted by the diffraction indices (hkl) in reciprocal space:

$$\boldsymbol{Q} = h\boldsymbol{b}_1 + k\boldsymbol{b}_2 + l\boldsymbol{b}_3. \tag{2}$$

The component of Q along the in-plane direction is called $Q_\parallel$, along the out-of-plane direction $Q_\perp$. Typically, the (reciprocal) lattice vectors are chosen such that the diffraction index l denotes the direction perpendicular to the surface. For such a choice, the unit cell will often differ from the conventional bulk unit cell. A bulk crystal has only intensity at sharp spots in reciprocal space, denoted by *integer* values of the diffraction indices (hkl). At these spots, the bulk contribution completely dominates the diffracted signal. Such bulk Bragg peaks should thus be avoided if one wants to obtain surface-sensitive information. 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

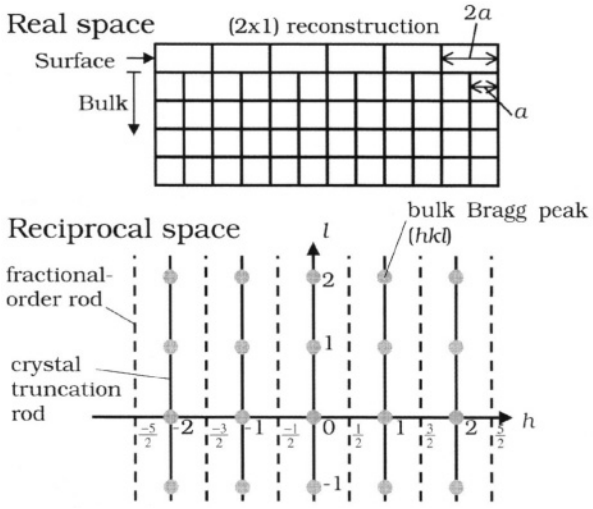


FIGURE 2.3. Schematic showing the position of surface sensitive information in reciprocal space. The bulk crystal yields only intensity at points (*hkl*) in reciprocal space. The surface is assumed to be (2 × 1) reconstructed, which gives rise to half-order diffraction rods in reciprocal space (dashed 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.

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.

For surfaces in vacuum, reconstruction is a common phenomenon. For solid–liquid interfaces this appears to be more rare, likely owing to the fact that the interaction of the surface with the liquid gives alternative ways to reduce the surface free energy. 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) [12,13]. On a crystal truncation rod the intensity varies continuously as a function of l and reaches a (high) maximum for integer values of l at the bulk Bragg peaks. Maximum surface sensitivity is obtained exactly midway between these bulk peaks. Each rod is labeled by the diffraction indices (hk). The (00) rod is a special one, because in that case there is no in-plane momentum transfer ($Q_{\parallel} = 0$) and therefore it is not sensitive to the in-plane structure of the interface. This rod is also known as the specular rod. Measuring X-ray reflectivity (without the other rods) is a field of its own, especially for non-crystalline samples [14].

2.2.2. *Experimental geometry*

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 allows accurate angular posi-

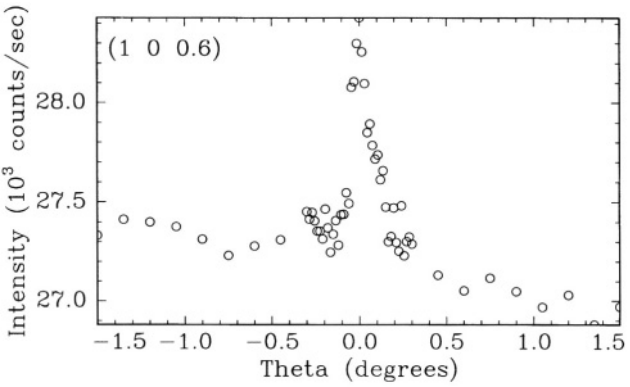


FIGURE 2.4. Rocking scan of the (1,0,0.6) reflection of KDP(101) measured in solution at the Troika beamline at the ESRF. The signal of ~1000 counts/sec is about 4% of the background of ~27000 counts/sec. For weaker reflections, the signal-to-background ratio can even drop below 1%. From ref. [17].

tioning and scanning of both the sample and the X-ray detector. In this way, any desired spot in reciprocal space can be observed. 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, and an example is shown in Figure 2.4. The integrated intensity in such a scan will depend on experimental parameters like counting time, step size and slit settings. Using standard procedures [15] this integrated intensity can be converted into the normalized amplitude of the so-called structure factor F_{hkl} for this location in reciprocal space. A structure determination using X-ray diffraction revolves around these structure factors. Other scan types are possible, and these will also yield the structure factor, but using a different normalization procedure [15].

One can use two different experimental geometries for *in situ* studies of a solid–liquid interface, see Figure 2.5 [16,17]. In a reflection geometry, the X-ray beam penetrates a thin liquid film from the top. In a transmission geometry, the X-ray beam travels through an entire column of liquid, that thus has to be fairly small in order not to attenuate the beam completely. This geometry is therefore mainly suitable for liquids like water and ethanol, but has the great advantage that also investigations during growth are possible. In the much thinner film in the reflection geometry, mass transport is severely hindered and only very limited growth is possible.

The signal from the surface is weak and it is therefore important to minimize the background scattering. When studying a solid–liquid interface, both the solid and the liquid are sources of background scattering. In order to minimize the background from the bulk crystal, diffraction experiments are typically performed at a small incoming or outgoing angle, because then the penetration depth is reduced. At very small angles, total reflection occurs and the penetration depth is reduced to the 100 Å range. The critical angle at which this occurs depends on the material and the X-ray energy, but is typical in the 0.1–0.2° range. Exactly at the critical angle the beam intensity at the surface is enhanced by a factor four. A disadvantage of such low angles is that the alignment becomes very critical, because a small variation of the incoming angle under total reflection conditions can lead to relatively large variations in the measured signal. Therefore, when the alignment is not very precise (as may easily be the case for a crystal in a liquid) it is better to use e.g. five times the

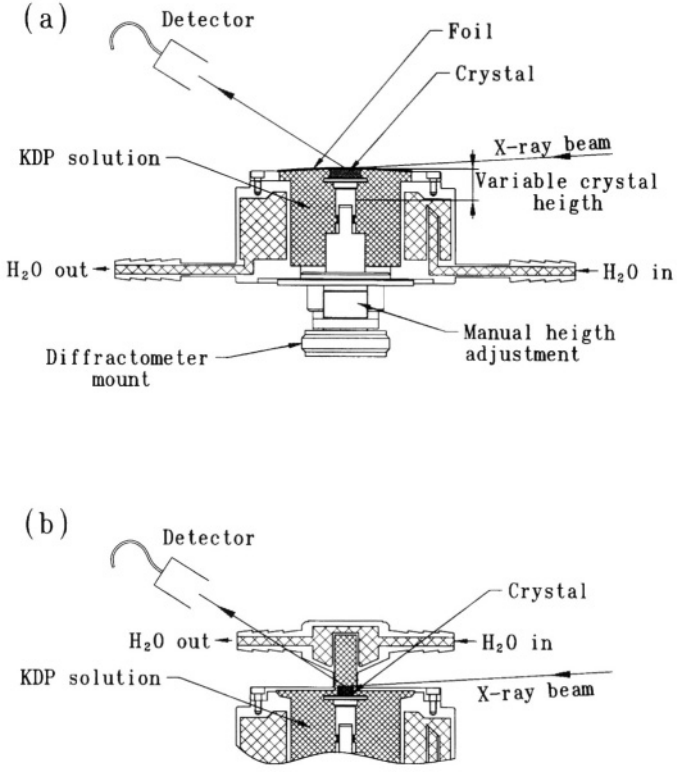


FIGURE 2.5. Two possible scattering geometries for *in situ* studies of a solid–liquid interface. This is illustrated for the case of KDP crystals in a saturated solution that are kept at a specific temperature by a water bath, (a) Reflection geometry, in which a thin layer of liquid is generated by pressing a foil against the crystal. (b) Transmission geometry, which is suitable for experiments during growth. From ref. [17].

critical angle of incidence. The surface sensitivity, after all, is predominantly obtained by the selection of the appropriate diffraction spots ("Fourier filtering").

In order to reduce the background from the liquid, it is important to limit the amount of penetrated and/or illuminated liquid. Compared to surfaces in vacuum, the signal-to-background ratio can easily be a factor 100 worse owing to the diffuse background signal from the liquid. A typical example is shown in Figure 2.4, where the surface signal of a high-quality, but fairly weakly scattering, KDP crystal is only about 4% of that of the background.

2.2.3. Structure factors

The structure factors that one measures in a surface diffraction experiment are the sum over all bulk and surface contributions from the crystal (Figure 2.6). The structure factor of a single bulk unit cell with j atoms is defined as [18]:

$$F_{hkl}^{u} = \sum_{j}^{\text{unit cell}} f_j e^{-M_j} e^{2\pi i(hx_j+ky_j+lz_j)}, \tag{3}$$

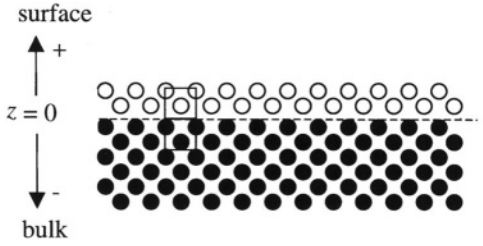


FIGURE 2.6. Schematic of a surface layer on top of a bulk crystal that is assumed to extend to minus infinity. Unit cells (of equal size) for both sides are indicated by outlined squares. In calculating a theoretical structure factor, all atoms that are not at bulk positions are included in the surface unit cell. The surface unit cell can thus also contain atoms in a liquid state.

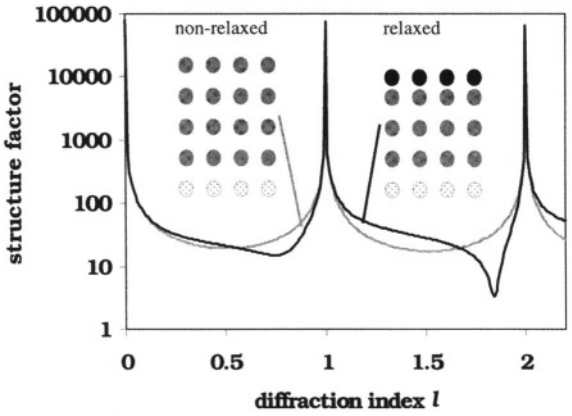


FIGURE 2.7. The variation of the structure factor amplitude as a function of the diffraction index l along a crystal-truncation rod for a simple cubic crystal. For integer values of l one finds diffraction spots from the bulk crystal with very high intensity. In between these spots, the structure factor 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.

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 atom j 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:

$$F_{hkl}^{\text{bulk}} = \sum_{j=-\infty}^{0} e^{2\pi i l j} e^{\alpha j} F_{hkl}^{\text{u}} = \frac{F_{hkl}^{\text{u}}}{1 - e^{-2\pi i l} e^{-\alpha}}, \tag{4}$$

where α is a (small) factor that describes the attenuation of the X-ray beam when it penetrates the bulk crystal. Equation (4) describes the structure factor of a crystal truncation rod [13]. Figure 2.7 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. At this minimum, the structure factor amplitude is equivalent (for a cubic crystal) to half a layer of the bulk crystal. This is easily seen by substituting $l = 0.5$ in eq. (4).

The surface contribution to the structure factor is given by an expression completely analogous to eq. (3), where now the summation is over all the atoms in the surface unit cell. In this surface unit cell all atoms are put that are not at bulk positions. The total structure factor is the interference sum of the bulk and the surface contributions:

$$F_{hkl} = F_{hkl}^{\text{bulk}} + F_{hkl}^{\text{surf}}. \tag{5}$$

For a fractional-order rod, $F^{\text{bulk}} = 0$ and only the surface contributes to the intensity. Such a rod contains of course no bulk Bragg peaks and typically has a much smaller intensity variation than a crystal truncation rod. The total structure factor is sensitive to surface relaxation, reconstruction and roughness [19]. Figure 2.7 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, showing the surface sensitive part of the rod.

When the crystal is immersed in a solution, also molecules from the liquid may contribute to the X-ray intensity in a rod from the substrate. The liquid is modulated by the periodic potential of the crystal surface and its density thus acquires Fourier components corresponding to the substrate. This leads to an extra term in eq. (5). In practice, the contributions of the liquid can simply be added to the surface unit cell. The surface unit cell thus contains all atoms that are not part of the bulk crystal.

Normally, the modulated liquid at the interface will not contribute equally strong to all crystal truncation rods. The interaction of a liquid with a crystal is illustrated in Figure 2.8. For clarity, in this case the fractional-order rods are ignored in reciprocal space. If the liquid layer is thick and has everywhere a bulk structure, there is no contribution to the diffraction rods of the substrate (Figure 2.8a). In reciprocal space, one would only observe extra intensity in a sphere around the origin, corresponding to the pair-correlation function of the liquid. This sphere can be observed using a so-called radial scan, which starts at a momentum transfer smaller than the location of the sphere and then moves radially outwards. If the liquid consists of a perfectly flat monolayer, without any in-plane ordering, this single layer would have a strong contribution to the (0,0) crystal truncation rod (i.e., the specular reflectivity), but none to the other rods (Figure 2.8b). One might still observe the (very weak) peak origination from the pair correlation function, but this would now have a cylindrical shape around the origin owing to the 2-dimensional nature of the layer. Again a radial scan is needed to detect this cylinder. Figure 2.8c illustrates the extreme case of a liquid monolayer that is perfectly ordered with respect to the underlying crystal. In that case it can of course not really be called a liquid. It would contribute strongly to all crystal truncation rods. What one expects for a real liquid, is a situation in between the extreme cases of perfect and no lateral ordering, thus only partial in-plane order (Figure 2.8d). In that case the liquid would contribute strongly to the specular rod, and would have rapidly decreasing contributions to rods with a higher in-plane momentum transfer. One expects the liquid side of a thick solid–liquid interface to consist of a few of such monolayers with partial order.

The lesson from this discussion is that for a full characterization of the solid–liquid interface, one needs to measure several rods. A layer that looks perfectly ordered from

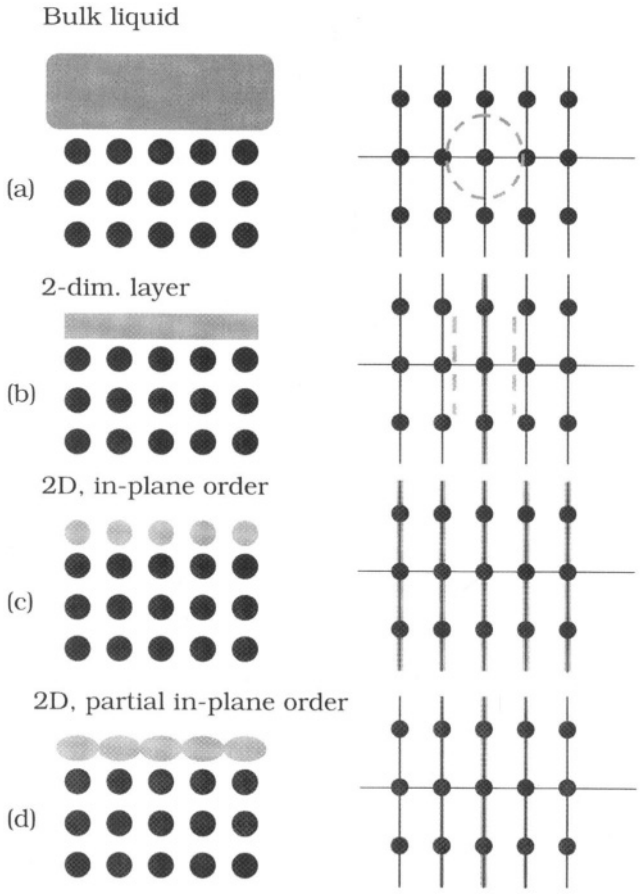


FIGURE 2.8. A schematic showing the location in reciprocal space of the contribution of a 'liquid' layer on top of a crystal, (a) No interaction between liquid and crystal; (b) single liquid monolayer without in-plane order; (c) fully ordered monolayer and (d) partially ordered monolayer.

the specular rod, may in fact be found to be laterally disordered from a rod with in-plane momentum transfer. As is always true in diffraction, the larger the measured Q-range, the higher the resolution of the structure determination.

Unfortunately, a measurement does not yield the complete structure factor, but only its amplitude. This is the famous phase problem in crystallography [18]. Instead of a simple Fourier transform, model calculations and fitting are therefore required to determine the structure [20]. 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 [19,21]. Measuring reflection profiles thus yields information about the surface morphology and can for example be used to measure the surface domain size as a function of the temperature.

2.3. RESULTS

2.3.1. Crystal growth

The growth of a crystal occurs at the interface with its growth environment: a vapor, 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, solvent 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 [22], 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 the crystal. X-ray diffraction is a good technique to determine the extent to which growing crystal surfaces deviate from a simple bulk termination. It is further of interest to study the role that solvents [23] and impurities have on the morphology of crystals. All these items require a structure determination of the solid-liquid interface [6].

The earliest attempts to study the interface structure of a growing crystal date from the early 1990's [24]. While these experiments used the penetrating power of X-rays to study ADP crystals in an aqueous growth environment, the geometry was not suited for the measurement of surface-sensitive reflections. Some years later, such reflections were observed from the $(10\bar{1}4)$ face of calcite by Chiarello and Sturchio [25]. They measured a full specular reflectivity curve, but could only conclude that the surface structure was not measurably different from that of the bulk. Only five years ago, the experimental data became of sufficient quality to obtain atomic-scale information about the interface structure. In 1997, Gidalevitz *et al.* determined the surface termination of β-alanine and α-glycine crystals [26]. For both crystals, two different surface terminations were in principle possible. Figure 2.9 shows the data obtained on the (0*k*4) rod of β-alanine. For each of the two possible surface terminations, it is straightforward to calculated the expected rod profile. These are drawn in the same figure as a solid and dashed curve. The data immediately show which of the two terminations occurs in reality. The agreement between theory and data is not perfect, indicating that the system deviates from a simple, bulk-terminated crystal without relaxation. The data set of Gidalevitz *et al.* was, however, not accurate/large enough to determine such deviations.

Measuring surface relaxations is relevant, as for example the structure determinations of the {101} faces of KDP (potassium dihydrogen phosphate, $K_2H_2PO_4$) [27] and ADP (ammonium dihydrogen phosphate, $NH_4H_2PO_4$) [28] have shown. The two crystals are isomorphous and their surfaces are compared in Figure 2.10. In both cases, the surface has two possible terminations: K^+ or PO_4^- for KDP and NH_4^+ or PO_4^- for ADP. From full X-ray diffraction data sets it was shown that in both cases the surface terminates in a positive layer, i.e. K^+ and NH_4^+ for KDP and ADP, respectively. This positive termination explains why the presence of positive impurities like Fe^{3+} and Cr^{3+}, has little effect on the growth speed of the {101} face. The {100} face, on the other hand, has both positive and negative

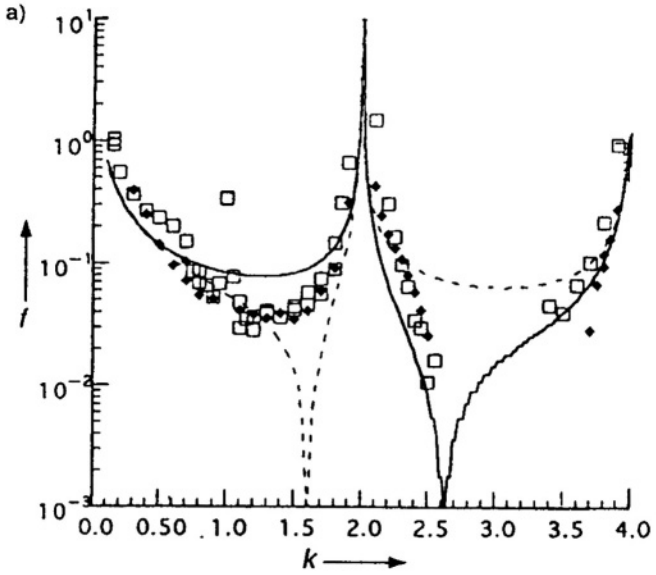


FIGURE 2.9. The (0k4) crystal truncation rod of β-alanine as measured by Gidalevitz *et al.* [26]. The points represent data measured under various circumstances. The two curves are calculations for the two possible surface terminations of this crystal. Note that Gidalevitz *et al.* do not follow the convention that l represents the out-of-plane direction. From ref. [26].

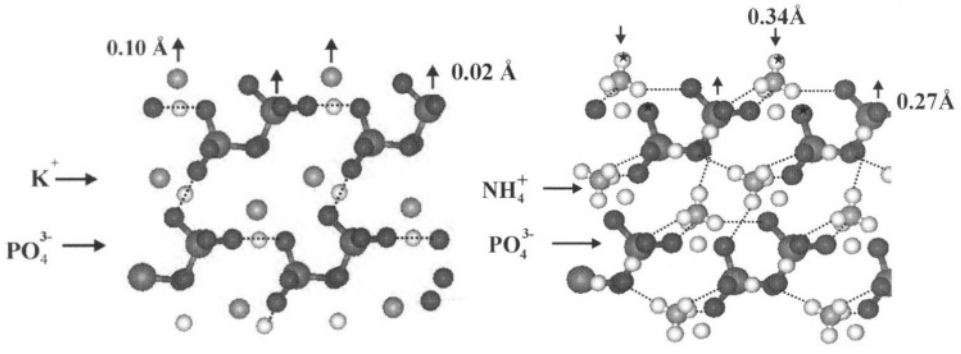


FIGURE 2.10. The surface structures of the {101} faces of KDP (left) and ADP (right), including the measured relaxations in the two top layers. Both surfaces terminate with a positive layer, but the relaxations differ significantly. The 'stars' in the ADP surface indicate the positions where a hydrogen bond is broken due to the termination of the crystal.

ion groups at its surface, and therefore positive impurities can block the growth there [29]. This leads to a change in growth morphology from more compact to more elongated for increasing concentration of impurities.

While the surface terminations are similar, the measured relaxations for KDP and ADP are very different. For KDP the outermost layers have a small outward relaxation, but for ADP the top NH_4^+ layer is relaxed inward and the PO_4^- is relaxed outward by a large

amount. The spacing between these two layers is therefore found to be contracted by approximately 50%! The hydrogen bonding of the ammonium in ADP may cause the difference between KDP and ADP. This particular hydrogen bonding is absent in the case of KDP. In the bulk, the four hydrogen atoms of the ammonium group form hydrogen bonds to the oxygen ions of four neighboring phosphate groups (Figure 2.10). At the ammonium-terminated surface, however, one of those hydrogen bonds is broken, leading to a broken bond in the top layer (ammonium) and a broken bond in the third layer (phosphate). Stars in Figure 2.10 indicate these bonds. The large inward relaxation of the top layer reduces the distance between the associated nitrogen and oxygen atoms from 3.17 Å to 2.97 Å. This is very close to the value of 2.91 Å for the N–H–O hydrogen bond in bulk ADP. The driving force for the large inward relaxation of the surface ammonium groups may therefore well be the formation of an extra hydrogen bond with the lower-lying phosphate groups, as a compensation for the bonds that are broken due to the crystal termination [28].

In the field of geochemistry, knowing the structure of mineral-fluid interfaces is important for understanding many environmental processes. Fenter, Sturchio and co-workers have investigated the structure of various adsorbates on calcite [30]. The *in situ* structure of feldspars has also been investigated, mainly using X-ray reflectivity. The consecutive studies demonstrate once more that the amount of information that one can derive, depends on the resolution, i.e. the amount of reciprocal space measured [31].

The total number of papers on the surface structure of crystals in solution is still very limited. On the other hand, there is a huge amount of literature on the interaction of water with various surfaces from the perspective of adsorption. In that case the amount of water is typically limited to (sub)monolayers. For a comprehensive review, see ref. [32]. X-ray diffraction appears not to have been used in this area.

2.3.2. Electrochemistry

Understanding the structure and dynamics of electrochemical interfaces is a long-term goal in the field of electrochemistry. Various techniques are used to study such interfaces at an atomic scale, including STM [33] and techniques that require vacuum [34]. It is clear that at such interfaces, X-ray diffraction has a lot to offer [7,35]. Compared to the systems discussed in the previous section, the electrochemical interfaces tend to be somewhat easier from an X-ray diffraction point of view, because the substrates are typically strongly scattering metals, while the liquid is mainly water. Signal-to-background ratios are therefore reasonable, provided a proper geometry is chosen [36].

The first experiment of this type was done in 1987 by Samant *et al.* on the structure of Pb monolayers on Ag(111) and Au(111) electrodes [37]. This was in fact the first X-ray diffraction experiment of a solid–liquid interface with proper surface sensitivity. Several other electrodeposited monolayers (e.g. Tl, I, Br) have since then been investigated [38]. Here we will discuss a few 'clean' surfaces, because they are the simplest electrochemical solid–liquid interfaces. Compared to surfaces in vacuum, the electrochemical potential is an additional and convenient thermodynamic parameter, which may thus be used to control the surface properties [39]. Several metals are known to reconstruct in vacuum, but the driving forces for this are usually quite small and could easily be modified by the potential.

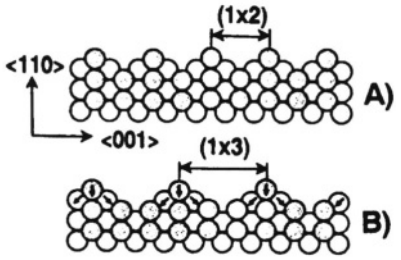


FIGURE 2.11. Cross-sectional view of the (a) (1 × 2) and (b) (1 × 3) missing-row reconstructions of the Au(110) surface. In the (1 × 2) phase every other row is missing, in the (1 × 3) phase two out of three rows are missing. From ref. [41].

Examples are the (110) surfaces of Au and Pt that have been studied using X-ray diffraction both in vacuum and in electrochemical environments. These surfaces are found to reconstruct in vacuum, forming a so-called missing-row structure with a (1 × 2) periodicity in which every other close-packed row is missing [40]. This is thus quite a drastic reconstruction, but it nevertheless leads to a lower free energy state of this surface. In an electrochemical environment, the free energy balance may vary as a function of the potential, and that is indeed observed for Au(110) in salt and acid solutions [41]. For decreasing potentials in an $HClO_4$ solution, the (1 × 2) reconstruction was found to continuously evolve into a (1 × 3) reconstruction, see Figure 2.11. In the (1 × 3) phase, two out of three rows are missing. If separate domains of the (1 × 2) and (1 × 3) phase were present, one would measure peaks at the corresponding fractional order positions (with the diffraction indices at values that are multiples of 1/2 and 1/3). Instead a single peak is found, at intermediate positions. This shows that the surface consists of irregular sequences of one and two missing rows. In salt solutions the potential can be made even more negative and then a pure (1 × 3) phase is found. At high potentials, on the other hand, no reconstruction is observed, so indeed the surface structure is strongly influenced by the potential. From a fit to the rod profiles, the top layer is found to contract [41], in agreement with the structure in vacuum [40].

In vacuum, the (110) surface of Pt is very similar to that of Au, but a different behavior is found as a function of the electrochemical potential [42]. On this surface the (1 × 2) reconstruction is found to be stable over a wide potential range. There are, nevertheless, significant structural changes as compared to the surface in vacuum, see Figure 2.12. In vacuum, the top layer is contracted [40], but in the electrolyte an expansion if found instead [42].

The effect of the potential is similar to the effect of the deposition of alkali metals in vacuum. For example, Ag(110), which is isoelectronic with Au(110) and Pt(110), does not reconstruct in vacuum. After the adsorption of ~0.1 monolayers of K, also this surface forms a (1 × 2) missing row reconstruction [43]. It therefore appears that in all cases a similar mechanism is operating and that the surface charge plays an important role. It seems likely that the Ag(110) surface can be made to reconstruct to a missing-row structure at an appropriate electrochemical potential, but this has not be reported yet.

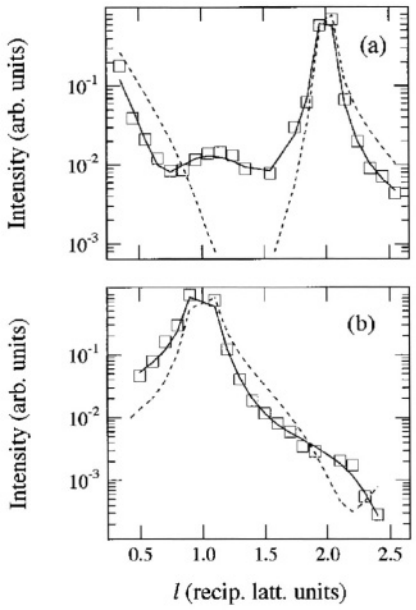


FIGURE 2.12. The measured (squares) X-ray intensity along the (a) (00) rod and (b) the (01) rod of Pt(110) at an electrode potential of 0.1 V. The solid curves are the best fit with a model containing an expansion in the top layer, the dashed curves use the relaxations as determined for the surface in vacuum. From ref. [42].

2.3.3. Liquid structure

In the previous sections we focussed our attention on the solid side of the solid–liquid interface. Now we will consider the liquid side as well. There are even less available techniques to study the liquid structure at the interface than the solid side, because the liquid is of course easily disturbed. There are some alternatives to X-ray diffraction, in particular transmission-electron microcopy [3,44] and even using AFM some features can be seen [2], but the most generally applicable probe appears to be X-rays.

Ordering of the liquid in the direction perpendicular to the surface ('layering') is expected to be the strongest, and it is therefore not surprising that the first experiments have probed this direction in particular [45–48]. One of the earliest results is from 1990, when Ocko observed the temperature dependent layering of cyanobiphenyl compounds at alkylsilane-coated silicon surfaces [45]. Such liquid crystals also show smectic layering in the bulk phase, but here the layering was already observed at temperatures such that the bulk material is in the isotropic phase. In 1994, Toney *et al.* [47] observed the voltage-dependent ordering of water molecules at a Ag(111) surface in an electrochemical environment. Figure 2.13 shows the oxygen distribution at the interface as derived from the specular rod. The distance to the crystal surface changes as a function of the voltage. For both the positive and the negative potential the water was found to have a higher density than that in bulk water. Somewhat surprisingly, despite this high density, no lateral order was found in the layers as derived from non-specular rods. The water therefore appears to be really liquid and not icelike in the lateral direction. Particular clear cases of layering were found by Huisman *et al.* for undercooled Ga in contact with a diamond surface [48]

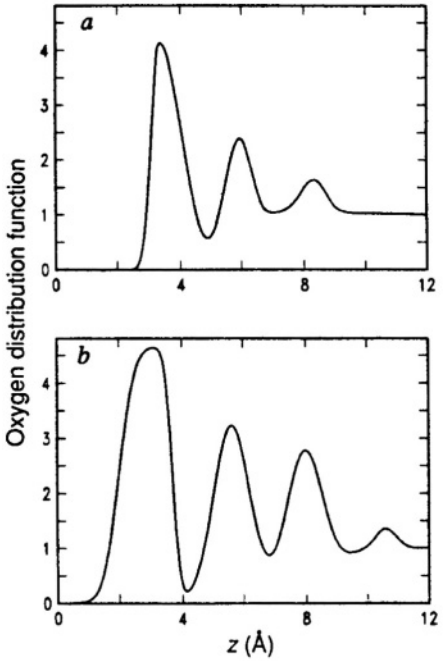


FIGURE 2.13. The oxygen distribution at the interface of water with Ag(111) at a voltage of (a) –0.23 V and (b) +0.52 V with respect to the potential of zero charge. Oxygen is equivalent to water in this case, because the hydrogen atoms are largely invisible for X-rays. The distribution is normalised to that in bulk water. Near the Ag(111) surface, the water density is enhanced by about a factor two for case (b). From ref. [47].

and by Yu *et al.* for molecular liquids in contact with a silicon surface [49]. In the latter case, the silicon substrates were treated such that they were essentially a hard wall, while the liquids, e.g. tetrakis(2-ethylhexoxy)silane, were nearly spherical. In that case one therefore closely mimics a model system of spheres that interact with a hard wall and in which consequently no lateral order is expected. Layering was found to extend over 3–6 layers.

More recently, the lateral ordering component in a liquid has also been investigated. It has been found that along this direction the water in contact with a crystal may in fact be fully ordered, i.e., that the first water layer behaves very much like ice. As an example, Cheng *et al.* [50] concluded from high-resolution specular X-ray reflectivity data that water layering occurs at the interface with mica(001). Without non-specular data, no conclusions about the lateral ordering can be drawn based on the X-ray data alone, but the results agree with other experiments [51] and calculations [52]. Layers not in direct contact with the crystal are thought to be more liquidlike.

A particularly nice and comprehensive study was done by Chu and Lister *et al.* [53, 54] on the $RuO_2(110)$–water interface under electrochemical conditions. From extensive data sets containing specular and non-specular rods, they could derive the changes at the interface as a function of the voltage, see Figure 2.14. All the visible water appeared to be well ordered, at anodic potentials the structure was in fact similar to ice X.

The number of published results is still very limited and it's therefore too early to draw general conclusions. It is somewhat surprising that in the careful experiments on

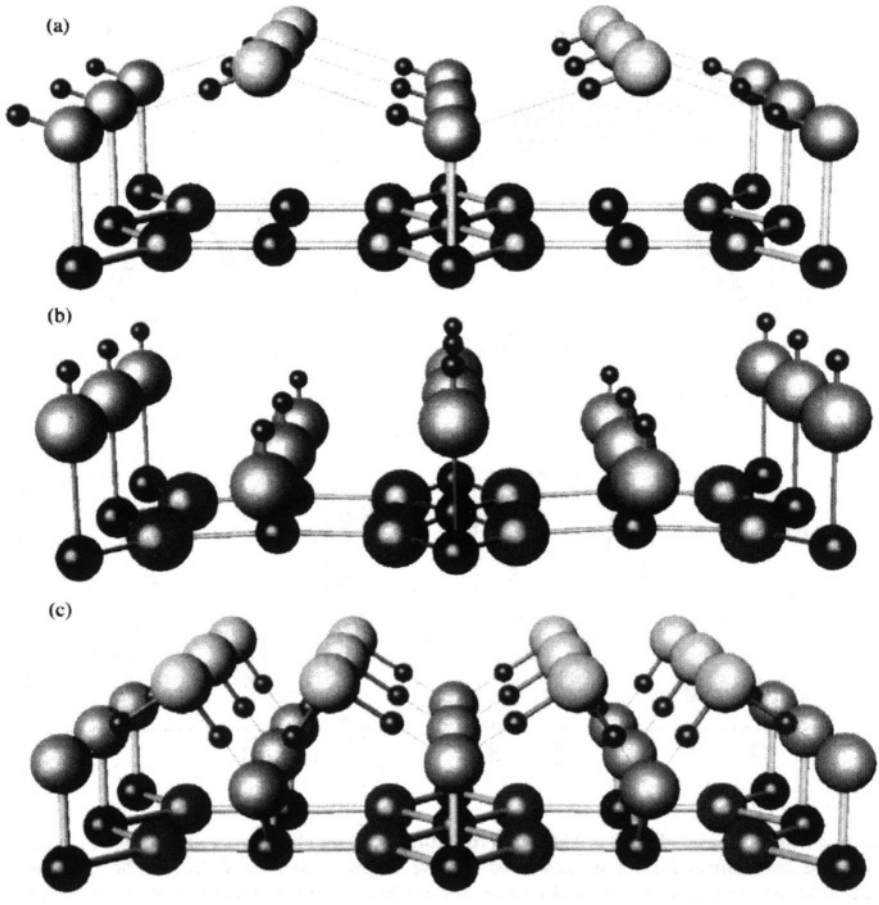


FIGURE 2.14. The structure of the $RuO_2(110)$–water interface for three different potentials. The bottom layer is the topmost RuO_2 layer, the other layers are the ordered water or oxygen layers. From ref. [54].

$RuO_2(110)$ no partly-ordered water was found, while in other cases the order was found to decay in a few layers away from the interface [47,48,50]. Also the role of the electrochemical potential is not clear in this respect.

Very recent experiments on the KDP{110}–water interface [55] found both icelike and disordered water layers. The crystal side of the interface was already discussed and was shown in Figure 2.10. This structure was derived from non-specular data that is far less sensitive to the liquid structure than the specular rod. This fact is clearly demonstrated in Figure 2.15, where specular reflectivity data is shown (circles) together with a calculation (dash-dotted curve) based on the model of Figure 2.10. The large discrepancy is entirely due to the structure in the liquid. If this is included, a proper fit is obtained (solid curve) in which the structure of the crystal side of the interface still agrees with the earlier result.

By using a geometry with a very thin water layer, Reedijk *et al.* [55] have been able to measure a full set of rods with sufficient accuracy such that the ordering in the liquid was revealed in the non-specular rods. Figure 2.16 shows the amount of liquid order probed for

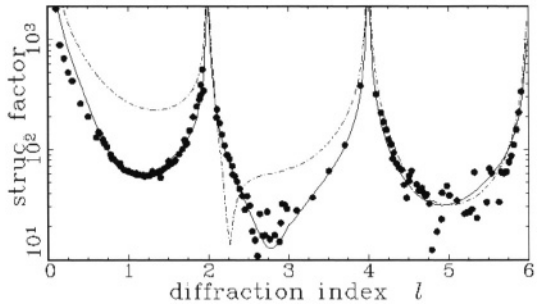


FIGURE 2.15. The structure factor amplitude of the specular rod for KDP{101} in aqueous solution. Circles are the experimental data points, the dashed and solid curves are model calculations in which the layering effects in the growth solution are ignored and included, respectively.

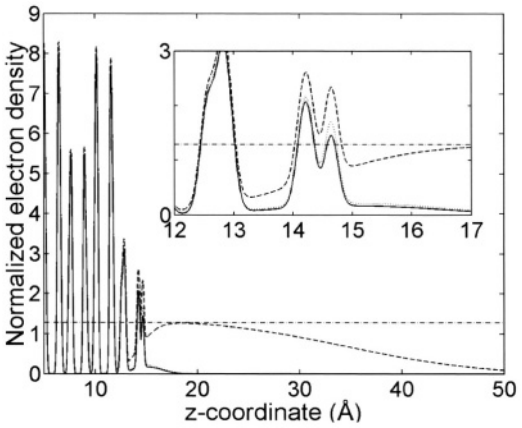


FIGURE 2.16. The electron density distribution across the interface of KDP{101} with water. The density is normalised to that of water and is shown for three different parallel momentum transfers: (00)-rod (dashed line), (10)-rod (dotted line) and (20)-rod (solid line). The (00)-rod shows all the water, the other two rods only the water that has (partial) in-plane order.

a number of parallel momentum transfers. The (00)-profile is the 'true' density distribution, and reveals two well-ordered water layers together with a film of about 25 Å thickness that has the bulk structure. In the (10) and (20) Fourier components, the bulk water is not visible, and only the two icelike layers and small amounts of the next layers are seen. This result follows the behavior expected for such an interface in which the liquid rapidly becomes bulklike.

2.3.4. Liquid monolayers

As a simple model for a 3D solid–liquid interface, it can be advantageous to study the structure of liquid monolayers on a crystal surface. In such systems, there is no need to penetrate a macroscopically thick film, there are no concentration gradients and there is no background signal from the film. Typically, such experiments are performed in the

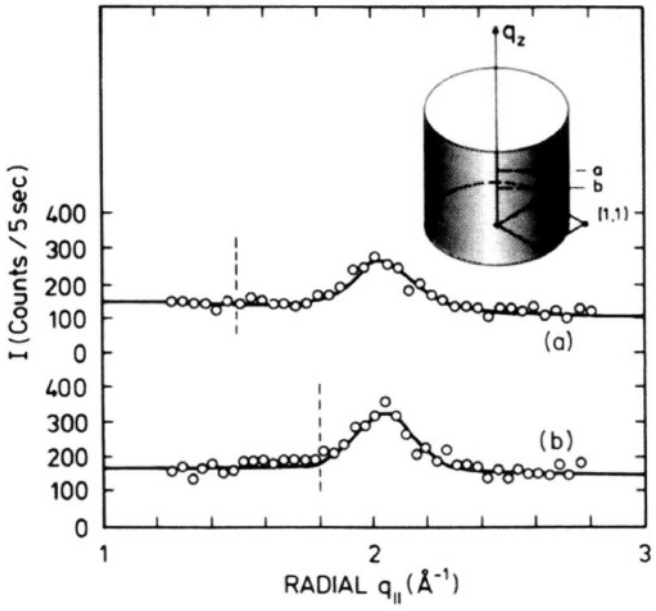


FIGURE 2.17. Radial scans for liquid Pb on Ge(111) revealing that the diffuse scattering from the quasi-liquid Pb has a cylindrical shape as expected for a 2D layer. The vertical dashed lines indicate where a diffuse peak is expected in case the liquid is 3D, i.e., when the diffuse scattering has a spherical shape. From ref. [62].

controlled environment of an ultra-high vacuum system. The simplicity of such a system often allows a more detailed structure determination than in the case of a macroscopically thick liquid. In particular it becomes possible to observe the scattering originating from the pair correlation function in the liquid. In addition to being a model for a 3D solid–liquid interface, the behavior and phase transitions of such 2D systems are also of fundamental interest [8].

In this case various techniques other than X-ray scattering can be used, since the penetration depth is no longer an issue. Examples are reflection high-energy electron diffraction (RHEED) [56], low-energy electron diffraction (LEED) [57] and STM [58]. We will limit ourselves here to X-ray diffraction. Note that the physics of such monolayers is similar to that of intercalation compounds, that have also been studied using X-ray diffraction [59].

Monolayers on a crystal are a popular subject in surface science, but actual melting has received far less attention. Pb and Sn are often used overlayers for this latter type of study, because they combine a convenient melting point with a low vapor pressure. The first X-ray investigation of this type was done by Marra *et al.* [60] in 1982, just one year after the very first surface X-ray diffraction experiment [61].

The system that has received most attention is Pb on Ge(111). This is an almost ideal model system, because Pb and Ge don't intermix (at the relevant temperatures), Pb scatters X-rays strongly and Ge provides a highly ordered substrate. Following the observation of liquid Pb using RHEED [56], Grey *et al.* performed careful X-ray diffraction experiments on this system [62]. Radial scans revealed the presence of the diffuse ring of scattering

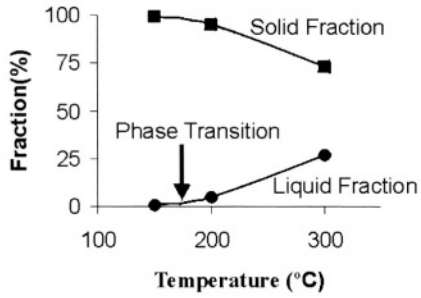


FIGURE 2.18. The fraction of Sn atoms on Ge(111) in the solid state (squares) and liquid state (circles) at different temperatures. Above the melting transition (indicated by an arrow), the system still continues to increase its liquid-like character. From ref. [66].

originating from the pair correlation function of the 2D-liquid. By performing such scans for different perpendicular momentum transfers, Grey *et al.* showed that the scattering indeed has the expected cylindrical from, see Figure 2.17. This is a realization of the situation sketched earlier in Figure 2.8b. In addition the liquid was found to be anisotropic, owing to the interaction with the substrate. This was the first observation of the diffuse ring from a monolayer film.

Despite the clear presence of a liquid ring, the structure of the molten Pb on the Ge(111) surface became controversial, because STM experiments indicated a well-ordered structure above the phase transition [58]. It appears that both observations reveal part of the properties of the system. From molecular dynamics simulations [63] and new X-ray diffraction experiments [64] a picture emerges in which the 'liquid' overlayer has both solid and liquid properties (a solid–liquid *layer*). We will illustrate this by discussing the similar system of Sn on Ge(111) [65,66]. At a coverage of 1.5 monolayer, the Sn layer melts at a temperature of 170°C. Below the melting point the Sn is well ordered, above the melting point the properties of the layer *gradually* change from more solidlike to more liquidlike. As a function of temperature, the system thus evolves from the situation sketched in Figure 2.8c to that in Figure 2.8d. At low temperatures the fully ordered Sn contributes to all diffraction rods. For increasing temperatures the non-specular rods contain a decreasing contribution from the Sn. Only the specular rod continues to reveal the total amount of Sn present. By measuring full data sets, Reedijk *et al.* were able to quantify the amount of liquidlike and solidlike Sn, see Figure 2.18. An important observation is that the liquid fraction continues to increase, even above the phase transition. The Sn layer consists thus of rapidly diffusing atoms that reside preferentially on sites imposed by the Ge(111) lattice. At this high coverage these sites are found to be not the high symmetry sites because of geometrical reasons. Figure 2.19 illustrates the relative occupancy of the various sites at 300°C; some sites are clearly more occupied than others. The figure does not show the 25% liquid fraction that could in principle be represented by a uniform gray background in the figure. At any moment there are thus liquidlike and solidlike atoms, but they are not separate species, but instead exchange their character as a function of time. The liquid fraction can be considered as the time that the atoms cannot be associated with a particular site. The atoms in this fraction cause the liquid ring that has also been observed on this system [66].

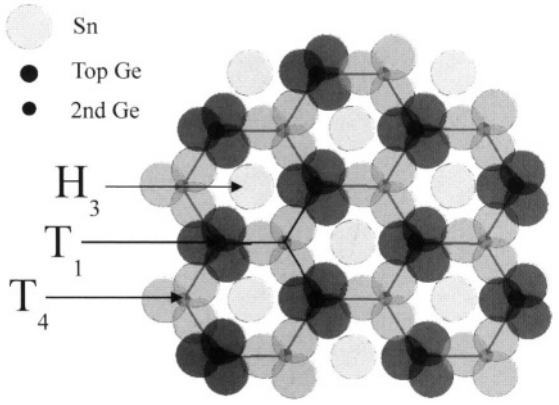


FIGURE 2.19. Schematic top view of the time-averaged structure of Sn on Ge(111) at 300°C. The grayscale of the Sn atoms approximately indicates the occupancy. Not all positions can be occupied at the same time, e.g., only one of the three positions near the $\mathbf{T_1}$ site can be occupied. Averaged over time, however, all equivalent positions will have the same occupancy. The indicated Sn atoms account for 75% of the occupancy, the remaining 25% is not associated with any particular lattice site and can be considered as the liquid fraction.

At very high temperatures one can view the system as a liquid that is modulated by the potential of the substrate [59,67]. The strength of this modulation depends on the depth of the well and the 'lattice' match with the substrate. A liquid ring will only be observed if the modulation is weak. In the context of surface melting, the presence of such a ring was reported on the surface of Pb(110) near its melting point [68], but later it was shown that the scattering was due to thermal diffuse scattering from the bulk crystal [69]. The fact that in this case no liquid ring has been observed, could well be caused by the perfect match between liquid and solid.

We need, finally, to mention the recent work of Reichert *et al.* [70] on the structure of thick layers of liquid lead in contact with the (100) surface of silicon. The aim of that work was less the structure at the interface itself, because the interactions between solid and liquid were found to be less important than those between the liquid lead atoms. The main purpose of the solid was to induce an orientation to liquid lead clusters such that the predicted, local five-fold symmetry could be observed. When the solid–liquid interactions are weak, it is thus possible that the solid does (almost) not influence the liquid.

2.4. CONCLUSIONS

Even though the study of solid–liquid interfaces using X-ray diffraction is fairly recent, several important results have been obtained. It is clear that the surface of a crystal in solution may have a structure that strongly deviates from that in the bulk. When describing processes such as crystal growth and electrochemical reactions, such deviations need to be taken into account. It has further been shown that the liquid layers in contact with the solid show extra ordering, typically over a total of three to five layers. The amount of this ordering varies from system to system: ice-like layers are found at a number of ionic

systems, while in other cases no lateral ordering is present and only layering occurs. In general one should assign both solid and liquid properties to such layers, so in fact one can say that such solid–liquid *layers* are part of the solid–liquid *interface.*

The coming years will undoubtedly see more results, including the role played by different solvents and by the match between crystal and liquid. X-ray diffraction will remain an important tool in order to obtain a better understanding of the solid–liquid interface.

REFERENCES

1. W.A. Curtin, Density-functional theory of the solid–liquid interface, *Phys. Rev. Lett.* **59**, 1228–1231 (1987); F. Spaepen, A structural model for the solid–liquid interface in monatomic systems, *Acta Met.* **23**, 729–743 (1975).
2. W. Han and S.M. Lindsay, Probing molecular ordering at a liquid-solid interface with a magnetically oscillated atomic force microscope, *Appl. Phys. Lett.* **72**, 1656–1658 (1998).
3. S.E. Donnelly, R.C. Birtcher, C.W. Allen, I. Morrison, K. Furuya, M. Song, K. Mitsuishi and U. Dahmen, Ordering in a fluid intert gas confined by flat surfaces, *Science* **296**, 507–510 (2002).
4. R. Feidenhans'l, Surface structure determination by X-ray diffraction, *Surf. Sci. Rep.* **10**, 105–188 (1989); I.K. Robinson and DJ. Tweet, Surface X-ray diffraction, *Rep. Prog. Phys.* **55**, 599–651 (1992); I.K. Robinson, in: *Handbook on Synchrotron Radiation,* Eds G.S. Brown and D.E. Moncton (North-Holland, Amsterdam, 1991), Vol. 3, pp. 221–266.
5. X.Y. Liu, E.S. Boek, W.J. Briels and P. Bennema, Prediction of crystal growth morphology based on structural analysis of the solid–fluid interface, *Nature* **374**, 342–345 (1995).
6. E. Vlieg, Understanding crystal growth in vacuum and beyond, *Surf. Sci.* **500**, 458–474 (2002).
7. Z. Nagy and H. You, Applications of surface X-ray scattering to electrochemistry problems, *Electrochim. Acta* **47**, 3037–3055 (2002).
8. M.A. Glaser and N.A. Clark, Melting and liquid structure in two dimensions, *Adv. Chem. Phys.* **83**, 543–709 (1993).
9. J.W.M. Frenken and H.M. van Pinxteren, in: *The Chemical Physics of Solid Surfaces,* Eds D.A. King and D.P. Woodruff (Elsevier, Amsterdam, 1993), Vol. 7, p. 259.
10. B. Bhushan, J.N. Israelachvili and U. Landman, Nanotribology: friction, wear and lubrication at the atomic scale, *Nature* **374**, 607–616 (1995).
11. M.J. Zwanenburg, J.H.H. Bongaerts, J.F. Peters, D.O. Riese and J.F. van der Veen, X-ray waveguiding studies of ordering phenomena in confined fluids, *Phys. Rev. Lett.* **85**, 5154–5157 (2000).
12. S.R. Andrews and R.A. Cowley, Scattering of X-rays from crystal surfaces, *J. Phys. C: Solid State Phys.* **18**, 6427–6439 (1985).
13. I.K. Robinson, Crystal truncation rods and surface roughness, *Phys. Rev. B* **33**, 3830–3836 (1986).
14. J. Als-Nielsen, in: *Handbook on Synchrotron Radiation,* Eds G.S. Brown and D.E. Moncton (Elsevier, Amsterdam, 1991), Vol. 3.
15. E. Vlieg, Integrated intensities using a six-circle surface X-ray diffractometer, *J. Appl. Cryst.* **30**, 532–543 (1997).
16. H. You, C.A. Melendres, Z. Nagy, V.A. Maroni, W. Yun and R.M. Yonco, X-ray reflectivity study of the copper–water interface in a transmission geometry under in situ electrochemical control, *Phys. Rev. B* **45**, 11288–11298 (1992).
17. 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, X-ray diffraction studies of potassium dihydrogen phosphate (KDP) crystal surfaces, *J. Cryst. Growth* **205**, 202–214 (1999).
18. B.E. Warren, *X-ray Diffraction* (Dover, New York, 1990).
19. E. Vlieg, J.F. van der Veen, S.J. Gurman, C. Norris and J.E. Macdonald, X-ray diffraction from rough, relaxed and reconstructed surfaces, *Surf. Sci.* **210**, 301–321 (1989).
20. E. Vlieg, ROD, a program for surface crystallography, *J. Appl. Cryst.* **33**, 401–405 (2000).
21. C.S. Lent and P.I. Cohen, Diffraction from stepped surfaces – I. Reversible surfaces, *Surf. Sci.* **139**, 121 (1984); S.K. Sinha, E.B. Sirota, S. Garoff and H.B. Stanley, X-ray and neutron-scattering from rough surfaces, *Phys. Rev. B* **38**, 2297–2311 (1988).

22. P. Hartman and W.G. Perdok, On the relations between structure and morphology of crystals I, *Acta Cryst.* **8**, 49–52 (1955).
23. M. Lahav and L. Leiserowitz, The effect of solvent on crystal growth and morphology, *Chem. Eng. Sci.* **56**, 2245–2253 (2001).
24. D. Cunningham, R.J. Davey, K.J. Roberts, J.N. Sherwood and T. Shripathi, Structural studies of the crystal/solution interface using synchrotron radiation, *J. Cryst. Growth* **99**, 1065–1069 (1990); K.J. Roberts, J.N. Sherwood, K. Wojciechowski, A.A. Chernov, Y.G. Kuznetsov and R.J. Davey, In situ surface X-ray diffraction using high-energy synchrotron radiation: studies of the ADP(100) and (101) growth interfaces under aqueous solution growth conditions, *J. Phys. D: Appl. Phys.* **26**, A107–A114 (1993).
25. R.P. Chiarello and N.C. Sturchio, The calcite (1014) cleavage surface in water: early results of a crystal truncation rod study, *Geochim. Cosmochim. Acta* **59**, 4557–4561 (1995).
26. D. Gidalevitz, R. Feidenhans'l, S. Matlis, D.-M. Smilgies, M.J. Christensen and L. Leiserowitz, Monitoring in situ growth and dissolution of molecular crystals: towards determination of the growth units, *Angew. Chem., Int. Ed. Engl.* **36**, 955–959 (1997).
27. 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, Surface atomic structure of KDP crystals in aqueous solution: an explanation of the growth shape, *Phys. Rev. Lett.* **80**, 2229–2232 (1998).
28. J. Arsic, M.F. Reedijk, A.J.R. Sweegers, Y.S. Wang and E. Vlieg, Compression versus expansion on ionic crystal surfaces, *Phys. Rev. B* **64**, 233402 (2001).
29. T.A. Land, T.L. Martin, S. Potapenko, G.T. Palmore and J.J. De Yoreo, Recovery of surfaces from impurity poisoning during crystal growth, *Nature* **399**, 442–145 (1999).
30. P. Fenter and N.C. Sturchio, Structure and growth of stearate monolayers on calcite: first results of an in situ X-ray reflectivity study, *Geochim. Cosmochim. Acta* **63**, 3134–3152 (1999); N.C. Sturchio, R.P. Chiarello, L. Cheng, P.F. Lyman, M.J. Bedzyk, Y. Qian, H. You, D. Yee, P. Geissbühler, L.B. Sorensen, Y. Liang and D.R. Baer, Lead adsorption at the calcite–water interface: synchrotron X-ray standing wave and X-ray reflectivity studies, *Geochim. Cosmochim. Acta* **61**, 251–263 (1997).
31. M.L. Farquhar, R.A. Wogelius and C.T. Tang, In situ synchrotron X-ray reflectivity study of the oligoclase feldspar mineral–fluid interface, *Geochim. Cosmochim. Ada* **63**, 1587–1594 (1999); P. Fenter, H. Teng, P. Geissbühler, J.M. Hanchar, K.L. Nagy and N.C. Sturchio, Atomic-scale structure of the orthoclase (001)–water interface measured with high-resolution X-ray reflectivity, *Geochim. Cosmochim. Acta* **64**, 3663–3673 (2000).
32. M.A. Henderson, The interaction of water with solid surfaces: fundamental aspects revisited, *Surf. Sci. Rep.* **46**, 1–308 (2002).
33. D.J. Trevor, C.E.D. Chidsey and D.N. Loiacono, In situ scanning-tunneling microscope observations of roughening, annealing, and dissolution of gold (111) in an electrochemical cell, *Phys. Rev. Lett.* **62**, 929–932 (1989).
34. E.M. Stuve and N. Kizhakevariam, Chemistry and phyiscs of the "liquid"/solid interface: a surface science perspective, *J. Vac. Sci. Technol., A* **11**, 2217–2224 (1993).
35. C.A. Lucas, Surface relaxation at the metal/electrolyte interface, *Electrochim. Acta* **47**, 3065–3074 (2002).
36. Z. Nagy, H. You and R.M. Yonco, Cell design for in-situ X-ray scattering studies of metal solution interfaces under electrochemical control, *Rev. Sci. Instrum.* **65**, 2199–2205 (1994).
37. M.G. Samant, M.F. Toney, G.L. Borges, L. Blum and O.R. Melroy, Grazing incidence X-ray diffraction of lead monolayers at a silver (111) and gold (111) electrode/electrolyte interface, *J. Phys. Chem.* **92**, 220–225 (1988).
38. B.M. Ocko, G.M. Watson and J. Wang, Structure and electrocompression of electrodeposited iodine monolayers on Au(111), *J. Phys. Chem.* **98**, 897–906 (1994); M.F. Toney, J.G. Gordon, M.G. Samant, G.L. Borges, O.R. Melroy, D. Yee and L.B. Sorensen, Underpotentially deposited thallium on silver (111) by in situ surface X-ray scattering, *Phys. Rev. B* **45**, 9362–9374 (1992); C.A. Lucas, N.M. Markovic and P.N. Ross, Observation of an ordered bromide monolayer at the Pt(111)-solution interface by in-situ surface X-ray scattering, *Surf. Sci.* **340**, L949–L954 (1995).
39. J. Wang, A. Davenport, H.S. Iscaacs and B.M. Ocko, Surface-charge induced ordering of the Au(111) surface, *Science* **255**, 1416–1418 (1992).
40. E. Vlieg, I.K. Robinson and K. Kern, Relaxations in the missing-row structure of the (1 × 2) reconstructed surfaces of Au(110) and Pt(110), *Surf. Sci.* **233**, 248–254 (1990).

41. B.M. Ocko, G. Helgesen, B. Schardt and J. Wang, Charge induced (1 × 3) reconstruction of the Au(110) surface: an X-ray scattering study, *Phys. Rev. Lett.* **69**, 3350–3353 (1992).
42. C.A. Lucas, N.M. Markovic and P.N. Ross, Surface structure and relaxation at the Pt(l10)/electrolyte interface, *Phys. Rev. Lett.* **77**, 4922–4925 (1996).
43. J.W.M. Frenken, R.L. Krans, J.F. van der Veen, E. Holub-Krappe and K. Horn, Missing-row surface reconstruction of Ag(l10) induced by potassium adsorption, *Phys. Rev. Lett.* **59**, 2307–2310 (1987).
44. H. Saka, A. Sakai, T. Kamino and T. Imura, *In situ* HREM observation of solid–liquid interface, *Philos. Mag. A* **52**, L29–L32 (1985); S. Arai, S. Tsukimoto, S. Muto and H. Saka, Direct observation of the atomic structure in a solid–liquid interface, *Microsc. Microanal.* **6**, 358–361 (2000).
45. B.M. Ocko, Smectic-layer growth at solid interfaces, *Phys. Rev. Lett.* **64**, 2160–2163 (1990).
46. J. Wang, B.M. Ocko, A.J. Davenport and H.S. Isaacs, In situ X-ray diffraction and reflectivity studies of the Au(111)/electrolyte interface: reconstruction and anion adsorption, *Phys. Rev. B* **46**, 10321–10338 (1992).
47. 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, Voltage-dependent ordering of water molecules at an electrodo–electrolyte interface, *Nature* **368**, 444–146(1994).
48. 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, Layering of a liquid metal in contact with a hard wall, *Nature* **390**, 379–381 (1997).
49. C.-J. Yu, A.G. Richter, A. Datta, M.K. Durbin and P. Dutta, Observation of molecular layering in thin liquid films using X-ray reflectivity, *Phys. Rev. Lett.* **82**, 2326–2329 (1999); C.-J. Yu, A.G. Richter, J. Kmetko, S.W. Dugan, A. Datta and P. Dutta, Structure of interfacial liquids: X-ray scattering studies, *Phys. Rev. E* **63**, 021205 (2001).
50. L. Cheng, P. Fenter, K.L. Nagy, M.L. Schlegel and N.C. Sturchio, Molecular-scale density oscillations in water adjacent to a mica surface, *Phys. Rev. Lett.* **87**, 156103 (2001).
51. P.B. Miranda, L. Xu, Y.R. Shen and M. Salmeron, Icelike water monolayer adsorbed on mica at room temperature, *Phys. Rev. Lett.* **81**, 5876–5879 (1998).
52. M. Odelius, M. Bernasconi and M. Parrinello, Two dimensional ice adsorbed on mica surface, *Phys. Rev. Lett.* **78**, 2855–2858 (1997); S.-H. Park and *G.* Sposito, Structure of water adsorbed on a mica surface, *Phys. Rev. Lett.* **89**, 085501 (2002).
53. Y.S. Chu, T.E. Lister, W.G. Cullen, H. You and Z. Nagy, Commensurate water monolayer at the RuO_2(110)/water interface, *Phys. Rev. Lett.* **86**, 3364–3367 (2001).
54. T.E. Lister, Y.S. Chu, W.G. Cullen, H. You, R.M. Yonco, J.F. Mitchell and Z. Nagy, Electrochemical and X-ray scattering study of well defined RuO_2 single crystal surfaces, *Journal of Electroanalytical Chemistry* **524–525**, 201–218 (2002).
55. M.F. Reedijk, J. Arsic, F.F.A. Hollander, S.A. de Vries and E. Vlieg, Liquid order at the interface of KDP crystals with water: evidence for icelike layers, *Phys. Rev. Lett.* **90**, 066103, 1–4 (2003).
56. T. Ichikawa, On the structure of monolayer liquid Pb on Ge(111) surfaces, *Solid State Commun.* **49**, 59–64 (1984).
57. B.H. Muller, T. Schmidth and M. Henzler, Growth and melting of a Pb monolayer on Cu(111), *Surf. Sci.* **376**, 123–132(1997).
58. I.-S. Hwang and J.A. Golovchenko, Phase transition of monolayer Pb/Ge(111): β-$(\sqrt{3} \times \sqrt{3})$R30° ↔ 1 × 1 at ~180°C, *Phys. Rev. B* **50**, 18535–18542 (1994).
59. S.C. Moss, G. Reiter, J.L. Robertson, C. Thompson and J.D. Fan, X-ray determination of the substrate modulation potential fora two-dimensional Rb liquid in graphite, *Phys. Rev. Lett.* **57**, 3191–3194 (1986).
60. W.C. Marra, P.H. Fuoss and P.E. Eisenberger, X-ray diffraction studies: melting of Pb monolayers on Cu(110) substrates, *Phys. Rev. Lett.* **49**, 1169–1172 (1982).
61. P. Eisenberger and W.C. Marra, X-ray diffraction study of the Ge(001) reconstructed surface, *Phys. Rev. Lett.* **46**, 1081–1084 (1981).
62. F. Grey, R. Feidenhans'l, J.S. Pedersen, M. Nielsen and R.L. Johnson, Pb/Ge(111)1 × 1: An anisotropic two-dimensional liquid, *Phys. Rev. B* **41**, 9519–9522 (1990).
63. F. Ancilotto, A. Selloni and R. Car, Beta phase of Pb on Ge(111): the competing roles of electronic bonding and thermal fluctuations, *Phys. Rev. Lett.* **71**, 3685–3688 (1993).
64. S.A. de Vries, P. Goedtkindt, P. Steadman and E. Vlieg, Phase transition of a Pb monolayer on Ge(111), *Phys. Rev. B* **59**, 13301–13308 (1999).
65. T. Ichikawa, Structure anomaly of liquid Sn layers adsorbed on Ge(111) surfaces, *Solid State Commun.* **63**, 1173–1177(1987).

66. M.F. Reedijk, J. Arsic, F.K. de Theije, M.T. McBride, K.F. Peters and E. Vlieg, Structure of liquid Sn on Ge(111), *Phys. Rev. B* **64**, 033403, 033404 (2001).
67. M.F. Toney, J.G. Gordon, M.G. Samant, G.L. Borges, O.R. Melroy, L.S. Kau, D.G. Wiesler, D. Yee and L.B. Sorensen, Surface X-ray scattering measurments of the substrate-induced spatial modulation of an incommensurate adsorbed monolayer, *Phys. Rev. B* **42**, 5594–5603 (1990); E. Vives and P.-A. Lingard, Substrate influence on two-dimensional solids and liquids: a Monte Carlo simulation study, *Phys. Rev. B* **44**, 1318–1328(1991).
68. P.H. Fuoss, L.J. Norton and S. Brennan, X-ray scattering studies of the melting of lead surfaces, *Phys. Rev. Lett.* **60**, 2046–2049 (1988).
69. H.M. van Pinxteren, S. Chandavarkar, W.J. Huisman, J.M. Gay and E. Vlieg, Thermal diffuse scattering from surface-melted Pb(110), *Phys. Rev. B* **51**, 14753–14755 (1995).
70. H. Reichert, O. Klein, H. Dosch, M. Denk, V. Honkimaki, T. Lippmann and G. Reiter, Observation of five-fold local symmetry in liquid lead, *Nature* **408**, 839–841 (2000).

This page intentionally left blank

3

Interactions at solid–fluid interfaces

Aleksandr Noy*
Chemistry and Materials Science, Lawrence Livermore National Laboratory, L-234, 7000 East Ave, Livermore, CA 94550, USA

3.1. INTRODUCTION

Solid–fluid interfaces occupy a special place in condensed matter science since most of the important events in chemistry, physics and biology happen at a these interfaces. Examples of such processes span many orders of magnitude on the length scale and extend from capillarity and lubrication on a macroscopic scale to self-assembly and colloidal interactions on the mesoscopic scale to molecular recognition and membrane interactions at the nanoscale. These events may seem very complex and diverse, yet their outcomes are largely determined by the interactions among molecular assemblies that comprise the surfaces and surrounding liquid phase.

Detailed knowledge of the range and magnitude of these interactions is required for developing a fundamental understanding of these processes and for gaining precise control over their outcome. Potential technological and scientific benefits are enormous. Much of the current progress in nanotechnology hinges upon the bottom-up approach that harnesses the power of self-assembly methods that utilize precisely tailored intermolecular interactions to form a particular structure [1]. Understanding the interactions that govern self-assembly can help us to realize many of nanotechnology promises. Solid–fluid interfaces are also central for conventional technology. For example, 2% of the gross domestic product of the UK is lost to unwanted friction and wear in the devices [2]. “Smart” lubricants that are precisely tailored to the task can drastically reduce such losses. The benefits to the life sciences are equally impressive. Knowledge of the interactions forces acting between distinct functional groups in proteins and membrane structures will be instrumental to molecular biology and rational drug design. Progress in proteomics is currently limited by the rate of obtaining protein crystals suitable or X-ray diffraction. Understanding and controlling the interactions between the solution phase particle and crystal surface will enable an enormous increase in the efficiency of protein crystallization efforts. This list can continue.

*E-mail address: noy1@llnl.gov

At present, our progress in understanding interactions at the interfaces is closely linked to the development of the physical techniques for measuring such interactions. Typical chemical bond strength ranges from 10 pN for a very weak van der Waals bond to 2–10 nN for a strong covalent bond. This interaction strength also varies significantly depending on the surrounding medium. Therefore, an ideal technique for measuring interactions at the interfaces must be capable of measuring forces in this range with sub-nanometer distance resolution in a variety of environments and solvent media. Recent two decades saw an emergence of several such techniques that gave researchers the capability to measure interaction strength on the relevant length scale and often on a single molecule level. For the first time, it became possible to probe the fundamental forces that determine nanoscale interactions. We will survey some of these forces while paying particular attention to the techniques that are used to probe them.

3.2. OVERVIEW OF THE INTERACTIONS AT SOLID–FLUID INTERFACES

We can divide intermolecular interactions in condensed phases loosely into three groups (in decreasing order of generality): van der Waals interactions; chemical interactions; Coulomb (electrostatic) interactions (inevitably complicated by the presence of the double layers); and "collective" interactions that comprise solvation, structural and hydrophobic forces. We will begin by describing each of these interactions.

3.2.1. Van der Waals forces

The most ubiquitous forces in molecular systems originate in the interactions that involve electrostatic interactions between dipoles or induced dipoles. These interactions that involve polarization of molecules comprise van der Waals interactions. Often, we can further subdivide these interactions into three categories, according to the nature of interacting species.

When two permanent dipoles interact with each other the resulting force tries to reorient the dipoles to maximize the interaction. This orientation force is counteracted by the thermal motion that tries to randomize molecular motion. If the energy of the dipole–dipole interaction is smaller than the thermal energies the molecules continue to rotate freely. Therefore the dipole–dipole interaction energy gets averaged over all the orientations of molecules. Such orientation-averaged interaction comprises *Keesom forces* or *orientation force*. The energy of the orientation interaction between two dipoles with dipole moments μ_1 and μ_2 separated by the distance r in the medium with permeability ε is given by:

$$W_{\text{orient}}(r) = -\frac{\mu_1^2 \mu_2^2}{3(4\pi\varepsilon_0\varepsilon)^2 k_B T} \cdot \frac{1}{r^6}. \tag{1}$$

Israelachvili noted that most of molecular systems would exhibit Keesom type interaction (i.e. the orientation interaction is weaker than thermal motion) [3]. However, in a few special cases, such as O–H, N–H and F–H bonds the dipole moments are large enough to counteract the thermal motion and lead to the permanent alignment of polar molecules in condensed state. This special case of orientation force plays a rather important role in

biological systems which leads most researchers to separate them into a special class of *hydrogen bonding* interaction.

Another component of van der Waals interactions arises when a permanent dipole interacts with a non-polar molecule. If the polar molecule has non-zero polarizability, then the field of the permanent dipole will induce a dipole moment. The interactions of a permanent dipole with the induced dipole are the *Debye forces* or *induction forces.* The energy of the induction interaction between a dipole μ_1 and a molecule with the polarizability α_{02} is given by

$$W_{\text{ind}}(r) = -\frac{\mu_1^2 \alpha_{02}}{(4\pi\varepsilon_0\varepsilon)^2} \cdot \frac{1}{r^6}. \tag{2}$$

The most general component of the van der Waals forces—*dispersion force*—arises from a quantum mechanical effect of the induced dipole interaction. Rigorous treatment of the dispersion force origin is beyond the scope of this work. Suffice to say that intuitively dispersion forces originate in the interaction of the transient dipoles that arise as electrons orbit around the nuclei. While the overall (average) dipole moment can be zero, these transient dipoles are not small (even for a hydrogen atom the transient dipole moment is about 2.4 Debye!); therefore the resulting force is not insignificant. In fact, dispersion forces most often make a dominating contribution to the overall van der Waals interaction. A classic model developed by London gives the energy of a dispersion interaction of two dissimilar atoms as:

$$W_{\text{disp}}(r) = -\frac{3\alpha_{01}\alpha_{02} h \nu_1 \nu_2}{2(\nu_1 + \nu_2)(4\pi\varepsilon_0\varepsilon)^2} \cdot \frac{1}{r^6}, \tag{3}$$

where ν_1 and ν_2 are orbiting frequencies of electrons. Therefore, an overall van der Waals force can be given as a sum of all three components of the interaction:

$$W_{\text{vdW}}(r) = -\left(\left(\mu_1^2 \alpha_{02} + \mu_2^2 \alpha_{01}\right) + \frac{\mu_1^2 \mu_2^2}{3k_B T} + \frac{3\alpha_{01}\alpha_{02} h \nu_1 \nu_2}{2(\nu_1 + \nu_2)} \right) \cdot \frac{1}{(4\pi\varepsilon_0\varepsilon)^2 r^6}. \tag{4}$$

Total van der Waals force is given by the derivative of the eq. (4) with respect to distance *r:*

$$F_{\text{vdW}}(r) = \frac{\mathrm{d}W}{\mathrm{d}r} = 6 \cdot \left(\left(\mu_1^2 \alpha_{02} + \mu_2^2 \alpha_{01}\right) + \frac{\mu_1^2 \mu_2^2}{3k_B T} + \frac{3\alpha_{01}\alpha_{02} h \nu_1 \nu_2}{2(\nu_1 + \nu_2)} \right) \cdot \frac{1}{(4\pi\varepsilon_0\varepsilon)^2 r^7}. \tag{5}$$

In general, van der Waals interactions are non-additive (i.e. the total interaction is not equal to the sum of pair-wise interactions), but most of the models assume interaction additivity as an approximation. The overall interaction potential decays as r^6, therefore the resulting forces are short-ranged in comparison with other non-covalent interactions (covalent bonding interactions are intrinsically quantum mechanical in nature and are even more short-ranged). A quick estimate shows that for two atoms sitting at a distance of 10 Å from one another in vaccum the attractive van der Waals force will be on the order of 0.02 pN

(we assumed zero dipole moment, orbital frequency of 3.03×10^{15} Hz and polarizability 1.36×10^{-20} F m^2). However, if the distance between the atoms is only 5 Å, then the van der Waals attraction becomes as big as 2.6 pN. When the distance between atoms approaches 2 Å then van der Waals attraction force given by the London formula reaches an enormous 1.5 nN. Of course, in dielectric medium screening weakens the interactions forces, yet this example does illustrate that van der Waals forces become formidably strong at close distances.

McLachlan [4] developed a generalized theory of van der Waals forces that corrected some of the notable shortcomings of the London model, namely its inability to treat interactions of molecules in solvents. McLachlan's model treatment is more realistic because it considers the whole absorption spectrum of a molecule instead of only a single frequency featured in London's model. The sum of a static dipolar contribution and a nonzero frequency dispersion contribution gives the total interaction energy between two small molecules 1 and 2 of radii a_1 and a_2 in the media 3:

$$W(r) = -\left[3k_B T a_1^3 a_2^3 \cdot \left(\frac{\varepsilon_1(0) - \varepsilon_3(0)}{\varepsilon_1(0) + 2\varepsilon_3(0)}\right) \cdot \left(\frac{\varepsilon_2(0) - \varepsilon_3(0)}{\varepsilon_2(0) + 2\varepsilon_3(0)}\right) + \frac{3h a_1^3 a_2^3}{\pi} \cdot \int_0^\infty \left(\frac{\varepsilon_1(i\nu) - \varepsilon_3(i\nu)}{\varepsilon_1(i\nu) + 2\varepsilon_3(i\nu)}\right) \cdot \left(\frac{\varepsilon_2(i\nu) - \varepsilon_3(i\nu)}{\varepsilon_2(i\nu) + 2\varepsilon_3(i\nu)}\right) \mathrm{d}\nu\right] \cdot \frac{1}{r^6}, \quad (6)$$

where $\varepsilon(0)$ is the static dielectric permittivity of medium and $\varepsilon(i\nu)$ is the dielectric permittivity of a medium at imaginary frequency $i\nu$. Ultimately, in case where all media have one strong absorption peak frequency ν_e, we can approximate the total van der Waals free energy as:

$$W(r) = -\left[3k_B T \left(\frac{\varepsilon_1(0) - \varepsilon_3(0)}{\varepsilon_1(0) + 2\varepsilon_3(0)}\right)\left(\frac{\varepsilon_2(0) - \varepsilon_3(0)}{\varepsilon_2(0) + 2\varepsilon_3(0)}\right) + \frac{\sqrt{3} h \nu_e}{4} \cdot \frac{(n_1^2 - n_3^2)(n_2^2 - n_3^2)}{(n_1^2 + 2n_3^2)^{1/2}(n_2^2 + 2n_3^2)^{1/2}((n_1^2 + 2n_3^2)^{1/2} + (n_2^2 + 2n_3^2)^{1/2})}\right] \cdot \frac{a_1^3 a_2^3}{r^6}. \quad (7)$$

Equation (7) shows several important characteristic of the van der Waals interaction. It predicts a large decrease in the dispersion forces in solvent medium with respect to the interaction strength in vacuum. Also, eq. (7) predicts that symmetric interactions (i.e. interactions between identical molecules) are always attractive. Yet, eq. (7) also shows that van der Waals forces can be repulsive, as they are when the refractive index of the medium n_3 lies between the refractive indices of the interacting species n_1 and n_2.Incidentally, this model shows the physical rationale behind the well-known empirical principle of "like likes like" and the phenomena of separation of immiscible liquids, such as water and oil.

In all cases that we discussed van der Waals forces follow a simple inverse power law. This property allows analytical solutions for a few simple practical cases, such as a molecule–surface interactions or a dielectric spherical particle interacting with the surface.

For all these cases we assume that van der Waals interactions are additive. To get the total interaction energy we simply sum up the interactions between the molecule and each molecule comprising the surface. In case of the molecule interacting at a distance D with the surface of density ρ, we arrive to a simple expression for the interaction energy:

$$W(D) = -\frac{\pi C\rho}{6D^3}. \tag{8}$$

Here we assume that the van der Waals interaction energy between our molecule and a molecule of the surface is proportional to inverse sixth power of the distance and C is the proportionality coefficient. For the case of large sphere or radius R interacting with the surface at a distance D the interaction energy is given by:

$$W(D) = -\frac{\pi^2 C\rho_1\rho_2}{6D} \quad \text{when } D \ll R, \tag{9}$$

$$W(D) = -\frac{\pi C\rho_2(4\pi R^3\rho_1/3)}{6D^3} \quad \text{when } D \gg R. \tag{10}$$

Note that eqs (10) and (8) predict the same functional dependence for the interaction energy, which is not surprising since we can effectively treat a sphere at a large distance as a point object.

Finally, a common way to present these equations is to introduce a Hamaker constant, $A \equiv \pi^2 C\rho_1\rho_2$. Then the interaction energy between the sphere and the surface at close distances will simply be:

$$W(D) = -\frac{AR}{6D} \tag{11}$$

and the interaction energy of two spheres of radii R_1 and R_2 will be:

$$W(D) = -\frac{A}{6D} \cdot \frac{R_1 R_2}{R_1 + R_2}. \tag{12}$$

When the interacting bodies are significantly large than the interaction distance, van der Waals interactions decay significantly slower than the inverse sixth power law would indicate. Finally we have to mention one more useful expression that connects the interacting force measured between two interacting macroscopic surfaces with the interaction energy per unit area between the two surfaces:

$$F(D) = 2\pi\left(\frac{R_1 R_2}{R_1 + R_2}\right) \cdot W_0(D). \tag{13}$$

This expression is known as Deryaguin approximation and it is often use in the interpretation of the interaction force measurements because it relates interaction force between two curved bodies to the standard interaction energy per unit area between the two surfaces. Often the last value is easy to obtain from contact angle measurements or another macroscopic measurement.

3.2.2. *Electrostatic and double-layer interactions*

The interactions of two charged molecules in absence of other movable charges obey a simple Coulomb law. A much more complicated situation arises when other movable charges are present. These charges also move in the electric field created by the interacting charges and effectively "screen" the interaction. This situation typically applies to the interactions of ionizable surfaces in aqueous solutions. Ions dissociated from the surfaces along with the ions already present in the solution interact with the surface charges and give rise to complex electric potential distributions. The potential distribution ultimately determines the interaction strength between the two surfaces, yet the presence of counterions in the gap between two charged surfaces makes the physical picture slightly more complicated than just simple electrostatic repulsion. The repulsive force between two surfaces of the same charge originates in the osmotic pressure of the counterion cloud rather than in the electrostatic repulsion between surfaces [3]. In fact, electric fields from the two identically charged surfaces cancel each other in the gap. However, the counterions between the two surfaces repel each other and try to maximize their configuration entropy. Bringing the surfaces together decreases configuration entropy and thus gives rise to the opposing force which we register as "electrostatic repulsion". One of the important consequences of this mechanism is that only the electrostatic repulsion of counterions and their entropy determines interactions of charged surfaces in electrolytes. Therefore, most of the quantitative models must focus on the counterions distribution in the gap.

Typically, the potential distribution $\Psi(x)$ can be described by the Poisson–Boltzmann equation:

$$\frac{\mathrm{d}^2\Psi}{\mathrm{d}x^2} = -\frac{e}{\varepsilon\varepsilon_0}\sum_i z_i \rho_{xi} e^{\frac{ez_i\Psi}{k_\mathrm{B}T}}, \tag{14}$$

where z_i and ρ_{xi} are counter-ion charge and local density, respectively. We can solve the Poisson–Boltzmann equation using several boundary conditions. We will discuss these cases in more details when we consider the interpretation of chemical force microscopy measurements. At this point we will simply list a few key results. For the case when a surface potential Ψ_0 of the interacting surfaces is constant, the interaction force between two charged spherical bodies separated by the distance R in a solution that has only one type of counterions is:

$$F(D) = -64\pi R k_\mathrm{B} T \lambda \rho_\infty \gamma^2 e^{-D/\lambda}. \tag{15}$$

In this equation ρ_∞ denotes the density of counterions far away from the charged surface, and

$$\gamma \equiv \tanh\left(\frac{ze\Psi_0}{4k_\mathrm{B}T}\right), \tag{16}$$

$$\lambda = \frac{1}{\sqrt{\frac{\rho_\infty e^2 z_i^2}{\varepsilon\varepsilon_0 k_\mathrm{B} T}}}. \tag{17}$$

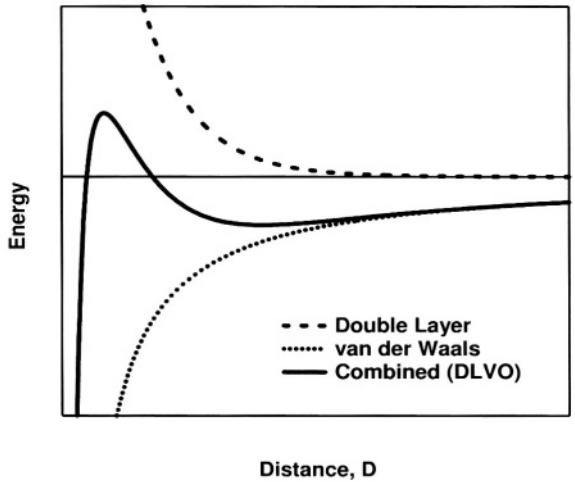


FIGURE 3.1. Energy profile of a generic interaction of charged surfaces in solution showing two main components of DLVO potential.

Equation (15) shows that in the electrolyte solutions interaction force between the charged surfaces decays exponentially. The characteristic length of such decay, given by the eq. (17) is often called *Debye length.* This parameter plays a central role in the electrostatic interaction description because it defines the *screening* of the electric field in the electrolyte solution that is caused by the presence of an ionic cloud. Finally, to give a realistic picture of interaction between two charged surfaces in an electrolyte solution, we must also take into account van der Waals attraction, which is always present, and which becomes significant at shorter distances. The interplay between repulsive double layer force and attractive van der Waals force leads to a fairly complex potential represented on the Figure 3.1. DLVO theory gives quantitative description of these interactions [5,6]. In this model the interaction potential between two spherical particles is given by

$$W(D) = 64\pi k_B T R \rho_\infty \gamma^2 \lambda^2 \cdot e^{-\frac{D}{\lambda}} - \frac{AR}{6D}. \tag{18}$$

Incidentally, DLVO equations show that if the electrostatic repulsion is strong enough then they lead to the creation of a secondary "outer" minimum on the potential energy profile. If the barrier between the local energy minimum at longer distances and a global minimum at contact ($D = 0$) is sufficiently high, it will prevent the particle from adhering to the surface. This is central argument that explains kinetic stability in colloid systems. DLVO treatment of the electrostatic interactions becomes increasingly important in nanoscale systems as the size shrinks down and van der Waals interactions become even more prominent.

3.2.3. Collective forces

Finally we will touch on the last major type of interactions that originates from collective behavior of molecules at the interfaces. It is difficult to pinpoint the mechanism of such

interactions to a single physical process. Rather, these interactions originate when other interactions induce some order in the collection of molecules near the interface. Due to their complexity, these *collective* forces are the least understood of all the other types of intermolecular interactions and often require numerical modeling. Nevertheless, the qualitative picture of some of these interactions can be quite simple. For example, structural forces arise when strongly-bound solvent layers form at the surface. When we press such surface against another surface, the system strongly favors the situation when the gap contains a whole number of solvent layers. This effect gives rise to an oscillatory force as the gap size becomes comparable to the solvent molecule size and each consecutive molecular layer is squeezed out of the gap.

A conceptually similar effect can give rise to a completely different interaction. If solvent molecules order at an interface, the configuration entropy of the system decreases. Therefore, such situation will favor the interactions of two such surfaces, since their binding releases solvent molecules into the bulk solution and raises the entropy. This mechanism can generate a surprisingly strong attractive force between two surfaces. A famous example of such entropic force is *hydrophobic effect,* which refers to the strong association of hydrophobic surfaces in aqueous solutions. In this case, the water molecules associate at the interface. Researchers typically attribute the orientation mechanism to strong hydrogen bonding (orientation interaction) between water molecules [3,7], however recent studies disputed this interpretation [8]. Entropic (hydrophobic) forces have several important distinctions. First, they arise only when ensembles of molecules is present, therefore they are absent in the interactions between individual molecules. Second, they become more prominent as the contact area between interacting surfaces increases. Third, entropic nature of the force leads to a curious temperature dependence of the interaction strength: entropic forces increase as the temperature increases.

3.3. EXPERIMENTAL MEASUREMENTS OF THE INTERACTIONS AT SOLID–LIQUID INTERFACES

There are almost no instances of a real interaction that consists of exclusively one type, on the contrary most interactions at the solid–fluid interfaces mix contributions from several types of forces. For example, a colloidal particle interacting with the surface may simultaneously experience van der Waals forces, electrostatic double layer interactions, and often significant solvation forces. Separating these contributions and accessing each one individually is a formidable experimental and theoretical challenge. Several experimental approaches have evolved to address this challenge.

Most of the classical techniques for experimental physical chemistry focus on determining energy characteristics of intermolecular processes using ensemble measurements of large numbers of molecular species. On the contrary, techniques for measuring interactions often try to limit the number of the interacting species. These techniques generally strive to measure two main characteristics of the interaction: interaction strength and the contact area. Ideally, these two quantities should be measured independently. The majority of experimental approaches rely on one of the three strategies:

(a) We can try to manipulate the molecules of interest by attaching large “handles” such as polystyrene beads to them.

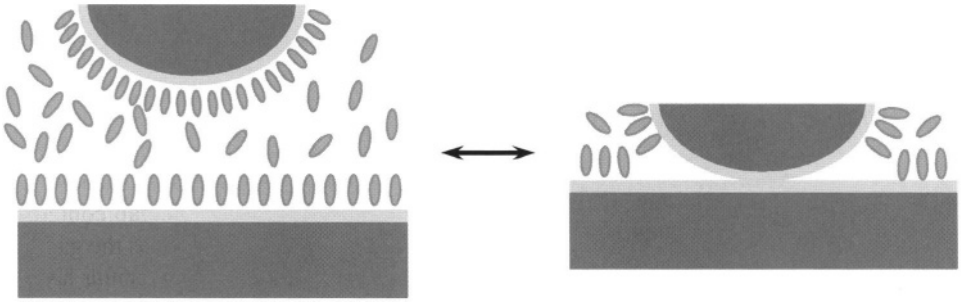


FIGURE 3.2. A cartoon illustrating the origin of the entropic force. Tightly bound solvent layers that form at the interacting surfaces are expelled out of the gap as the surfaces come into contact.

(b) We can try to measure forces originating form multiple identical interaction events. The success of this approach requires very uniform interacting surfaces in the contact zone.
(c) We can try to measure intermolecular interactions by using sharp probes that bear specific functional groups.

Figure 3.3 presents several techniques for interaction measurements that realize these approaches. Optical trapping techniques present the best example of the first approach. They turned out to be extremely useful for studying and manipulating long chain biopolymers, i.e. DNA. However optical trapping does require attaching the interacting species to micron-sized beads and the requirement of having long tethers does limit their usefulness for surface interaction studies.

Surface forces apparatus (SFA) technique realizes the second approach. SFA uses cross cylinders of atomic smooth mica surfaces to create a uniform macroscopic interface. Different monolayers physisorbed on the mica surface provide access to different chemical interactions. One of the main advantages of the SFA is its ability to measure the contact area independently using an optical interference setup. However the macroscopic size of the contact area leads to significant averaging and severely limits SFA usability for site-specific measurements.

The third approach, represented by atomic force microscopy (AFM) [9], is equally suitable to measuring single molecule level interactions and interactions between molecular assemblies. An atomic force microscope uses a sharp (typically microfabricated) probe [10] mounted on a flexible cantilever. The probe tip contacts a sample mounted on a piezo scanner, which can control the sample position to better than 1 Å accuracy. AFM instruments typically measure forces exerted on the probe by monitoring the deflection of the cantilever via an optical lever system [11]. Modern cantilevers permit routine detection of forces in the picoNewton (10^{-12} N) range and researches even reported measurement of attoNewton (10^{-18}) forces in vacuum with a specially-designed cantilever probes [12]. A wide variety of cantilevers in different stiffness ranges is now available commercially, which gives access to the techniques to scientists from a broad spectrum of disciplines. The list would be incomplete without mentioning a specific variation of AFM—Colloidal Probe Microscopy, in which researchers place microscopic size particles on the end of AFM cantilevers and use the cantilever to study interactions of this particle with the surface [13]. As the name suggests, this technique is especially suited for studying interactions in colloidal systems.

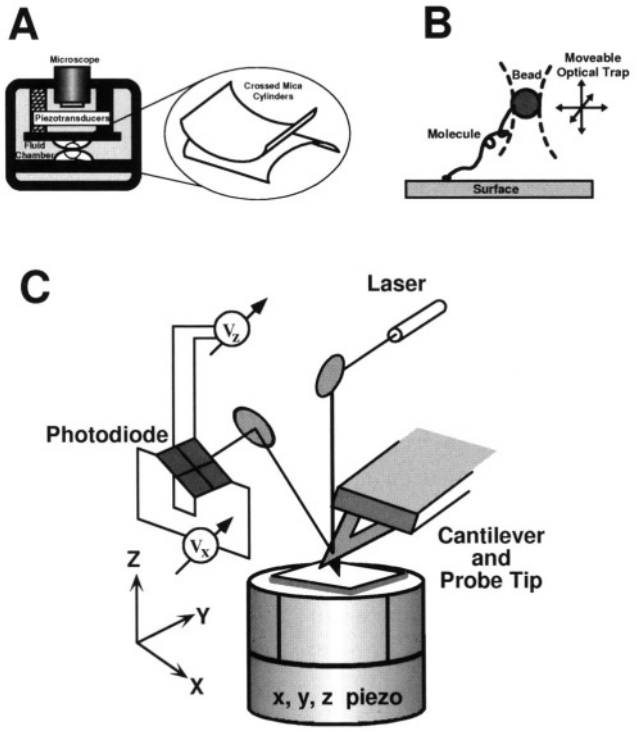


FIGURE 3.3. Experimental techniques for measuring intermolecular interactions. (A) Surface Forces Apparatus. (B) Optical Tweezers. (C) Atomic Force Microscopy.

Table 3.1 compares the main features of interaction force measurement techniques. In recent years AFM emerged as a method of choice for measuring intermolecular interactions, therefore we are going to pay particular attention to the results of AFM-based measurements.

A typical interaction force measurement in an AFM experiment consists of force curve cycle (Figure 3.5), in which the sample approaches the surface and then withdraws after making contact. As the cantilever restoring force overcomes the tip-sample interaction force, the probe jumps away from the surface. The magnitude of such jump provides a measure of the interaction strength. Atomic force microscope design is fully compatible with measurements in liquids, which gives researchers access to a wide range of the interactions in chemical and biological systems. Typical AFM probes are made from silicon or silicon nitride. As manufactured, their surface is poorly characterized and often is contaminated by residues from manufacturing process. Therefore, it is critical to control and modify the chemical functionality of the AFM probe, which researchers typically achieve by covering the probe and sample surfaces with well-defined functional groups (Figure 3.4). Interaction force measurements using such chemically-modified probes, Chemical Force Microscopy (CFM), will be our main subject.

On of the most straightforward probe functionalization approaches involves coating of the probe tips with a thin layer of gold and followed by a self-assembled alkanethiol monolayer [14]. Crystalline monolayers of long-chain alkanethiols provide robust and uni-

TABLE 3.1.
Comparison of the main experimental methods for measuring intermolecular interactions.

	Force resolution	Maximum applied force	Typical contact area	Contact area measured independently	Distance resolution	Operating environment	Typical experimental systems	Substrate requirements	Spatial mapping capability	Shear force measurement possible?
Surface forces apparatus	10 nN	$>1\ \mu$N	$\sim 1\ mm^2$	YES	1 Å	Ambient liquid UHV	Atomically smooth surfaces, uniform molecular layers, polymers	Requires two flat contacting surfaces	NO	YES
Optical/ magnetic trapping	1 pN	<1 nN	N/A	N/A	30 nm	Liquid only	Individual long-chained polymer and biopolymer molecules	Requires a way to attach a bead to the molecule of interest	NO	NO
Colloidal probe microscopy	<10 pN[a]	>100 nN[a]	$>500\ nm^2$	NO	0.1 Å	Ambient liquid UHV	Colloidal beads interacting with a surface	No specific substrate requirements	NO	YES
Atomic force microscopy	<10 pN[a]	>100 nN[a]	$1–5\ nm^2$	NO	0.1 Å	Ambient liquid UHV	Any substrates that can be immobilized on surfaces	No specific substrate constraints	YES	YES

[a] Depends on the cantilever stiffness.

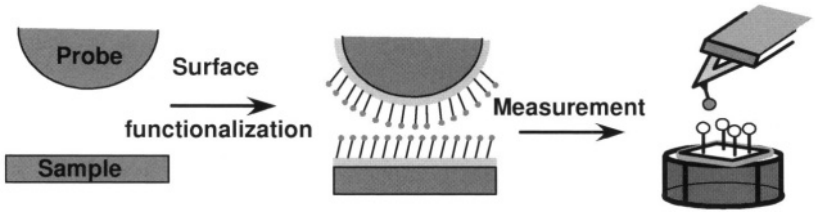


FIGURE 3.4. Chemical Force Microscopy.

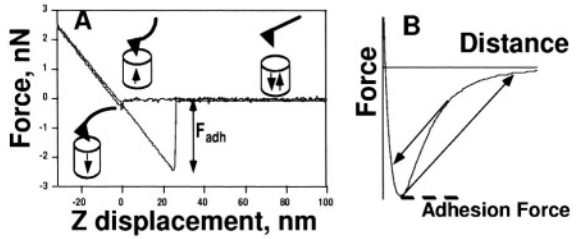


FIGURE 3.5. (A) Representative force curve obtained in an AFM experiment. As the probe tip approaches sample surface and makes contact the cantilever deflects upwards. As the sample direction reverses adhesion forces hold the cantilever bound to the surface until cantilever restoring force causes it to snap back. (B) A corresponding force profile showing two jumps caused by cantilever mechanical instabilities.

form coatings. In addition, researchers can vary the terminal chemical functionality of those monolayers without changing the structural characteristics of the coating. Thus, alkanethiol modification of the AFM tips provides a flexible and versatile system for studying intermolecular interactions in condensed phases.

Lieber and co-workers used CFM to discriminate between basic types of chemical interactions [14,15]. Histograms of adhesion force observed in repeated pull-off measurements for the tips and samples terminating with basic functional groups such as COOH and CH_3 in ethanol show that such surface chemical functionality indeed influences the interactions strength (Figure 3.6). The trends observed in these experiments follow our intuitive expectations—hydrogen bonding COOH groups interact stronger than the CH_3 groups that can only interact through van der Waals interactions. Further studies extended this approach to a large number of different functionalities and solvents [16–18]. Some of these measurements are summarized in the Table 3.2. These results clearly show that the simplistic model that is based solely on the predominant type of forces that exist between the functional groups on the surfaces cannot adequately describe the trends that govern the interaction strength. Instead a detailed analysis of the thermodynamics and kinetics of the interactions is necessary.

3.3.1. Thermodynamic models of interaction strength: contact mechanics approach

We can analyze the thermodynamics of the breaking of an adhesive junction using a continuum approach based on a contact mechanics model developed by Johnson, Kendall and Roberts (JKR model) [23]. This model treats the probe and the sample as two elastic bodies that can interact with each other and the surrounding solvent. As we consider the

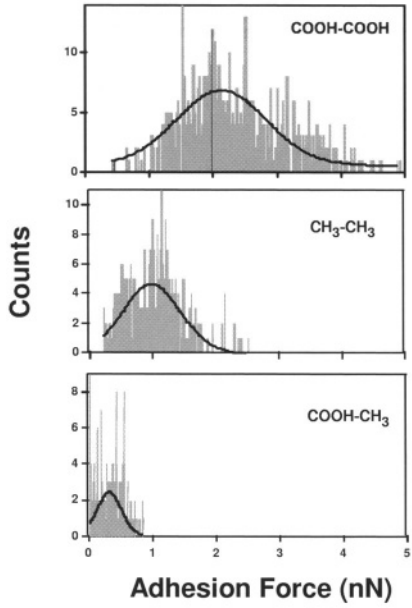


FIGURE 3.6. Histograms of the interactions forces measured between different functional groups in ethanol. From [15].

break-up of the adhesive junction between the probe and the sample under external force, we need to balance the external load energy, elastic energy stored in the deformed bodies and the adhesion energy. JKR model assumes that surface forces are acting only within the contact area of the two bodies and that those bodies deform elastically under the applied force. The model predicts that the pull-off force, F_{ad}, required to separate a tip of radius R from a planar surface (Figure 3.7) will be given by

$$F_{ad} = \frac{3}{2}\pi R W_{SMT}, \tag{19}$$

where

$$W_{SMT} = \gamma_{SM} + \gamma_{TM} - \gamma_{ST} \tag{20}$$

is the thermodynamic work of adhesion for separating the sample and tip, γ_{SM} and γ_{TM} are the surface free energies of the sample (S) and tip (T), in contact with the medium M (vacuum, vapor (V) or liquid (L)) and γ_{ST} is the interfacial free energy of the two contacting solid surfaces.

If the sample and tip bear identical functional groups (e.g. CH_3/CH_3 interactions), then $\gamma_{ST} = 0$ and $\gamma_{SL} = \gamma_{TL}$, and eq. (20) simplifies to $W_{SMT} = 2\gamma$, where γ corresponds to the free energy of the surface in equilibrium with vapor phase or solvent. Therefore, it is the solid–vapor or solid–liquid surface free energy that determines the adhesive force between tip and sample pairs modified with the same molecular groups!

TABLE 3.2.
Adhesion forces between functional groups in liquid measured in CFM experiments. From [17].

Functional group pair (tip–surface)	Monolayer, chain length	Medium	Adhesion (nN)	Ref.	Tip radius (nm)
CH_3–CH_3	silane, C_2	EtOH	0.4 ± 0.3	[19]	50
CH_3–CH_3	silane, C_9	EtOH	0.7 ± 0.6	[19]	50
CH_3–CH_3	silane, C_{14}	EtOH	2.4 ± 1.2	[19]	50
CH_3–CH_3	silane, C_{18}	EtOH	3.5 ± 2.3	[19]	50
CH_3–CH_3	thiol, C_{18}	EtOH	1.0 ± 0.4	[15]	60
CH_3–CH_3	thiol, C_{12}	EtOH	2.3 ± 1.1	[16]	30
CF_3–CF_3	silane, C_{10}	EtOH	15.4	[20]	NR(20–40)
CH_3–CF_3; CF_3–CH_3	silane, C_{18}, C_{10}	EtOH	repulsive	[20]	NR(20–40)
CH_3–CH_3	silane, C_{18}	$CF_3(CF_2)_6CF_3$	52	[20]	NR(20–40)
CF_3–CF_3	silane, C_{10}	$CH_3(CH_2)_6CH_3$	21	[20]	NR(20–40)
CH_3–CH_3	thiol, C_{12}	$CH_3(CH_2)_{14}CH_3$	0.07 ± 0.05	[16]	30
COOH–COOH	thiol, C_{11}	$CH_3(CH_2)_{14}CH_3$	0.11 ± 0.02	[16]	30
COOH–COOH	thiol, C_{11}	$CH_3(CH_2)_4CH_3$	0.95 ± 0.26	[21]	NR(20–40)[a]
COOH–CH_3	thiol, C_{11}, C_{18}	EtOH	0.3 ± 0.2	[15]	60
COOH–COOH	thiol, C_{11}	EtOH	2.3 ± 0.8	[15]	60
COOH–COOH	thiol, C_{11}	EtOH	0.27 ± 0.04	[16]	30
COOH–COOH	thiol, C_{11}	PrOH	1.37 ± 0.26	[21]	NR(20–40)[a]
CH_2OH–CH_2OH	thiol, C_{11}	EtOH	0.18 ± 0.18	[16]	30
COOH–COOH	thiol, C_{11}	H_2O	2.80 ± 0.20	[21]	NR(20–40)[a]
COOH–COOH	thiol, C_{11}	H_2O, pH < 5[b]	7.0 ± 0.2	[22]	60
COOH–COOH	thiol, C_{11}	H_2O, DI	2.3 ± 1.1	[16]	30
COOH–CH_2OH	thiol, C_{11}, C_{11}	H_2O, pH < 5[b]	1.1 ± 0.5	[22]	30
CH_2OH–CH_2OH	thiol, C_{11}	H_2O[b]	1.0 ± 0.2	[22]	20
CH_2OH–CH_2OH	thiol, C_{11}	H_2O, DI	0.30 ± 0.05	[16]	30
CH_3–CH_3	thiol, C_{18}	H_2O	60 ± 5	[22]	60
CH_3–CH_3	thiol, C_{12}	H_2O	12.5 ± 4.4	[16]	30

[a] Tips were coated with 50 nm of Au.
[b] Ionic strength IS = 0.01 M.

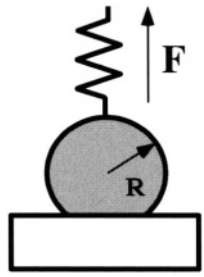


FIGURE 3.7. Sphere-on-a-plane configuration in JKR model.

Liber and coworkers checked this approach by comparing the measured value of the adhesion force between CH_3-terminated surfaces and tips with the calculated value [15]. The calculated value of $F_{ad} = 1.2$ nN was based on the previously determined value of $\gamma = 2.5$ mJ/m^2 [24,25], and the experimentally determined probe tip radius; and provided

a very good agreement with the experimental value of 1.0 ± 0.4 nN. Thus, the continuum approach of JKR model provides a reasonable interpretation of microscopic CFM measurements. Further experiments using different solvents provided additional verification for this approach [18].

In addition, JKR model analysis of the CFM measurements can yield important information about the surface free energies, which is especially important in cases where conventional contact angle measurements fail. For example, contact angle measurements cannot be used to probe high free energy surfaces, since most liquids wet them. Adhesion measurements are free of this limitation, and, moreover, high surface energies lead to larger (easily measured) forces. For example, for COOH-terminated SAMs, CFM data yields a value of $\gamma = 4.5$ mJ/m^2 in ethanol [15].

Equations (19) and (20) highlight the central role that interfacial free energy plays in determining the strength of the intermolecular interactions between surfaces. In addition they show why the naïve interpretation based on the type of the interaction does not always work. The interfacial free energy term includes not only the surface–surface energy term, but also two surface–solvent energy terms; therefore, the interaction strength depends on the interplay of those terms. The surface–surface interactions term dominates the overall interaction energy only if the solvent–surface interactions are very small, which is not always true. If the surface–solvent interactions are large, then the interaction between surface chemical functionalities can be obscured by the solvent interactions with the surface [26]. Therefore, extremely strong dependence of the surface forces on the solvent composition is not surprising [18]. Such dependence can even be advantageous for fine-tuning the interactions between surfaces. Thus, all components of the system play important roles in determining the strength of the interactions between surfaces in liquids. Skulason and Frisbie recently developed quantitative criteria for choosing a proper solvent medium for observing true intermolecular interactions in CFM experiments [26].

3.3.2. Interactions in aqueous solutions: measurement of double-layer forces

Water is by far the most important solvent due to its natural abundance and the role that water plays in biological systems. A large number of surface functional groups ionize in water giving rise to the electrostatic interactions. Ionization of the surface functional groups is almost always accompanied by the presence of significant amount of counterions in solution; therefore pure (unshielded) electrostatic interactions are replaced by the double layer interactions. Often, the dissociation constants of the surface groups in these systems differ from those of monomer analogs in solution. Several factors contribute to these differences, including: (i) a low dielectric permittivity of an adjacent hydrocarbon region; (ii) fewer degrees of freedom for the immobilized species; (iii) an excess electrostatic free energy of the supporting surface; and (iv) changes in the solution dielectric constant in the vicinity of a charged surface [27]. These factors make direct measurements of the double layer forces all the more useful.

The surface free energy depends on the ionization state of functional groups and reflects their degree of ionization. In principle, researchers can monitor this change in free energy on the macroscopic scale by measuring contact angle versus solution pH. At the pH value equal to the pK of the surface group the contact angle goes through a large change [28]. Alternatively, Lieber and co-workers showed that chemically-modified AFM tips and samples can probe the changes in solid–liquid surface free energies with pH [22]. In that work

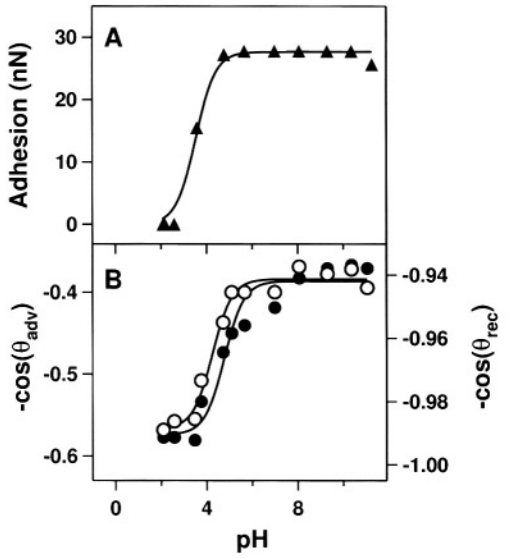


FIGURE 3.8. (A) Force titration for tip and surface terminated with NH_2 groups. (B) Contact angle titration of NH_2-terminated surface. From [22].

researchers detected changes in surface charge induced by the dissociation of acidic/basic groups by monitoring the adhesive force with an AFM probe sensitive to electrostatic interactions. Variations in the sign and magnitude of the force indicated changes in the surface charge; moreover, an abrupt transition in the force occurred at a $pH \approx pK$ of the functional group on the surface.

Specifically, Lieber and coworkers measured adhesion force values at different solution pH values for tips and samples terminating with amine groups. Their experiments showed a sharp drop to zero (indicating a repulsive interaction) below a pH of 4 (Figure 3.8A), which was consistent with protonation of the amine groups on these two surfaces at lower pH values. Contact angles values measured using buffered solution droplets on this same surface (Figure 3.8B) also showed a sharp transition (an increase in wetability) as the droplet pH is reduced below pH of 4.5. Remarkably, local force microscopy measurements using a modified probe tip and macroscopic wetting studies provide very similar values for the pK of the surface amine group for the amine-terminated SAMs, even if those values were 6–7 pK units lower than typical bulk solution values. Vezenov *et al.* attributed large shifts in pK values to then local hydrophobic environment [22]. Such difference underscores the value of the AFM-based approach ("*force titrations*") to determining local pK values in chemical and biological systems.

Vezenov *et al.* also used force titrations to determine pK values for COOH-terminated surfaces (Figure 3.9A) [22]. A prominent feature in this plot is the sharp transition from positive adhesion forces at low pH to zero (indicating repulsion) at high pH. The observed repulsion at pH > 6 originates in the electrostatic repulsion between negatively charged carboxylate groups, while the adhesive interaction at low pH values is indicative of hydrogen bonding between uncharged COOH groups. These data show that the pK_a of the surface-confined carboxylic acid is 5.5 ± 0.5, which is very similar to the pK_a

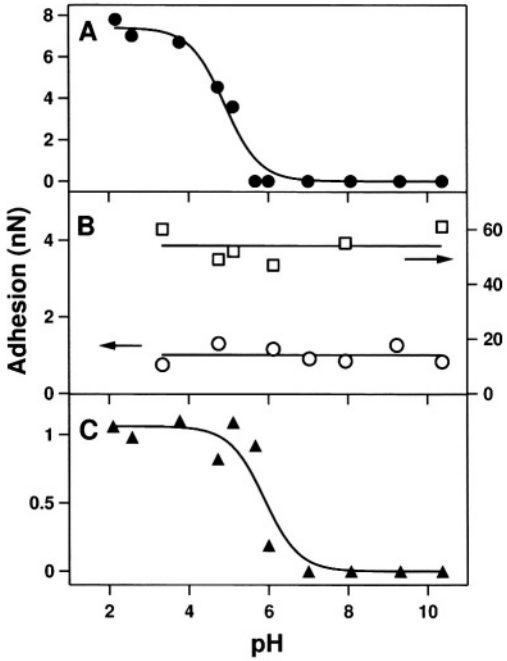


FIGURE 3.9. Force titrations on surfaces ternminated with different functional groups. (A) COOH/COOH pair; (B) CH_3/CH_3 pair (right axis; OH/OH pair (left axis); (C) OH (tip)/COOH (sample) pair. From [22].

for COOH functionality in aqueous solution. It is reasonable to expect that non-ionizing functional groups would not interact through double-layer forces. Indeed, CFM titration curves for tip/sample SAMs terminating in non-ionizing OH/OH and CH_3/CH_3 functionalities showed an approximately constant, finite adhesive interaction throughout the whole pH range (Figure 3.9B). Vezenov *et al.* also demonstrated the use of probes terminating in hydrophilic non-ionizing groups (such as OH-groups) as pK probes for unknown surface functionalities. They used an OH-terminated tip to determine the pK of the surface terminated in COOH groups (Figure 3.9C). The pK value determined by this method was indeed identical to the pK values determined by other techniques [22].

The JKR theory of contact mechanics can also provide quantitative interpretation for interaction force measurements in aqueous medium. However, to interpret pH dependent adhesion data in electrolyte solutions we also have to include long-range electrostatic interactions between the tip and sample surface into the overall energy balance. When the free energy of a double layer per unit area w_{DL} balances γ_{SL} we do not expect to measure any adhesion. Quantitatively, the pull-off force, $P_{\text{pull-off}}$, is related to these two terms:

$$P_{\text{pull-off}} = -\frac{3}{2}\pi R W_{SLS} + \frac{5}{2} P_{DL}, \tag{21}$$

where $P_{DL} \approx 2\pi R w_{DL}$ is an additional load that we have to apply to a spherical tip due to the presence of a double layer. Thus, repulsion between similarly-charged surfaces ($P_{DL} > 0$) decrease the magnitude of the pull-off force compared to that given by the clas-

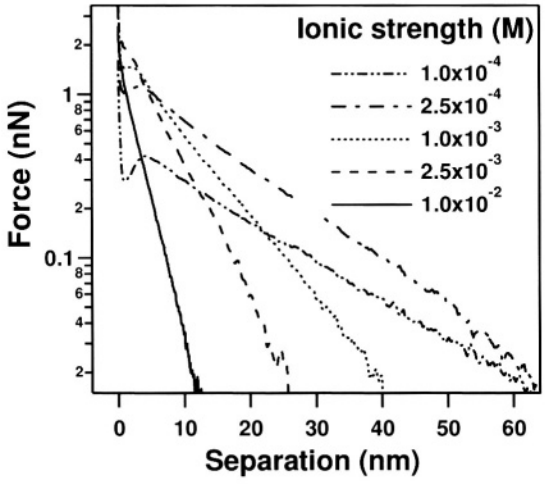


FIGURE 3.10. Repulsive double layer interaction recorded between tip and surface terminated in COOH functional groups at pH = 7.2 for different solution ionic strengths. From [22].

sic JKR model. Beyond a threshold value of the repulsive electrical double layer force $P_{DL} = P_{pull\text{-}off}$, the deformation of a spherical tip becomes fully reversible and the pull-off force vanishes. The corresponding surface potential

$$\psi = \sqrt{\frac{\lambda}{\varepsilon\varepsilon_0}\gamma_{SL}} \tag{22}$$

is independent of the tip radius for $\lambda \ll R$ (λ is the Debye length). This equation shows that a change from adhesive to repulsive behavior does signal a change in the ionization state of the interacting surfaces.

Researchers also verified the electrostatic origin of the pH-dependent repulsive forces by changing the Debye screening length λ; that is the solution ionic strength. Figure 3.10 shows that the repulsive interaction becomes progressively longer-ranged as the solution ionic strength (IS) decreases. A detailed analysis of the electrostatic force entails taking into account the surface charge–potential regulation imposed by the potential dependent binding of ions at the interface [29–31]. Vezenov *et al.* developed a model [22] that uses linearized charge–potential regulation condition [32]. Fits of this model to the experimental data indicated that the measurements occur in the constant potential (vs. constant charge) regime independent of ionic strength, i.e. the charge–potential regulation is a significant effect in these systems. However, Shestakov *et al.* also obtained reasonable agreement between simulation and experiment when they modeled a similar measurement by solving non-linear Poisson–Boltzmann equation in constant charge case [33]. Their simulations reproduced the experimental data reasonably well for a partially-ionized surface (Figure 3.11).

Finally, the modified JKR model provides an interesting estimate of the number of species involved in double layer interactions. For a tip with a radius of R = 30 nm in contact with the sample surface in solution of 0.01 M ionic strength, there are more than 200 ions and 1×10^6 water molecules involved in the interaction. In contrast, in organic

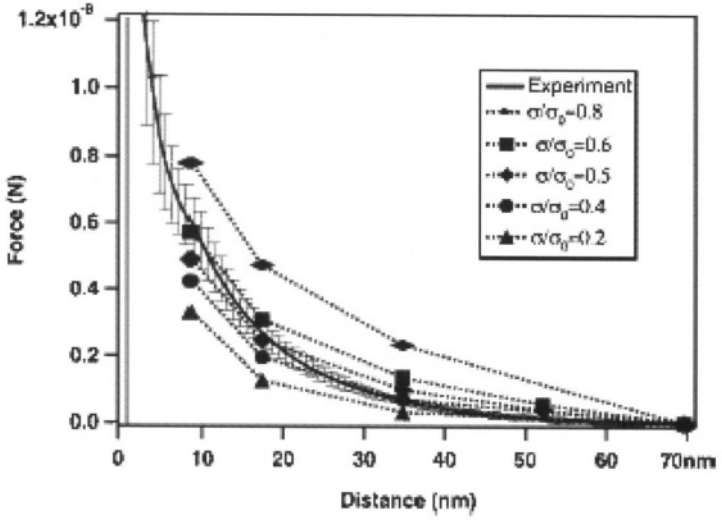


FIGURE 3.11. Comparison of the Poisson–Boltzmann simulation with the CFM experimental results (solid line). Each simulation (designated by a symbol) is a separate solution of the PB equation for a particular separation distance and imposed surface charge density. The situation when only half of the surface groups ionize ($\sigma/\sigma_0 = 0.5$) provides the best agreement. From [33].

solvents we typically have to deal with only 15–25 molecular contacts. Thus, unlike the van der Waals and hydrogen bonding interactions the electrostatic forces are always averaged over relatively large areas.

3.3.3. Kinetic model of intermolecular interactions

So far, contact mechanics models formed the basis of our theoretical description. However, these thermodynamic approaches present only phenomenological picture of the interaction. We describe the interaction using an overall energy parameter (W) that does not fully represent the nature of the physical processes occurring in the system. The value of W is certainly related to the interaction potential parameters, yet this relationship is not straightforward. Moreover, contact mechanics models are continuum theories; therefore we expect them to break down as the size of the contact area shrinks to molecular dimensions. Therefore, it is unlikely that such models will give us the answer to the ultimate question: what is the strength of a single bond? To answer this question and to make a more direct connection between the measured interaction strength and the interaction potential parameters we need to consider the kinetics of the adhesive junction break-up under an external load. Kinetic model of bond dissociation, developed by Evans and coworkers in 1997 provided a breakthrough in such analysis and uncovered several fascinating details about the behavior of a single bond under external load [34].

3.3.3.1. Strength of a single bond Evan's model is based on the Kramers' theory of thermally-assisted barrier crossing in liquids [35]. The model describes an irreversible escape from a potential well ("bound state") under an influence of external loading force. In the absence of loading force thermal fluctuations magnitude and the interactions potential parameters define a finite lifetime of the adhesive bond. In other words, given enough

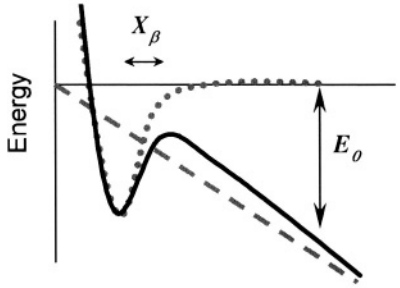


FIGURE 3.12. A cartoon illustrating the effect of the external loading force on the interaction potential.

time any adhesive bond will eventually break under the influence of even an infinitesimal loading force. When we apply the loading force, we tilt the potential, effectively reducing the barrier height (Figure 3.12). As the result, the system driven by thermal fluctuations has a higher probability to reach the top of the barrier and escape. An AFM experiment brings an important modification to this intuitive picture—the loading force is no longer constant; instead it changes constantly during the force curve cycle (typically, the loading force increases linearly with time). Therefore, during the force measurement the potential landscape that the system sees is constantly changing as the loading force tilts the barrier more and more. Evans and colleagues showed that such external force leads to the exponential amplification of the escape rate, which in tern defines the bond survival time and bond strength. Qualitatively, at lower applied forces the barrier is still too high for the thermally-activated transition to happen and at higher applied forces the transition has most likely happened already. In other words, the unbinding of a chemical bond under external load does happen in a fairly narrow range of the applied forces, which ultimately defines the bond strength that we register in the experiment. Importantly, the kinetic nature of the process leads to a fascinating result: the force required for unbinding (i.e. bond strength) is not unique. Rather, the bond strength depends on the bond loading history!

Quantitative analysis verifies this intuitive picture. Evans and Ritchie [34] postulated that the rate of escape from the bound state dP/dt under applied load $f(t)$ obeys first order kinetics:

$$\frac{dP}{dt} = -k_{off} P(t), \tag{23}$$

where P is the population of the bound state and the rate constant k_{off} is inversely proportional to the average bound state survival time. Since application of the external load force $f(t)$ lowers the energy barrier, the rate constant k_{off} becomes a function of the external force:

$$k_{off} = \frac{1}{\tau_D} \exp\left[-\frac{E_0 - f(t)x_\beta}{k_B T}\right]. \tag{24}$$

Here E_0 is the activation energy barrier, x_β is the distance between the bound state and the transition state, and τ_D is the characteristic diffusion time of motion in the system. If the external load force changes with time, the problem becomes a first order kinetic

process with time-dependent rate constant. AFM pull-off force measurements fall in this category, since they feature constant rate loading $f(t) = r_f t$. According to the eq. (23), the likelihood of detachment (i.e. probability density $p(t) \equiv \mathrm{d}P/\mathrm{d}t$) is proportional to rate constant multiplied by the population of the bound state. As the loading force increases, the rate constant increases while the population of the bound state diminishes. Therefore, the probability of detachment will go through a maximum at a certain value of the applied force, which defines the bond strength, $f_{\text{pull-off}}$ [34,36]. Evans solved the eqs (23) and (24) to find the position of the detachment probability maximum in the case of a single energy barrier loaded at constant rate, r_f, and obtained the following expression for the pull-off force [36]:

$$f_{\text{pull-off}} = \frac{k_B T}{x_\beta} \ln\left(\frac{r_f}{r_0}\right), \tag{25}$$

where r_0 is defined as:

$$r_0 = \frac{k_B T}{x_\beta} \cdot \frac{1}{\tau_D \exp\left(\frac{E_0}{k_B T}\right)}. \tag{26}$$

Immediately, we can see from eq. (25) that the pull-off force increases logarithmically as the loading rate r_f increases. Thus, the strength of an individual bond can vary quite significantly over a wide range of loading rates. Moreover, the slope of this dependence is determined by the potential width, x_β. If the system potential energy surface has several barriers at different values of x_β, then further increase in the loading rate will eventually expose the inner barriers and the slope of the force versus $\log(r_f)$ will switch to progressively higher values determined by the widths of the inner barriers. Evans and coworkers demonstrated this behavior in a *Dynamic Force Spectroscopy* experiment by measuring the interaction between biotin ligands and avidin and streptavidin proteins over many decades of loading rate [37]. Their data (Figure 3.13) clearly show the existence of several barriers on the potential energy surface for these interactions and provide a powerful demonstration of the Dynamic Force Spectroscopy potential for mapping such barriers.

3.3.3.2. Bond strength in multiple-bond systems Measurements utilizing a single bond represent an ideal experimental system. Yet, in most cases single bond regimes are difficult to achieve. In addition, a significant portion of biological interactions involve a number of distinct bonds that do not necessarily break simultaneously. Recently, Evans and Williams considered the kinetics of the bond rupture in such systems [38,39]. They identified several general cases of loading in the multiple-bond systems: (1) serial loading when we apply the force to a chain of bonds, (2) parallel loading when the loading force is shared by several bonds. The detailed analysis is beyond the scope of this review, so we simply summarize the main results of Evans and Williams for the idealized cases of N identical bonds in different configurations. We can simplify the analysis by assuming that all the bonds are correlated, i.e. by assuming that they share a reaction coordinate. Then the system can be analyzed as a single "macro-bond" with the potential equal to the sum of the potentials of individual

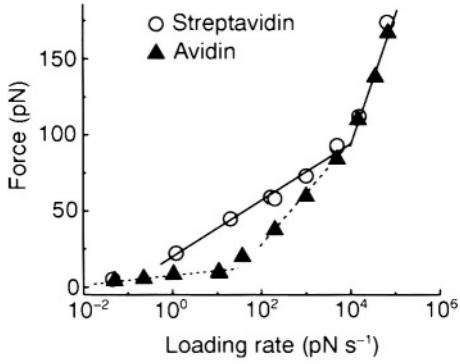


FIGURE 3.13. Dynamic force spectroscopy measurement of the strength of biotin–streptavidin (circles) and biotin–avidin (triangles) bonds. Biotin–streptavidin data show activation barriers at 0.5 nm and 0.12 nm. Biotin–avidin pair also shows an inner barrier at 0.12 nm; but the outer barrier shifts to 0.3 nm (dashed line). In addition, at very low loading rates the biotin–avidin force spectrum spectrum exhibits a low-strength regime (dashed line) that corresponds to a barrier at 3 nm. From [37].

components [39]. For the serial loading of N identical bonds Evans and Williams obtained the following expressions for the unbinding force:

$$f_{\text{pull-off}} = \frac{k_B T}{N x_\beta} \ln\left(\frac{r_f}{r_0} N^2 \exp\left(\frac{(N-1)E}{k_B T}\right)\right). \tag{27}$$

In case of parallel loading they obtained:

$$f_{\text{pull-off}} = \frac{k_B T}{x_\beta} \ln\left(\frac{r_f}{r_0} N \exp\left(\frac{(N-1)E}{k_B T}\right)\right). \tag{28}$$

Qualitatively, in case of parallel loading of N bonds the binding force is only slightly smaller than N times single bond strength, but in the serial loading case, the binding force is much smaller than that value. The parallel loading case is by far the most important for analysis of real experimental systems, so we will consider it further. One of the most important features of this case is that the distance scale of the interactions is unchanged, i.e. the width of the potential for the "macro-bond" is still equal to the width of the potential for a single bond. Moreover, the scaling of the bond strength with the loading rate predicted by the kinetic model for a single bond case is still valid. Therefore, we can use the dynamic force spectroscopy measurement to determine the width of the interaction potential using CFM experiments that provide an almost ideal parallel loading case due to the configuration constraints imposed by the rigid self-assembled monolayers. Zepeda *et al.* [40] demonstrated such measurement for the interactions of COOH-terminated surfaces and for interactions of mica surface with Si_3N_4-terminated probe (Figure 3.14). In both cases they observed the scaling predicted by the kinetic model and determined the distance scale parameters for both interactions. Interestingly, in case of mica-Si_3N_4 interactions they observed two distinct energy barriers—the tight inner barrier at 1.3 Å followed by the outer barrier at ~11 Å which pointed out to a rather large role solvation plays in structuring the energy landscape of the interactions between those surfaces.

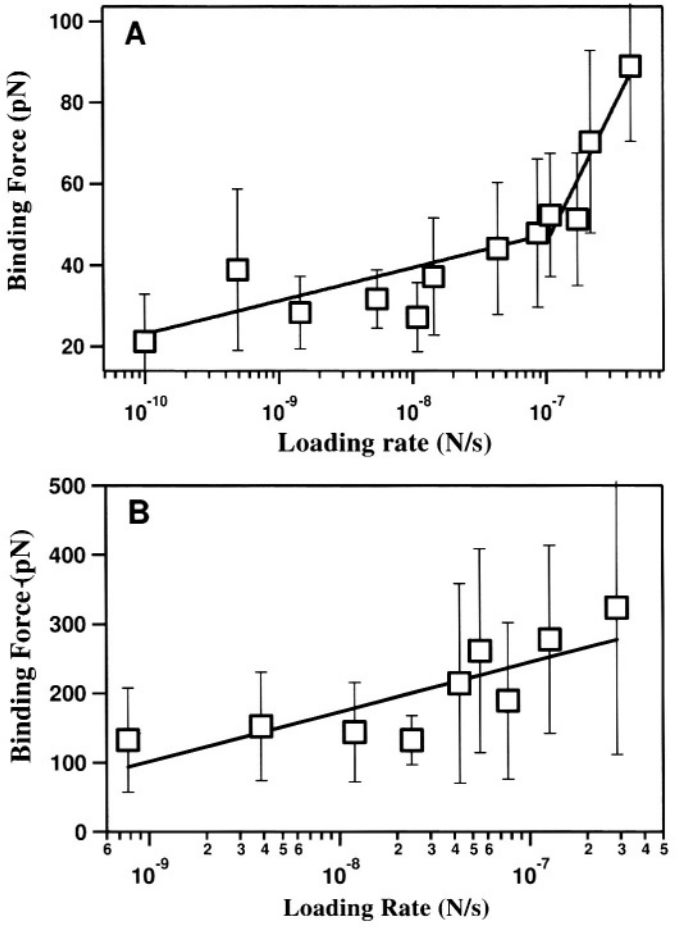


FIGURE 3.14. Interaction forces measured as a function of loading rate for (A) Si_3N_4 tip and mica surface, and (B) COOH-terminated tip and surface. From [40].

If we allow the intermolecular bonds to be uncorrelated, the analysis of the multiple-bond interaction becomes rather complicated, yet, qualitatively most of the physical features of the simple "macro-bond" model remain valid. One interesting case of the uncorrelated unbinding is the "zipper loading" in which N bonds are broken consecutively one after another (such as when unzipping a DNA double helix). Kinetic model analysis predicts that for N bonds in the zipper binding force will scale as $\log(N)$. This unusual prediction should be an interesting subject for the experimental verification.

3.3.3.3. Temperature dependence of the intermolecular bond strength Lastly, we consider the temperature dependence of the intermolecular bond strength. We can represent eq. (25) in a much more revealing form, if we separate the energy barrier into enthalpic and entropic components, $E_0 = \Delta H - T\Delta S$, and substitute eq. (26) into eq. (25),

$$f_{\text{pull-off}} = \frac{\Delta H}{x_\beta} - \frac{\Delta S}{x_\beta} T - \frac{k_B T}{x_\beta} \ln\left[\frac{k_B T}{r_f \tau_D x_\beta}\right]. \tag{29}$$

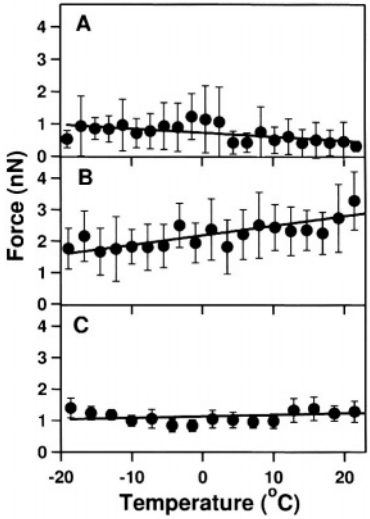


FIGURE 3.15. Interaction forces measured as a function of temperature for COOH-terminated surfaces in different solvents (A) hexane (B) ethanol, and (C) 1:1 ethanol hexane mixture. From [41].

The first two terms in eq. (29) describe the enthalpic and the entropic contribution to the bond strength and the third term describes the contribution of thermal motion to the bond strength. In other words, the first two components describe the true energy-barrier contribution and the (always negative) third component describes the "thermal weakening" of a bond caused by the thermal fluctuations helping the system to get over the activation barrier.

Equation (29) highlights another non-trivial physics of a single bond behavior under external load. The third term in the eq. (29) ("thermal weakening") always increases in magnitude as the temperature increases, leading to the overall decrease in the observed force, in full agreement with the intuitive picture of bond "loosening". Yet, the entropic term can lead to either increase or decrease in the overall interaction force depending on the sign on the entropy change for the unbinding process. Therefore, for the cases when the energy barrier has a large entropy component (i.e. such as in cases of entropic interactions) we expect the bond strength to increase with the temperature! Noy and co-workers observed this behavior when they measured the interaction strength between probe and sample terminated in the carboxylic functional groups in polar solvents (Figure 3.15B) and attributed this effect to the large negative entropy accompanying ordering of solvent molecules at the interfaces [41]. When they switched to a non-polar solvent, the interaction force showed the expected decrease with the increase in temperature (Figure 3.15A).

The relative magnitude of the entropic and the kinetic terms in the eq. (29) defines two regimes of bond rupture: (a) thermally-dominated kinetics where the kinetic weakening leads to decrease in the observed bond strength with the increase in temperature and (b) barrier-dominated kinetics where the entropic term overwhelms the kinetic term and leads to an increase in interaction strength with increase in temperature. Further analysis of the eq. (29) also indicates that the entropic regime of unbinding must exist only in the limited range of temperatures. As the temperature increases further, kinetic term which increases as $T \cdot \ln T$ will always overwhelm the entropic term which increases only linearly.

For the entropic forces caused by the ordering of the solvent molecules at the surface this cross-over point simply corresponds to the situation when the thermal motion becomes so strong that it overwhelms molecular ordering in the solvent layers.

3.4. OUTLOOK

Recent developments in experimental methods of measuring intermolecular interactions are remarkable for several reasons. First, they allowed direct exploration of the role that different functionalities, solvents and environmental variables play in shaping the strength of intermolecular interactions. Chemical Force Microscopy approach in particular, turned out to be very effective in exploring some of these factors. Moreover, CFM studies clearly debunked the naïve notion that intermolecular interaction strength is determined only by the nature of the interacting groups. These studies showed that the interaction strength between two chemical species must always considered in context of the environment surrounding these species. In addition CFM studies highlighted the critical role solvent plays in shaping intermolecular interactions in condensed phases.

Furthermore, the emerging kinetic view of the intermolecular interactions introduced a completely new paradigm for understanding these interactions. Kinetic model showed that the measured interactions strength depends not only on the energy landscape of the system, but also on the loading history prior to the bond break-up. This new paradigm refocused our attention to the energy landscape as a fundamental characteristic of the interaction. Moreover, the Dynamic Force Spectroscopy approach derived from kinetic model allowed direct characterization of the potential energy barrier geometry. Further investigations of the interactions in different systems, especially interactions between biomolecules, will uncover many interesting characteristics of intermolecular potentials. These studies have the potential to reveal for the first time a true picture of the energy landscapes in complex chemical and biological systems.

REFERENCES

1. D. Philp, J.F. Stoddart, *Angew. Chem. Int. Ed. Engl.* **35**, 1154 (1996).
2. A. Stoneham, M. Ramos and A. Sutton, *Phylos. Mag. A* **67**, 797–811 (1993).
3. J. Israelachvili, *Intermolecular and Surface Forces* (Academic Press, New York, 1992).
4. A. McLachlan, *Discuss. Farad. Soc.* **40**, 239–245 (1965).
5. B. Derjaguin and L. Landau, *Acta Physicochim. USSR* **14**, 633–662 (1941).
6. E. Verwey and J. Overbeek, *Theory of Stability of Liophobic Colloids* (Elseiver, Amsterdam, 1948).
7. N. Muller, *Account Chem. Res.* **23**, 23–28 (1990).
8. L.F. Scatena, M.G. Brown and G.L. Richmond, *Science* **292**, 908–912 (2001).
9. G. Binnig, C.F. Quate and C. Gerber, *Phys. Rev. Lett.* **56**, 930–933 (1986).
10. T.R. Albrecht, S. Akamine, T.E. Carver and C.F. Quate, *J. Vac. Sci. Technol. A* **8**, 3386–3396 (1990).
11. G. Meyer and N.M. Amer, *Appl. Phys. Lett.* **53**, 1045–1047 (1988).
12. O. Zuger, H. St, C. Yannoni and D. Rugar, *J. Appl. Phys.* **79**, 1881–1884 (1996).
13. W.A. Ducker, T.J. Senden and R.M. Pashley, *Nature* **353**, 239–241 (1991).
14. C.D. Frisbie, L.F. Rozsnyai, A. Noy, M.S. Wrighton and C.M. Lieber, *Science* **265**, 2071 (1994).
15. A. Noy, C.D. Frisbie, L.F. Rosznyai, M.S. Wrighton and C.M. Lieber, *J. Am. Chem. Soc.* **117**, 7943 (1995).
16. S.K. Sinniah, A.B. Steel, C.J. Miller and J.E. Reutt-Robey, *J. Am. Chem. Soc.* **118**, 8925–8931 (1996).
17. A. Noy, D. Vezenov and C. Lieber, *Ann. Rev. Mat. Sci.* **27**, 381–421 (1997).

18. D.V. Vezenov, A.V. Zhuk, G.M. Whitesides and C.M. Lieber, *J. Am. Chem. Soc.* **124**, 10578–10588 (2002).
19. T. Nakagawa, K. Ogawa, T. Kurumizawa and S. Ozaki, *Japan. Jour. Appl. Phys., Pt. 2* **32**, L294–L296 (1993).
20. T. Nakagawa, K. Ogawa and T. Kurumizawa, *J. Vac. Sci. Technol. B* **12**, 2215–2218 (1994).
21. T. Han, J.M. Williams and T.P. Beebe, Jr., *Anal. Chim. Acta* **307**, 365–376 (1995).
22. D.V. Vezenov, A. Noy, L.F. Rosznyai and C.M. Lieber, *J. Am. Chem. Soc.* **119**, 2006–2015 (1997).
23. K.L. Johnson, K. Kendall and A.D. Roberts, *Proc. Royal Soc. London A* **324**, 301–313 (1971).
24. C.D. Bain, J. Evall and G.M. Whitesides, *J. Am. Chem. Soc.* **111**, 7155 (1989).
25. C. Bain and G. Whitesides, *Angew. Chem. Int. Ed. Engl.* **28**, 506–512 (1989).
26. H. Skulason and C.D. Frisbie, *Anal. Chem.* **74**, 3096–3104 (2002).
27. B.V. Zhmud and A.A. Golub, *J. Colloid Interface Sci.* **167**, 186 (1994).
28. S.R. Holmes-Farley, R.H. Reamey, T.J. McCarthy, J. Deutch and G.M. Whitesides, *Langmuir* **4**, 921 (1988).
29. D. Larson, C.J. Drummond, D.Y.C. Chan and F. Grieser, *J. Phys. Chem.* **99**, 2114 (1995).
30. L. Meagher and R.M. Pashley, *Langmuir* **11,** 4019 (1995).
31. R.M. Pashley, *J. Colloid Interface Sci.* **83**, 531 (1981).
32. E.S. Reiner and C.J. Radke, *Adv. Colloid Interface Sci.* **58**, 87 (1993).
33. A.I. Shestakov, J.L. Milovich and A. Noy, *J Colloid Interface Sci.* **247**, 62–79 (2002).
34. E. Evans and K. Ritchie, *Biophys. Journal* **72**, 1541–1555 (1997).
35. P. Hanggi, P. Talkner and M. Borkovec, *Rev. Mod. Phys.* **62**, 251–341 (1990).
36. E. Evans, *Faraday Discuss.,* 1–16 (1999).
37. R. Merkel, P. Nassoy, A. Leung, K. Ritchie and E. Evans, *Nature* **397**, 50–53 (1999).
38. E. Evans and P. Williams, in: *Physics of Bio-Molecules and Cells,* Eds H. Flyvbjerg, F. Jülicher, P. Ormos and F. David (Springer and EDP Sciences, Heidelberg, 2002), Vol. 75, pp. 147–185.
39. E. Evans and P. Williams, in: *Physics of Bio-Molecules and Cells,* Eds H. Flyvbjerg, F. Jülicher, P. Ormos and F. David (Springer and EDP Sciences, Heidelberg, 2002), Vol. 75, pp. 187–203.
40. S. Zepeda, Y. Yeh and A. Noy, *Langmuir,* in press (2003).
41. A. Noy, S. Zepeda, C.A. Orme, Y. Yeh and J.J. De Yoreo, *J. Am. Chem. Soc.* **125**, 1356–1362 (2003).

4

Optical consequences of chemistry at growing crystal interfaces

Bart Kahr*, Miki Kurimoto, Werner Kaminsky, Sei-Hum Jang, Jason B. Benedict
Department of Chemistry, Box 351700, University of Washington, Seattle, WA 98195-1700, USA

4.1. INTRODUCTION

The chemical states and distributions of impurities in crystals contain information, in principle, about the dynamic interactions of the impurities with the growing crystal interfaces through which they become entrapped. However, extracting this information is by no means straightforward. To ascertain the state of a guest molecule in a crystal by X-ray diffraction, the quantity of the guest must be greater than ~10 mole % as X-ray scattering increases with the square of the number of electrons. Contributions from impurities are contained in the diffuse scattering between the Bragg reflections but this has been difficult to treat systematically and is often ignored [1]. Minor components in crystals that may well determine chemical and physical properties are invisible in most crystal structure determinations, restricting X-ray studies of mixed crystals to those rare examples of isomorphous host/guest pairs that are miscible in great proportion. For those mixed crystals in which guests are less than ~1 mole %, optical measurements and imaging technologies may take the place of X-ray analysis. Optical microscopies and spectroscopies are often comparatively quick, easy, and cheap while being very sensitive to small quantities of guest molecules or ions. Here, we illustrate how optical methods can be used to extract information about dynamic processes at growing crystal interfaces so long as the guests are sufficiently complex that they can encode and report information about these encounters at surfaces. A cartoon representing the logic of this course of study is shown in Figure 4.1. Here, the optical signature of a large guest molecule changes upon adsorption at a kink site, and overgrowth by the host crystal.

In order for this program to succeed we need to prepare optically labeled crystals—mixed crystals—almost at will. However, it is widely assumed that mixed crystals only grow when there is a homology of size and shape between the additive and host molecules

*E-mail address: kahr@chem.washington.edu

or ions. Statements of the presumed restrictions of the principle of isomorphism [2] are not hard to find.

> Trying to predict solubility, we should compare, first of all, the shapes of the host and impurity molecules… Similarity of the shapes of molecules is a necessary condition of solubility [3].

> Unfortunately, it is frequently difficult to find a suitable host crystal into which the solute of interest enters in an isomorphous fashion and which is transparent throughout the spectral region of interest [4].

> Solid solutions in organic systems are rare and only expected where the host and impurity are similar in size and shape (isomorphous) [5].

The study of mixed crystals with *anisomorphous* guests has been lagged because of broad trends in solid state chemistry in the past century, especially the ascension of X-ray scattering experiments. Diffraction has reinforced the notion that crystals clean themselves up as they grow, excluding those square pegs that will not fit in the round holes vacated by host molecules. If there is not anything in the ORTEP diagram of a single crystal structure other than the principle component it is easy to be lulled into thinking that it is the only significant part of the composition.

The principle of isomorphism makes sense on thermodynamic grounds. However, crystals grow from solution under conditions of supersaturation far from equilibrium. Under such conditions, crystals are more flexible in their ability to orient and overgrow impurities than was previously thought. With this in mind, we set out to generalize the process of single crystal matrix isolation for a number of reasons, not least of which is to understand crystal growth mechanisms and the role of interfaces therein.

Given a polar, chiral, optically responsive molecule mono-dispersed in an otherwise single crystal, we would like to deduce as much as possible about the growth processes from the optical properties of the included guest or its optical consequences on the host. In principle, optical measurements should allow us to determine the spatial disposition of impurities, their solvation and conformation states, as well as their orientation, symmetry, polarity, and chirality. And, in cases where we can detect polarity and enantioselectivity through nonlinear optical or chiroptical measurements we also will want to establish the sense of the polar axis and absolute configuration. Shown in Scheme 4.1 is a generalized optical probe represented as a chiral olefin with conjugated donor and acceptor that render the molecule dipolar and light absorbing. This representation is further generalized as an arrow appended to a chiral moiety indicated as *R* or *S*. Figure 4.2 then illustrates possi-

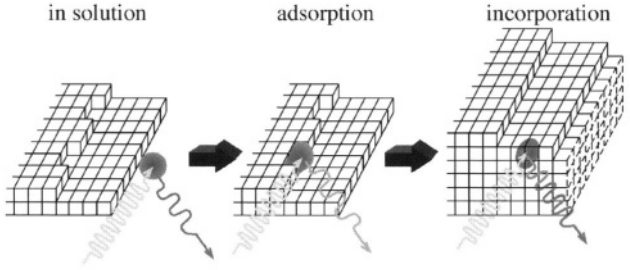
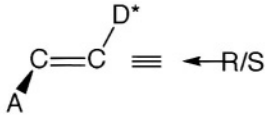


FIGURE 4.1. Cartoon of optical changes associated with the adsorption of an impurity at the kink site of a growing crystal, and the subsequent overgrowth. It is emblematic of the general strategy of using optical probes to study crystal growth mechanisms elaborated on herein.

ble guest distributions that we will describe throughout. (However, not all of the guests discussed herein are polar and/or chiral.)



Scheme 4.1.

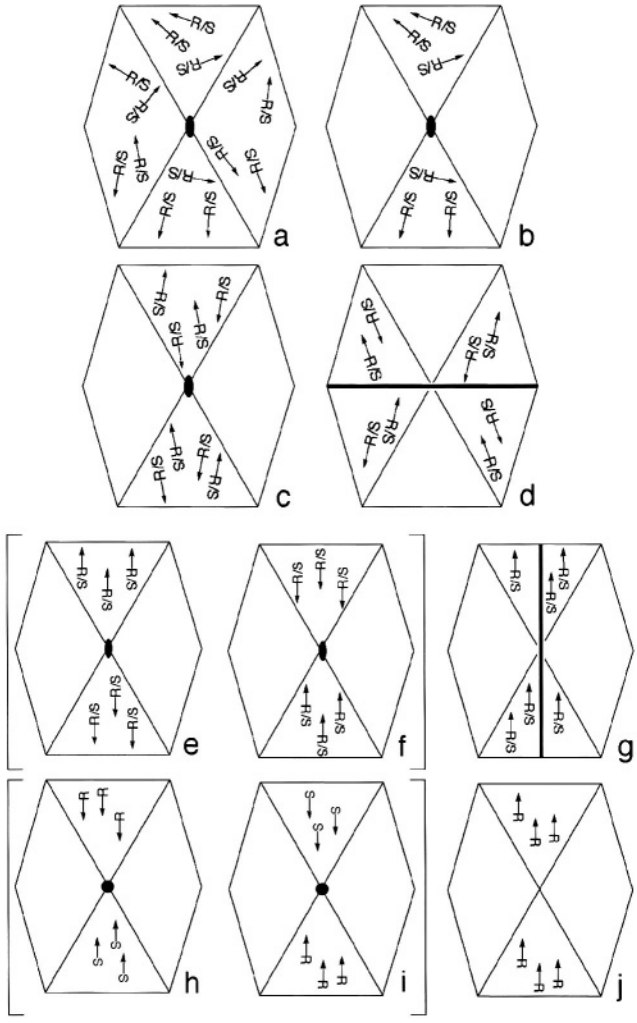


FIGURE 4.2. Schematic representation of two dimensional mixed crystals containing various distributions of guest molecules. (a) Random homogeneous distribution, (b) random growth sector specific distribtution, (c) oriented growth sector specific distributions, (d) oriented growth sector specific distribution with reduced symmetry, (e, f) polar guest distributions in crystals with diad symmetry, (g) polar guest distribution in a polar crystal, (h, i) oriented, polar, enantioselective distributions in a centric crystal, (j) enantioselective distribution in a polar crystal.

The optical methods employed here include measurements of the absorption and emission energies and anisotropies (linear dichroism and fluorescence anisotropy) [4], measurements of luminescence lifetimes of both singlet and triplet excited states [6], reflected light microscopy using the differential interference contrast method [7], and polarized light microscopy [8]. These are common techniques and we need not describe them here. Methods for measuring chiroptical properties of crystals are still in development and these we will discuss more fully in Sections 4.8 and 4.10. With the exception of measurements of second harmonic generation and two-photon luminescence, we restrict ourselves to linear crystal optics and single photon processes. There are numerous nonlinear, multi-photon spectroscopies currently in development that may well contribute to the characterization of some of the materials described herein [9].

Materials containing preformed layers, channels, or pores that serve as hosts for dyes will not be discussed here because, while of great interest, they do less to challenge our intuition about mixed crystal structure and crystal growth as free or loosely solvated spaces may naturally be filled. And, they have not been the subjects of the authors' research. We also exclude optical studies of species merely adsorbed to crystallographic surfaces. These are legion and space does not permit discussion here.

4.2. SPATIAL DISPOSITION

Impurities in crystals are rarely distributed uniformly. The heterogeneous incorporation of impurities among crystal growth sectors, volumes of a crystal that have grown in a particular direction through a specific face, is termed *intersectoral zoning* (Figure 4.3a). Intersectoral zoning is a general feature of impure crystals in which symmetry distinct

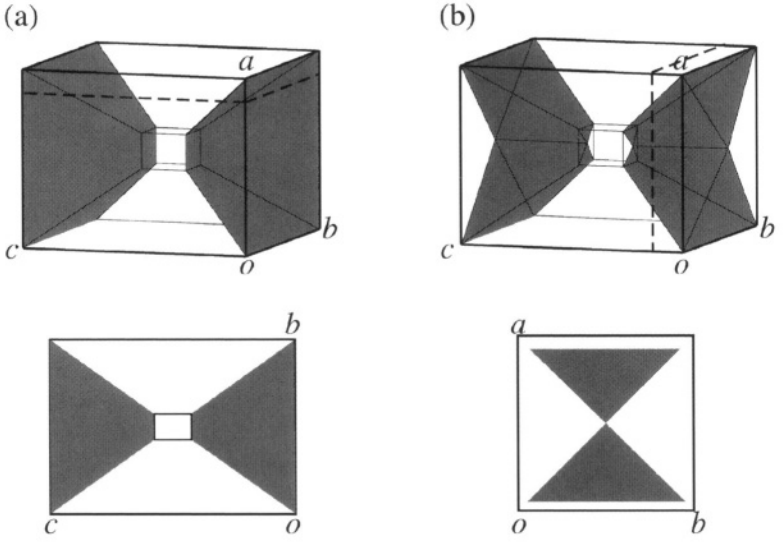


FIGURE 4.3. Intersectoral zoning (a) vs. intrasectoral zoning (b). The shaded regions represent the location of an impurity. The lower figures represent cross sections indicated by the dotted slices above.

facets express different affinities for additives. Figures 4.2a and 4.2b contrast crystals with diad symmetry (point group 2) containing a generalized impurity as described above. In both cases, the impurity is randomly oriented, but in 2a it is uniformly distributed while in 2b it is adsorbed and overgrown by only two of the six growth sectors. The latter case provides evidence of face-selective recognition. Intersectoral zoning of molecular guests is dramatically revealed in transparent crystals marked by dyes [10-13].

Impurities may inhomogeneously deposit *within* a single growth sector depending on the crystal's surface topography. Surfaces of crystals grown in the lower supersaturation regime often propagate through dislocations that produce growth spirals or hillocks [14,15]. Polygonization of hillocks partitions a face into vicinal regions, each having slightly different inclinations. Impurity partitioning among vicinal regions, *intrasectoral zoning* (Figure 4.3b), results from the selective interactions of additives with stepped hillock slopes. Intrasectoral zoning therefore provides more detailed information about recognition mechanisms than does intersectoral zoning because the active growth surfaces at the time of incorporation are more fully defined.

Figure 4.4 shows an especially rich example of intersectoral zoning in a crystal of potassium hydrogen phthalate (commonly abbreviated as KAP for potassium acid phthalate) grown from a millimolar solution of *trans*-4-[4-(dimethylamino)styryl]-1-methylpyridinium iodide (**1**) (Scheme 4.2). The luminescence was multicolored suggesting an illuminated quilt of sorts with at least 7 distinct regions in cross section. The differences in photophysical properties were consequences of acid/base chemistry, photoisomerizations and dimerizations, as well as aggregation. We will not describe the full characterization of these crystals in this review. Nonetheless, they best serve to illustrate the surprising range of large, optical reporters, that can selectively and specifically substitute for host crystal molecules or ions bearing no size, shape, nor compositional similarly to the guest molecules or ions.

Intrasectoral zoning in KAP was manifest in mixed crystals with the red, luminescing dye 1,1′,3,3,3′,3′-hexamethyldicarbocyanine (**2**). Compound **2** recognized (010) surface, but not uniformly. The pattern of luminescence was consistent with the recognition of the hillock steps growing in the $-c$ direction (Figure 4.5b) that were also evident by differential interference contrast microscopy (Figure 4.5a). While **2** did not change the shape of the polygonal hillocks, the triarylmethyl dye acid violet 17 (**3**, C.I. #42650) resulted in hillocks that were rounded on one side. Such hillock morphology changes have been observed previously [16], and were attributed to the selective poisoning of specific kink sites as illustrated in Figure 4.6.

Zaitseva and coworkers dyed the hillocks on the {101} faces of KH_2PO_4 thereby demonstrating both inter- and intrasectoral zoning [17]. This observation required introduction of sulfonated azo dyes during late growth thereby coloring only a thin surface layer so that patterns of color were not confounded by moving dislocation cores. Gurney *et al.* saw striking luminescent blue bands from *ortho*-aminobenzenesulfonate in K_2SO_4 {021} sectors [18]. These stripes corresponded to the slowly advancing steps of the macrospirals, α-Lactose monohydrate crystals grown in the presence of green fluorescent protein (GFP) luminesced exclusively from the (010) sector [19]. Differential interference contrast microscopy of a pure α-lactose monohydrate crystal revealed a single polygonized hillock [20,21] that partitioned the (010) surface into four vicinal faces pair-wise related by two-fold symmetry. A comparison of the interference contrast micrographs showed clearly

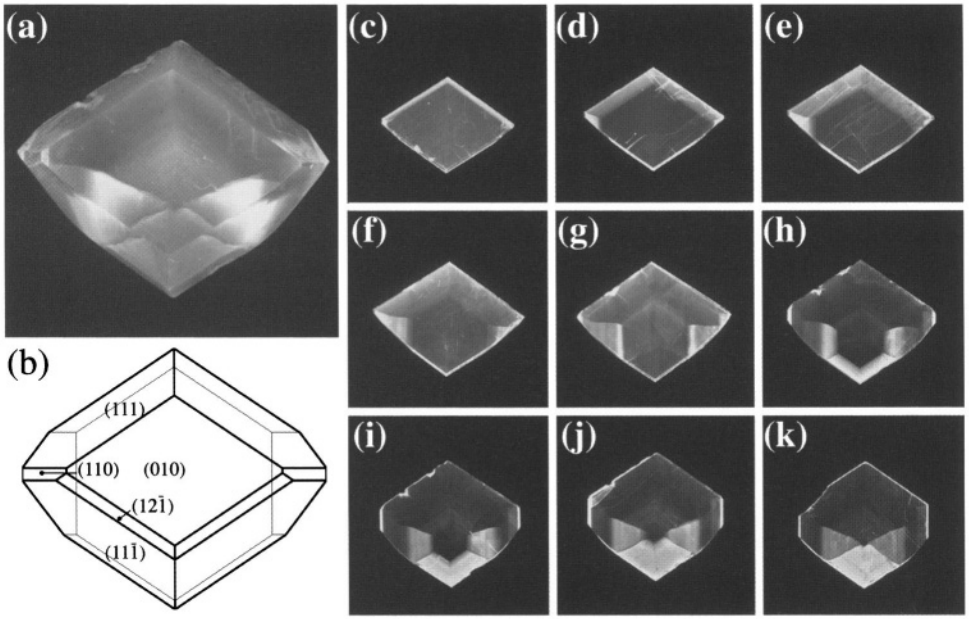
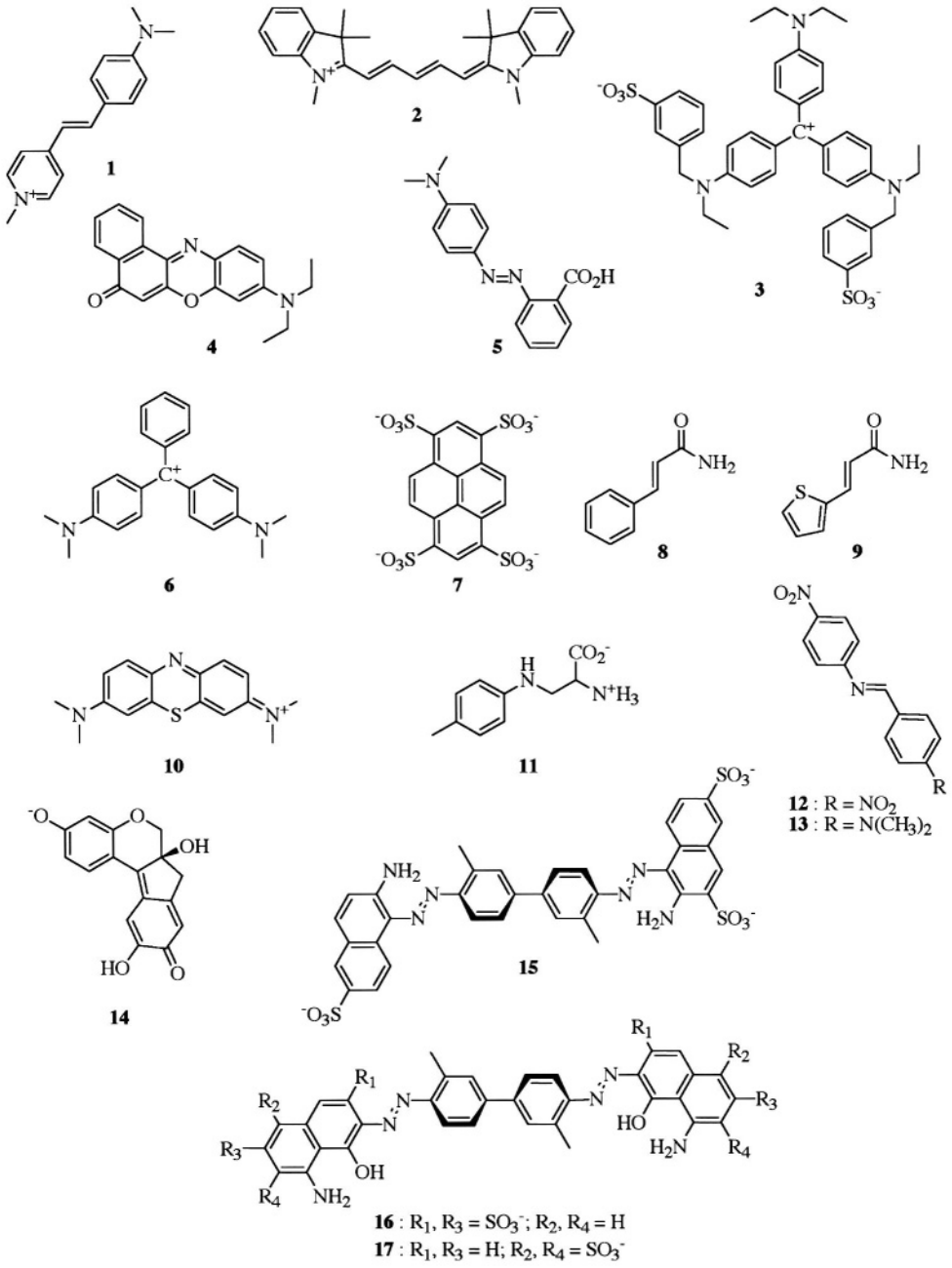


FIGURE 4.4. Intersectoral zoning in potassium hydrogen phthalate crystal grown in the presence of 4-[4-(dimethylamino)styryl]-1-methylpyridium iodide (**1**). (a) The crystal as grown, and (b) schematic representation of the habit, (c–k) are 100 μm slices of (a) from the bottom of the crystal (k) to the top (c). The slices reveal an astonishing range of photophysical properties associated with the guests in five different growth sectors and at different times during growth.

that GFP only recognized the lateral slopes with the greatest step advancement velocity. These examples illustrate how colored or luminescent molecules can be used to quickly image dislocations, even long after they have ceased to be active and have been overgrown. Imaging dislocations within crystals is usually accomplished only by much more laborious technique of X-ray topography [22].

4.3. SOLVATION AND CONFORMATION

The photophysical properties of chromophores inside of single crystals reflect interactions on growing crystal surfaces. These encounters can be detected as solvatochromism or more particularly in the protonation states of included guests that respond to surface charge and acidity. For example, the same dyes included in K_2SO_4 and KH_2PO_4 [23,24] were typically red and blue shifted respectively, compared with saturated solutions of these salts. The red shifts were comparatively slight and most likely a consequence of the fact that in an ionic medium the ground states of the dyes were closer in character to charge-transfer transition states [25]. The blue shifts undoubtedly arose because lone electron pairs that form hydrogen bonds in KH_2PO_4 raised the transition energies by lowering the energies of the ground states. For example, benzamide dyed in the $\{10,\bar{2}\}$ sectors with Nile red (**4**) showed an excitation polarization consistent with the pre-organized hydrogen bonding of **4**



Scheme 4.2.

to three benzamide molecules in a crystal chain. The energy of the luminescence was consistent with H-bonding found in polar protic solvents where the $-N(CH_2CH_3)_2$ group has turned out of plane [26].

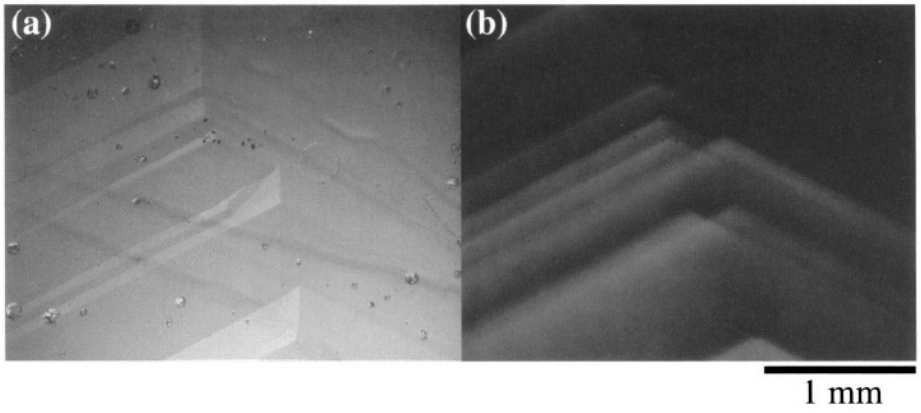


FIGURE 4.5. Differential interference contrast micrographs of KAP (010) face grown in the presence of 2.5 μM 1,1′,3,3,3′,3′-hexamethylindodicarbocyanine iodide (**2**) (a) compared with (b), the luminescence micrograph. The luminophores reveal individual hillocks in the disposition of luminescent chevrons associated with intrasectoral zoning.

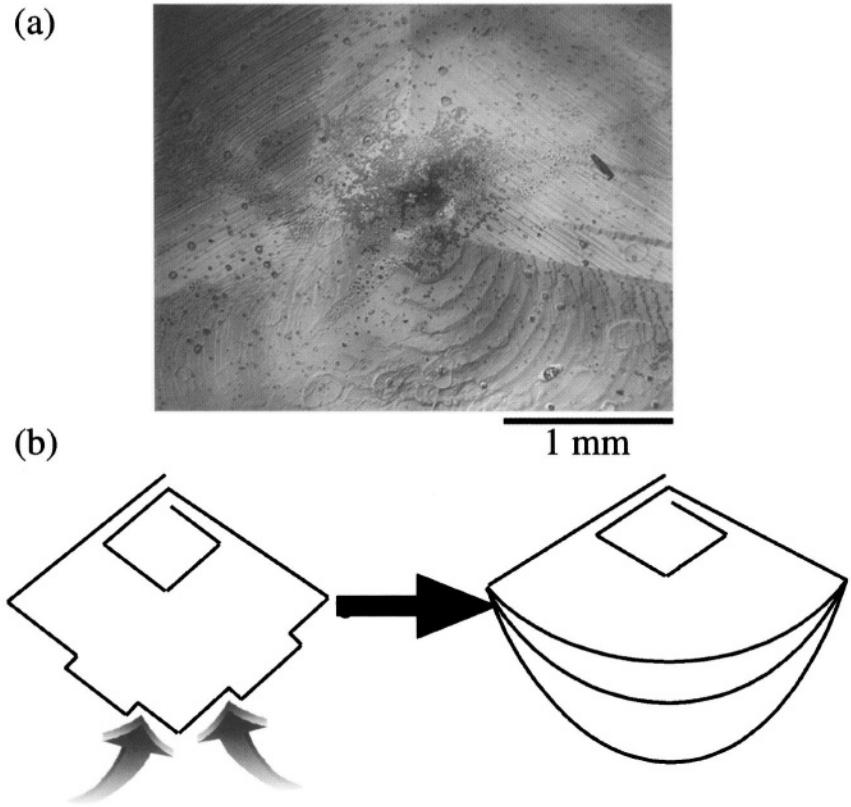


FIGURE 4.6. Differential interference contrast micrograph of potassium acid phthalate (KAP) (010) face grown in the presence of 17.2 μM acid violet 17 (**3**) (a) and schematic illustration of the hillock morphology change on selectively binding to a subset of kink sites (b).

Nile red (**4**) and methyl red (**5**, C.I. #13020) showed different colors in phthalic acid depending upon the faces through which they had adsorbed [27]. Benedict *et al.* found that a variety of acid/base indicators could be oriented in specific growth sectors or aromatic acid crystals in various protonation states [28].

Phthalic acid oriented a great variety of cationic dyes, especially in its {021} growth sectors [29]. Red shifts in the absorption spectra of the crystals indicated the possibility of hydrogen phthalate/dye charge-transfer complexes [30]. Such associations were confirmed in crystal structures of stoichiometric hydrogenphthalate/malachite green (**6**) co-crystals. This co-crystal structure was used in the interpretation of the linear dichroism of the corresponding mixed crystal.

Gurney *et al.* studied the excited state lifetimes of room temperature phosphorescent *ortho*-aminosulfonated benzenes and naphthalenes from which they found evidence of cation–π interactions [31] through the observation of heavy atom effects [32]. The triplet lifetime of *ortho*-aminobenzenesulfonate was shorter in Rb_2SO_4 as compared with isomorphous K_2SO_4. Similarly, phosphorescence lifetimes of *para*-aminobenzoate were shorter in barium acetate crystals as compared to sodium acetate trihydrate crystals. It had been demonstrated by optical detection of magnetic resonance that intersystem crossing rates for aromatic compounds are only affected by heavy atoms if they are interacting with the faces of the π-systems [35].

Flexible molecules may respond to surfaces by changing their conformations. Distinct facets of a single crystal may express different affinities for guests that can exist in more than one conformation. Crystals of K_2SO_4 incorporating *ortho*-aminobenzenesulfonate showed an astonishing range of photophysical responses in different growth sectors [32]. The luminescence energies and lifetimes were inversely correlated; the more energetic the emission, the shorter the lifetime [34]. Such a correlation suggested that the differences among the molecules in the sectors were a consequence of a progressive rotation of the $-NH_2$ lone electron pair out of conjugation with the π system [35,36]. Similar observations were obtained for a variety of *ortho*-aminoarenesulfonates [37]. Other examples of conformational selectivity have been published previously [38]. The observations and measurements described herein speak to the nature of the first coordination sphere surrounding the impurities that must reflect the dynamic processes at the growing crystal interfaces.

4.4. ORIENTATION

In crystals, oriented guests are the rule rather than the exception. Orientation can be evidenced by polarized light spectroscopy when the host is transparent in a region of the electromagnetic spectrum where the guest is absorbing, or luminescing. Schematic Figure 4.2c represents an oriented, growth sector specific impurity distribution. Because the guest, while oriented, can adopt more than one orientation, the diad symmetry of the host crystal is not broken in any of the growth sectors.

A single guest orientation can preserve the crystallographic symmetry if the host and guest have the same point symmetry. A simple example is that of 1,3,6,8-pyrenetetrasulfonate in K_2SO_4 crystals (**7**) (Figure 4.7) [39]. Here, both the crystal and the guest molecules have *mmm* symmetry. The dichroism and fluorescence anisotropy were very high. This

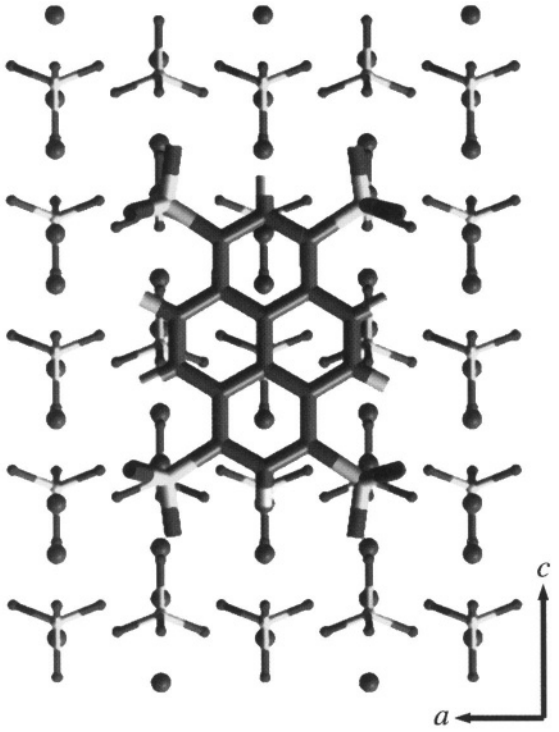


FIGURE 4.7. 1,3,6,8-Pyrenetetrasulfonate (**7**) superimposed within the K_2SO_4 lattice such that sulfur atoms in sulfonate groups were best fit to the sulfur atoms of the sulfates ions in the lattice related to one another by the *a* and *c* lattice constants.

is because the molecules recognized the host lattice by substituting their coplanar sulfonate substituents for sulfate ion pairs that were parallel to the basis vectors and extinction directions of the host. Symmetry requires that the dominant transition moment of **7** is parallel with the long molecular axis. Given the recognition mechanism, the moment was aligned with extinction directions in the birefringent crystal, thereby resulting in the large anisotropies.

The sulfonate–sulfate substitution mechanism, while generalizable, can not always be relied as other factors may play a decisive role [25]. In other dyed crystals, the orientation of the dye was most likely determined by the lamellar structure of the crystal as opposed to very particular non-covalent interactions [40].

We have measured the polarized absorption or excitation spectra of hundreds of dyed crystals. When the anisotropy is high, orientations can be proposed that suggest stereospecific recognition mechanisms giving rise to these orientations at surfaces. This mass of data needs to be approached cautiously. The appeal of simple, intuitive, specific non-covalent interactions between host and guest can be overextended. When the anisotropy is very high it is safe to assume narrow orientation distributions. However, when the anisotropy is low, one can not distinguish by polarized light microscopy whether there is a distribution of orientations or whether there is a narrow distribution with the relevant dipoles oriented between extinction directions. Furthermore, it is often impossible to distinguish multiple symme-

try related or electronically equivalent orientations from a single member of such a set. Single molecule spectroscopy [41] will enable us to determine polarization distributions of individual chromophores. In this way, we can resolve a number of the aforementioned ambiguities.

4.5. SYMMETRY REDUCTION

Shown in Figure 4.2d is an example of two-dimensional crystal with *m* symmetry. Here, reflections within the four symmetry-related growth sectors are obviated by the unidirectional growth of any one sector.

We now know that solid solutions are routinely less symmetric than pure counterparts because guests will be unequally distributed among surface sites corresponding to symmetry-related positions in the lattice when these sites have different presentations on a growing crystal face [42–44]. The resultant crystal will be an assembly of growth sectors. Researchers from the Weizmann Institute showed that, for example, cinnamamide (**8**), which crystallizes in the centrosymmetric space group $P2_1/c$, when grown with **2**-thienylacrylamide (**9**), formed mixed crystals (~7.5 mol% **9**) with plate-like habits composed of six principal growth sectors of reduced symmetries. (Thienyl and phenyl rings are isomorphous.) The {011} sectors had triclinic symmetry, $P1$, by X-ray diffraction, whereas the {100} sectors maintained their mirrors (Pc).

Symmetry reductions can be detected optically. In fact, the study of crystals whose optical symmetries were not in accord with their morphological symmetries by virtue of desymmetrization was a vigorous aspect of crystallographic research prior to the inception of X-ray diffraction [45]. These materials that seemingly violated the Neumann–Curie principle [46], termed *optically anomalous,* were ubiquitous in the literature of the Nineteenth Century [45]. In most instances highly symmetric forms had lower symmetries as evidenced by the appearance of linear birefringence where otherwise there would be none.

Examples of optically anomalous crystals whose properties could be explained in terms of the selective adsorption of impurities to symmetry related sites include bromoundecanoylperoxide containing a variety of other long-chain acyl peroxides [47], the disordered dichlorodinitrobenzene [45,48] and $Ba_xPb_{1-x}(NO_3)_2$ [49]. Other examples of optically anomalous crystals not readily explained in this way include $NaCl_xBr_{1-x}O_3$ [50]; here the photoelastic consequences of strain affect the measured retardation.

Gaubert was the first to observe anomalous linear birefringence [47] in dyed crystals during his investigations of sodium and ammonium halides [51]. Slavnova also detected anomalous linear birefringence in dyed alkaline earth nitrates [52].

Symmetry reduction can be detected in measurements of linear dichroism as well as linear birefringence. France and Davis observed anomalous linear dichroism in cubic alum crystals containing various dyes [52]. Frondel observed anomalous linear dichroism in the cubic alkali fluorides [53]. Cubic barium nitrate crystals grown in the presence of methylene blue (**10**) typically exhibited strong linear dichroism as shown in Figure 4.8, where the angles between dyes in adjacent sectors equal the angles between the corresponding growth faces [55].

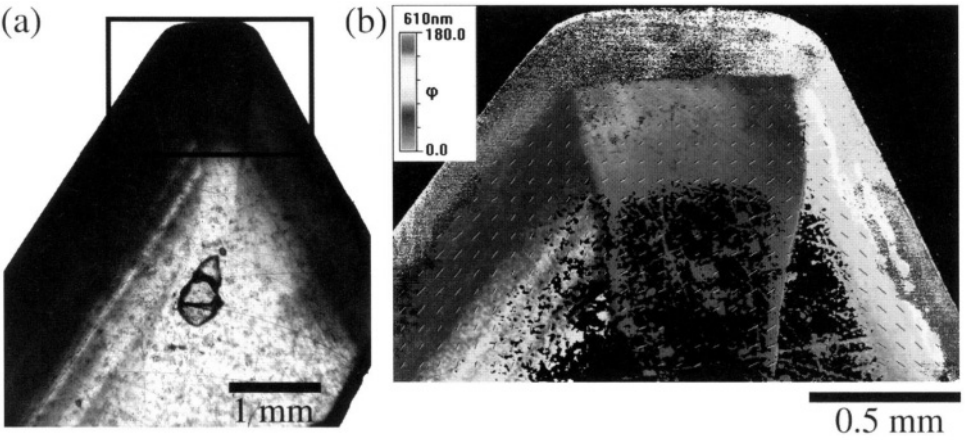


FIGURE 4.8. Linear dichroism image of $Ba(NO_3)_2$ grown in the presence of methylene blue (**10**, C.I. #52015) (a). The digital micrograph (b), made with a linear birefringence and linear dichroism imaging system, shows the polarization of the absorption as a false-color map plotted on the image.

4.6. POLARITY

Figures 4.2d and 4.2e contrast crystals with oriented sublattices. In each case, the symmetry of the host is broken. In Figure 4.2d, each sector loses its mirror symmetry but the reflection between sectors is preserved. In Figure 4.2e (as well as 4.2f) we see the loss of a two-fold rotation resulting from a polar distribution of dipolar guest molecules.

Many physical properties of crystals depend on a polar axis, and thus there are many ways to establish whether symmetry reduction results in a polar sublattice [56]. As no crystal surface—or any surface for that matter—can be centrosymmetric, it is self-evident that mixed crystal growth must often lead to polar sublattices. Nevertheless, the extent of the selectivity will determine the magnitude of physical properties that are associated with polarity.

With the knowledge that symmetry is often broken in the act of growing a mixed crystal, Weissbuch *et al.* set out to design a crystal with second order non-linear optical properties by reducing the symmetry of a centrosymmetric host with carefully chosen impurities having large molecular hyperpolarizabilities [57]. For example, N-γ-(*p*-nitrophenyl)-α,γ-diaminobutyric acid (**11**), or other comparable amino acids bearing charge transfer chromophores were included in growing crystals of (*S*)-glutamic acid, or glycine. Mixed crystals containing <0.1% weight percent **11** showed easily measurable phase matched second harmonic generation (SHG) signals. In some cases, the directional dependence of the SHG signal specified the nature of the symmetry reduction accompanying mixed crystal growth. We have observed that centrosymmetric crystals such as phthalic acid that were dyed with organic compounds having large hyperpolarizabilities became SHG active [58].

The growth of polar clathrate crystals containing dipolar chromophores was systematically studied by Hulliger and coworkers [59]. They developed phase sensitive second harmonic microscopy to investigate systems such as perhydrotriphenylenecontaining 1-(4-nitrophenyl)piperazine. In this method, contrast occurs because oppositely oriented macroscopic polar domains cause a reference 532 nm beam to interfere constructively, or destruc-

tively, with the second harmonic signal generated from the fundamental at 1064 nm [60]. Whether this can be made to work with dilute mixed crystals remains to be determined, but it appears to be a promising method for analyzing induced crystal polarity.

4.7. SENSE OF THE POLAR AXIS

It is one thing to establish that a guest sublattice is polar and it is another to determine the sense of the polar axis. In our schematic Figure 4.2 this is tantamount to distinguishing between the arrangements in 4.2e and 4.2f.

Previously, the sense of a polar sub-lattice was assumed by Weissbuch *et al.* given the expectation of substitution of host and guest zwitterionic amino acid substituents (see Section 4.6). They also reported the generation of SHG by including 10–30 wt% of isomorphous impurities such as of *p, p′*-dinitrobenzylideneaniline (**12**) in *p*-(N-dimethylamino)benzylidene-*p′*-nitroaniline (**13**) crystals. Here, the sense of the polar axis could in principle be determined by careful X-ray scattering experiments of individual growth sectors. The sense of the polar axis was nevertheless assumed given the reasonable expectation that nitro–nitro host–guest interactions were repulsive, thereby biasing the guest orientation [57]. In admixtures with 0.01 mole % of guest (Section 4.6), the determination of the sense of the polar axis could not be determined by X-ray diffraction even though such small quantities were sufficient to produce significant non-linear optical activity.

One strategy for determining the sense of the polar axis in mixed crystals is to extrapolate features of stoichiometric co-crystals. For example, *trans*-4-[4-(dimethylamino)styryl] pyridine (**1**) stained the {021} and {001} sectors of phthalic acid. It also formed cocrystals with phthalic acid in a ratio of 2:3 (Figure 4.9). The pyridines were hydrogen bound to carboxylate groups on the ends of the phthalic acid triad. The arrangement of the phthalic acid molecules in the mixed crystals mirrored that in the crystal structure of pure phthalic acid [61]. Thus, the intermolecular interactions that were manifest in the co-crystals might well carry over to the mixed crystals. Confirmation of this proposition awaits the determination of a series of crystal structures and the measurement of the linear dichroism in a comparable series of mixed crystals.

Malachite green (**6**) was previously co-crystallized with phthalic acid. In this way we proposed a mixed crystal structure [29]. Figure 4.10 shows the optical properties of phthalic acid crystals dyed with **6**. Figure 4.10a is an optical micrograph in transmitted light. The corresponding false color images result from a Fourier analysis of intensity data collected as a function of a forced linear polarizer [62]. In this way, we have separated the contributions to the intensity (4.10b) that result from absorption (expressed as % transmittance) the linear dichroism (hyperbolic tangent of ε, where $\varepsilon = (\varepsilon_1 - \varepsilon_2)/(\varepsilon_1 + \varepsilon_2)$ and ε_1 and ε_2 are the extinction coefficients in orthogonal directions, Figure 4.10c) and dipole orientation (ϕ, Figure 4.10d). These new optical imaging techniques provide quick and detailed insights into the orientational distributions of absorbing components of mixed crystals.

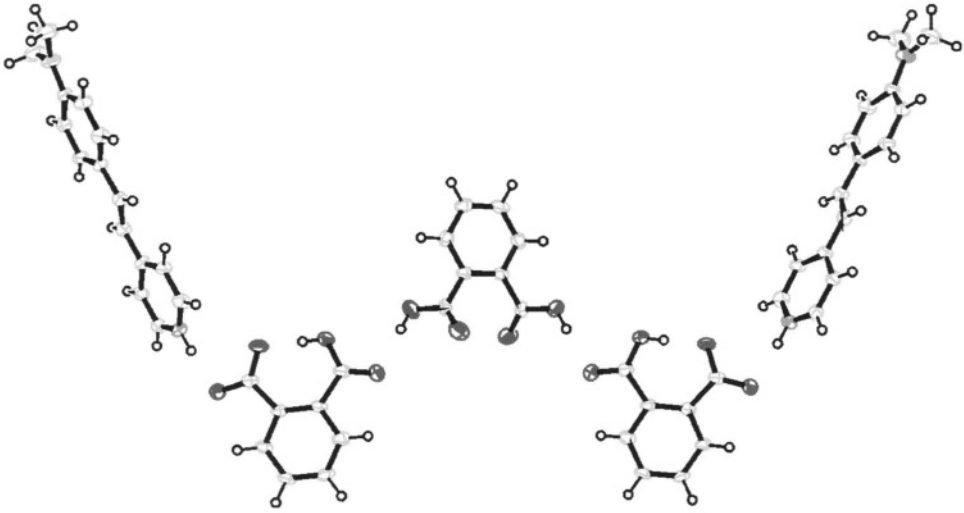


FIGURE 4.9. Co-crystal structure of **1** and phthalic acid in a ratio of 2:3. The triad of phthalic acid molecules mimics the arrangement of phthalic acid molecules in crystals of pure phthalic acid.

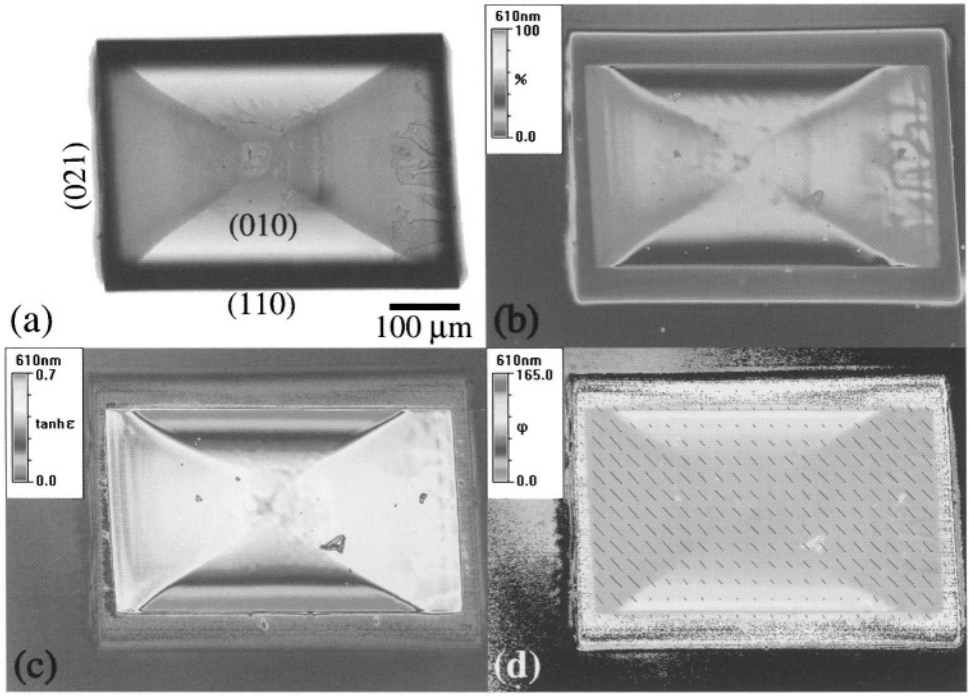


FIGURE 4.10. Crystal of phthalic acid grown in the presence of malachite green (**6**). (a) Micrograph in transmitted light and (b, c, d) digital images made with linear dichroism imaging system. (b) Transmittance in %; (c) hyperbolic tangent of the linear dichroism; and (d) the angle of maximum absorbance measured from the horizontal axis.

4.8. ENANTIOSELECTIVITY

In Figures 4.2g and 4.2h our general racemic guest, $\leftarrow R/S$, is now separated as $\leftarrow R$ or $\leftarrow S$ in crystallographic faces that are related to one another only by a center of symmetry. In other words, the guests recognize the faces enantioselectively.

Chiral discrimination of organic compounds by minerals has been a rich subject for speculation since the early suggestions by Goldschmidt [63] and Bernal [64] that chiral minerals or quartz may have been responsible for the origin of the asymmetry in biopolymers [65]. The Weizmann group [66] and others [67] have subsequently worked out many mechanisms of chiral discrimination by crystal surfaces. These studies used racemic mixtures of enantiomers with fixed configurations (e.g., amino acids) with homochiral crystallites (e.g., quartz) [68] or resolved chiral compounds with fixed configurations with mirror image crystal facets (Figure 4.11) [69].

Enantioselective adsorption was also detected optically. Blattner *et al.* reported that solutions of hematein (**14**) stained the {100} growth sectors of KH_2PO_4 [70]. Recent reinvestigation of their work revealed that hematein stained every other prism face [71]. Hematein (**14**) is chiral; adjacent prism faces of KH_2PO_4 are mirror symmetric in the space group $I\bar{4}2d$. The adsorption therefore must have been enantioselective. Adenosine phosphates also adsorbed selectively to the {100} and {010} growth sectors as evidenced by differences in the intensities of the luminescence of growth sectors that would otherwise have been related by symmetry [72].

Enantioselective adsorption is evident in intrasectoral as well as intersectoral zoning. Orme *et al.* presumed that racemic mixtures such as *D*- and *L*-aspartic acid interacted enantioselectively with the mirror symmetric growth hillocks of calcite because the right or left handed curving of the step fronts depended upon which enantiomer was added during crystallization [73]. Unfortunately, the authors were unable to detect aspartic acid in or on the crystals. Therefore, the enantioselectivity was deduced from the results of computational modeling with empirical force fields.

When the configuration of chiral impurities is fixed, enantioselectivity can be judged by removing individual growth sectors and liberating the molecules upon dissolution of the crystal. Subtler, however, are those instances of enantioselectivity that result from the adsorption of equilibrium racemic mixtures on enantiomorphous crystal surfaces.

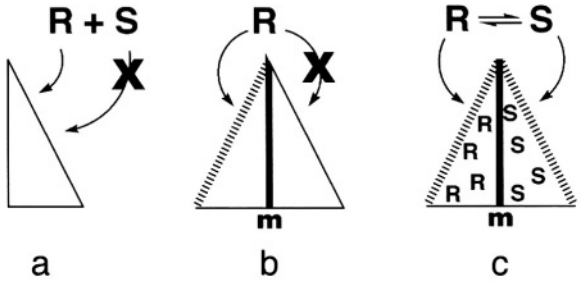


FIGURE 4.11. Three different examples of chiral discrimination by crystals: (a) The resolution of a racemate with fixed configurations by a chiral crystal. (b) The assignment of the absolute configuration of enantiomorphous faces using a resolved chiral molecule. (c) The direct detection by optical rotation imaging of chiral conformations in rapid equilibrium oriented in enantiomorphous crystal faces.

We attempted to use chiroptical methods to reveal selectivity of this kind. We sought a guest with a large intrinsic optical rotation that when enantioselectively adsorbed onto mirror image vicinal facets of a hillock should give opposite signs of optical rotation on other side of the hillock. These conditions were met by triarylmethyl dyes that adsorbed, presumably enantioselectively, to the mirror image slopes of achiral hillocks on the {010} faces of KAP. This is in essence the calcite/asapartic acid system recast with a host/guest pair that can be investigated optically. However, such measurements press the limits of current analytical methods. The measurement of chiroptical properties of anisotropic media has been a centuries-old problem.

The difficulty with measuring optical rotation in crystals was first encountered by Arago, who discovered optical rotation in 1811, when he passed linearly polarized light along the high-symmetry *c*-axis of quartz [74]. However, the determination of the anisotropy of optical rotation—the complete gyration tensor—by making off-axis measurements was impossible for Arago, and a challenging area of research as late as 1988 [75]. This is because linear birefringence is often $\sim 10^3$–10^4 as large as circular birefringence (optical rotation) and the minor chiroptical perturbation to the polarization state of light in anisotropic media is of the same order of magnitude as parasitic ellipticities from imperfect samples and the polarimeter components.

In order to measure optical rotation along a general direction of a crystal one has to understand how the combination of linear birefringence ($\Delta n = n'' - n'$) and circular birefringence (or optical rotation, $\Delta n = n_R - n_L$) affects the polarization state of light. The key insight is that the *intensity* of light passing through a polarizer, chiral anisotropic sample, and analyzer, contains all of the necessary information, in principle, for extracting optical rotation parameters [76]. However, the implementation of this idea prior to the invention of electrophotometry and stable, high-intensity light sources was impossible.

In 1983, Uesu and Kobayashi used photon counting techniques, lasers, and computerized modulation of polarizer and the analyzer orientations to determine optical rotation in crystals for directions off of the optic axes. This so-called HAUP (high accuracy universal polarimetry) method applies reliably only to small values of linear birefringence; the contribution to the intensity from optical rotation varies as does $\sin\delta/\delta$, where δ is the optical retardation ($=2\pi\Delta nL/\lambda$, where Δn is the birefringence and L is the sample thickness) [77]. In HAUP, chiroptical contributions to intensity measurements must be large enough such that parasitic effects that result from the misalignment and the imperfect quality of optical components and samples can be neglected [78]. Moreover, when the phase difference between propagating modes is $\sim\pi/2$ or a multiple thereof, it is not always possible to derive the polarimetric properties by the HAUP method, as the governing equations become indeterminate.

Nevertheless, we set out to try to measure optical rotation in KAP dyed with **3** by using the HAUP method in an imaging mode whereby the sample, mounted on a translation stage, was scanned. HAUP was implemented with a polarimeter that was described elsewhere [79]. The hillocks on the {010} faces of KAP are mirror symmetric in keeping with the crystallographic point symmetry, *mm*2. **3** is an equilibrium racemic mixture of propellers in solution at room temperature. In principle, mirror image propellers should partition between the mirror image slopes of the hillocks. A HAUP map should show dextro and levorotatory domains from the crystal in Figure 4.12 with one dominant hillock. Unfortunately, we were unable to measure the optical rotation because the birefringence of KAP

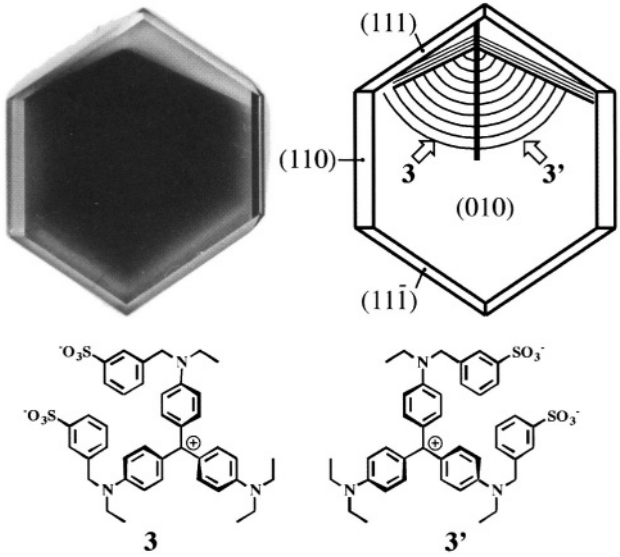


FIGURE 4.12. Intrasectoral zoning manifest in a crystal of potassium hydrogen phthalate grown in the presence of the triarylmethane dye acid violet 17 (**3**). The photograph in transmitted light (left) shows a crystal measuring 3 cm in length. The blue dye only recognized the fast growing steps of hillocks propagating towards the bottom of the page in the $-c$ direction. The crystal shown contains one dominant growth hillock centered at the upper left. The chiral dye forms an equilibrium racemic mixture in solution at room temperature. The hillocks are mirror symmetric. It therefore must mean that the dyes recognize the sides of the mirror image hillocks enantioselectively on the steps growing to the right or left of the mirror plane.

along [010] was very large (0.26) [80]. Even for thin sections ($100\ \mu m$) the retardation was insurmountable. (A related attempt is described in Section 4.10.)

4.9. ABSOLUTE CONFIGURATION

Determination of absolute configuration is arguably the most difficult task of the stereochemist or crystallographer. When and if we can use chiroptical methods to resolve the enantioselective adsoption of equilibrium racemic mixtures on enantiomorphous crystal facets we will then be faced with the challenge of assigning configuration from the sign of the measured effect. This assignment is essential for understanding the recognition processes at interfaces because the diastereomeric alternatives have different energies. In principle, computations can be a valuable guide. However, it has been notoriously difficult to calculate optical rotation ever since Rosenfeld expressed the phenomenon as the sum of electric (m) and magnetic dipole (μ) couplings for all of the electronic transitions: $R = \sum \mathrm{Im}\langle\Psi_0|m|\Psi_f\rangle\langle\Psi_f|\mu|\Psi_0\rangle$ [81]. Progress in the quantum mechanical calculations of optical rotation have been made recently [82,83]. (Classical, semi-empirical dipole–dipole methods work well in inorganic materials but they have yet to be tested for organic dyes [84].) It is likely that computations will be needed to guide the interpretation of chiroptical effects of impurities inside of crystalline hosts.

4.10. ANOMALOUS AZIMUTHAL ROTATION

In order to subvert the large birefringence in KAP that prevented the measurement of optical rotation, we set out to recast the problem with a weakly anisotropic host. As K_2SO_4 birefringence along [010] is small, 0.0038 [85] and crystals were polished to $<100\ \mu m$, HAUP was, in principle, appropriate for the measurement of the optical rotation because the retardation would be first order.

In 1934, Buckley described a crystal of K_2SO_4 dyed with brilliant Congo R (**15**, CI #23570) in the {111} sectors, thereby forming a Maltese cross-like arrangement of colored sectors (Figure 4.13) [86]. In K_2SO_4, with *mmm* (D_{2h}) symmetry [87] any face with a Miller index equal to 0 must be parallel to a mirror plane and achiral. Thus, the eight {111} faces are chiral and pair-wise enantiomorphously related to one another by reflections through the three orthogonal mirror planes. **15** is a biaryl dye that is chiral in its ground state although like **3**, it exists as an equilibrium racemic mixture of rapidly interconverting enantiomers in solution. Nevertheless, the chiral surfaces should adsorb the enantiomers selectively. We set out to determine whether the extent of the enantioselection would lend itself to measurement.

We were not able to stain K_2SO_4 with **15**, but did succeed with the related compounds trypan blue (**16**) and Evans blue (**17**). Thin (010) plates were cut and polished. Maps of K_2SO_4/**16** comprising $100 \times 100\ 30\ mm^2$ pixels were made of the optical rotation using the HAUP technique (Figure 4.14) [88]. Indeed, we measured signals associated with the oriented dyes that were consistent with optical rotation, but surprisingly, they did not transform like optical rotation. While, as required by symmetry, the signs of the effect were opposite for oscillators adsorbed through mirror image crystallographic facets, thereby suggesting enantioselective recognition, the signal changed sign with rotation of the sample

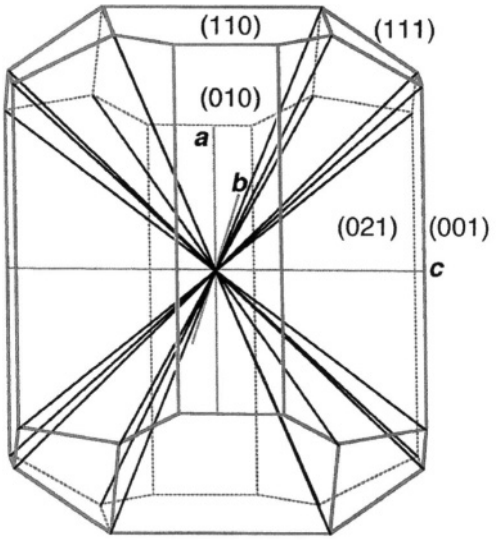


FIGURE 4.13. Idealized representation of K_2SO_4 habit (red). Black lines delinerate the {111} growth sectors.

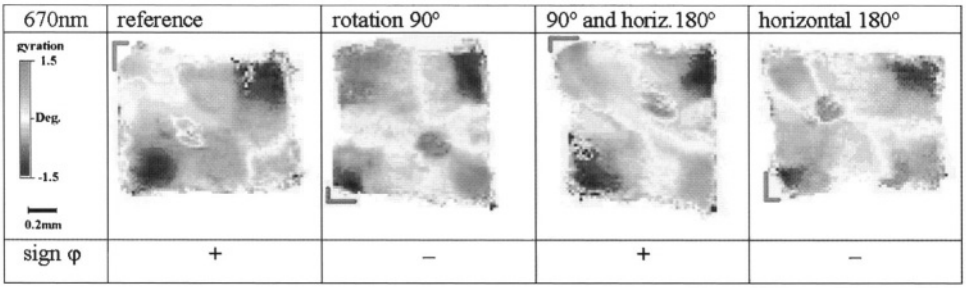


FIGURE 4.14. Optical rotatory effect (gyration in deg) of K_2SO_4 (010) slice containing **16**. Reference orientation compared with the same slice rotated, exchanged side for side and rotated, and exchanged side for side. Light and dark colors represent opposite signs of azimuthal rotation.

through 90° about the wave vector of the incident light. This was inconsistent with intrinsic optical rotation.

An alternative explanation relates the observed azimuthal rotation to the linear dichroism of the dye molecules. In the dyed samples, the anisotropy of the absorption does not coincide with that of refraction, however, even though the eigen ray directions are preserved the induced dipoles cause a perturbation of them. A polarized light wave along an eigenmode of the host passes a dye molecule but gets anisotropically absorbed. On the length scale of the dyes this leads to a rotation of polarization. Dyes **16** and **17** are ~2 nm long, and at a concentration of about 1 part in 10,000. With such dilute solutions there is no further interaction between the dyes. As a result we see anomalous azimuthal rotation (AAR).

If the dipoles in a {111} growth sector are similarly inclined with respect to host mirror symmetry, the resultant electromagnetic wave in the forward direction will undergo an azimuthal rotation about the wave vector. We call this anomalous azimuthal rotation (previously thought falsely to be a consequence largely of Rayleigh scattering and called optical rotatory scattering) [89].

AAR, unlike optical rotation, depends on the initial polarization direction. The effective angle of the superposition of the initial and scattered waves carries different signs for orthogonal incident polarizations. Because AAR is proportional to the induced electric dipole moment **R** it should be a much larger effect than intrinsic optical rotation which is given as the product of the magnetic and electric dipoles according to Rosenfeld's formulation.

As a rule of thumb, the ratio of magnetic and electric dipoles can be equated with the ratio of molecular size to the wavelength of light, which in this case is about 10^{-3} [4]. Thus, the contribution to an optical rotation signal due to scattering exceeds the intrinsic molecular contribution by about 1000 times. The observation of AAR does not require a chiral dye and may be observed with rigid, achiral oscillators so long as they are inclined *in the same sense* relative to the initial polarization. To the best of our knowledge, this effect has not been described previously because our mixed crystals have features that are unmatched elsewhere. The anisotropy of growth gives rise to oriented gases of strong oscillators inclined with respect to mirror planes of the medium but always with the same angular sign.

4.11. CIRCULAR DICHROISM

Of the two spontaneous chiro-optical phenomena, circular dichroism is more important in the structural analysis, but it has been the more difficult to measure in anisotropic media [90–93]. The complex interaction of optical rotation, linear birefringence, and linear dichroism, and also systematic instrumental errors can be devastating [94]. Nordén and coworkers showed that a major source of error in circular dichroism on oriented systems was the coupling of the residual strain induced linear birefringence of the modulator, and/or imperfect incident circularly polarized light, with the sample linear dichroism to give an apparent circular dichroism spectrum [95,96]. Because linear dichroism is large compared with CD, the slightest ellipticities can have disastrous consequences. Extraordinarily thorough analyses of these and other artifacts that one can encounter in the measurement of circular dichroism on oriented systems with imperfect optical components have been given by Schellman, Jensen and coworkers [97], Shindo and coworkers [98], and Kuball and coworkers, among others [99].

In 1982, Maestre and Katz built a micro-CD spectrometer [100] for measuring the average circular dichroism spectra of individual cells, chromatin, and chromosomes but were still stymied by instrumental artifacts [101]. They also conceived differential circular polarization imaging in which maps of individual Mueller matrix elements, including CD, would be produced independently [102]. They devised a method of image analysis and constructed a microscope for experimentally separating the various optical effects. However, Maestre and coworkers ultimately concluded that defects in the optical train and imperfect polarization modulation caused inseparable mixing of the circular dichroism with the much larger linear anisotropies.

The circular dichroism and induced-circular dichroism of guests in crystals should be very informative. A circular dichroism imaging microscope could be applied to the crystals in Figures 4.12 and 4.13. We recently described such a device. Future work will be directed toward circular dichroism imaging [103].

4.12. CONCLUSIONS

Optical probes are commonly used in biochemistry to study the specificity of non-covalent interactions among biopolymers. On the other hand, optical reporters have not been used to study the specificity on non-covalent interactions governing crystal growth from solution. This is undoubtedly because of the widely presumed restrictions on the preparation of mixed crystals laid out in the introduction. How far can this analogy between crystals and tissues carried? What information can be obtained about crystal growth mechanisms from the optical consequences of mixed crystal growth?

The set of experiments described herein obviously constitutes a work in progress. We aimed to make optical measurements to determine the spatial disposition of impurities, their solvation and conformation states, as well as their orientation, symmetry, polarity, and chirality. We also set out to establish the sense of polar impurity sublattices and the absolute configuration of species adsorbed enantioselectively. In so doing, we needed to stretch the optical analysis of mixed crystals by embracing the latest developments in polarized light microscopy and polarimetry. Successes were described, but also the failures, particularly

with respect to defining the sense of polar axes and absolute stererochemistry. However, in failing we have revealed new phenomenon, such as AAR.

The study of the optical consequences of molecular interactions at growing crystal interfaces is still in its infancy. There is every reason to believe that the subject will richly develop as other researchers bring their expertise to bear on the generalization of single crystal matrix isolation.

ACKNOWLEDGMENT

We are grateful for support from the National Science Foundation, the Petroleum Research Fund of the American Chemical Society, and the University of Washington Center for Nanotechnology.

REFERENCES

1. T.R. Welberry and B.D. Butler, Diffuse X-ray scattering from disordered crystals, *Chem. Rev.* **95**(7), 2369–2403 (1995).
2. I. Freund, *The Study of Chemical Composition: An Account of Its Method And Historical Development* (Dover, New York, 1968) pp. 385–53.
3. A.I. Kitaigorodsky, *Springer Series in Solid-State Sciences, Vol. 33: Solid Solutions* (Springer-Verlag, Berlin, 1984) p. 201.
4. J. Michl and E.W. Thulstrup, *Spectroscopy with Polarized Light. Solute Alignment by Photoselection, in Liquid Crystals, Polymers, And Membranes* (VCH Verlagsgesellschaft, Weinheim, 1986) p. 130.
5. D.L. Klug, in: *Handbook of Industrial Crystallization,* Ed. A.S. Meyerson (Butterworth-Heinemann, Boston, 1993).
6. J.R. Lackowitz, *Principles of Flourescence Spectroscopy,* 2nd edn (Kluwer Academic/Plenum, 1999).
7. M. Pluta, in: *Specialized Methods: Advanced Light Microscopy* (Elsevier, New York, 1989) Vol. 2, p. 146.
8. N.H. Hartshorne and A. Stuart, *Crystals and the Polarizing Microscope,* 4th edn (Edward Arnold, London, 1970).
9. Y.-P. Svirko, *Polarization of Light in Nonlinear Optics* (John Wiley, New York, 1999).
10. O. Lehmann, Über künsliche Färbung von Krystallen und amorphen, *Ann. Phys. Chem.* **51**, 47–76 (1894).
11. A. Neuhaus, Partly isomorphous systems, *Zeit. Krist.* **105**, 161–219 (1944).
12. H.E. Buckley, *Crystal Growth* (John Wiley & Sons, Inc., New York, 1951).
13. B. Kahr and R.W. Gurney, Dyeing crystals, *Chem. Rev.* **101**(4), 893–951 (2001).
14. W.J.P. Van Enckevort, Surface microtopography of aqueous solution grown crystals, *Prog. Cryst. Growth Charact.* **9**(1–2), 1–50 (1984).
15. I. Sunagawa, in: *Materials Science of the Earth's Interior,* Ed. I. Sunagawa (Terra Scientific Publishing Co., Tokyo, Japan, 1984) pp. 63–105.
16. M.H.J. Hottenhuis and C.B.J. Lucasius, The role of impurities on the process of growing potassium hydrogen phthalate crystals from solution; a quantitative approach, *J. Cryst. Growth* **91**(4), 623–631 (1988); M.H.J. Hottenhuis and A. Oudenampsen, $\mathbf{Iron(3^+)}$ and $\mathbf{cerium(3^+)}$ acting as competitive impurities in the crystallization process of potassium hydrogen phthalate from solutions, *J. Cryst. Growth* **92**(3–4), 513–529 (1988); M.H.J. Hottenhuis and C.B. Lucasius, The influence of impurities on crystal growth; *in situ* observation of the {010} face of potassium hydrogen phthalate, *J. Cryst. Growth* **78**(2), 379–388 (1986).
17. N. Zaitseva, L. Carman, I. Smolsky, R. Torres and M. Yan, The effect of impurities and supersaturation on the rapid growth of KDP crystals, *J. Cryst. Growth* **204**(4), 512–524 (1999).
18. R.W. Gurney, M. Kurimoto, J.A. Subramony, L.D. Bastin and B. Kahr, in: *Anisotropic Organic Materials: American Chemical Society Symposium Series,* Eds R. Glaser and P. Kaszynski (American Chemical Society, Washington, DC, 2001).

19. M. Kurimoto, P. Subramony, R.W. Gurney, S. Lovell, J. Chmielewski and B. Kahr, Kinetic stabilization of biopolymers in single crystal hosts: green fluorescent protein in α-lactose monohydrate, *J. Am. Chem. Soc.* **121**(29), 6952–6953 (1999).
20. T.D. Dincer, G.M. Parkinson, A.L. Rohl and M.I. Ogden, in: *Crystal Growth Org. Mater. 4th International Workshop,* Ed. J. Ulrich (Shaker Verlag, Aachen, Germany, 1997) pp. 25–32.
21. T.D. Dincer, G.M. Parkinson, A.L. Rohl and M.I. Ogden, Crystallisation of alpha-lactose monohydrate from dimethyl sulfoxide (DMSO) solutions: influence of beta-lactose, *J. Cryst. Growth* **205**(3), 368–374 (1999).
22. B.K. Tanner and D.K. Bowen, Synchrotron X-radiation topography, *Mat. Sci. Rep.* **8**(8), 369–407 (1992).
23. J.A. Subramony, S.-H. Jang and B. Kahr, Dyeing KDP, *Ferroelectrics* **191**(1–4), 292–300 (1997).
24. B. Kahr, S.-H. Jang, J.A. Subramony, M.P. Kelley and L.D. Bastin, Dyeing salt crystals for optical applications, *Adv. Mater.* **8**(11), 941–944 (1996).
25. L.D. Bastin and B. Kahr, Engineering oriented gases: the mechanism of dyeing potassium sulfate. *Tetrahedron* **56**(36), 6633–6643 (2000).
26. N. Sarkar, K. Das, D.N. Nath and K. Bhattacharvya, Twisted charge transfer processes of nile red in homogeneous solutions and in faujasite zeolite, *Langmuir* **10**(1), 326–329 (1994).
27. S. Lovell, Orientation of organic, inorganic, and biomolecular chromophores in aromatic carboxylic acid crystals, *Ph.D. Dissertation* (Purdue University, 2000).
28. A. Barbon, M. Bellinazzi, J.B. Benedict, M. Brustolon, S.D. Fleming, S.-H. Jang, B. Kahr and A.L. Rohl, Luminescent probes of crystal growth: Surface charge and polar axis sense in dye-doped potassium hydrogen phthalate, submitted for publication.
29. C.A. Mitchell, S. Lovell, K. Thomas, P. Savickas and B. Kahr, Charge transfer interactions in dyed crystals of aromatic carboxylic acids and their relevance to MALDI mass spectrometry, *Angew. Chem., Int. Ed. Engl.* **35**(9), 1021–1023 (1996).
30. S. Hamai, Complex formation in cationic dye organic anion systems in aqueous solution, *Bull. Chem. Soc. Japan* **58**(7), 2099–2106 (1985).
31. J.C. Ma and D.A. Dougherty, The cation–π interaction, *Chem. Rev.* **97**(5), 1303–1324 (1997).
32. R.W. Gurney, C.A. Mitchell, S. Ham, L.D. Bastin and B. Kahr, Salting benzenes, *J. Phys. Chem. B* **104**(5), 878–892 (2000).
33. S. Ghosh, M. Petrin, A.H. Maki and L.R. Sousa, Dependence of the triplet state properties on the orientation of metal ion perturbers in naphthalene–crown ether metal ion complexes. I. External heavy atom effect, *J. Chem. Phys.* **87**(8), 4315–4323 (1987); A.H. Maki, Optically detected magnetic resonance of photoexcited triplet states, *Methods Enzymol.* **246**, 610–638 (1995).
34. S.K. Sarkar and G.S. Kastha, Environmental effects on the absorption and luminescence properties of aromatic amines: aniline, *Spectrochim. Acta, Part A* **48A**(11–12), 1611–1624 (1992).
35. G. Köhler, Solvent effects on the fluorescence properties of anilines, *J. Photochem.* **38**, 217–238 (1987).
36. H. Shimamori and A. Sato, Dipole moments and lifetimes of excited triplet states of aniline and its derivatives in nonpolar solvents, *J. Phys. Chem.* **98**(51), 13481–13485 (1994).
37. C.A. Mitchell, Salting open shell organic molecules, *Ph.D. Dissertation* (Purdue University, 1998).
38. M.P. Kelley, B. Janssens, B. Kahr and W.M. Vetter, Recognition of dyes by K_2SO_4 crystals: choosing organic guests for simple salts, *J. Am. Chem. Soc.* **116**(12), 5519–5520 (1994).
39. M. Rifani, Y.-Y. Yin, D.S. Elliott, M.J. Jay, S.-H. Jang, M.P. Kelley, L.D. Bastin and B. Kahr, Solid state dye lasers from stereospecific host–guest interactions, *J. Am. Chem. Soc.* **117**(28), 7572–7573 (1995).
40. S. Lovell, P. Subramony and B. Kahr, Poppy acid: synthesis and crystal chemistry, *J. Am. Chem. Soc.* **121**(30), 7020–7025(1999).
41. O. Krichevsky and G. Bonnet, Fluorescence correlation spectroscopy: the technique and its applications, *Rep. Prog. Phys.* **65**(2), 251–297 (2002).
42. M. Vaida, I. Weissbuch, M. Lahav and L. Leiserowitz, Mixed crystals as host–guest systems for probing molecular interactions, *Isr. J. Chem.* **32**(1), 15–21 (1992).
43. I. Weissbuch, R. Popovitz-Biro, M. Lahav and L. Leiserowitz, Understanding and control of nucleation, growth, habit, dissolution and structure of two- and three-dimensional crystals using 'tailor-made' auxiliaries, *Acta Crystallogr., Sect. B: Struct. Sci.* **B51**(2), 115–148 (1995).
44. J. Aizenberg, J. Hanson, T.F. Koetzle, L. Leiserowitz, S. Weiner and L. Addadi, Biologically induced reduction in symmetry: a study of crystal texture of calcitic sponge spicules, *Chem. Eur. J.* **1**(7), 414–422 (1995).
45. B. Kahr and J.M. McBride, Optically anomalous crystals, *Angew. Chem., Int. Ed. Engl.* **31**(1), 1–26 (1992).

46. W. Kaminsky and B. Kahr, Crystal optics and the symmetry principle: an update, in: *Symmetry 2000* (Wenner-Gren Foundation, Stockholm, 2001) pp. 309–316.
47. J.M. McBride and S.B. Bertman, Using crystal birefringence to study molecular recognition, *Angew. Chem., Int. Ed. Engl.* **28**(3), 330–333 (1989).
48. R.W. Munn, Calculation of the refractive indices in the optically anomalous crystal of l,5-dichloro-2,3-dinitrobenzene, *J. Chem. Phys.* **113**(19), 8774–8782 (2000).
49. P. Gopalan and B. Kahr, Reevaluating structures for mixed crystals of simple isomorphous salts: $Ba_xPb_{1-x}(NO_3)_2$, *J. Solid State Chem.* **107**(2), 563–567 (1993).
50. G. Crundwell, P. Gopalan, A. Bakulin, M.L. Peterson and B. Kahr, Effect of habit modification on optical and X-ray structures of sodium halate mixed crystals: the etiology of anomalous double refraction, *Acta Crystallogr., Sect. B: Struct. Sci.* **B53**(2), 189–202 (1997); P. Gopalan, M.L. Peterson, G. Crundwell and B. Kahr, Reevaluating structures for mixed crystals of simple isomorphous salts: sodium chlorate bromate $(NaCl_xBr_{1-x}O_3)$, *J. Am. Chem. Soc.* **115**(8), 3366–3367 (1993).
51. P. Gaubert, Recherches sur les cristaux de salmiac et de sel imbibès de matières étrangères pendant leur accroissement, *Bull. Soc. Min. Fr.* **38**, 149–182 (1915).
52. I.M. Melankholin and E.N. Slavnova, Uptake of capri blue by growing lead nitrate crystals, *Sov. Phys. Cryst.* **4**, 529–534 (1959).
53. W.G. France and P.P. Davis, Adsorption at crystal–solution interfaces. IX: The concentration of foreign substances in solution relative to the quantity adsorbed by the host crystal, *J. Phys. Chem.* **40**, 177–185 (1936).
54. C. Frondel, Effect of dyes on the crystal habit and optical of NaF, LiF, NaCl, KCl, KBr, and KI, *Am. Mineral.* **25**, 91–110(1940).
55. B.G. Khlopin and M.A. Tolstaya, Structure of pseudo-mixed crystals of the methylene blue barium nitrate type, *Zh. Fiz. Khim.* **14**, 941–952 (1940).
56. D.Y. Curtin and I.C. Paul, Chemical consequences of the polar axis in organic solid-state chemistry, *Chem. Rev.* **81**(6), 525–541 (1981).
57. I. Weissbuch, M. Lahav, L. Leiserowitz, G.R. Meredith and H. Vanherzeele, Centrosymmetric crystals as host matrices for second-order optical nonlinear effects, *Chem. Mater.* **1**(1), 114–118 (1989).
58. J. Hulliger, M. Kurimoto and B. Kahr, unpublished results.
59. O. König, H.-B. Bürgi, T. Armbruster, J. Hulliger and T. Weber, A study in crystal engineering: structure, crystal growth, and physical properties of a polar perhydrotriphenylene inclusion compound, *J. Am. Chem. Soc.* **119**(44), 10632–10640 (1997).
60. J. Hulliger, New physical methods for a space resolved mapping of the macroscopic polarization in molecular crystals and a stochastic theory for understanding, *Chimia* **55**(6), 554–561 (2001); P. Rechsteiner, J. Hulliger and M. Flörsheimer, Phase-sensitive second harmonic microscopy reveals bipolar twinning of Markov-type molecular crystals, *Chem. Mater.* **12**(11), 3296–3300 (2000).
61. O. Ermer, Unusual structural feature of crystalline phthalic acid, *Helv. Chim. Acta* **64**(6), 1902–1909 (1981).
62. M.A. Geday, W. Kaminsky, J.G. Lewis and A.M. Glazer, Images of absolute retardance $L\Delta n$, using the rotating polariser method, *J. Microsc. (Oxford)* **198**(1), 1–9 (2000); A.M. Glazer, J.G. Lewis and W. Kaminsky, An automatic optical imaging system for birefringent media, *Proc. Roy. Soc. London, Series A—Math. Phys. Eng. Sci.* **452**, 2751–2765 (1996).
63. V.M. Goldschmidt, Geochemical aspects of the origin of complex organic molecules on the earth, as precursors to organic life, *New. Biol.* **12**, 97–105 (1952).
64. J.D. Bernal, *The Physical Basis of Life* (Routledge and Paul, London, 1951).
65. A.G. Cairns-Smith, *Genetic Takeover and The Mineral Origins of Life* (Cambridge University Press, Cambridge, 1987).
66. L. Addadi, Z. Berkovitch-Yellin, I. Weissbuch, M. Lahav and L. Leiserowitz, A link between macroscopic phenomena and molecular chirality: crystals as probes for the direct assignment of the absolute configuration of chiral molecules, *Top. Stereochem.* **16**, 1–85 (1986).
67. S.C. Bondy and M.E. Harrington, L-amino acids and D-glucose bind stereospecifically to a colloidal clay, *Science* **23**(4386), 1243–1244 (1979); A.M. Cody and R.D. Cody, Chrial habit modifications of gypsum from epitaxial-like adsorption of sterespecific growth inhibitors, *J. Cryst. Growth* **113**(3–4), 508–519 (1991).
68. W.A. Bonner, P.R. Kavasmaneck, F.S. Martin and J.J. Flores, Asymmetric adsorption by quartz: a model for the prebiotic origin of optical activity, *Origins Life* **6**(3), 367–376 (1975); P.R. Kavasmaneck and W.A. Bonner, Adsorption of amino acid derivatives by d- and l-quartz, *J. Am. Chem. Soc.* **99**(1), 44–50 (1977).

69. P. Cintas, Chirality of living systems: a helping hand from crystals and oligopeptides, *Angew. Chem. Int. Ed. Engl.* **41**(7), 1139–1145(2002).
70. H. Blattner, B. Matthias and W. Merz, Characteristic dye enclosures in seignettoelectric crystals, *Helv. Phys. Acta* **19,** 415–417(1946).
71. B. Kahr, S. Lovell and J.A. Subramony, The progress of logwood extract, *Chirality* **10**(1/2), 66–77 (1998).
72. W. Kaminsky, E. Haussühl, B. Kahr and L.D. Bastin, Correlation of chiral KH_2PO_4 growth hillocks with absolute configuration of the faces, *J. Cryst. Growth* **234**(2–3), 523–528 (2002).
73. C.A. Orme, A. Noy, A. Wierzbicki, M.T. McBride, M. Grantham, H.H. Teng, P.M. Dove and J.J. De Yoreo, Formation of chiral morphologies through selective binding of amino acids to calcite surface steps, *Nature* **411**(6839), 775–779 (2001); L. Addadi and S. Weiner, Biomineralization: crystals, asymmetry and life, *Nature* **411**(6839), 753–755 (2001).
74. F. Arago, Memoire sur une modification remaquable qu'eprouvent les rayons lumineux dans leur passage a travers certains corps diaphenes, et sur quelques autres nouveaux phenomenes d'optique, *Mem. Inst. Fr.* **1**, 93–134 (1811).
75. J. Kobayashi, T. Asahi, S. Takahashi and A.M. Glazer, Evaluation of the systematic errors of polarimetric measurements: application to measurements of the gyration tensors of α-quartz by the HAUP, *J. Appl. Crystallogr.* **21**(5), 479–484 (1988).
76. G. Szivessy and C. Münster, Über die prüfung der gitteroptik bei aktiven kristallen, *Ann. Phys.* **20**, 703–736 (1934).
77. J.R.L. Moxon and A.R. Renshaw, The simultaneous measurement of optical-activity and circular-dichroism in birefringent linearly dichroic crystal sections. 1. Introduction and description of the method, *J. Phys.: Condens. Matter* **2**(32), 6807–6836 (1990).
78. J. Kobayashi and Y. Uesu, A new optical method and apparatus 'HAUP' for measuring simultaneously optical activity and birefringence of crystals. I. Principles and construction, *J. Appl. Crystallogr.* **16**(2), 204–211 (1983); J. Kobayashi, Y. Uesu and H. Takahara, A new optical method and apparatus 'HAUP' for measuring simultaneously optical activity and birefringence of crystals. II. Application to triglycine-sulfuric acid ($(NH_2CH_2CO_2H)_3 \cdot H_2SO_4$), *J.Appl. Crystallogr.* **16**(2), 212–219 (1983); W. Kaminsky and A.M. Glazer, Comparision of experimental optical properties of TGS with calculations using the DES model, *Phase Transitions* **66**(1–4), 1–21 (1998).
79. W. Kaminsky and A.M. Glazer, Measurement of optical rotation in crystals, *Ferroelectrics* **183**(1–4), 133–**141** (1996).
80. A.N. Winchell, *Optical Properties of Organic Compounds* (McCrone Research Institute, Chicago. 1989) p. 122.
81. W. Kaminsky, Experimental and phenomenological aspects of circular birefringence and related properties in transparent crystals, *Reports on Progress in Physics* **63**(10), 1575–1640 (2000).
82. R.K. Kondru, P. Wipf and D.N. Beratan, Atomic contributions to the optical rotation angle as a quantitative probe of molecular chirality. *Science* **282**(5397), 2247–2250 (1998).
83. P.L. Polavarapu and D.K. Chakraborty, *Ab initio* theoretical optical rotations of small molecules, *Chem. Phys.* **240**(1–2), 1–8 (1999); P.L. Polavarapu and D.K. Chakraborty, Absolute stereochemistry of chiral molecules from *ab initio* theoretical and experimental molecular optical rotations, *J. Am. Chem. Soc.* **120**(24), 6160–6164 (1998).
84. V. Devarajan and A.M. Glazer, Theory and computation of optical rotatory power in inorganic crystals, *Acta Crystallogr., Sect. A: Found. Crystallogr.* **A42**(6), 560–569 (1986).
85. A.N. Winchell and H. Winchell, *The Microscopical Charcters of Artifical Inorganic Solid Substances: Optical Properties of Artificial Minerals* (McCrone Research Institute, Chicago, 1989) p. 125.
86. *Colour Index,* 3rd edn (Society of Dyers and Colourists, Bradford, 1984).
87. K. Ojima, Y. Nishihata and A. Sawada, Structure of potassium-sulfate at temperatures from 296K down to 15K, *Acta Crystallogr., Sect. B: Struct. Sci.* **B51**(3), 287–293 (1995).
88. W. Kaminsky, Reinvestigation of optical activity in the course of the ferroelastic phase transition in cadmium-langbeinite, $K_2Cd_2(SO_4)_3$, *Phase Transitions* **59**(1–3), 121–133 (1996); W. Kaminsky. Topographies of chiral and associated optical properties in $FeBO_3$ using a novel polarimeter, the tilter, *Ferroelectrics* **204**(1–4), 233–246 (1997).
89. W. Kaminsky, M. Geday, J. Herreros-Cedres and B. Kahr, Optical rotatory and circular dichroic scattering, **107**(16), 2800–2807 (2003).

90. J. Schellman and H.P. Jensen, Optical spectroscopy of oriented molecules, *Chem, Rev.* **87**(6), 1359–1399 (1987).
91. M.J.B. Tunis-Schneider and M.F. Maestre, Circular dichroism spectra of oriented and unoriented deoxyribonucleic acid films—a preliminary study, *J. Mol. Biol.* **52**(3), 521–541 (1970).
92. G.A. Olah and H.W. Huang, Circular dichroism of oriented α helices. I. Proof of the exciton theory, *J. Chem. Phys.* **89**(4), 2531–2538 (1988); Y. Wu, H.W. Huang and G.A. Olah, Method of oriented circular dichroism, *Biophys. J.* **57**(4), 797–806 (1990).
93. H.-G. Kuball and T. Höfer, in: *Circular Dichroism: Principles and Applications,* Eds K. Nakanishi, N. Berova and R.W. Woody (Wiley-VCH, New York, 2000) pp. 133–158.
94. R.L. Disch and D.I. Sverdlik, Apparent circular dichroism of oriented systems, *Anal. Chem.* **41**(1), 82–86 (1969).
95. Å. Davidsson, B. Nordén and S. Seth, Measurement of oriented circular dichroism, *Chem. Phys. Lett.* **70**(2), 313–316(1980).
96. Å. Davidsson and B. Nordén, On the problem of obtaining accurate circular dichroism. Calibration of circular dichroism spectrometers, *Spectrochim. Acta, Pt. A* **32A**(4), 717–721 (1976).
97. H.P. Jensen, J.A. Schellman and T. Troxell, Modulation techniques in polarization spectroscopy, *Appl. Spec.* **32**, 192–200 (1978).
98. Y. Shindo and Y. Ohmi, Problems of CD spectrometers. 3. Critical comments on liquid crystal induced circular dichroism, *J. Am. Chem. Soc.* **107**(1), 91–97 (1985).
99. A. Schönhofer, H.-G. Kuball and C. Puebla, Optical activity of oriented molecules. IX. Phenomenological Mueller matrix description of thick samples and of optical elements, *Chem. Phys.* **76**(3), 453–467 (1983).
100. M.F. Maestre and A. Katz, A circular dichroism microspectrophotometer, *Biopolymers* **21**(9), 1899–1908 (1982).
101. M.F. Maestre, G.C. Salzman, R.A. Tobey and C. Bustamante, Circular dichroism studies on single Chinese hamster cells, *Biochemistry* **24**(19), 5152–5157 (1985); F. Livolant, W. Mickols and M.F. Maestre, Differential polarization microscopy (CD and Linear dichroism) of polytene chromosomes and nucleoli from the dipteran Sarcophaga footpad, *Biopolymers* **27**(11), 1761–1769 (1988); F. Livolant and M.F. Maestre, Circular dichroism microscopy of compact forms of DNA and chromatin *in vivo* and *in vitro:* cholesteric liquid-crystalline phases of DNA and single dinoflagellate nuclei, *Biochemistry* **27**(8), 3056–3068 (1988).
102. D. Keller, C. Bustamante, M.F. Maestre and I. Tinoco, Jr., Imaging of optically active biological structures by use of circularly polarized light, *Proc. Natl. Acad. Sci. USA* **82**(2), 401–405 (1985).
103. K. Claborn, E. Puklin-Faucher, M. Kurimoto, W. Kamirsky and B. Kahr, Circular dichroism imaging microscopy: Application to enantiomorphous twinning in 1,8-dihydroxyanthrquinone, *J. Am. Chem. Soc.* **125**(48), 14825–14831 (2003).

This page intentionally left blank

5

From solid–fluid interfacial structure to nucleation kinetics: principles and strategies for micro/nanostructure engineering

Xiang Yang Liu*
Department of Physics, Faculty of Sci., National University of Singapore, 2 Science Drive 3, Singapore 117542

5.1. INTRODUCTION

Crystallization, in particular nucleation, plays an essential role in the preparation of essential materials [1–4], the structural characterization of natural and synthetic molecules [5,6], and the development of future technologies [7,8]. Up to now, crystallization has in many ways continued to be more like an art than a science [5], mainly because there is not sufficient knowledge on its critical early stages, except for some local events of crystallization/quasi-crystallization of large species, namely proteins [8] and colloidal particles [9]. From the point of view of nano sciences and technologies, it is of critical importance to develop a quantitative understanding of crystallization systems at their early stages. The reason is that many nano materials are crystalline phases and the essential structures and utmost important properties of the systems are determined by nucleation and the correlation between the nucleating nano phase and the substrate, etc.

From the technological point of view, the knowledge of nucleation and of the effect of the substrate is also practically very important in identifying robust technologies in electronic, photonic and life sciences and technologies. For instance, in the case of growing self-assembled quantum dots systems on certain substrates (such as IV (Ge/Si), lI/VI (CdSe/ZnSe) and III-V (InAs/GaAs)) [10], the nucleation of semiconductors on the substrates plays a key role in controlling the size and the self-organized nano structure. In the human body, a process called bone turnover or remodeling, continues throughout adult life, which involves resorbing primary osteons, and replacing them by succeeding generations

*E-mail address: phyliuxy@nus.edu.sg

of osteons [11]. This process will affect the essential micro/nano structure of human bones. As one of the key steps in the generation of osteons, biomineralization/remineralization occurs at and around the substrates containing loose collagen fibers [12,13]. This involves the initial heterogeneous nucleation of Ca minerals at the substrates (collagen fibers) and the further mineralization achieved by crystal growth [12–14]. The nature of nucleation will also determine the orientation of Ca minerals and the structural synergy between substrates and biominerals, consequently the basic structure and the mechanical properties of bones. If nucleation is more heterogeneous in nature, the coherence of biomimerals to biosubstrates will be tough, and the essential bone structure is expected to be compact and strong. Conversely, if nucleation is more homogeneous in nature, the effect of the substrates on the nucleation of biominerals will be negligible since substrates play no role in this case. Nucleation will then occur randomly in the bulk fluid phase. This may significantly reduce the correlation between the substrates and biominerals, and give rise to a loose and porous structure.

In many cases of crystallization, we are dealing with crystal networks rather than individual crystals (for instance chemicals or foods processing). Since crystal networks possess certain self-organized structures, many physical properties of such network systems will be very much different from individual crystals, and determined to a large extent by the network structures. For instance, as functional materials of an important class, the supramolecular materials [15] having 3D fibrous network structures formed by interconnecting nano-sized fibers are found to be important in drug delivery, coatings, lithography, catalyst supporters, or as scaffolds for tissue engineering [15–20]. Such materials are also employed in the engineering of nanostructural materials and self-supporting porous materials in the novel separation for macromolecules [15–21]. Macroscopic properties, in particular the rheological properties of this type of materials, are determined solely by the micro/nanostructure of fiber networks. Fibrous networks with permanent interconnection will effectively entrap and immobilize liquid in the meshes, consequently possess both the elastic properties of ideal solids and the viscosity properties of Newtonian liquids. This leads to the formation of self-supporting supramolecular materials [16,18,21]. In contrast, those consisting of non-permanent or transient interconnecting (or entangled) fibers or needles can only form weak and viscous paste at low concentrations [16].

It is believed [16, 18] that the formation of interconnecting fiber networks, which leads to the formation of supramolecular materials, takes place via the self assembly of fibers. Nevertheless, the latest progress (cf. Section 5.4) indicates that the 3D self-organized micro/nanostructure of supramolecular functional materials is controlled by the so-called crystallographic mismatch branching. This is essentially a special case of heterogeneous nucleation. This result indicates that even for areas such as the formation of supramolecular functional materials where conventionally crystallization was regarded unimportant, the knowledge of nucleation also plays a very crucial role.

It is the purpose of this chapter to introduce the most recent developments in the kinetics of nucleation under the influence of substrates and additives. Based on the acquired knowledge, the principles and strategies for the engineering of micro/nano structures of various systems, in particular supramolecular functional materials, will be discussed.

The chapter will be arranged as follows: in Section 5.2, our attention will be first devoted to the thermodynamics and the kinetics of nucleation. Then, the effects of foreign bodies and additives on nucleation will be treated in detail. In Section 5.3 we will focus

on the interfacial processes of nucleation, in particular, some newly identified interfacial effects for nano structure engineering, which include non-classical epitaxial interfacial effects. In Section 5.4, we will discuss the mechanisms of supersaturation and impurity driven interfacial structure mismatch and the crystallographic structural mismatch branching. The principles and the strategies of micro/nano engineering of functional materials will be introduced based on these mechanisms. Finally, some general remarks and conclusions will be given in Section 5.5.

5.2. NUCLEATION UNDER THE INFLUENCE OF SUBSTRATE

The formation of a new phase in the body of the ambient phase, such as gas or liquid, is one of the most fundamental aspects of phase transitions, in particular, crystallization. The potential barrier which a system must overcome in order to create a (crystalline) nucleus in the ambient phase, and which determines the rate of nucleation, is defined by the interfacial energy. If under a certain condition the probability of creating a nucleus is uniform throughout the system, nucleation is defined as homogeneous nucleation. Otherwise, it is defined as heterogeneous nucleation. Heterogeneous nucleation normally occurs on solid or liquid surfaces, microclusters, dusts, macromolecules or other foreign bodies. Since these foreign bodies occur inevitably in most systems, nucleation has in most cases a heterogeneous nature rather than a homogeneous nature [5]. During heterogeneous nucleation, the properties of these foreign bodies form an additional factor upon which the nucleation barrier and nucleation rate depend. In this section, heterogeneous nucleation will be discussed from both the macroscopic and microscopic points of view. Within this 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 panicles, forming dimers, trimers, tetramers, etc. The kinetics of nucleation is described by the nucleation rate *J*. The nucleation rate *J* is defined as the number of nuclei created per unit volume–time, and it is determined by the nucleation barrier, the kink integration rate, the transport of growth units and other factors. Therefore, the central problem in nucleation theory and experiment is to find 7 as a function of the parameters controlling the process. After the thermodynamic results of Gibbs [22], the paper by Volmer and Weber [23] in 1926 was the very first work devoted to *J* and it was followed by the pioneering studies of Farkas [24], Kaischew, Stranski and others [25–27]. Up to now, many theories have been published to describe the kinetics of nucleation [28–48]. We will look at the latest progress in modern heterogeneous nucleation theories, in particular the impact on nanostructure engineering and the architecture of functional materials.

5.2.1. Thermodynamic driving force

Nucleation is a process commonly shared during the beginning stages of first-order phase transitions. The thermodynamic driving force for the nucleation of a new phase (e.g., a crystal) is $\Delta\mu$, which is defined as the difference between the chemical potentials μ_i^{ambient} and μ_{crystal} of a growth unit in the ambient mother and in the crystalline phase:

$$\Delta\mu = \mu_i^{\text{ambient}} - \mu_{\text{crystal}}. \tag{1}$$

(Subscript i denotes the solute in the ambient phase.) When $\Delta\mu > 0$, the system is said to be 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 a condition, crystals will dissolve. When $\Delta\mu = 0$, the ambient phase is in equilibrium with the crystalline phase. This implies that for temperature T and pressure P, one has

$$\left(\mu_i^{\text{ambient}}\right)_{\text{eq}} = \mu_{\text{crystal}}, \tag{2}$$

where $(\mu_i^{\text{ambient}})_{\text{eq}}$ is the chemical potential of a solute molecule in a state of phase equilibrium (or coexistence) between the mother and the crystalline phases. It follows from eq. (2) that μ_{crystal} can be expressed as $(\mu_i^{\text{ambient}})_{\text{eq}}$ for given T and P. Therefore $\Delta\mu$ as given in eq. (1) can be rewritten as

$$\Delta\mu = \mu_i^{\text{ambient}} - \left(\mu_i^{\text{ambient}}\right)_{\text{eq}}. \tag{3}$$

For the condensation of vapor, the chemical potential of species i is given by [47,49]

$$\mu_i = \mu_i^0 + kT \ln p_i, \tag{4}$$

where p_i is the actual partial vapor pressure of species i, μ_i^0 denotes the chemical potential of the standard state ($p_i = 1$) of species i, k is the Boltzmann constant, and T is the absolute temperature. Based on eqs (2) and (4), the dimensionless thermodynamic driving force is given by

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

(p_i^{eq} is the partial pressure of vapor i in the equilibrium state). Based on the same principles, the thermodynamic driving force for the deposition of thin films from vapor is given by [47, 49]

$$\frac{\Delta\mu}{kT} = \ln(R/R_{\text{e}}) \tag{6}$$

(R and $R_{\text{e}}(T)$ are, respectively, the actual rate of impingement of molecules to the substrate, and at equilibrium, respectively).

For crystallization from solution, the chemical potential of species i is given by [47,49]

$$\mu_i = \mu_i^0 + kT \ln a_i \approx \mu_i^0 + kT \ln C_i, \tag{7}$$

where a_i, and C_i denote the activity and concentration of species i, respectively, and μ_i^0 denotes the standard state ($a_i = 1$) of the chemical potential of species i. This then gives rise to the thermodynamic driving force

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

(a_i^{eq}, C_i^{eq} are, respectively, the equilibrium activity and concentration of species i).

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

$$\sigma = (a_i - a_i^{eq})/a_i^{eq} \approx (C_i - C_i^{eq})/C_i^{eq}. \tag{9}$$

Equation (8) can then be rewritten as

$$\frac{\Delta\mu}{kT} = \ln(1+\sigma). \tag{10}$$

In the case of $\sigma < 1$, eq. (10) can be approximated, after applying a Taylor series expansion, by

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

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

$$mA^{n+} + nB^{m-} \rightarrow A_mB_n \tag{12}$$

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

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

where K_{sp} ($= (a_{A^{n+}}^{eq})^m(a_{B^{m-}}^{eq})^n$) is the precipitation constant. Based on eqs (10) and (13), the supersaturation can be expressed in this case by

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

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

For crystallization from the melt at temperatures not far below the melting or equilibrium temperature, the thermodynamic driving force can be obtained by applying the van't Hoff equation and eq. (3) [47,49], as

$$\frac{\Delta\mu}{kT} = \frac{\Delta h_m \Delta T}{kTT_e}, \tag{15}$$

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

(Δh_m is the enthalpy of melting per molecule, T_e is the equilibrium temperature, ΔT is the supercooling).

For electro-deposition, the thermodynamic driving force is given by [49]

$$\frac{\Delta\mu}{kT} = \frac{z_i e}{kT}\left(\phi^{e} - \phi^{e}_{eq}\right) \tag{17}$$

(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).

5.2.2. Nucleation barrier and impact of the substrate

A characteristic feature of any nucleation process is that the substance having the properties of the new phase is fluctuating and 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 remaining in equilibrium with the surrounding ambient phase are the critical nuclei. The smaller or larger clusters are subnuclei or supernuclei, respectively. Only supernuclei can grow spontaneously to reach macroscopic sizes. For simplicity, we call hereafter subnuclei "clusters", and supernuclei "nuclei".

The nucleation rate J is determined by the height of the free energy barrier, the so-called nucleation barrier. The occurrence of a nucleation barrier is attributed to the following two contradictory effects:

(1) since the crystalline phase is a stable phase, the occurrence of the new phase from the ambient phase will result in a lowering of the (Gibbs) free energy of the system;
(2) due to the interfacial (or surface) free energy, an increase in the size of the crystalline new phase leads to an increase of the interface (or surface) area, and consequently this enhances the interface (or surface) free energy. This will cause an increase in the Gibbs free energy of the system.

The combination of these two effects results in the formation of the nucleation barrier. In the following discussion, a general description for the nucleation barrier will be given.

The free energy change due to the formation of a cluster of $n = 1, 2, 3\ldots$ molecules can be found from 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} denotes the Gibbs free energies of the system in the initial and final states before and after cluster formation, respectively). If M is the number of solute molecules in the system, G_{ini} is simply given by

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

To find G_{fin} one usually employs the Gibbs method [22] of introducing a surface which divides the system into a new phase (or a cluster) of n molecules, and an old phase of the remaining $M - n$ solute molecules. Then G_{fin} is written as

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

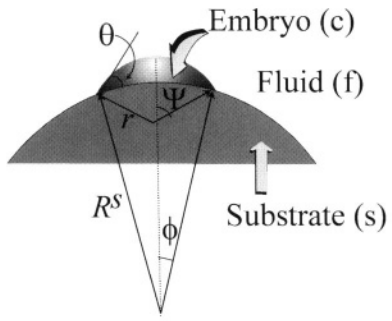


FIGURE 5.1. Illustration of nucleation on a foreign particle. θ is the virtual contact angle between nucleating phase and substrate.

where Φ_n is the total surface energy of the n-sized cluster (except for the 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

$$\Delta G = -n\Delta\mu + \Phi_n. \tag{21}$$

The function ΔG reaches its maximum ΔG^* at $r = r_c$, or $n = n^*$ (r and r_c are the radius and the critical radius of the cluster, respectively). A cluster of n^* molecules is a critical nucleus, r_c is the radius of curvature of that critical nucleus, and ΔG^* is the nucleation barrier. One of the major problems in nucleation theory is to find ΔG^* which, physically, is the energy barrier of nucleation.

The occurrence of a foreign body in the system normally reduces the interfacial (or surface) free energy; therefore it will also lower the nucleation barrier. Let ΔG^*_{homo} be the homogeneous nucleation barrier (the nucleation barrier in the absence of any influence from a foreign body), and let $\Delta G^*_{\text{heter}}$ be the heterogeneous nucleation barrier (the nucleation barrier in the presence of an influence from foreign body). In the following discussion, we will focus on the nucleation barrier under the influence of a foreign body (or $\Delta G^*_{\text{heter}}$). Let us define an interfacial correlation factor f describing the lowering of the nucleation barrier due to the action of a foreign body:

$$f = \Delta G^*_{\text{heter}}/\Delta G^*_{\text{homo}}. \tag{22}$$

It will be shown in the following discussion how ΔG^*_{homo} and f can be derived [46,47].

As shown in Figure 5.1, we assume that nucleation occurs on a foreign body with a radius of R^{s}. In the following derivations, the ambient phase is represented by subscript f, the crystalline phase by c and the foreign body by s. If we denote the volume by V and the surface area of the foreign body by S, then the free energy of forming a cluster of radius r on a foreign particle of radius R^{s} is given, according to eq. (21), by

$$\Delta G = -\Delta\mu V_{\text{c}}/\Omega + \gamma_{\text{cf}}S_{\text{cf}} + (\gamma_{\text{sf}} - \gamma_{\text{sc}})S_{\text{sc}}, \tag{23}$$

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 a contact angle θ (i.e. Figure 5.1) can still be applied in this case. We have then

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

Referring again to Figure 5.1, we have,

$$S_{sc} = 2\pi\left(R^s\right)^2(1 - \cos\phi), \qquad S_{cf} = 2\pi r^2(1 - \cos\Psi), \tag{25}$$

and

$$V_c = \frac{1}{3}\pi r^3\left(2 - 3\cos\Psi + \cos^3\Psi\right) - \frac{1}{3}\pi\left(R^s\right)^3\left(2 - 3\cos\phi + \cos^3\phi\right), \tag{26}$$

with

$$\cos\phi = \left(R^s - r\cos\theta\right)/l = \left(R^s - rm\right)/l, \tag{27}$$

$$\cos\Psi = -\left(r - R^s\cos\theta\right)/l = -\left(r - R^s m\right)/l, \tag{28}$$

and

$$l = \left[\left(R^s\right)^2 + r^2 - 2R^s rm\right]^{1/2}. \tag{29}$$

To evaluate the critical free energy ΔG^*_{heter}, we can substitute expression (26) into (23) and require that

$$(\partial\Delta G/\partial r) = 0. \tag{30}$$

Considering that a critical nucleus is a stable nucleus with a maximum curvature for a given thermodynamic condition, it follows from thermodynamic principles [47,49–51] that such a maximum curvature (or the minimum radius—the critical radius r_c) is only determined by γ_{cf} and the driving force $\Delta\mu$. In other words, given a certain experimental condition, the size of the critical nucleus is the same for homogeneous and heterogeneous nucleation. This implies that we adopt r_c for homogeneous nucleation and for heterogeneous nucleation under the same condition [47,49–51] which gives rise to

$$r_c = 2\Omega\gamma_{cf}/\Delta\mu. \tag{31}$$

This is also the thermodynamic principle behind the Gibbs–Thomson effect [51].

Note that in the case of epitaxial growth, a strain will develop due to the structural mismatch at the crystal–substrate interface. That strain will affect both the bulk free energy of the nuclei and the interfacial free energy γ_{cf}. In this case, the occurrence of a substrate exerts a direct impact on r_c. Nevertheless, in many cases, such as the nucleation of biominerals on soft biological substrates, the strain effect is very low and may be ignored. Therefore, the effect of the strain on r_c at the interface is neglected in the above treatment.

Now substituting expressions (24)–(31) into eq. (23) and writing

$$R' = R^s/r_c = R^s\Delta\mu/\Omega\gamma_{cf} = R^s kT\ln(1+\sigma)/\Omega\gamma_{cf}, \tag{32}$$

the free energy of formation of a critical nucleus is, according to eq. (22), given by

$$\Delta G^*_{heter} = \Delta G^*_{homo} f \tag{33}$$

with

$$\Delta G^*_{homo} = \frac{16\pi\gamma_{cf}^3\Omega^2}{3[kT\ln(1+\sigma)]}, \tag{34}$$

$$f(m, R') = \frac{1}{2} + \frac{1}{2}\left(\frac{1-mR'}{w}\right)^2 + \frac{1}{2}R'^3\left[2 - 3\left(\frac{R'-m}{w}\right) + \left(\frac{R'-m}{w}\right)^2\right] + \frac{3}{2}mR^2\left(\frac{R'-m}{w} - 1\right) \tag{35}$$

and

$$w = \left[1 + (R')^2 - 2R'm\right]^{1/2}. \tag{36}$$

Here R' is actually the dimensionless radius of curvature of the substrate in reference to the radius of the critical nucleus r_c.

Substituting appropriate values of R^s, m, γ_{cf} and $\Delta\mu$ into eqs (31)–(36), 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 the determination of the heterogeneous nucleation barrier ΔG^*_{heter}. One can see from eq. (22) that the influence of foreign particles on the nucleation barrier can be fully characterized by this factor.

Figure 5.2a shows $f(m, R')$ as a function of R' for a given m. When $R' \to 0$, $f(m, R') = 1$, implying that the foreign body "vanishes" completely as a nucleating substrate. In practice, if foreign bodies are too small, e.g. 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 $R' \gg 1$, the foreign body can be treated as a flat substrate with respect to the critical nuclei. In this case, $f(m, R') = f(m)$ is solely a function of m, and the curvature of the foreign body has no effect on the nucleation kinetics. Equation (35) is then reduced to

$$f(m, R') = f(m) = \frac{1}{4}(2 - 3m + m^3). \tag{37}$$

In many cases, one needs to express ΔG^* as a function of n^*. Geometrically, in many cases of heterogeneous nucleation [28,33],

$$f \cong V^*/V^*_{homo} = n^*/n^*_{homo}, \tag{38}$$

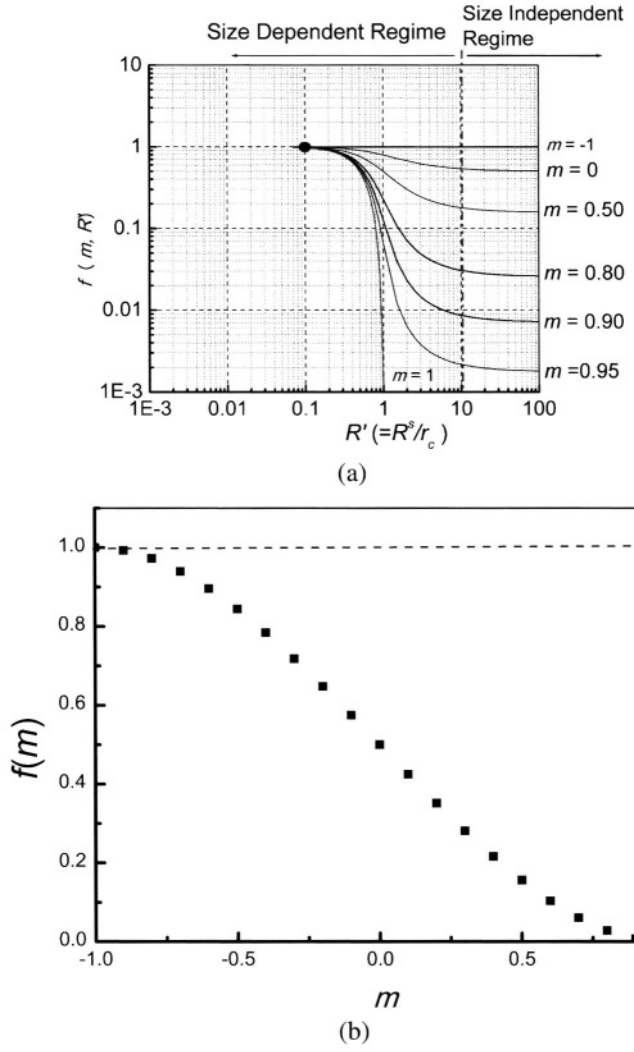


FIGURE 5.2. (a) Dependence of the interfacial correlation function $f(m, R')$ on m and $R' \ll 10$. (b) Dependence of the interfacial correlation function $f(m, R')$ on m at $R' \gg 10$.

where V^*, V^*_{homo}, n^* and n^*_{homo}, are volumes and the numbers of heterogeneously and homogeneously formed nuclei at the given supersaturation. It follows from eqs (22), (35) and (38) that

$$n^* = n^*_{\text{homo}} f_{n^*}\left(m, R^s, n^*_{\text{homo}}\right), \quad n^*_{\text{homo}} = \frac{32}{3(\Delta\mu)^3}\Omega^2\gamma^3_{\text{cf}}, \tag{39}$$

$$\Delta G^*_{\text{heter}}(\Delta\mu) = \frac{16}{3(\Delta\mu)^2}\Omega^2\gamma^3_{\text{cf}} f\left(m, R^s, n^*_{\text{homo}}\right) \tag{40}$$

with

$$f_{n^*}(m, R', n^*_{\text{homo}}) = \frac{1}{2} + \frac{1}{2}\left(\frac{1 - mR^s}{w_{n^*}(3n^*_{\text{homo}}\Omega/4\pi)^{1/2}}\right)^2 + \frac{1}{2}\left(\frac{R^s}{(3n^*_{\text{homo}}\Omega/4\pi)^{1/2}}\right)^3 \times \left[2 - 3\left(\frac{R^s/(3n^*_{\text{homo}}\Omega/4\pi)^{1/2} - m}{w_{n^*}}\right) + \left(\frac{R^s/(3n^*_{\text{homo}}\Omega/4\pi)^{1/2} - m}{w_{n^*}}\right)^2\right] + \frac{3}{2}m\left(\frac{R^s}{(3n^*_{\text{homo}}\Omega/4\pi)^{1/2}}\right)^2\left(\frac{R^s/(3n^*_{\text{homo}}\Omega/4\pi)^{1/2} - m}{w_{n^*}} - 1\right) \tag{41}$$

and

$$w_{n^*} = \left\{1 + \left[R^s/\left(3n^*_{\text{homo}}\Omega/4\pi\right)^{1/2}\right]^2 - 2R^s m/\left(3n^*_{\text{homo}}\Omega/4\pi\right)^{1/2}\right\}^{1/2}. \tag{42}$$

n^*_{homo} and r_c given in eqs (31) and (41) are often referred to as the Gibbs–Thomson equations for the size of the critical nucleus.

5.2.3. *Kinetics of nucleation: the influence of the substrate*

5.2.3.1. Nucleation kinetics Let us recapture the picture of heterogeneous nucleation on a substrate. A variety of molecular processes take place on the substrate surface, 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 (cf. Figure 5.3). The change in the free energy associated with the transition when n molecules adsorb to form an n-mer of size 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. (31).

The widely accepted kinetic model of nucleation (within the cluster approach) was first used by Farkas [24] in 1927. It is based on the Szilard scheme of successive "chain reactions" between monomer molecules and n-sized clusters:

$$\text{Monomer} \rightleftharpoons \text{Dimer} \ldots \rightleftharpoons (n-1)\text{-mer} \rightleftharpoons n\text{-mer} \rightleftharpoons (n+1)\text{-mer} \ldots \tag{43}$$

A master equation for the concentration $Z_n(t)$ of the n-sized clusters at time t is written in the form of a continuity equation [33,38]

$$dZ_n/dt = J_{n-1} - J_n, \tag{44}$$

where J_n is the flux through point n on the size axis. Thus in this formulation J is the flux through the nucleus size n^*, i.e., $J \sim J_{n^*}(t)$, so that so that a time dependence is introduced in the nucleation rate.

The basic problem in nucleation kinetics is to solve the master eqs (43) and (44) for the unknown cluster size distribution $Z_n(t)$, since knowledge of $Z_n(t)$ enables the determination of the nucleation rate. There exist three physically distinct states of the system which

are of particular interest: (i) the equilibrium, (ii) the stationary (or steady) state, and (iii) the nonstationary state. Since the physically realistic states are the stationary (or steady) state and the nonstationary state, in the sequel we will focus on the steady state, and briefly discuss the nonstationary state.

In the stationary state, $dZ_n/dt = 0$. Because $J_n = \text{constant} = J_{n^*} = J$, $Z_n \sim Z'_n$ is the steady-state cluster size distribution. The stationary nucleation rate for homogeneous nucleation is given by the Becker–Doering formula [26]

$$J = \frac{zK^*}{\nu_m} \exp\left(-\frac{\Delta G^*_{\text{homo}}}{kT}\right) \tag{45}$$

with

$$z = Z'_{n^*}/C_{n^*} - Z'_{n^*+1}/C_{n^*+1}, \tag{46}$$

where z is the so-called Zeldovich factor [29,34], $K^* = K_{n^*}$ is the frequency of monomer attachment to the critical nucleus, ν_m denotes the average volume of structural units in the ambient phase, and C_n is the equilibrium concentration of n-sized clusters given by [46,47]

$$C_n \cong C_1 \exp(-\Delta G_n/kT). \tag{47}$$

Based on the definition of J_n, one has

$$J_n = k_n Z'_n. \tag{48}$$

At the steady state, J', that is the formation rate of critical nuclei per unit volume–time around a foreign particle, is equal to the steady state growth of clusters on the surface of the particle. The nucleation rate can then be expressed in terms of

$$\begin{aligned} J' &= J_{n^*} = J_n = \text{constant} = \text{critical sized nuclei formed/unit volume–time} \\ &\equiv \cdots = k_n Z'_{n+1} - k_n Z'_{n+1} = \text{constant}. \end{aligned} \tag{49}$$

Before further analyzing eq. (49), we introduce the following two boundary conditions:

$$(1) \quad Z'_{n^*} = 0; \tag{50}$$

$$(2) \quad \lim_{n \to 1} (Z'_n/C_n) = 1. \tag{51}$$

Taking into account the effect of the substrate on both the nucleation barrier and the transport process, and the fact that the average nucleation rate in the fluid phase depends on the density and size of the foreign particles present in the system, the nucleation rate is given by [46,47]

$$J = 4\pi a \left(R^s\right)^2 N^0 f''(m, R') \left[f(m, R')\right]^{1/2} \times B \exp\left[-\frac{16\pi \gamma_{cf}^3 \Omega^2}{3kT[kT \ln(1+\sigma)]^2} f(m, R')\right] \tag{52}$$

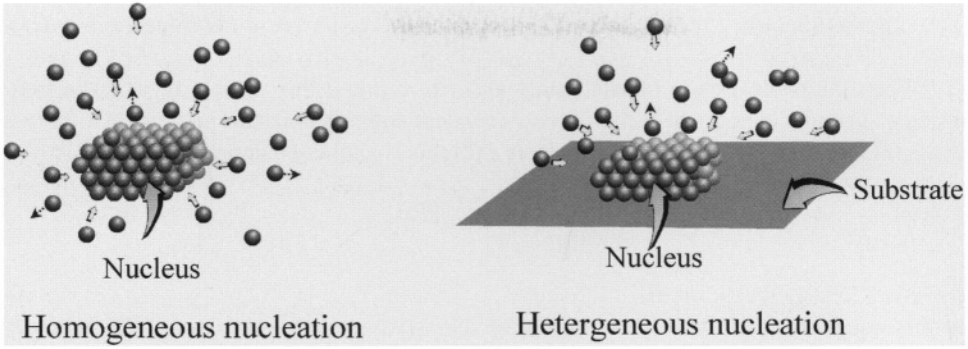


FIGURE 5.3. Schematic illustration of *the shadow effect* of the substrate in heterogonous nucleation. The presence of the substrate blocks the collision of growth units onto the surface of the nucleus.

with

$$B = (C_1)^2 4 D \beta_{\text{kink}} \Omega \left(\frac{\gamma_{\text{cf}}}{kT} \right)^{1/2}, \tag{53}$$

$$f''(m, R') = \frac{1 + (1 - R'm)/w}{2} \tag{54}$$

and

$$f''(m, R') = f''(m) = \frac{1}{2}(1 - m) \quad \text{at } R' \gg 1, \tag{55}$$

where B is the kinetic constant and N^0 denotes the number density of the substrates (or "seeds"). The growth of nuclei is subject to the effective collision and incorporation of growth units into the surfaces of the nuclei (cf. Figure 5.3). In the case of homogeneous nucleation, the growth units can be incorporated into the nuclei from all directions. However, in the case of heterogeneous nucleation, the presence of the substrate will block the collision path of the growth units to the surfaces of these nuclei from the side of the substrate (cf. Figure 5.3). This is comparable to the "shadow" of the substrate cast on the surface of the nuclei. $f''(m, R')$ in the pre-exponential factor, which is the ratio between the average effective collision in the presence of substrates and that of homogeneous nucleation (i.e., in the absence of a substrate), describes this effect.

Both $f(m, R')$ and $f''(m, R')$ are functions of m and R'. When $R' \to 0$ or $m \to -1$, $f(m, R'),\ f''(m, R') = 1$. This is equivalent to the case of homogeneous nucleation. In the case where $m \to 1$ and $R \gg 1$, one has $f(m, R'),\ f''(m, R') = 0$. Normally, heterogeneous nucleation occurs in the range between 1 and – 1, or $f(m, R')$ between 0 and 1, depending on the nature of the substrate surface and the supersaturation.

Note that for homogeneous nucleation, one has $f''(m, R') = f(m, R') = 1$, and $4\pi a (R^s)^2 N^0 \to 1$. In this case, eq. (52) is converted to

$$J = B \exp\left[-\frac{16\pi \gamma_{\text{cf}}^3 \Omega^2}{3kT[kT \ln(1 + \sigma)]^2} \right]. \tag{56}$$

This implies that eq. (52) is applicable to both homogenous and heterogeneous nucleation.

When the nucleation is nonstationary, $dZ_n/dt \neq 0$ and flux J_n is a function of both n and t. The nucleation rate is then time-dependent and this nonstationary nucleation rate $J_{\text{nonst}}(t) = J_{n^*}(t)$. According to Kashchiev [52] this is expressed approximately as

$$J_{\text{nonst}}(t) = J\left[1 + 2\sum_{i=1}^{\infty}(-1)^i \exp\left(k^2 t/\tau\right)\right]. \tag{57}$$

In this equation, J is the nucleation rate of the steady state given by eq. (52), whereas $J_{\text{nonst}}(t)$ at $t = 0$ and gradually increases to the stationary nucleation rate J. It turns out [52] that $J_{\text{nonst}}(t) \sim J$ for $t > 5\tau$, where τ is the nucleation time-lag given by

$$\tau = 4/\pi^3 z^2 k_{n^*}. \tag{58}$$

Physically, τ is a measure of the time needed for the transformation of the initial cluster size distribution into the steady-state one [53–55]. Since $z = 0.1$ [57], τ is mainly determined by k_{n^*}, which is a function of diffusivity D. With typical values of D, the time lag τ varies from microseconds for nucleation in vapor and less viscous solutions to days for nucleation in solids and glass-forming melts [56].

5.2.3.2. Induction time in nucleation One of the most common ways to describe the kinetics of nucleation is to measure the induction time t_s of nucleation at different supersaturations. What we normally acquire from experiments is the induction time t_i of crystallization, which is defined as the mean time lapse before an observable amount of the new phase is detected. As mentioned before, crystallization can be regarded as consisting of the two key steps: *nucleation* and *growth*. Therefore, t_i includes the time t_g for the growth of crystals to the observable size, and t_{nucl}. As mentioned in Section 5.2.3.1, a certain amount of time is required to establish nucleation from time zero to the steady state. This is the transient period t_{nonst} which is associated with the nucleation of the non-stationary state. Denoting the induction time for the nucleation of the steady state by t_s one has then

$$t_i = t_g + t_{\text{nonst}} + t_s. \tag{59}$$

Since the free energy barrier for 3D nucleation is much higher than for 2D nucleation associated directly with crystal growth [47,51], the growth of crystals is much easier than nucleation in most cases. If crystals with a sufficiently small size can be detected with certain techniques, then we can have $t_g \ll t_{\text{nucl}}(t_{\text{nonst}} + t_s)$.

Nowadays, the laser light scattering method [4,12] promises the detection of particles from several nm to several μm. This has already been very close to the critical size of nuclei in many cases. Under such a situation, we can even assume $t_g \to 0$. Apart from this, t_{nonst}, according to the previous section, is determined to a large extent by the diffusivity. If the ambient phase is not too viscous, as is the case in aqueous solutions, t_{nonst} is about a few microseconds [56]. This implies that $t_{\text{nonst}} \ll t_s$. Therefore, we can approximate eq. (59) by

$$t_i \cong t_{\text{nucl}} \cong t_s. \tag{60}$$

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

$$J = 1/(t_s V), \tag{61}$$

where V is the volume of the system. It follows from eq. (52) that

$$\ln t_s = \frac{\rho f(m, R')}{[\ln(1+\sigma)]^2} - \ln\left\{ V f''(m, R') \left[f(m, R') \right]^{1/2} \beta_{\text{kink}} (C_1)^2 B' \right\} \tag{62}$$

with

$$\rho = 16\gamma_{\text{cf}}^3 \Omega^2 / 3(kT)^3, \tag{63}$$

where $B' = B\lfloor (R^s)^2 N^0 \rfloor$. For a given system, since ρ, C_1 and B' are constant, the 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}. On the other hand, if the ambient phase is very viscous, such as the glass phase, t_{nonst} cannot be ignored. In this case, eqs (60) and (62) cannot be applied directly.

Note that in applying eqs (53) and (62) to examine the kinetics of nucleation, the total volume of the system should remain constant. In the case of nucleation in micro-sized droplets or emulsions, the condition of constancy is difficult to achieve. In such a case, we should consider $\ln(t_s V) \sim 1/[\ln(1+\sigma)]^2$ rather than $\ln t_s \sim 1/[\ln(1+\sigma)]^2$, as given by eq. (62) [48].

5.2.3.3. Size effect and the surface roughness of the substrate In the exponential term of eq. (52), $f(m, R')$ describes the reduction of the nucleation barrier from a genuine homogeneous nucleation ΔG^*_{homo} to the actual heterogeneous nucleation $\Delta G^*_{\text{heter}}$, due to the presence of foreign bodies (cf. eq. (22)). Figure 5.2a shows that for a given m, $f(m, R')$ will decrease with m when $R' \geqslant 0.1$, and gradually approach a constant value when $R' \geqslant 10$. The regime where $f(m, R')$ is independent of R' is referred to as the "size independent regime", and the regime where $f(m, R')$ varies with R' is referred to as the "size dependent regime".

According to eq. (32), the relative radius R' of a foreign body is determined by the radius of curvature of the substrates and of the critical nuclei. Obviously, R' will increase with R^s. On the other hand, for a given system, the average size or the curvature of the substrate should be fixed. Therefore R' should in that case depend on the critical radius of the nuclei r_c which is inversely proportional to $\Delta\mu/kT$ $(= \ln(1+\sigma))$ (eq. (31)). It follows that R' is proportional to $\Delta\mu/kT$.

At high supersaturations, r_c can be much smaller than R^s. This gives rise to a large R' (eq. (32)). In the case where nucleation occurs in the size independent regime, $f(m, R')$ is independent of the supersaturation. It follows that the $\ln(t_s) \sim 1/[\ln(1+\sigma)]^2$ plot should be linear according to eq. (62). Conversely, r_c increases as the supersaturation decreases. If supersaturation is sufficiently low so that $R' \leqslant 10$ (in the size dependent regime) is obtained, $f(m, R')$ will drastically increase with decreasing supersaturation (cf. Figure 5.2a.) It follows from eq. (62) that a drastic increase on the slope of $\ln(t_s) \sim 1/[\ln(1+\sigma)]^2$ plot will be obtained, and the plot will become non-linear.

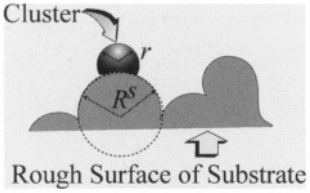


FIGURE 5.4. The roughness of the substrate surface determines the local curvature of the substrate.

At very low supersaturations or very small R^s, we have $R' \leqslant 0.1$. 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 have no effect on the nucleation kinetics. Any nucleation occurring in the neighborhood of foreign bodies having much smaller radii than the critical nucleus radius r_c would not be able to "feel" any influence on the part of the foreign bodies.

We notice that in most cases, the surface of the substrate is not atomically smooth (cf. Figure 5.4). The curvature of the substrate is then determined by the local curvature of the surface R^s (cf. Figure 5.4). In other words, the curvature of the substrate is the local curvature (or roughness) of the substrate surface.

5.2.3.4. Interfacial correlation between substrate and nucleating phase As shown by Figure 5.2a, in the case of $R' \gg 1$, the substrate can be regarded as being essentially flat, and $f(m, R')$ is then solely dependent on m. This implies that $f(m, R') = f(m)$ is independent of the supersaturation. According to eq. (62), the plot of $\ln(t_s) \sim 1/[\ln(1+\sigma)]^2$ should give rise to a straight line with the slope of $\rho f(m)$. Obviously, for a given system (ρ, $B' = \text{const.}$), the slope of the straight line will change according to $f(m)$. In this sense, the slope of the $\ln(t_s) \sim 1/[\ln(1+\sigma)]^2$ plot gives the relative $f(m)$ for the given system. One can analyze the change of the correlation between the substrate and the crystalline phase in terms of the variation of the slope.

As given by eq. (24), m is directly associated with γ_{cs}, which is determined by the interaction and/or structural match between the nucleating phase and the substrate. For a given crystalline phase and a given substrate, the optimal structural match at the crystallographic orientation $\{hkl\}$, corresponds to the strongest *average* interaction or the lowest interfacial energy difference between the crystalline phase and the substrate between the two phases. This orientation corresponds to the (minimal) cusp in the γ-plot (Figure 5.5a).

In the neighborhood of the minimal γ_{cs} (denoted by γ_{cs}^{min}) $\gamma_{cs}(\alpha)$ is given by [57]

$$\gamma_{cs}(\alpha) \approx \gamma_{cs}^{min} + \frac{\varepsilon b \alpha}{4\pi(1-p)}\left(1 - \frac{\alpha - \alpha_{max}}{\alpha_{max}}\right), \tag{64}$$

where: γ_{cs}^{min} = minimal specific interfacial free energy at a given orientation α; ε = elastic modulus; p = Poisson constant; b = Burgers vector; α = misorientation angle.

Combining eqs (24) and (64) yields

$$m \approx \frac{\gamma_{sf}}{\gamma_{cf}}\left\{1 - \frac{1}{\gamma_{sf}}\left[\gamma_{cs}^{min} + \frac{\varepsilon b \alpha}{4\pi(1-p)}\left(1 - \frac{\alpha - \alpha_{max}}{\alpha_{max}}\right)\right]\right\}. \tag{65}$$

Evidently, an excellent structural match ($\gamma_{cs}(\alpha) \to 0$, at $\alpha \to 0$) between the nucleating phase and the substrate leads to $m \to \gamma_{sf}/\gamma_{cf}$. Given that $\gamma_{sf} \approx \gamma_{cf}$, one has then $m \to 1$,

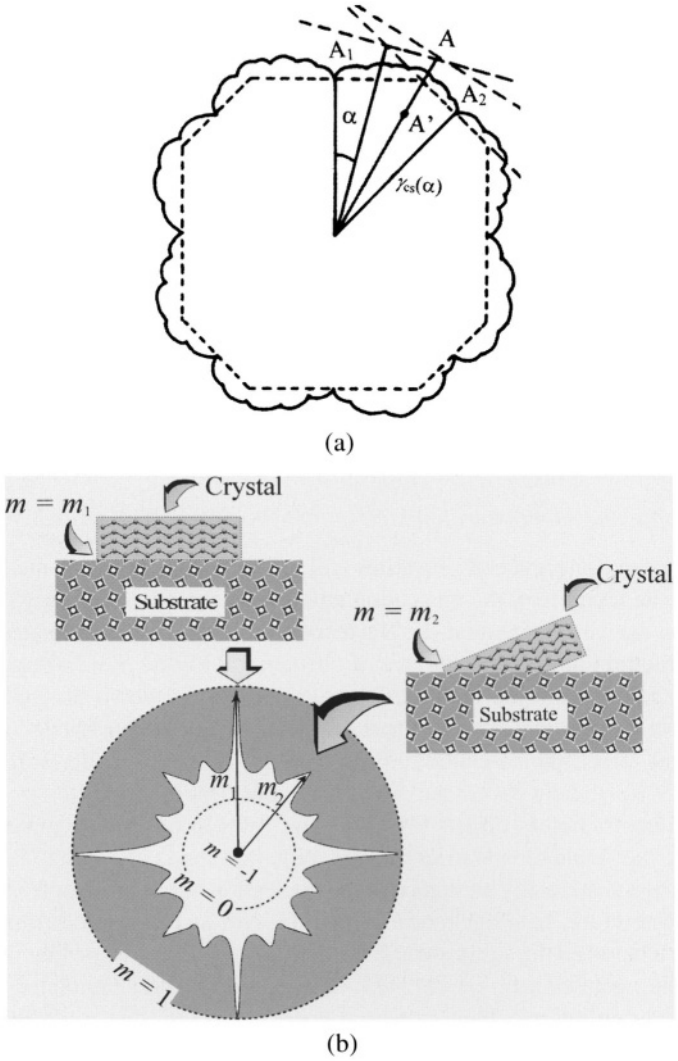


FIGURE 5.5. (a) Schematic plot of $\gamma_{cs}(\alpha)$ in two dimensions. (b) Illustration of m as a function of the crystallographic orientation. The structural match between the nucleating phase and the substrate can be described by m: $m = 1$ corresponds to the best structure match and $m = -1$ corresponds to the poorest structure match.

and $f(m) \to 0$ (cf. eq. (37)). This implies that $\Delta G^*_{\text{heter}}$ vanishes almost completely (cf. eq. (22)), which occurs during the growth of crystals, bringing about an excellent epitaxial relation.

As the structural match deviates from a perfect match, tending towards a poor match, m decreases from 1 to 0, and then to – 1. The extreme case will be $m \to -1$, corresponding to the situation where no crystal–substrate correlation exits. This is the case where substrates exert almost no influence on nucleation, and nucleation is controlled by the kinetics of

homogeneous nucleation. Nuclei emerging in this case are completely disordered, bearing no correlation to the substrate. One has then $f(m, R') = 1$.

In general, as $f(m, R')$ varies from 0 to 1 (or m from 1 to – 1), the interfacial structure correlation between the nucleating phase and the substrate changes from a completely correlated and ordered state to a completely uncorrelated and disordered state.

If the structural match between the nucleating phase and the substrate deviates from the global optimal structural match position, the secondary optimal structural match (the second lowest $\gamma_{cs}(\alpha)$) will be adopted, which is the second lowest minimum of γ_{cs}, associated with the orientation $\{h', k', l'\}$. Due to the anisotropy of the crystalline phase, the available $\gamma_{cs}(\alpha)$ values belong to a discrete set of values rather than a continuous range. A similar principle holds for subsequent deviations.

5.3. INTERFACIAL EFFECTS

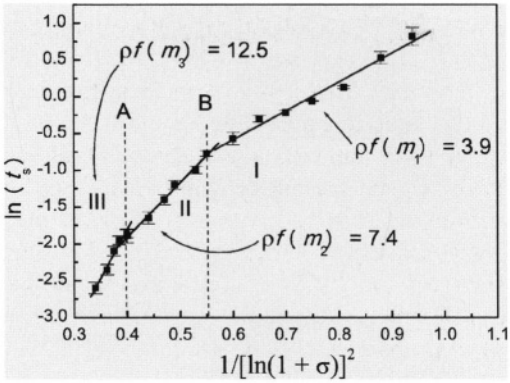
5.3.1. Templating and supersaturation driven interfacial structure mismatch

Kinetically, the occurrence of substrates will, on one hand, lower the nucleation barrier thus leading to an increase in the nucleation rate, but on the other hand it will exert a negative impact on the surface integration. Nucleation on a substrate will reduce the effective collision of structural units to the surface of clusters (cf. Figure 5.3), where the structural units are incorporated into the crystal phase. This will slow down the nucleation kinetics, which effect is contrary to the effect of lowering the nucleation barrier. As mentioned above, the latter effect, the so-called *"shadow effect"* of the substrate, is reflected in the appearance of $f''(m)$ and $f(m)$ in the pre-exponential term of eq. (52).

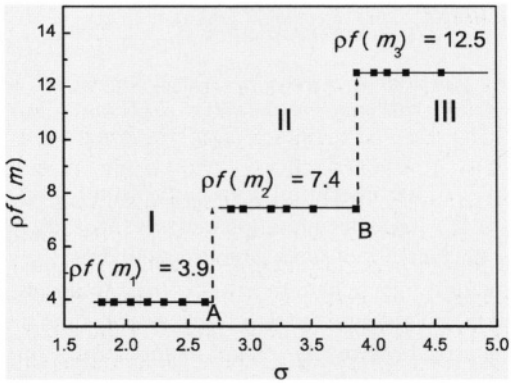
These two contradictory effects play different roles in different regimes. At low supersaturations, the nucleation barrier is very high (cf. eqs (33) and (34)). The nucleation rate will be substantially enhanced if the nucleation barrier is suppressed effectively ($f(m) \to 0$). Therefore, heterogeneous nucleation with a strong interaction and optimal structural match between the substrate and the nucleating phase will be kinetically favored. In this case, the nucleation of crystalline materials will be best templated by substrates capable of providing the excellent structural correlation with the crystalline phase. The structural synergy between the nucleating phase and the substrate will be optimal under this condition.

Epitaxy is an important technology that has been widely employed in producing semiconductor chips and engineering materials with complex structures [51,58]. Epitaxial growth, which utilizes the similarity in the structure of the host substrate to template the growth of the required crystalline material in a desired orientation [51,58], will achieve the best epitaxial relation.

At higher supersaturations, the exponential term associated with the nucleation barrier becomes less important. Instead, *the shadow effect* of the substrate, described by the pre-exponential factors $f(m)$ and $f'(m)$, dominates the kinetics more strongly (cf. Figure 5.3). Nucleation on substrates having larger $f(m)$ and $f''(m)$ (or $m \to 0, -1$) corresponds to a higher degree of orientational freedom (or a larger entropy). This will reduce *the shadow effect* of the substrate, and therefore it will become kinetically more favorable. This implies that the epitaxial template relationship between substrate and the nucleating phase cannot



(a)



(b)

FIGURE 5.6. (a) The plot of $\ln t_s$ vs $1/[\ln(1+\sigma)]^2$ for $CaCO_3$ (calcite) nucleation. Within the range of supersaturations where experiments were carried out, three straight lines with different slopes intercept one another, dividing the space into three regimes. (b) Supersaturation driven interfacial structural mismatch: with the increase of the supersaturation, the interfacial correlation factor $f(m)$ will increase abruptly at a certain supersaturation, such as A, B,..., corresponding to the transition from an ordered and structurally matched to a less ordered and structurally mismatched biomineral/substrate interface. Reproduced with permission from *J. Am. Chem. Soc.* **125,** 888–995 (2003). Copyright 2003 Am. Chem. Soc.

be maintained even for substrates having an excellent structural match with crystalline materials, if the supersaturation is too high.

If σ progressively increases from low supersaturations to high supersaturations, nucleation will be governed by a sequence of progressive heterogeneous processes associated with increasing $f(m)$. In analogy with the above analysis, we should obtain a set of pairwise interceding straight lines if $\ln t_s$ is plotted against $1/[\ln(1+\sigma)]^2$ (cf. Figure 5.8). Since for the crystalline phase, m and $f(m)$ take on only those values which correspond to some crystallographically preferential orientations (cf. Figure 5.5), $f(m)$ or the slope of the straight lines will take on discrete values, and $f(m)$ will increase as σ increases. A typical example is given in Figure 5.6, where the measured induction time is shown as a function of supersaturation for $CaCO_3$ nucleating from aqueous solutions [12].

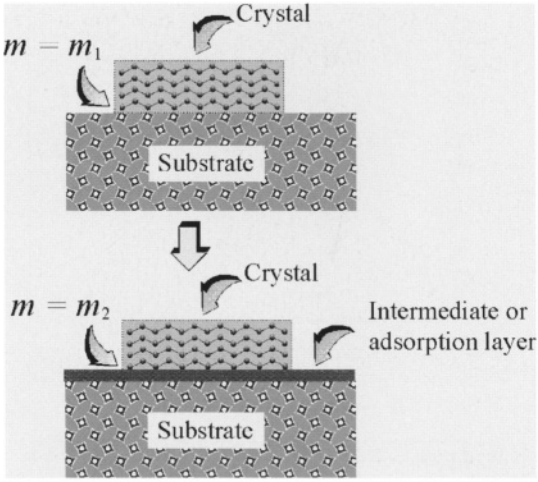


FIGURE 5.7. The adsorption of additives or other foreign materials between the nucleating phase and the substrate will change the interfacial structural match between the two phases ($m_1 \rightarrow m_2$).

As shown in Figure 5.6, the interfacial correlation factor $f(m, R')$ subsequently increases from $3.9/\rho$ to $12.5/\rho$, as supersaturation increases from 1.5 to 5. This result unambiguously confirms that the increase of supersaturation will drive the substrates/biominerals from an interfacial structural match state (a lower $f(m)$) to an interfacial structure mismatch state (a higher $f(m)$). This is referred to as *supersaturation driven interfacial structural mismatch.* As mentioned above, the abrupt changes from one state to the other at certain supersaturations (such as A, B, … in Figure 5.6) are due to the anisotropy of the crystalline phase.

Note that the templating and the supersaturation driven interfacial structure mismatch are the two contradictory effects. By carefully adjusting these two effects, we should be able to engineer and fabricate the complex structures of functional materials at the micro/nano structural level.

5.3.2. *Modification of interfacial correlation between substrate and nucleating phase*

The correlation between a substrate and the nuclei can be modified by introducing additives (or impurities), depositing an adsorption layer on the substrate [47,58]. If the adsorbed additives or the intermediate layers improve the interfacial correlation, we then have a nucleation promoting effect. Otherwise, we have a counter nucleation effect. 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, one has then $m \rightarrow \gamma_{sf}/\gamma_{cf} \sim 1$ and $f(m, x) \rightarrow 0$ (cf. eq. (37)). This can

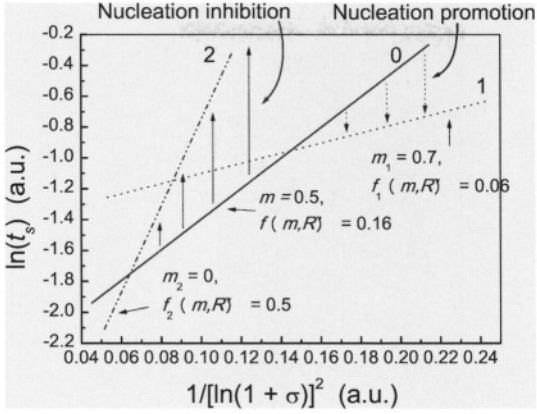


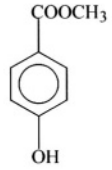
FIGURE 5.8. Illustration of the dependence of the nucleation rate on supersaturation and m (or $f(m)$). Foreign particles with a large m (strong interaction and better structural match with the nucleating phase) will control the kinetics at low supersaturations while those with a small m will control the kinetics at high supersaturations. Reproduced with permission from *J. Phys. Chem.* **B105,** 11550–11558 (2001). Copyright 2001 Am. Chem. Soc.

be identified from the lowering of the slope of $\ln(t_s) \sim \ln[(1+\sigma)]^2$, and the increase of the intercept due to the negative change in $\ln\{f''(m,x)[f(m,x)]^{1/2}\}$, which is illustrated by the shift of curve 0 to curve 1 in Figure 5.8. The introduction of surfactants to promote nucleation turns out to be one of several typical examples [59].

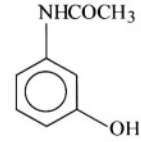
Conversely, the adsorption of additives leading to repulsion and the interfacial structure mismatch between the substrate and the nucleation crystalline phase will cause a substantial increase in γ_{sc}. It follows from eq. (37) that $f(m,x) \to 1$ at $m \to -1$. This enhances the nucleation barrier 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 the decrease of the intercept (illustrated by the shift from line 0 to line 2 in Figure 5.8.)

As an example, the influence of *m*-acetamidophenol (metacetamol), *p*-acetoxyacetanilide (PAA) on the nucleation of paracetamol is discussed below [58]. The molecular structures of paracetamol and of the additives are given as follows:

Paracetamol



p-Hydroxybenzoic acid methyl ester (methylparaben)



m-Acetamidophenol

The additive molecules share the acetanilide group, which strongly interacts with the surface of paracetamol embryos and/or the substrates. On the other hand, some functional groups attached to these additive molecules serve as disrupters to prevent the further ap-

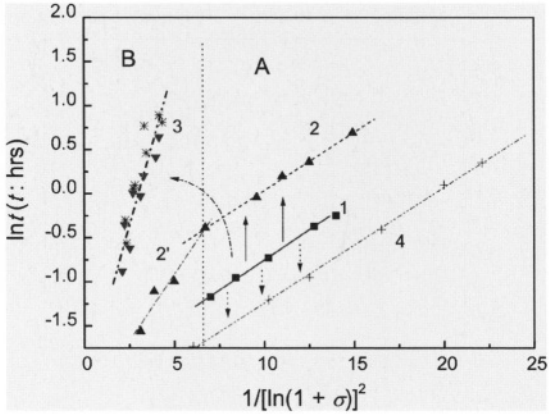


FIGURE 5.9. The effect of additives on the nucleation kinetics of paracetamol from aqueous solutions. (The equilibrium temperature is 44.1 °C.) Line 1 (■), no additives; line 2 (▲): 0.4 mol% methylparaben; line 3: (▼) 4 mol% *p*-acetoxyacetanilide (PAA), (∗): 4 mol% metacasaniol; line 4 (+): 0.13 wt% polysaccharide. In our experiments, polysaccharide (grade 1, Oxoid (UK)), *p*-hydroxybenzoic acid methyl ester (methylparaben) (97%, Aldrich (Gillingh, UK)), *m*-acetamidophenol (metacetamol) (97%, Adrich (Gillingham, UK)), *p*-acetoxyacetanilide (PAA) (99%, Aldrich (Gilingham, UK)) were chosen to examine the influence on paracetamol nucleation. Polysaccharide was added at 0.13 wt%, but all other substances were added at 4 mol%. based on the initial concentration of paracetamol. The experiments of paracetamol nucleation were carried out in a polarized microscope system [58]. Reproduced with permission from *J. Phys. Chem.* **B105,** 11550–11558 (2001). Copyright 2001 Am. Chem. Soc.

TABLE 5.1.
The effect of additives on the interfacial effect parameter and kink kinetic energy barrier for the nucleation of paracetamol from aqueous solutions.

Nucleation systems	$\rho f(m, x)$
1: Paracetamol (controlled)	0.14
2: Paracetamol + 4 mol% methylparaben	0.14
2′: Paracetamol + 4 mol% methylparaben	0.30
3: Paracetamol + 4 mol% metacetamol	0.65
4: Paracetamol + 0.13 wt% polysaccharide	0.14

proach of the paracetamol molecules. When the additive molecules adsorb onto the surface of the substrate, these functional groups will disrupt the substrate–nucleus structural match.

The results given in Figure 5.9 show that both metacetamol and PAA inhibit the nucleation of paracetamol in similar ways. As regards metacetamol, an OH group attaches to a "wrong" position at the benzene ring, in comparison with paracetamol. The adsorption of the molecules leads to the steric repulsion against the approach of paracetamol molecules. For PAA, a $COOCH_3$ group replaces the OH group in a paracetamol molecule. The bulky $COOCH_3$ group also causes a significant repulsion against the approaching paracetamol molecules. The adsorption of these additives on foreign bodies significantly interrupts the nucleation of paracetamol, which is characterized by an increase in $f(m, R')$ by a factor of 4.3 (cf. Table 5.1).

5.3.3. *Non-classical interfacial effect*

The nucleation related processes, such as epitaxy, templating, etc., are those processes that are associated with changes in the nucleation barrier. Nevertheless, in many cases, the presence of some impurities or additives only gives rise to a change in the interfacial kinetics without affecting significantly the nucleation barrier. Such effects are referred to hereafter as *non-classical interfacial effects.*

Let us take a close look at the interfacial aspects of the nucleation process. During the "growth" of clusters, structural units should be transported from the bulk phase to the surface of the clusters, and then become incorporated into kink positions on the clusters (see Figure 5.10). Before structural units can become incorporated into crystalline embryos (clusters), 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 (cf. Figure 5.10) [60–64]. Associated with this process, there is a free energy barrier $\Delta G^{\#}$ to be overcome. Actually, the kink integration coefficient β_{kink} occurring in eq. (62) is a factor describing this process, which is defined as [51,63,64]

$$\beta_{\text{kink}} = \nu \frac{a}{\lambda_0} \exp\left(-\Delta G^{\#}/kT\right) \tag{66}$$

(ν denotes the vibration frequency of structural units in the neighborhood of the surface, λ_0 is the average distance between two kinks at the surface and $\Delta G^{\#}$ is the activation free energy for kink integration).

Although this interfacial process has long been recognized [51], unfortunately it has never been defined unambiguously. Recently, the effect on the kinetics of crystal growth has been examined quantitatively [60–65]. It follows from recent computer simulations and theoretical analyses [60–64] that $\Delta G^{\#}$ should include:

(a) the desolvation energy barrier $\Delta G^{\#}_{\text{desolv}}$, i.e. the energy required to remove solvent or impurity molecules adsorbed at the kink sites;
(b) the orientational and conformational entropy barrier $\Delta G^{\#}_{\text{order}}$, which must be overcome in order to enable the molecules to adopt the crystalline structure at the crystal surface.

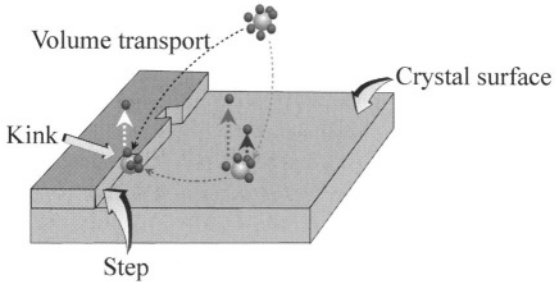


FIGURE 5.10. Schematic illustration of the surface integration occurring on the surface of a cluster. Reproduced with permission from *J. Phys. Chem.* **B105,** 11550–11558 (2001). Copyright 2001 Am. Chem. Soc.

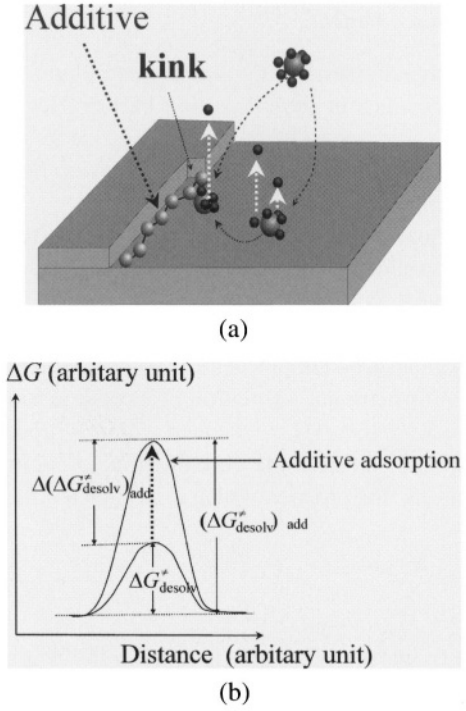
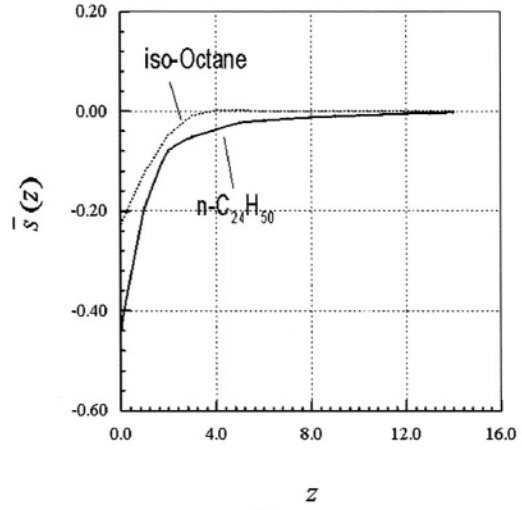


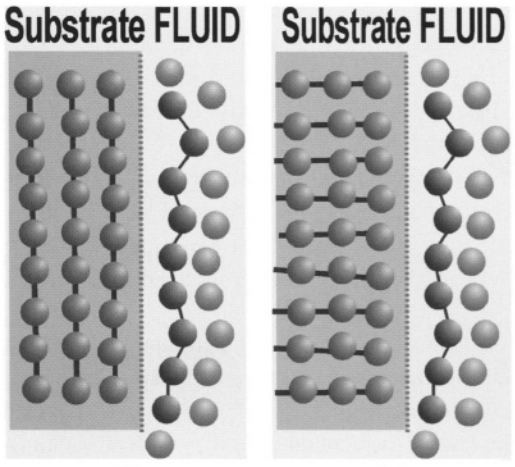
FIGURE 5.11. (a) Schematic illustration of the adsorption of additives at a kink site and (b) the enhancement of the desolvation free energy barrier.

In many cases, a stronger adsorption of additives at the surface of clusters will enhance $\Delta G^{\#}_{\text{desolv}}$, and therefore it will substantially reduce β_{kink} and consequently it will suppress the nucleation rate (cf. Figure 5.11).

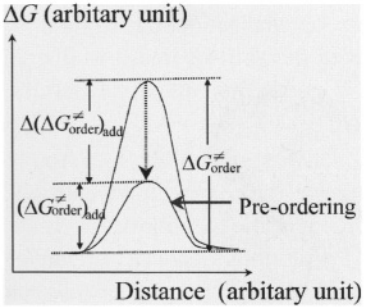
In the following discussion, we will focus on $\Delta G^{\#}_{\text{order}}$ since this effect will play an important role in the nucleation of complicated macromolecules. The crystallization of complex molecules is very different from the crystallization of simple molecules [60–64]. During the crystallization process, before such complex molecules can become completely converted into solid molecules, the molecules in the fluid, having statistically random orientations and conformations, should acquire an orientation and a conformation similar to the solid molecules in the crystal structure [60–64]. In practice, the substrates may induce the pre-ordering at the crystal-fluid interface. Figure 5.12a shows the preordering of chain-like molecules at a flat substrate [60–67]. For such a system, if the ordered solute molecule has the same orientation as the solid molecules at the surface (Figure 5.12b), it becomes much easier for it to acquire the crystalline order from the pre-ordered liquid molecules. The result will be a decrease in $\Delta G^{\#}_{\text{order}}$, and consequently a promotion of the nucleation (cf. Figure 5.12d). On the other hand, if the ordered solute molecules are in a significantly different orientation from that of the solid molecules at the surface (Figure 5.12b), it becomes difficult for them to acquire the crystalline order (or conformation) from the pre-ordered liquid molecules. It follows that this pre-ordering of the liquid molecule will raise $\Delta G^{\#}_{\text{order}}$, and consequently hinder nucleation.



(a)



(b) (c)



(d)

FIGURE 5.12. (a) The average order parameter $s(z)$ of n-$C_{24}H_{50}$ molecules as a function of z, generated by density functional theory calculations [65]. The density of segments in the bulk is $\phi_A = 0.1$; the torsion energy of the chains is $\varepsilon^{tor} = 1.5kT$. The order parameter is $s = 1/2(3\langle\cos^2\alpha'\rangle 1)$, where α' is the angle between a bond connecting two segment units and the normal. When the molecules are completely parallel to the surface, $s(z) = -1/2$. A random bond distribution will result in an order parameter $s(z) = 0$ [60,63,65]. (b) Illustration of the ordering of paraffin molecules at the solid–fluid interface. The ordered solute molecules are in the same orientation as the solid molecules on the surface. The pre-ordering will then facilitate the surface integration of solute molecules into the crystal surface. (c) The ordering of paraffin molecules at the solid–fluid interface. The ordered solute molecules have different orientations from the solid molecules on the surface. The pre-ordering will hinder the surface integration of the solute molecules into the crystal surface. (d) The pre-ordering of solute molecules as illustrated in (b) may lower $\Delta G^{\#}_{order}/kT$, and therefore promote the nucleation kinetics.

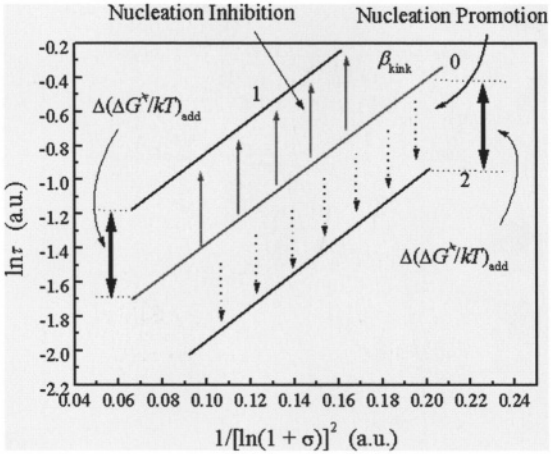


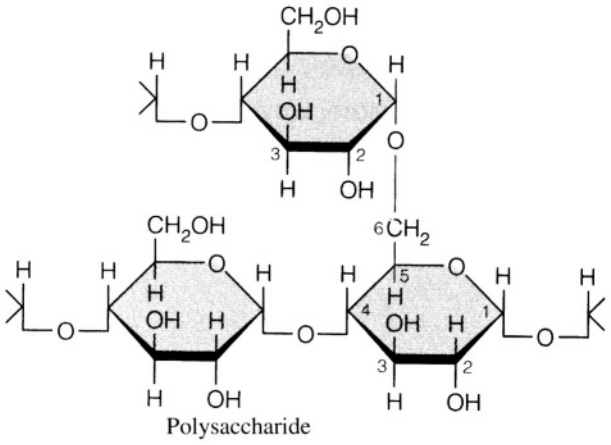
FIGURE 5.13. Illustration of the change in kink kinetics and the corresponding shift in the $\ln t_s \sim 1/[\ln(1+\sigma)]^2$ plot. Reproduced with permission from *J. Phys. Chem.* **B105,** 11550–11558 (2001). Copyright 2001 Am. Chem. Soc.

The change in $\Delta G^{\#}_{order}$ due to the molecular ordering can be quantified from the $\ln t_s \sim 1/[\ln(1+\sigma)]^2$ plot, as illustrated by Figure 5.13. If the pre-ordering of molecules promotes nucleation, the straight line of $\ln t_s \sim 1/[\ln(1+\sigma)]^2$ will shift downward (parallel to the shift of line 0 to line 2 in Figure 5.13). According to eqs (62) and (66), the difference in the intercept equals the change in the entropic barrier $\Delta(\Delta G^{\#}_{order}/kT)_{add}$ [$= (\Delta G^{\#}_{order}/kT)_{add} - \Delta G^{\#}_{order}/kT$] (see Figure 5.12c). The non-classical nucleation inhibition will shift the $\ln t_s \sim 1/[\ln(1+\sigma)]^2$ plot upward (the shift from line 0 to line 1). Similarly, the change in the entropic barrier can also be obtained from the difference in the intercept of the two lines.

Let us look again at the nucleation of paracetamol under the influence of polysaccharide, and *p*-hydroxybenzoic acid methyl ester (methylparaben), *m*-acetamidophenol (metacetamol), *p*-acetoxyacetanilide (PAA), respectively.

As shown in Figure 5.9, *m*-acetamidophenol (metacetamol) and *p*-acetoxyacetanilide behave classically. These two additives inhibit the nucleation of paracetamol by raising the slope from 0.14 to 0.65 (line 1 to line 3 in Figure 5.9). However, in the case of methylparaben, the inhibition of nucleation causes only a change in the intercept but not in the slope (from line 1 to line 3 in Figure 5.9) where the slopes remain unchanged (the slopes for both lines are 0.14). This implies that the occurrence of the molecules at the embryo surface does not change the interfacial free energy, but it does raise the desolvation energy in the kink integration by $0.84kT$ (the difference between the intercept of line 3 and line 1) [58].

The polysaccharide also gives rise to a surprising result. The shift from line 1 to line 4 in Figure 5.9 does not change the slope ($f(m, R')$ is approximately constant), meaning that polysaccharide promotes the nucleation of paracetamol crystals without lowering the nucleation barrier. This is completely different from the classical nucleation promotion and epitaxial growth.



Polysaccharide

The effect of polysaccharide is novel and has significant implications for the promotion of nonepitaxial nucleation. Results obtained from computer simulations, STM and AFM imaging [63–66] show that due to the influence of the solid molecules at the substrate surface, fluid units at the interface between the substrate and the fluid always acquire a certain degree of pre-ordering. The pre-ordered liquid molecules may become much more easily incorporated into the crystal structure at kink sites if the pre-ordering facilitates the transformation to the favored orientation and conformation at the kinks (lower $\Delta G^{\#}_{\text{order}}$) [61,63]. This leads to the enhancement of β_{kink}.

Notice that the occurrence of a substrate will inevitably change the nucleation barrier in most cases. This means that in most cases the non-classical epitaxial effect is always coupled with the classical epitaxial effect. Nevertheless, when $f(m, R') \to 1$, the nucleation barrier $\Delta G^{*}_{\text{heter}}$ reaches its uppermost limit, $\Delta G^{*}_{\text{homo}}$, in which case the classic interfacial effect is eliminated. According to Figure 5.2a, this elimination occurs only when the radius of the substrate R^{s} is smaller than 1/10 of the radius of critical nuclei r_c, ($R' < 0.1$). Because the substrate looks as though it has effectively vanished, this effect is named the "zero size" effect. Obviously, some polymers or very thin fibers can be considered as substrates providing the "zero" classic epitaxial effect since the thickness of polymers is much smaller than r_c in most cases. In fact, the polysaccharide turns out to be one of such examples, and the effect of nucleation promotion is characterized as a non-classical epitaxial molecular pre-ordering effect. As can be seen by Figure 5.9, the change in (difference between) the intercept of line 1 and 4, the kink integration barrier $\Delta G^{*}_{\text{kink}}$ is lowered by $0.42kT$.

When using additives with a structure somewhat different from that of paracetamol, we may expect a weak interaction between the additive molecules and the surface of the paracetamol embryos. This implies that a strong adsorption of the additive on the substrate is very unlikely. According to the above analysis, this will lower slightly β_{kink} or enhance $\Delta G^{\#}_{\text{desolv}}/kT$, $\Delta(\Delta G^{\#}_{\text{desolv}}/kT)_{\text{add}} > 0$. Figure 5.9 shows that methylparaben (line 2) changes the intercept only at relatively low supersaturations (regime A) relatively to the controlled (line 1), meaning that in this region the adsorption of methylparaben molecules on the surface of paracetamol embryos and the intervention for the kink kinetics is the major effect. The increase in the desolvation barrier is about $0.84kT$ (cf. Table 5.1). Further into the high supersaturation regime (line $2'$ in regime B), both the slope and the intercept are altered. $f(m, R')$ increases by a factor of 2 (cf. Table 5.1).

TABLE 5.2.
The effect of additives on the interfacial effect parameter and kink kinetic energy barrier for the nucleation of paracetamol from aqueous solutions.

Nucleation systems	B	$\Delta(\Delta G^{\#}_{\text{kink}}/kT)$	$\Delta(\Delta G^{\#}_{\text{kink}}/kT)_{\text{add}}$
1: Paracetamol (controlled)	−2.12	–	–
2: Paracetamol + 4 mol% methylparaben	−1.28	0.84	←
2′: Paracetamol + 4 mol% methylparaben	−2.42	0.81	←
3: Paracetamol + 4 mol% metacetamol	−1.91	2.64	←
4: Paracetamol + 0.13 wt% polysaccharide	−2.54	→	−0.42

If $f(m)$ and m can be obtained from the $\ln t_s \sim 1/|\ln(1+\sigma)|^2$ plot (cf. eqs (37), (62)), $f''(m)$ can also be obtained from eq. (55). Then $\Delta(\Delta G^{\#}_{\text{desolv}}/kT)_{\text{add}}$ can be calculated from the slopes and intercepts of lines 1 and 2 by solving simultaneously eqs (62) and (66). The resulting $\Delta(\Delta G^{\#}_{\text{desolv}}/kT)_{\text{add}} \sim 0.81kT$ is almost the same as the value obtained from regime A (cf. Table 5.2). This means that the methylparaben molecules still exert a comparable impact on the kink kinetics at low supersaturations.

The promotion of the nucleation caused by the pre-ordering of liquid molecules at the solid–fluid interface has been experimentally identified in the nucleation of paracetamol from aqueous solutions. The non-epitaxial molecular pre-ordering effect implies that macromolecules such as polysaccharide, can be used as molecular nucleation templators. Such an undertaking would have been impossible within the framework of the classic epitaxy. This effect may be utilized to template nucleation of crystalline materials along fiber or polymer networks to achieve microstructure engineering.

5.4. CRYSTALLOGRAPHIC MISMATCH BRANCHING AND MICRO/NANO STRUCTURAL ARCHITECTURE

Crystalline materials reveal a variety of patterns when they grow in different conditions. Pattern formation is the subject of significant fundamental and practical interest, and has been increasingly attracting a broad attention across various areas from materials science, biology, health, mineralogy, etc. [67–70]. The reason is that rheological and other physical properties of such crystal network systems are to a large extent determined by the self-organized structure of these patterns, rather than the building materials themselves. As mentioned at the beginning, a new type of functional materials, the so-called supramolecular materials, possess some unique properties and enable applications, which are directly related to the micro/nano structure of these self-organized networks. In this section we will focus on the mechanisms of self-organized crystal network formation, to identify principles for the architecture of crystal networks. The strategies for micro/nano structure engineering will also be discussed later in this section.

One of the most common patterns we encounter in our daily life is dendrites. When a liquid is supersaturated and begins to crystallize, the crystals grow and penetrate into the metastable liquid phase. The resulting crystallites often appear as needle-like dendrites [77, 78]. A common example of dendritic growth is snowflakes (see Figure 5.14).

Dendritic morphology is commonly observed in diffusion-controlled crystal growth in the presence of anisotropy [69,70,79–81]. For the growth of crystals, impurities and/or the

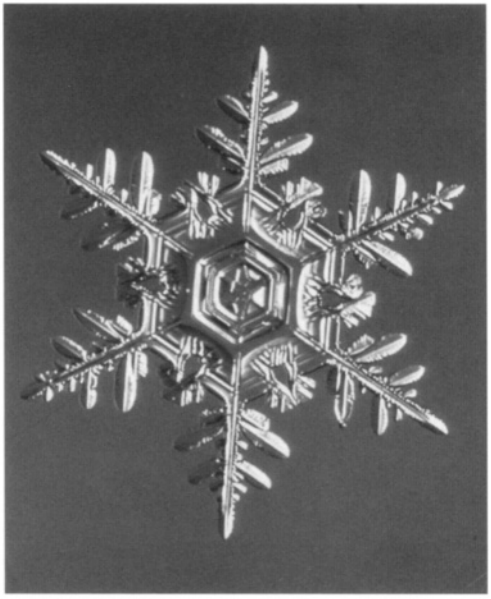


FIGURE 5.14. Snowflake: a typical example of dendritic growth. (The author wishes to thank Prof. Dr. Yoshinori Furukawa for his permission to use this photo.)

latent heat of crystallization that are generated at the interface between the crystal and the melt during freezing, must diffuse away from that interface. When the temperature gradient on the liquid side of the interface is very low or negative, like fins on a thermal radiator, the sharp tip of the dendrite may promote the diffusion of heat away from itself, allowing the crystal to grow rapidly in the direction in which the tip is pointing. When the ratio between the temperature gradient in the liquid phase G and the growth velocity of the crystal surface R_g is lower than a certain value i.e.

$$G/R_g < C \tag{67}$$

a surface instability will occur [77,78,86]. (Here C is a constant depending on the given crystallization system.) In general, the instability of growing crystal faces is due to a certain type of perturbation at the growth front, which triggers a local growth that is more rapid than the rest of the crystal surface.

Strictly speaking, a dendritic pattern is a crystal network. However from a theoretical point of view a dendritic pattern is developed from a single crystal, whereas the crystal networks under study are polycrystalline, with many crystallites being associated with each other by a certain linkage. Therefore, dendritic pattern will not be treated further.

The formation of certain patterns of crystalline materials can also occur via an aggregation process. It is believed [68–70,82] that the formation of a fractal pattern occurs via a diffusion-limited aggregation, a reaction control or a cluster–cluster aggregation process. Among these three patterns, the anisotropy in the interfacial development plays an important role in the pattern transformation [68–70,82]. The microscopic interfacial anisotropy

is essential to stabilize the dendritic growth, whereas insufficient anisotropy encourages the formation of fractal patterns [68–70,82].

In the following discussion, the crystal network formation is controlled by a completely new mechanism, the so-called crystallographic mismatch branching mechanism. The patterns created by this mechanism can have either the microscopic interfacial anisotropy of the dendrites or the characteristics of fractal patterns, or both. In particular, the evolution of a pattern depends on the supersaturation and impurity content.

5.4.1. Crystallographic mismatch nucleation and growth

In Section 5.3.1, we have shown that as supersaturation increases, the interface structural match between the substrate and the nucleating phase will deviate from the optimal structural match position. For normal growth, the substrate is a growing crystal itself, and the optimal structural match corresponds to the case that new crystalline layers match perfectly the crystallographic orientation of the parent crystal [71–75]. On the other hand, if crystallographic mismatch nucleation and growth take place, new crystalline domains will occur on the surface in different orientations with respect to the parent crystal (cf. Figure 5.15). (At relatively low supersaturations, the mismatch angle α is determined by the local minimum of the specific mismatch free energy γ_{mis} (see Figure 5.15).)

Actually, during the growth of crystals, there is a tendency for the occurrence of a mismatch of new layers at high supersaturations due to the supersaturation driven mismatch nucleation. Similarly to normal 3D nucleation, the mismatched domain should first nucleate on the growing crystal surface. If the energy cost to create a mismatch domain per area on the parent crystal is defined as the specific mismatch free energy, γ_{mis}, the barrier of mismatch nucleation is determined by the surface supersaturation and the specific mismatch free energy γ_{mis} (cf. eq. (34)). Obviously, the crystallographic mismatch nucleation is a special case of heterogeneous nucleation, where $\gamma_{\text{mis}} = \gamma_{\text{cs}}$. If the mismatched growth does not deviate much from the orientation of the parent crystal, we can in principle have $\gamma_{\text{cf}} \sim \gamma_{\text{sf}}$. It then follows from eq. (24) that

$$m = 1 - \gamma_{\text{mis}}/\gamma_{\text{cf}}. \tag{68}$$

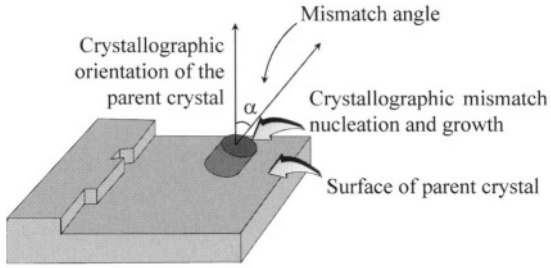


FIGURE 5.15. Schematic illustration of the crystallographic mismatch nucleation of a crystallite on a growing crystal surface. The crystallite occurs on the surface at an angle deviating from the crystallographic orientation of the parent crystal. At relatively low supersaturations, the mismatch angle α is determined by the local minimum of the specific mismatch free energy γ_{mis} (cf. Figure 5.5).

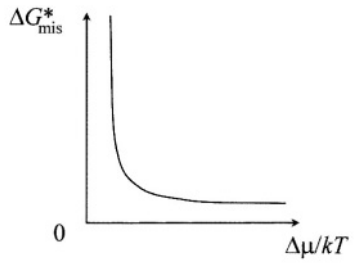


FIGURE 5.16. The change of the mismatch nucleation barrier ΔG^*_{mis} with supersaturation (cf. eq. (34)).

Likewise, the nucleation barrier and the rate of crystallographic mismatch nucleation are given by eqs (33), (34) and (52). As indicated by eq. (34), ΔG^*_{mis} decreases as the surface supersaturation increases (see Figure 5.16).

Here, we will discuss briefly some key factors controlling the crystallographic mismatch nucleation and growth:

(a) *Supersaturation.* The crystallographic mismatch nucleation occurs on the surface of a growing crystal. Similarly to normal nucleation and growth, the kinetics of crystallographic mismatch nucleation and growth depends additionally on the supersaturation at the surface of the growing crystal, which increases monotonically with bulk supersaturation. Therefore, at low bulk supersaturations, ΔG^*_{mis} is very high (cf. Figure 5.16). Therefore, it is difficult for the crystallographic mismatch nucleation to occur at low σ. As bulk supersaturation increases, ΔG^*_{mis} will drop rapidly (cf. Figure 5.16). ΔG^*_{mis} becomes relatively low at high supersaturations, thus facilitating the interfacial mismatch nucleation.

(b) *Impurities.* Adsorbed impurities may disturb the interfacial structural match between nucleating layers and the parent crystal surfaces. This gives rise to a lowering of *m,* hence promoting the crystallographic mismatch nucleation (cf. Section 5.4.4.3). For that reason, the adsorption of additives on the surface of a growing crystal will also promote the crystallographic mismatch nucleation by changing the surface supersaturation. Actually, the growth of a crystal can be simply regarded as a process by which growth units are transported to the surface of the crystal and incorporated into the crystal structure at the crystal surface (cf. Figure 5.10). When surface kinetics is a much faster process than volume transport, the growth units which reach the crystal surface will be readily incorporated into the crystal structure. Consequently, the concentration of growth units on the surface is then very close to the equilibrium concentration; in other words, the surface supersaturation is very small or close to zero.

On the other hand, the rate of surface integration will be significantly reduced if additives adsorb onto the growing crystal surface. When surface kinetics becomes much slower than volume transport, the growth units which reach the crystal surface will accumulate on the surface. This will then raise the surface supersaturation to the bulk supersaturation. This implies that the adsorption of additives will give rise to a high surface supersaturation, which will trigger the crystallographic mismatch nucleation and growth.

(c) *Slow surface integration.* As mentioned above, the faces with a low surface integration rate correspond to higher surface supersaturation, i.e., much closer to the bulk supersaturation. It follows that at low supersaturations, crystal faces with slow surface integration kinetics can take advantage of the highest possible supersaturation—the bulk supersaturation in a given system, and can trigger much more easily crystallographic mismatch nucleation (cf. Figure 5.19).

(d) *Low specific mismatch free energy.* According to eq. (34), a low specific mismatch free energy γ_{mis} corresponds to a low ΔG_{mis}, according to eq. (34). This implies that crystallographic mismatch nucleation can occur much more easily on crystal surfaces with low values of γ_{mis}. Note that normally, the crystal surfaces with low γ_{mis} values often coincide with those with slow surface integration kinetics. Therefore, criteria (c) and (d) are likely applicable to the same crystallographic orientation for a given crystalline material.

Based on (a), (c) and (d), we can expect that at low supersaturation crystallographic mismatch nucleation may take place very easily in slowly growing crystallographic orientations (cf. Figure 5.17a), whereas at high supersaturations crystallographic mismatch

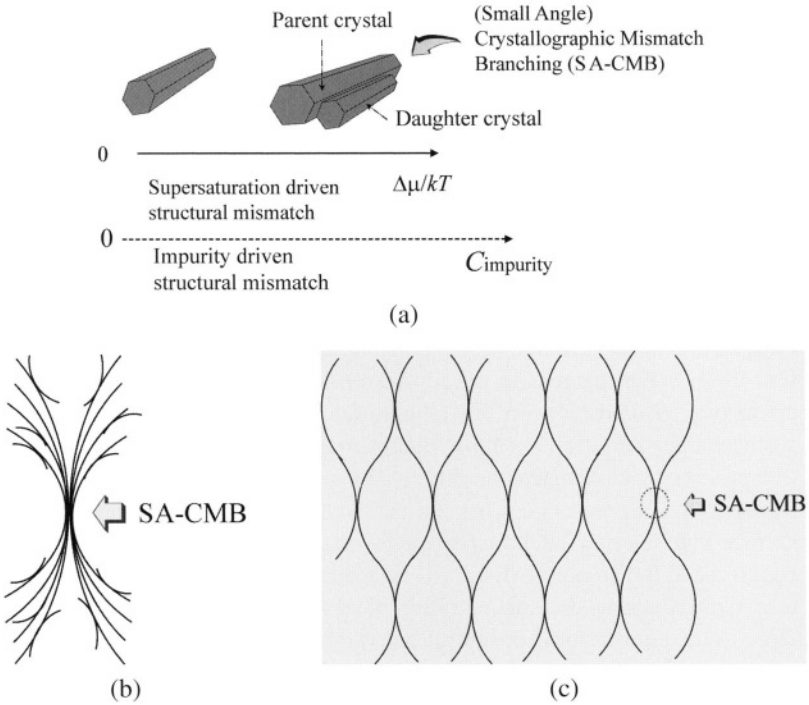


FIGURE 5.17. Crystallographic mismatch nucleation and growth: the newly created crystal differs from the parent crystal due to the (homo-epitaxial) crystallographic mismatch nucleation and growth. (a) Crystallographic mismatch nucleation and growth on the side face of a needle crystal, resulting in small angle crystallographic mismatch branching (SA-CMB). This normally occurs at relatively low supersaturations. (b) Spherulitic pattern resulting from SA-CMB. (c) Open interconnecting fiber network resulting from SA-CMB. In comparison with the SA-CMB spherulitic pattern shown in (b), the open network structure shown in (c) occurs at higher supersaturations.

nucleation may occur in faster growing crystallographic orientations (cf. Figure 5.18a). Note that another reason for the occurrence of crystallographic mismatch nucleation in faster growing crystallographic orientations at high supersaturations is that the faster growing crystallographic orientations can easily penetrate into the bulk, and "feel" effectively much higher supersaturations in the bulk. According to criterion (a), this will trigger crystallographic mismatch nucleation at the tips.

Once crystallographic mismatch nucleation and growth take place in the needle-like crystals, a special *branching* will occur (cf. Figures 5.17a and 5.18a). Unlike dendritic branching, in this type of branching, the daughter branches of fibers cannot be correlated strictly to the crystallographic orientation of their parent fibers. It is for this reason that this type of branching is referred to as *crystallographic mismatch branching*. (See Figures 5.17 and 5.18.)

At low supersaturations, crystallographic mismatch nucleation and growth will occur only on the side faces of needle-shaped crystals, leading to the so-called "small angle" crystallographic mismatch branching. Typical crystalline networks resulting from "small angle" crystallographic mismatch branching are illustrated in Figures 5.17b and 5.17c. Figure 5.17c shows a more open structure occurring at relatively high supersaturations, in contrast to the structure shown in Figure 5.17b.

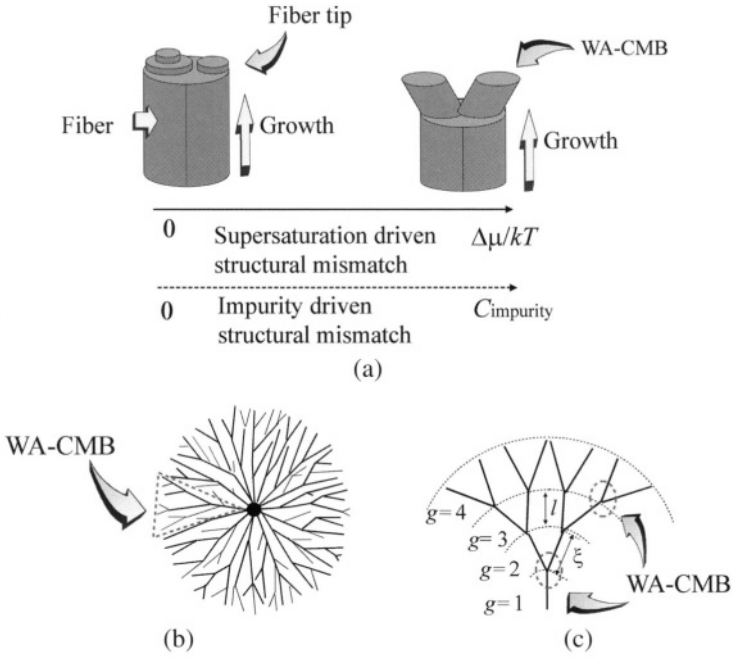


FIGURE 5.18. (a) Supersaturation driven structural mismatch at the tip of fiber crystals leads to wide angle crystallographic mismatch branching (WA-CMB). This occurs at relatively high supersaturations, or in the presence of additives (including self-impurities.) (b) Spherulitic pattern resulting from WA-CMB. (c) Open interconnecting fiber network resulting from WA-CMB. In comparison with the WA-CMB spherulitic pattern shown in (b), the open network structure shown in (c) occurs at much higher supersaturations.

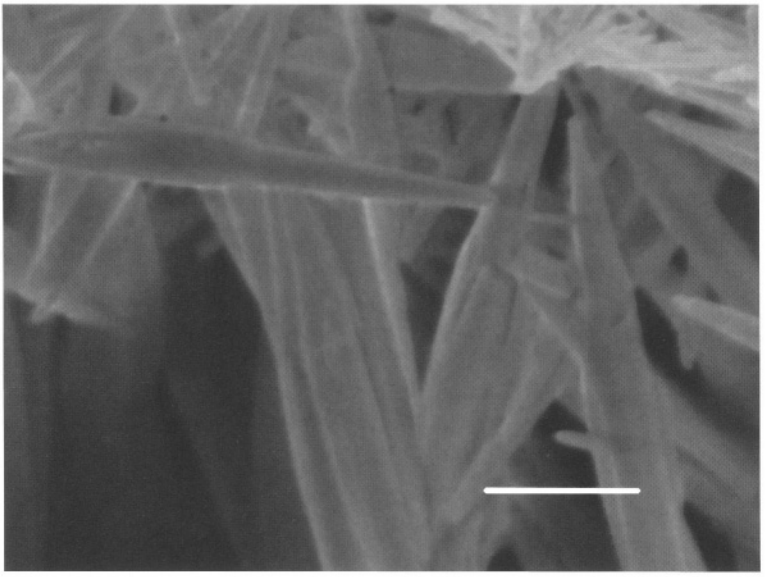
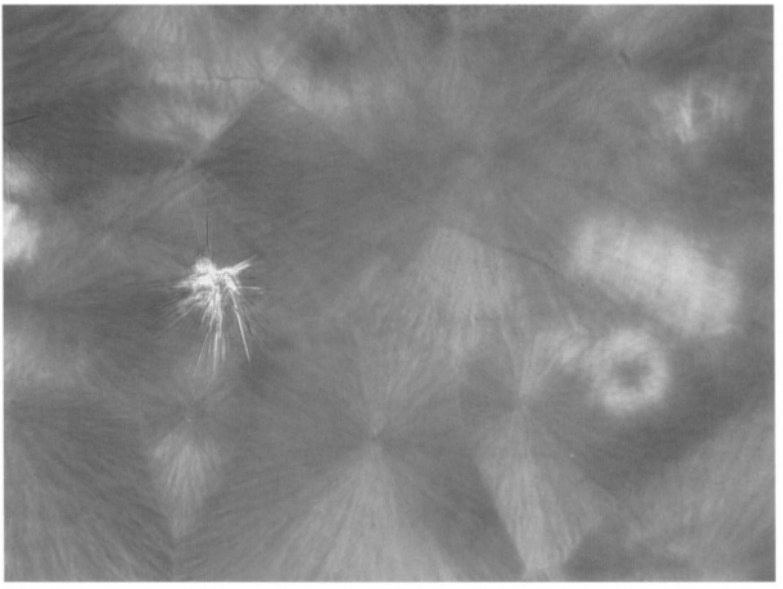


FIGURE 5.19. SEM micrograph showing compact hydroxyl apatite (HAP, $Ca_{10}(PO_4)_6(OH)_2$) crystals obtained from a relatively low supersaturation. Due to the (small angle) crystallographic mismatch branching (SA-CMB), the crystallites show a very good structural synergy. SA-CMB in this regime could lead to an ordered, compact and tough structure. Scale bars: 500 nm. Reproduced with permission from *J. Am. Chem. Soc.* **125**, 888–995 (2003). Copyright 2003 Am. Chem. Soc.

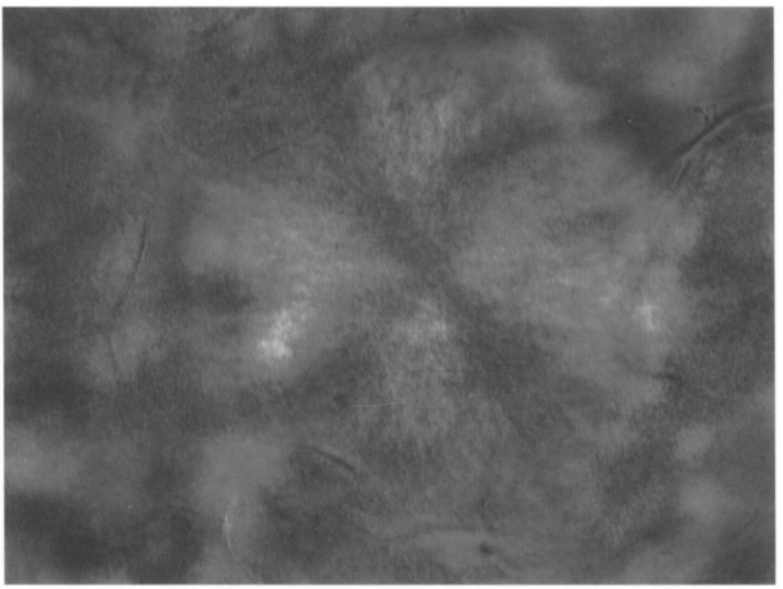
At high supersaturations, crystallographic mismatch nucleation and growth will take place at the tips of needle-shaped crystals, leading to the so-called "wide angle" crystallographic mismatch branching. Typical crystalline networks or patterns resulting from "wide angle" crystallographic mismatch branching are illustrated in Figures 5.18b and 5.18c. Actually, the more open structure illustrated in Figure 5.18c occurs at much higher supersaturations than that shown in Figure 5.18b. As will be discussed in Section 5.4.4, the crystallographic mismatch branching can be utilized to construct self-organized interconnecting fiber network structures or patterns, so as to engineer supramolecular functional soft materials.

Figure 5.19 shows self-assembled hydroxyapatite ($Ca_{10}(PO_4)_6(OH)_2$, HAP) crystals caused by (small angle) crystallographic mismatch branching on the prism faces. The prism side faces are the slowest growing faces in HAP crystals. Therefore crystallographic mismatch nucleation and growth will occur primarily on these faces at relatively low supersaturations. This crystallization process of HAP will lead to an ordered, compact and tough structure.

Two typical examples of spherulites resulting from wide angle crystallographic mismatch (cf. Figure 5.18b) are shown in Figures 5.20a and 5.20b. Figure 5.20a shows the spherulitic networks of 12-hydroxyoctadecanoic acid (5 wt%) formed in a dodecane solution. Similarly, the topology of spherulitic networks of N-lauroyl-*L*-glutamic acid di-*n*-butylamide (GP-1) in an *iso*-stearyl alcohol (ISA) organogel (3.6%) is shown in Figure 5.20b.



(a)



(b)

FIGURE 5.20. (a) Optical microscope graph of 12-hydroxyoctadecanoic acid spherulitic fiber networks obtained from dodecane gel. The resolution of a polarized light microscope is insufficient to recognize single nanofibers and fiber networks. But due to the radial distribution of fibers in the spherulitic form of networks, the micrograph will show the black extinction cross when the samples are viewed between crossed polarizers in such an optical microscope. C = 5 wt%, T = 20°C. (b) Optical microscope graph of N-lauroyl-*L*-glutamic acid di-*n*-butylamide (GP-1) fiber networks in *iso*-stearyl alcohol (ISA) gel. C = 3.6 wt%, T = 20°C.

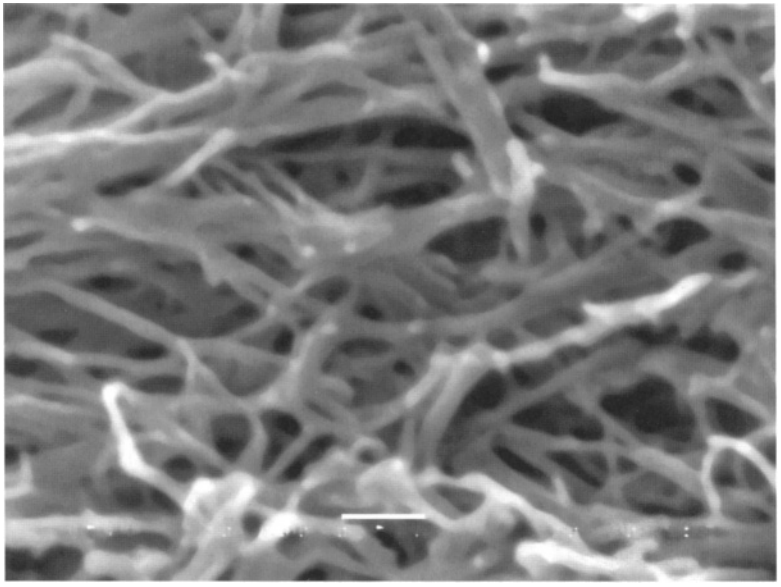


FIGURE 5.21. SEM microscope graph of fibrous networks of N-lauroyl-*L*-glutamic acid di-*n*-butylamide (GP-1) grown from a 7 wt% *iso*-stearyl alcohol (ISA) solution at 20°C, resulting from the mechanism of (wide angle) crystallographic mismatch branching [73–75]. To examine the nano/micro structure of fibrous networks in supramolecular materials, SEM, coupled with a CO_2 super critical fluid extraction technique (Thar Design) was applied. The latter is to remove the liquid captured in the networks, without disturbing the essential structure of the networks [73]. (Scale bar: 2 μm.) Reproduced with permission from *Chemistry of Materials* **14,** 3793–3798 (2002). Copyright 2003 Am. Chem. Soc.

An example of the open structure resulting from the wider angle crystallographic mismatch branching as illustrated by Figure 5.18c is given by Figure 5.21, where the interconnecting fibrous network of N-lauroyl-*L*-glutamic acid di-*n*-butylamide (GP-1) is grown from a 7 wt% *iso*-stearyl alcohol (ISA) solution at 20°C [73–75]. As mentioned above, the crystallographic mismatch branching of this type often occurs at high supersaturations and will give rise to an open and porous structure. Such structures are often used as a medium to deliver different benefits (cf. Section 5.4.4).

Apart from needle-like crystals, crystallographic mismatch nucleation and growth are also frequently observed on the surfaces of plate-like crystals. Most *n*-paraffin crystals have a thin platy and rhombic shape bounded by the large {001} and the narrow {110} faces. The large top and bottom {001} faces correspond to the slowest growing directions. As mentioned above, crystallographic mismatch nucleation and growth may easily occur on these faces at relatively low supersaturations. Figure 5.22 shows the (small angle) crystallographic mismatch branching occurring on the large {001} faces of n-$C_{24}H_{50}$ crystals at low supersaturation ($\sigma \leqslant 0.2$). Some daughter crystals are observed on the {001} face of the parent crystal.

Crystallographic mismatch branching on the plate-like crystals will also lead to the formation of crystal networks. One of the typical examples of such networks is the paraffin network formed in crude oil in severe winter conditions. Only a few percent of the paraffin crystals occurring in the form of interconnecting networks will be sufficient to gel crude

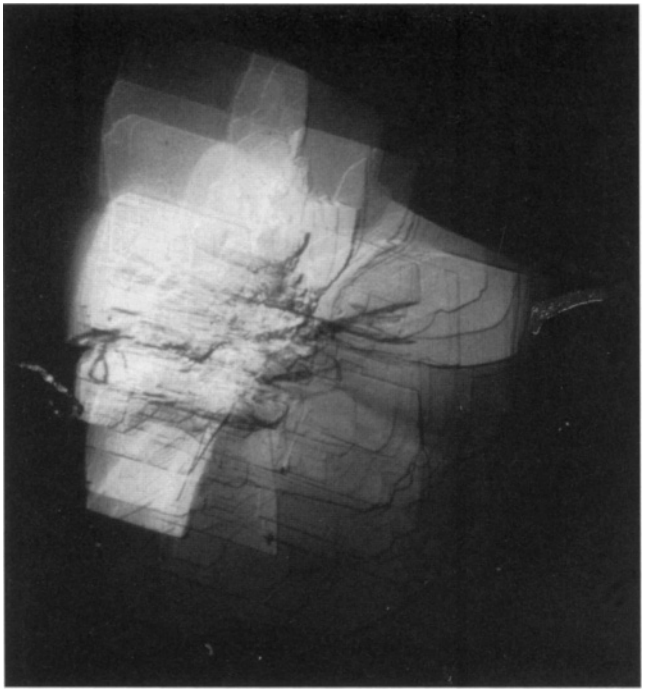


FIGURE 5.22. n-$C_{24}H_{50}$ crystal grown from n-hexane solutions. The large top and bottom faces are {001} and the side faces are {100}and {010} [76]. Reproduced with permission from *J. Crystal Growth* **135,** 209–223 (1994). Copyright 1994 North Holland.

oil. This will cause the blockage of oil pipelines, which turns out to be a serious issue for the petroleum industry [76]. On the other hand, crystal networks formed by platy crystals can also be used to advantage. Fat crystals have a platy shape. The crystal network formed by fat crystals is widely used as a structuring agent of margarines.

5.4.2. Fibrous network formation via crystallographic mismatch branching

We will focus on the formation of fibrous networks/patterns by means of wide angle crystallographic mismatch branching, in particular, on the typical patterns shown in Figures 5.17b and 5.17c. In this case, the crystal networks consist of radial arms originating in a core. As shown in Figure 5.20, although the fibers are too fine to be identified by optical microscopy, the spherulitic arrangement of fibers directed from the center can be recognized from the black extinction cross of the sample viewed between two crossed polarizers of an optical microscope. The radial arms are often found to be branched with the Cayley tree structure (cf. Figure 5.18c) [83,84].

The crystal networks are produced by 3D nucleation. This can be verified by the following experiments. Because the networks immobilize the liquid captured within the meshes, the formation of such crystal networks will change the rheological properties of the system (cf. Figure 5.23 and Section 5.4.3). The fact that the gelation time t_g corresponds to the induction time of the crystal network formation (cf. Figure 5.23a), is reflected in the linear

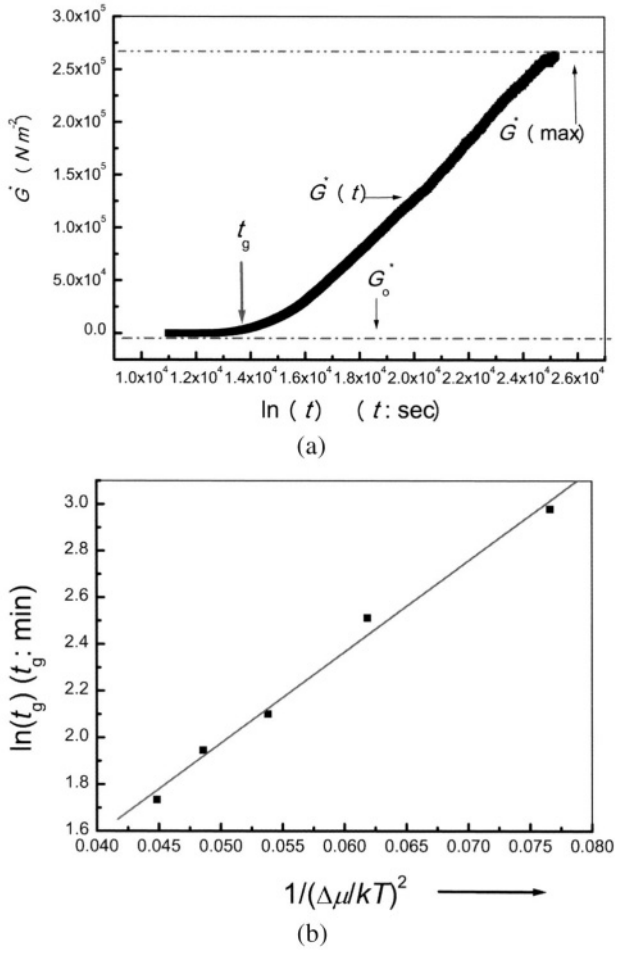


FIGURE 5.23. (a) The variation of the complex modulus G^* before and after fibrous network formation. The induction time t_g of fiber formation is identified from the change of G^*. T = 40°C, C = 6.7 wt%. (b) The relationship between t_g and $\Delta\mu/kT$. The linear relation of $\ln(t_g) \sim 1/(\Delta\mu/kT)^2$ indicates the nucleation and growth control during the fibrous network formation [51,87]. For rheology, we will focus on the storage modulus G' (describing the elastic property), the loss modulus G'' (describing the viscosity property), and the complex modulus G^* $(= [(G')^2 + (G')^2]^{1/2})$ [73–75]. Reproduced with permission from *Chem. Phys. Chem.* **3**, 374–377 (2002). Copyright 2002 John Willey & Sons, Inc.

plot $\ln t_g \sim 1/\ln[1+\sigma]^2$ obtained in Figure 5.23b. According to eq. (62), this confirms that the formation of a spherulitic pattern is controlled by 3D nucleation.

Since the self-organized crystal networks possess the structural characteristics of a Cayley tree (cf. Section 5.4.2.2), the process of network formation can be regarded as a sequence of the following steps: *initial nucleation*–fiber growth–fiber branching–fiber growth … (see Figure 5.24). Obviously, one of the key steps in constructing the Cayley tree resides in the branching of the growing nano fibers.

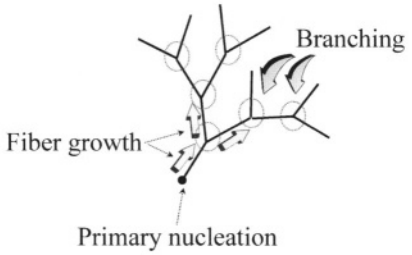


FIGURE 5.24. Crystallographic mismatch branching: *initial nucleation*–growth–branching–growth–branching and the self epitaxial nucleation of daughter branches on a growing parent branch [73–75].

5.4.2.1. Kinetics of crystallographic mismatch branching In this section, we will discuss first how Crystallographic mismatch branching takes place via *supersaturation driven* crystallographic mismatch nucleation and growth. The Crystallographic mismatch branching resulting from *impurity driven* Crystallographic mismatch nucleation and growth will be covered in Section 5.4.4.

The occurrence of Crystallographic mismatch branching is controlled by the following two steps: (1) growth on the surface of parent crystals; (2) *Crystallographic mismatch nucleation on the surface.*

As mentioned earlier, the growth of crystals can be regarded as a process of delivering growth units from the bulk to the crystal surface and incorporating these units into the kinks (cf. Figure 5.10). In the case of faceted growth, the crystal face is atomically smooth and the kinks occur only at the steps. In this case, the steps can be regarded as "sinks" for growth units to enter the crystals [51,86–88]. As shown by Figure 5.10, each advancing step will disappear when it spreads over the surface and reaches the edge of the surface. In order to continue the growth of the crystal surface, a subsequent crystal layer needs to be generated on the existing crystal surface. Therefore, the step source for the creation of new layers will determine the growth rate of the crystal surface. Due to the presence of a step free energy, the creation of a new layer on the existing layer of the crystal surface requires overcoming that free energy barrier, which is the so-called two-dimensional nucleation barrier [51,86–88]. Normally, for the growth of flat or faceted crystal surfaces, the screw dislocations occurring on the surface will provide uninterrupted step sources for a layer-by-layer growth [51,86], and in such cases the growth is controlled by screw dislocation mechanisms [51,86].

The growth of crystals that are free from screw dislocations is governed by the so-called 2D nucleation growth mechanism [87,88]. This implies that the crystal faces grow by depositing one crystal layer on top of the previous layer [88]. In the 2D nucleation model, the growth rate of the fibers is expressed as [88],

$$R_g = C_1 \left(\frac{\Delta\mu}{kT} \right)^{5/6} \exp\left[\frac{\Omega \gamma_{\text{step}}^2 \pi h}{3(kT)^2(\Delta\mu/kT)} \right], \tag{69}$$

where h, and γ_{step} denote the height of the step, and the step free energy of the growing surface, respectively. C_1 is a coefficient associated with the volume transport, and should be a constant for a given condition [88]. The growth of small crystallites such as crystalline

fibers is normally controlled by 2D nucleation, because it is difficult to accommodate screw dislocations in small crystalline domains.

When the growth of crystal surfaces occurs on the tips of fibrous or needle crystals, it is then likely to occur in a rough or a kinetically rough mode. In this case, the 2D nucleation energy barrier vanishes or is very small [51,87,88], and new layers or steps can be "generated" by thermal fluctuations. Therefore, the overall crystallographic orientation of the crystal surface will not be maintained [51]. The growth rate of rough (or normal) growth is expressed by [51,89]

$$R_g = A_1 \beta_{st} \left(\frac{\Delta\mu}{kT} \right), \tag{70}$$

where A_1 is a coefficient associated with volume transport, and should be a constant for a given condition [51,89].

As mentioned in Section 5.4.1, the crystallographic mismatch nucleation can be regarded as a special case of heterogeneous 3D nucleation caused by supersaturation driven interracial structural mismatch. Therefore, the nucleation rate can also be described by eq. (52) with m given by eq. (47).

The crystallographic mismatch nucleation and the growth of fibers can be approximately regarded as two independent physical processes. Given that the induction time for the crystallographic mismatch nucleation of new fibers on the host libers is τ ($\tau \sim 1/J$; J is the rate of crystallographic mismatch nucleation), and that the growth rate of fibers in the fibril axis direction is R_g, it follows then that the average branching distance can be expressed as

$$\langle \xi \rangle \sim R_g \tau \sim R_g / J. \tag{71}$$

For growth controlled by 2D nucleation, substituting eqs (52) and (69) into (71) yields

$$\langle \xi \rangle \sim C_2 \frac{(\Delta\mu/kT)^{5/6}}{f'' f^{1/2}} \exp\left\{ \frac{16\pi f}{3(\Delta\mu/kT)^2} \left(\frac{\gamma_{cf}\Omega^{2/3}}{kT} \right)^3 \right\}, \tag{72}$$

where $C_2 = C_1/B$. To check this mechanism, a set of experiments have been performed at different supersaturations.

In the case of rough growth (such as the growth on the tips of fibers or needles), substituting eqs (52) and (70) into (71) yields

$$\langle \xi \rangle = A_2 \frac{(\Delta\mu/kT)}{f'' f^{1/2}} \exp\left\{ \frac{16\pi f}{3(\Delta\mu/kT)^2} \left(\frac{\gamma_{cf}\Omega^{2/3}}{kT} \right)^3 \right\}, \tag{73}$$

where $A_2 = A_1 \beta_{st}/B$. Actually, eqs (72) and (73) are very similar apart from the kinetic coefficients C_2 and A_2. In addition, the power of the factor $(\Delta\mu/(kT))$ occurring in the pre-exponential expression is (5/6) for 2D nucleation growth and 1 for rough growth, which are very close to each other. Such a similarity may be attributed to the fact that the 2D nucleation model can also be applied to describe rough growth, in particular, the growth governed by kinetic roughening.

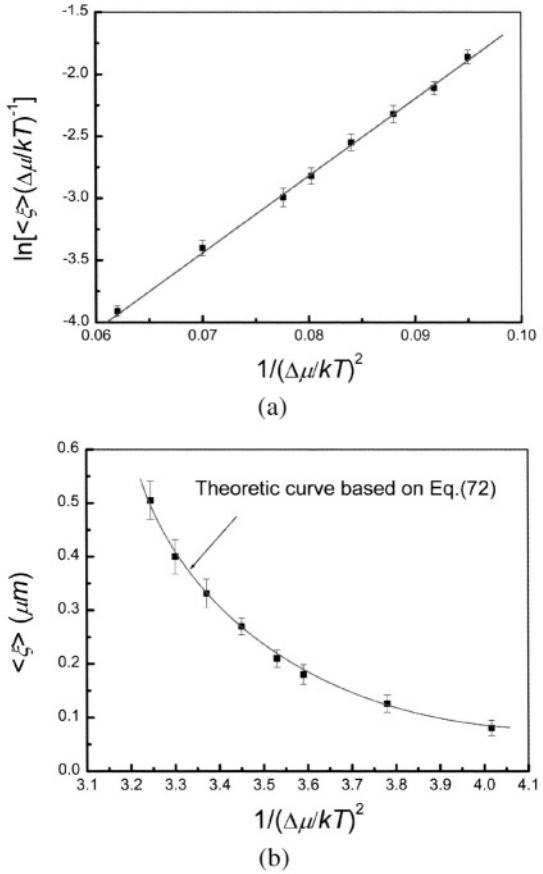


FIGURE 5.25. (a) The correlation between $\ln[\langle\xi\rangle(\Delta\mu/kT)^{-1}]$ and $1/(\Delta\mu/kT)^2$ for an L-DHL/DIOP system with 0.01% EVACP at $T = 20°C$. The linear relationship confirms the governing role of the crystallographic mismatch branching mechanism in the formation of organized interconnecting fiber networks. (b) Dependence of the average branching distance on supersaturation for the same system. $T = 298.15$ K, $X = 0.026$. Supersaturation difference $\Delta\mu/kT$ is obtained by changing the molar fraction of the solute in the solutions [75]. Reproduced with permission from *J. Am. Chem. Soc.* **124,** 15055–15063 (2002). Copyright 2002 Am. Chem. Soc.

The average length $\langle\xi\rangle$ of branching, which is proportional to the mesh size in the fiber networks, was examined by SEM, for an L-DHL (lanosta-8,24-dien-3β-ol: 24,25-dihydrolanosterol = 56:44)/DIOP (di-(2-ethyl-hexyl phthalate $(C_8H_{17}COO)_2\,C_6H_4$)) system with 0.03% EVACP (ethylene/vinyl acetate copolymer, $(C_4H_6O_2)_x(C_2H_4)_y$, approx. MW ~ 100,000). A linear relation is obtained between $\ln[\langle\xi\rangle(\Delta\mu/kT)^{-1}]$ and $1/(\Delta\mu/kT)^2$ (see Figure 5.25b). This confirms the above branching mechanism.

An important implication of this mechanism is that the branching density $(1/\langle\xi\rangle)$ of fibers will increase with $\Delta\mu/kT$ (or ΔT) (cf. eq. (73)). For a given concentration, T_{eq} is constant. It follows from eq. (15) that a lower T (or a large ΔT) corresponds to a larger $\Delta\mu/kT$. This means that a high supercooling or supersaturation will give rise to a highly branched and more open structure (cf. Figure 5.30b).

5.4.2.2. Structural characteristics of 3D network of supramolecular materials and correlation to rheological properties Although the structure of fibrous networks of supramole-

cular materials is very important with respect to the macroscopic properties [15–18] and has the typical characteristic of the Cayley fractal tree, the measurement of the fractal dimension D_f remains a challenging issue due to the lack of an effective characterization method. Due to the fact that the average length of the branches is often above μm, and the problem of multiple scattering, it is difficult to measure the fractal structure of this type of fibrous networks by the conventional light scattering method [93,94]. Here, we introduce a newly developed method [85] capable of measuring the *in-situ* fractal growth of fibrous networks of supramolecular materials.

Fractal structures are self-similar in that the two-point density-density correlation function, as well as their essential geometrical properties are independent of a length scale [90–93]. In d dimensional space, they can be characterized by a fractal or Hausdorff–Besicovitch dimension D_f [91–93]. The radius of gyration of such a crystal network R is related to the number N of particles or segments it contains, by [93,95]

$$N \sim R^{D_f}. \tag{74}$$

For a uniform object we have $D_f = d$, while for more open structures in which the density decreases with the distance from the center we have $D_f < d$.

Taking into account that fact that the formation of fibrous networks is controlled by nucleation and growth [73–75,85], the new method to be introduced is based on the famous Avrami equation [94], which was developed to describe the nucleation and growth of bulk crystals, taking into account the Poisson distribution of the crystal size [94]. It has the form

$$\ln[1 - X_{cr}] = -k^0 t^d, \tag{75}$$

where k^0 is a constant; t is the time; X denotes the crystallinity of the system, equal to $\phi(t)/\phi(\infty)$ (where $\phi(t)$ is the volume fraction of crystal materials at time t, and $\phi(\infty)$ is $\phi(t)$ at $t \to \infty$). For the growth of spherical crystals, one has $-k^0 t^d = \frac{4}{3} N_0 \pi (\nu_g t)^3$ (ν_g: the growth rate of bulk crystals; N_g: the number of crystals per unit volume) [94]. In the case of non-spherical crystals, this expression should be modified to $-k^0 t^d = K \frac{4}{3} N_0 \pi (\nu_g t)^3$ (where K describes the geometrical deviation of the crystals from the spherical shape). Generally, the one-dimensional or rod-like growth, the two-dimensional or plate-like growth, and the three-dimensional growth will lead to d = 1, 2, 3, respectively. In the case of fractal growth, one has growing fractal aggregates instead of uniform crystals. The length of bulk crystals (as a function of time) in the above consideration should then be replaced by the radius of gyration of fibrous networks, and d by the fractal dimension D_f. This implies that the above expression for fractal growth takes the form

$$\ln[1 - X_{cr}(t)] = -k^0 t^{D_f}. \tag{76}$$

The key step in applying eq. (76) is to measure X_{cr} as a function of t. This can be obtained from the correlation between $\phi(t)$ (the crystallinity) and such properties as the viscosity of the system [91]. Einstein's relation has been applied to examine the dependence of $\phi(t)$ on the viscosity of suspended particles and of polymer networks [91,95]. This relation allows the volume fraction of fibrous networks to be correlated to the specific viscosity η_{sp} by

$$\eta_{sp} \approx F\phi \tag{77}$$

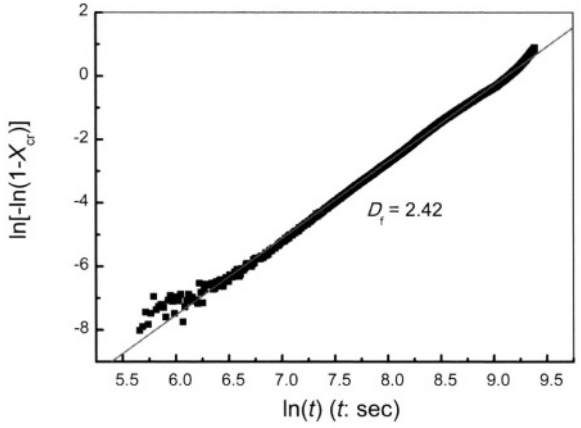


FIGURE 5.26. Dependence of $\ln\{-\ln[1 - X_{cr}(t)]\}$ on $\ln(t)$ for GP-1 fibrous networks occurring in an ISA solution. The fractal dimension is given by the slope ($D_f = 2.42$). The system is the same as in Figure 5.23; T = 40°C, C = 6.7 wt% [73]. Reproduced with permission from *Chem Phys Chem* **3**, 374–377 (2002). Copyright 2002 John Willey & Sons, Inc.

with

$$\eta_{sp} = \frac{\eta^* - \eta_0}{\eta_0}, \tag{78}$$

where η^* and η_0 are the complex viscosity of the system and the viscosity of the solvent, respectively; F denotes a factor depending on the shape of the particles [91]. The crystallinity can be given by

$$X_{cr}(t) = \frac{\varphi(t)}{\varphi(\infty)} = \frac{\eta^*(t) - \eta_0}{\eta^*(\infty) - \eta_0} = \frac{G^*(t) - G_0^*}{G^*(\infty) - G_0^*}, \tag{79}$$

where $\eta_{sp}(\infty)$ and $\eta^*(\infty)$ denote η_{sp} and η^* at $t \to \infty$, respectively; $\eta^*(t) = G^*(t)/\omega$, ($\omega$: angular frequency), $\eta^*(\infty) = G^*(\max)/\omega$ and $\eta_0^* \sim G_0^*/\omega$. Figure 5.23a shows the change of G^* as a function of time during the formation of GP-1 fibrous networks from an ISA solution (6.7 wt%). $X_{cr}(t)$ can be obtained from this graph by using eq. (79).

In Figure 5.26, $\ln\{-\ln[1 - X_{cr}(t)]\}$ is plotted versus $\ln(t)$. The curve at the beginning of network formation shows a certain scattering behavior and a slight deviation from a straight line. This is attributed to the 3D nucleation of fibers at the initial stage (cf. Figure 5.23b). As the networks grow, the power-law domain dominates the entire observable regime. This indicates that the essential geometrical properties are independent of the size increase of the gyration of fibrous networks during growth, and demonstrates the self-similarity of the fiber fractal patterns [91–93]. According to eq. (76), the slope is then legitimately interpreted as the fractal dimension D_f. A fractal dimension of 2.42 was observed for the above system, which is close to the value of which $D_f \sim 2.5$ for diffusion-limited aggregation (DLA) [91–93]. Nevertheless this does not imply that the formation of GP-1 networks is controlled by the DLA mechanism. Our experiments show that D_f varies with the experimental temperature T within the range of 20–40°C as shown in Table 5.3.

TABLE 5.3.
Dependence of the fractal dimension of GP-I fibrous networks on temperature.

Temperature (°C)	20	30	40
D_f	2.10	2.40	2.42
G'(max) (N m^{-2})	8.97×10^5	3.62×10^5	2.63×10^5
G''(max) (N m^{-2})	6.53×10^4	2.79×10^4	2.65×10^4

In Table 5.3 D_f varies from 2.42 to 2.10 as T decreases from 40 to 20°C. Since a denser structure adopts a higher Hausdorff dimension, this result implies that the fibrous networks will adopt a more open structure at lower temperatures [73].

We notice that some rheological approaches have also been applied to examine the fractal structure of networks of polymers and suspended particles [96–99]. For instance, Winter and his colleagues [98,99] utilized the correlation between the frequency and viscosity $\eta^* \sim (\omega)^{-2/(D_f+2)}$ to measure the fractal dimension of cross-linking polymers. However, these approaches are not likely to be applicable to the point where the gel occurs. As pointed out, the above relationship no longer holds at the gel point, because the molecular weight of a growing polymer cluster diverges at that point. It is worth noting that, unlike most other methods which are only suitable for the measurement of the fractal dimension, the approach described here allows one to measure fractal growth *in-situ,* which turns out to be very important in studying the kinetics of fractal network formation.

5.4.3. Evolution of a 3D crystal network

As shown in Section 5.4.2, the crystallographic mismatch branching leads to the formation of self-organized interconnecting networks. The two key elements associated with the crystallographic mismatch branching are the critical surface supersaturation σ_{tr} and the average branching period $\langle \xi \rangle$. The critical surface supersaturation is the supersaturation threshold at which crystallographic mismatch branching sets in. Evidently, it is crucial to find out how the transition from normal crystallization to the formation of self-organized interconnecting networks occurs, and how the crystal networks evolve. These are the key questions to be addressed in this section, and they are illustrated qualitatively in Figure 5.27a.

Two typical cases are characterized by low surface supersaturation. Firstly, when bulk supersaturation is low, since normal surface supersaturation is lower than bulk supersaturation, so is surface supersaturation. Secondly, when the surface kinetics of crystal growth is much faster than volume transport, surface supersaturation can also be very low even when a high bulk supersaturation is applied (cf. Section 5.4.1). Low surface supersaturation results in a long branching period $\langle \xi \rangle$. Let L be the average size of the individual crystals in the branching orientation. At $\sigma \leqslant \sigma_{tr}$ (regime A in Figure 5.27a), $\langle \xi \rangle > L$. Because the branching period is larger than the average crystal size, crystallographic mismatch branching will never take place. One will then have normal crystal growth. Figure 5.27b shows some sal-ammoniac crystals obtained under normal crystallization conditions.

As bulk supersaturation increases and the surface kinetics becomes a relatively slow process, surface supersaturation rises accordingly. This causes a decrease in $\langle \xi \rangle$ according

to eqs (72) and (73). Crystal network formation sets in once the elemental branching period becomes comparable to the average crystal size at $\sigma \geqslant \sigma_{tr}$. Figure 5.27c shows a case where dendritic growth and crystallographic mismatch of sal-ammoniac crystals occur simultaneously.

Note that at σ slightly higher than σ_{tr}, a relatively strict structural match between the parent crystal and the branching crystals (or a relatively low $f(m)$) will be present in order to facilitate the crystallographic mismatch nucleation and growth, considering in this case ΔG^*_{mis} is not sufficiently low (cf. Figure 5.16). In this case, a regular branching can be obtained. A further increase in the surface supersaturation reduces the average branching period $\langle \xi \rangle$ (cf. eqs (72) and (73)) and ΔG^*_{mis}. The requirement of a structural match between parent and daughter crystals becomes less strict due to a sharp drop in ΔG^*_{mis} (cf. Figure 5.16). It follows that random and highly branched crystal networks will be obtained.

The crystal networks will become more open as supersaturation increases (cf. Table 5.3). This will then lead to a decrease in the fractal dimension. (See Section 5.4.4.2.)

5.4.4. Micro/nano structural architecture of a self-organized three-dimensional nanofiber network of supramolecular functional materials

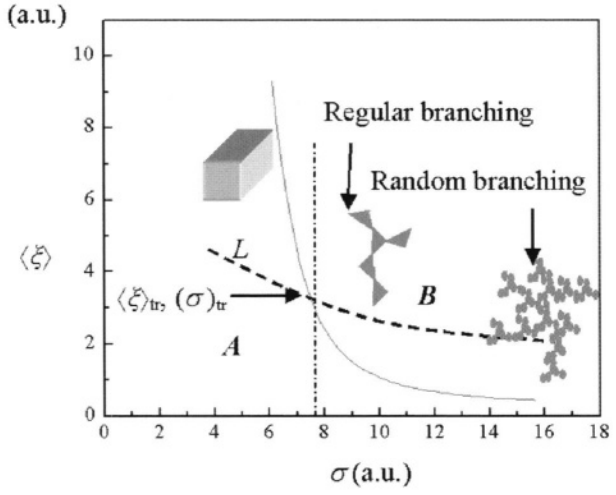
Well studied supramolecular functional materials having 3D fibrous network structures are produced for instance, by dilute solutions of polymers, proteins, and inorganic substances like silica or clays in water and organic solvents. In recent years there has been a rapidly growing interest in such materials, which is motivated by the many potential applications in the photographic, cosmetics, food and petroleum industries, in drug delivery, lithography, catalyst supporters, scaffolds for tissue engineering, novel separation methods for macromolecules, etc. [15–20].

For example, Figure 5.28 illustrates how supramolecular functional materials with 3D fibrous network structures can be utilized as a medium for drug delivery and controlled release. The specific cellular structure of interconnecting fiber networks will allow the delivery of certain drugs. The mesh size of 3D fiber networks will determine the rate of drug release.

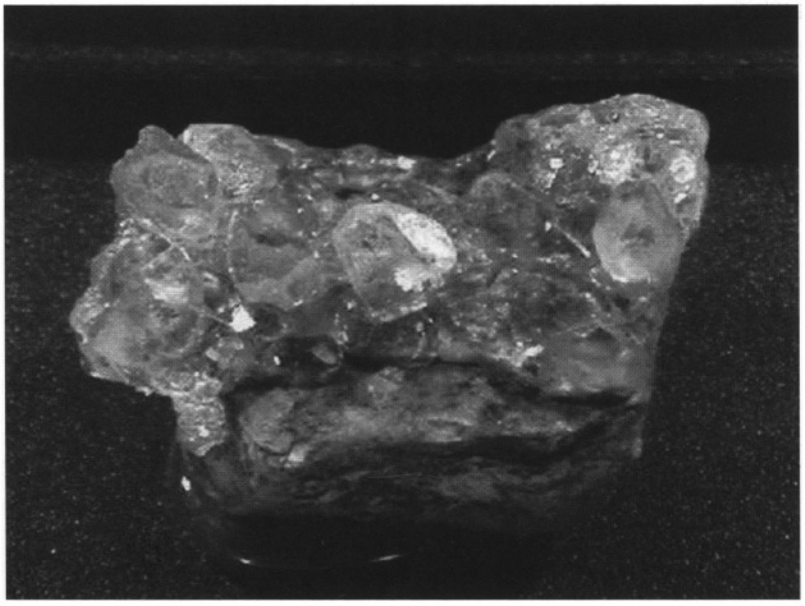
Another example is the use of organogelators as templates for the preparation of nano-structured materials. First, Möller, Weiss, and later Nolte [100–102] prepared membranes with man-sized pores by the gel-template leaching process. This process takes advantage of the distinct feature of organogelators to *reversibly* form elongated fibers with well defined dimensions and geometry. In the gel-template leaching process, gels of various gelling agents are prepared in polymerizable solvents like methacrylates or styrene in the presence of a cross-linking agent. After polymerization of the matrix, the organic gelling agent is removed again by extracting it using the appropriate solvent.

Consequently, the organogel fiber network is “imprinted” in the cross-linked polymer matrix, resulting in porous membranes with channels of a submicrometer scale and in some cases even nanometer dimensions (cf. Figure 5.29).

Apart from this, highly porous and self-supporting monolithic foams, which are found to have many important applications including novel bio-separation [106], can be obtained from the above materials after applying the CO_2 supercritical extraction.

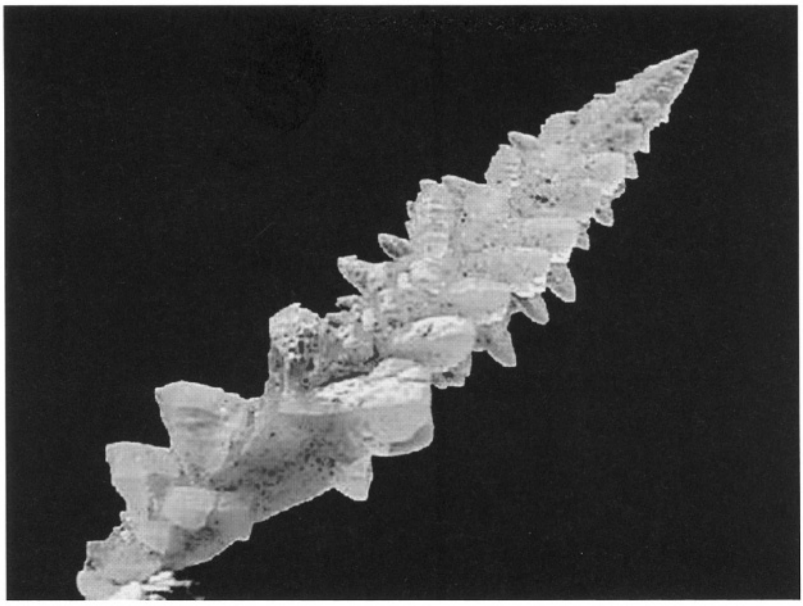


(a)



(b)

FIGURE 5.27. (a) Dependence of the average length of the crystals and the period of crystallographic mismatch branching on surface supersaturation, and the implications for pattern morphological/topological transformation. L: average size of crystals when no aggregation takes place. A regime $[(\sigma) < (\sigma)_{tr}, \langle\xi\rangle > L]$: normal crystal growth or dendritic growth regime; B $[(\sigma) \geqslant (\sigma)_{tr}, \langle\xi\rangle \leqslant L]$ regime: branching regime. $(\sigma)_{tr}$ and $\langle\xi\rangle_{tr}$ $(\sim L)$ are the widow parameters for branching. (b) Natural specimens of sal-ammoniac. (c) Natural dendritic like growth of sal-ammoniac. This pattern is also the case of the initial stage of the crystallographic mismatch branching. Many sub branches and the main branch belong to different crystals rather than the same crystal. [(b) and (c) were provided by Dr. T.N. Kerestedjian of the Geological Institute, Bulgarian Academy of Sciences.]



(c)

FIGURE 5.27. (Continued).

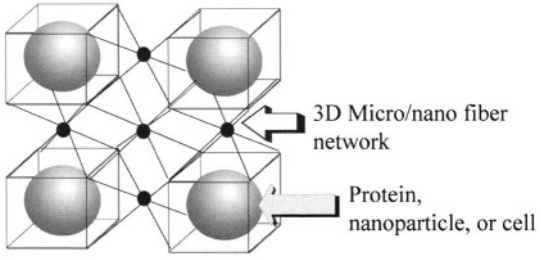


FIGURE 5.28. Supramolecular materials with interconnecting 3D fiber networks are applied to achieve drug delivery and controlled release. The specific cellular structure will allow the delivery of certain drugs. The mesh size of 3D fiber networks will determine the rate of drug release.

These materials have continuous three-dimensional interconnecting networks in the liquid, thereby preventing the liquid from flowing due to the capillary force. Among these materials, those formed from small molecules are a special class. In contrast to their macromolecular and inorganic counterparts, the network structure formed by gelators of a low molecular weight is held together solely by non-covalent forces. The macroscopic properties, in particular the rheological properties of supramolecular functional materials, are determined by the micro/nanostructure characteristics of the fiber networks. Fibrous networks with permanent interconnections will effectively entrap and immobilize liquid in the meshes, and possess both the elastic properties of ideal solids and the viscosity properties of Newtonian liquids, leading to the formation of self-supporting supramolecular materials [17–19,73–76]. In contrast, systems consisting of non-permanent or transient in-

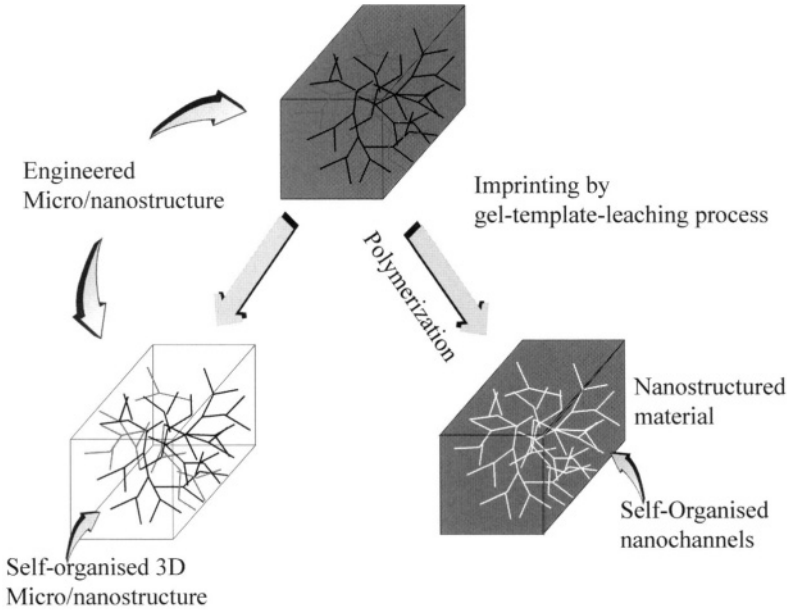


FIGURE 5.29. Schematic illustration for the gel-template leaching process to create nanostructured materials by nanofiber network imprinting.

terconnecting (or entangled) fibers or needles can only form a weak and viscous paste at low concentrations [16].

Due to the current lack of sufficient understanding of these materials, significant efforts have been devoted to identifying novel systems having a three-dimensional (3D) interconnecting self-organized network structure, in order to obtain *such functional materials* with the desired properties [17–19]. This includes the screening of a large number of compounds and solvent molecules, as well as suitable solvents capable of forming these structures [17–19,73–76]. In some cases, screening may become impossible due the limited choice of suitable materials. On the other hand, if interconnecting 3D micro/nano fiber networks with the required organization can be constructed, new functional materials can be produced having the required functionalities. Obviously, this is a completely new route in producing new functional materials. In this section, we will discuss a completely new approach to the production of new macromolecular functional materials.

5.4.4.1. Branching creation and micro/nano network architecture The strategy of this approach is to construct 3D permanent interconnecting nanocrystal fibrous networks from a system originally consisting of separate crystal needles via the promotion of crystallographic mismatch branching (CMB) (cf. Figure 5.30a), and consequently to engineer the materials with the required micro/nano structure (Figure 5.30b).

The system under study was obtained by precipitating L-DHL (lanosta-8,24-dien-3β-ol:24,25-dihydrolanosterol = 56:44) out from its DIOP (di-(2-ethyl-hexyl phthalate ($C_8H_{17}COO)_2$ C_6H_4)) solution at room temperature. The L-DHL powders used in our experiments are in a crystalline state (cf. the X-ray diffraction results in Figure 5.31). During the process of natural cooling of the aforementioned system (the concentration of L-DHL:

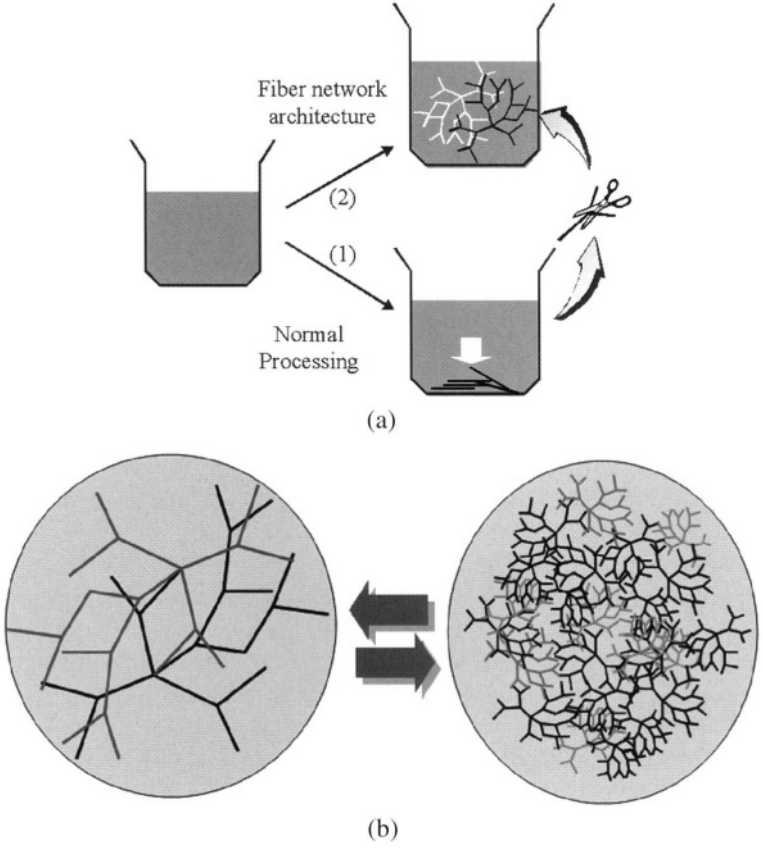


FIGURE 5.30. Illustration of the micro/nano structure engineering of functional materials. (a) Network architecture (route (2)) with respect to normal fiber formation. (b) Modification of micro/nanostructure of 3D interconnecting fiber network. Reproduced with permission from *J. Am. Chem. Soc.* **124,** 15055–15063 (2002). Copyright 2002 Am. Chem. Soc.

10 wt%) from 120°C to room temperature (~20°C), an opaque and viscous paste was obtained (see the right corner of Figure 5.32a). This shows that the system consists of only separated L-DHL *needles* (short and thick fibers), which are in temporary contact with each other (see Figure 5.32a). The aforementioned process of natural cooling is referred to as the "normal processing", illustrated by Figure 5.30a.

As illustrated by the route (2) in Figure 5.30a, networks with permanent interlinking can be created from such a system by introducing an additive, the so-called "branching creator". EVACP (ethylene/vinyl acetate copolymer, $(C_4H_6O_2)_x(C_2H_4)_y$, approx. MW ~ 100,000; from SP^2 Scientific Polymer Products Inc) was introduced to an identical L-DHL/DIOP solution as described above at 120°C. The molecular structure of EVACP is as follows:

$$-(CH_2{-}CH_2)_x-(CH_2{-}CH(-O{-}C(=O){-}CH_3))_y-$$

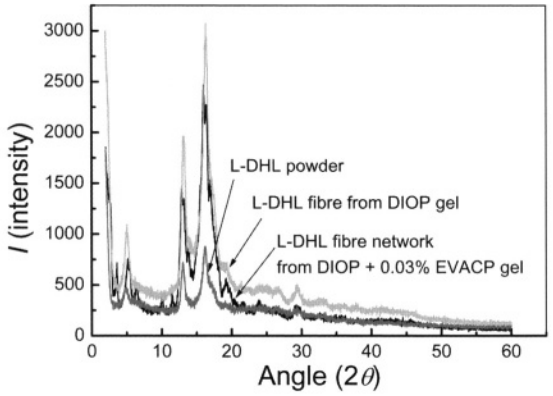


FIGURE 5.31. X-ray diffraction analysis of L-DHL crystalline powders and fibers. They all reveal identical crystalline structures [75]. Reproduced with permission from *J. Am. Chem. Soc.* **124**, 15055–15063 (2002). Copyright 2002 Am. Chem. Soc.

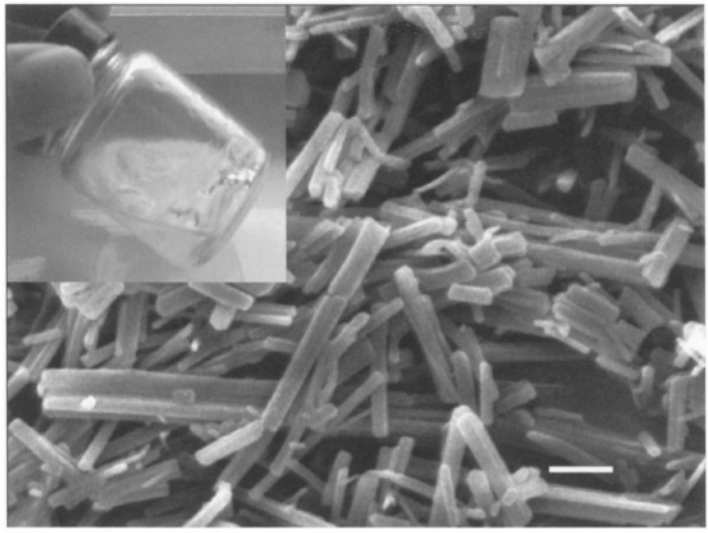
The addition of a tiny amount of EVACP (0.004 wt%) is sufficient to create completely different micro/nano networks of interconnecting self-organized L-DHL nanocrystal fibers (Figure 5.32b) at the experimental temperature. This addition completely changes the rheological properties of the materials, and leads to the formation of a self-supporting soft solid-like supramolecular material (see the upper corner of Figure 5.32b).

Figure 5.33a shows the change of G^* as a function of time t for the L-DHL/DIOP system. G^* increases abruptly at $t \geqslant t_g$ (t_g: gelation time, cf. Figure 5.33a), indicating the formation of L-DHL crystalline fibers. G^* approaches gradually its maximum value G^*_{max} as the formation of the fiber networks approaches completion at $t \to \infty$. In comparison with the controlled sample, G^*_{max} is almost doubled when a tiny amount of EVACP is added. (Similar changes were also obtained for G' and G'', respectively.)

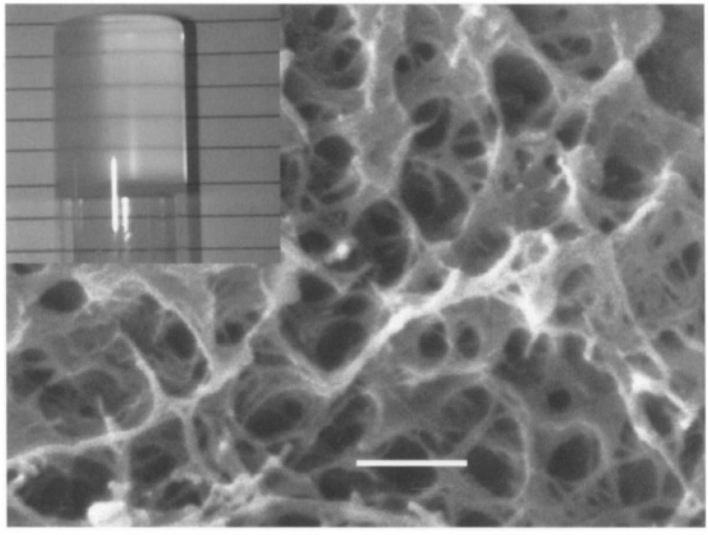
Figure 5.33b shows the change of the storage modulus G' of the above system as a function of various oscillating strain amplitudes γ with angular frequency $\omega = 0.2\pi$ rad/s. The strain corresponds to the deformation of the networks caused by the applied shear stress. G' remains constant (linear) at small strains and decreases abruptly when γ exceeds a certain value γ_0. The onset of the non linearity (decrease in G') at γ_0 corresponds to the breakage of the junctions in the networks [106]. The result given in Figure 5.33b indicates that the addition of EVACP leads to a significant enhancement of the limit of linearity γ_0 (~100 times cf. Figure 5.33b).

Figures 5.33a and 5.33b show that a new macromolecular material with significantly modified macroscopic properties is formed after the addition of EVACP. To the best of our knowledge, such a novel approach of creating new supramolecular materials by an additive has never been reported previously.

5.4.4.2. Mechanism of 3D interlinking network formation An adequate understanding fn the cross-linking of fiber networks is required to explore the network promotion mechanism of EVACP. The key issue to be addressed here is whether or not the formation of 3D interconnecting networks of small molecular organic gelling agents, such as L-DHL, is controlled by the self-assembly of fibers as suggested by most workers in this field [18,19]. It is extremely unlikely that the occurrence of a trace amount of EVACP would cause the



(a)



(b)

FIGURE 5.32. (a) Separate fibers occurring in the 10 wt% L-DHL/DIOP system. This system gives rise to an opaque gel as shown in the picture on the right upper corner. The length of the bar: 1 μm. (b) Interconnected fiber networks in 10 wt% L-DHL/DIOP system after adding 0.004 wt% EVACP. The length of the bar: 1 μm. This system gives rise to a clear and tough gel as shown in the picture on the right upper corner. The materials under study were obtained from L-DHL (lanosta-8,24-dien-3β-ol:24,25-dihydrolanosterol = 56:44, from Sigma) in DIOP (di-(2-ethyl-hexyl)phthalate $(C_8H_{17}COO)_2C_6H_4$ liquid) (99%, Cognis) at ~120°C, forming an L-DHL/DIOP solution, and lowering the sample temperature to around room temperature, the so-called experimental temperature [73–75]. In order to examine the nano/micro structure of fibrous networks in supramolecular materials, Scanning Electronic Microscopy (SEM), coupled with a CO_2 super critical fluid extraction technique (Thar Design), was applied. The purpose of implementing the latter is to remove the liquid captured in the networks, without disturbing the essential structure of networks [73–75]. Scale bar: 1 μm. Reproduced with permission from *J. Am. Chem. Soc.* **124**, 15055–15063 (2002). Copyright 2002 Am. Chem. Soc.

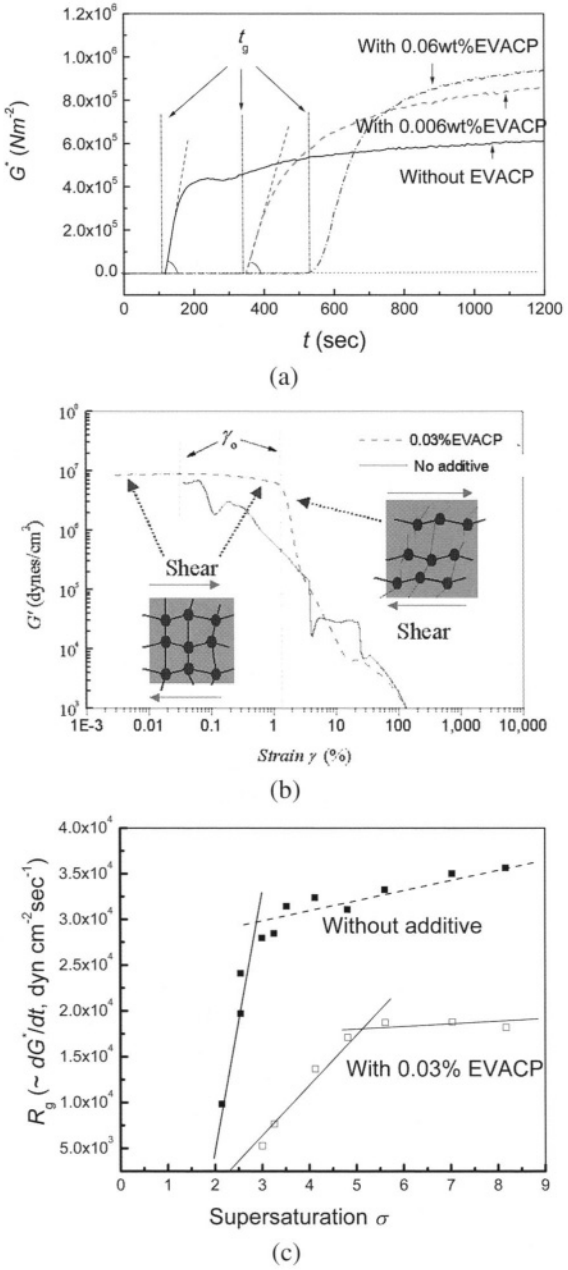


FIGURE 5.33. The changes of the rheological properties of the 10 wt% L-DHL/DIOP system in the absence and in the presence of EVACP as functions of time or strain. The rheological properties of the above system were measured by an Advanced Rheological Expansion System [75]. (a) Dependence of the complex modulus G^* on time. The maximal G^* is almost doubled after adding EVACP. (b) Dependence of the storage modulus G^* on strain. The linearity limit γ_0 is enhanced almost 100 times after adding EVACP. (c) R_g ($\sim dG^*/dt$) vs. supersaturation σ for the system without and with EVACP. Reproduced with permission from *J. Am. Chem. Soc.* **124**, 15055–15063 (2002). Copyright 2002 Am. Chem. Soc.

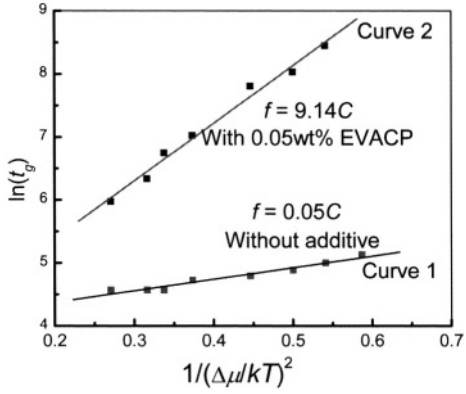


FIGURE 5.34. $\ln t_g \sim 1/(\Delta\mu/kT)^2$ plots for the system without and with EVACP (0.03 wt%). Both conditions give rise to a linear relation, indicating the occurrence of a nucleation-control mechanism in fiber formation [46, 47]. The slope of the straight lines represents the specific f [46,47]. After introducing EVACP, f is raised by a factor of more than 183, confirming the significant mismatch between the nucleating fibers and the substrate caused by the additive. Reproduced with permission from *J. Am. Chem. Soc.* **124**, 15055–15063 (2002). Copyright 2002 Am. Chem. Soc.

lash-shape of L-DHL crystal needles shown in Figure 5.32a to become assembled into the self-organized, and regularly branched nanofiber networks shown in Figure 5.32b.

As indicated in Figure 5.31, the X-ray diffraction (XRD) analyses show that L-DHL powders, needles and 3D interconnecting fibers have identical crystalline structures. This implies that the formation of L-DHL needles and 3D interconnecting fibers is essentially controlled by a new type of kinetics associated with the nucleation-growth process. As far as the primary fiber formation is concerned, it should be controlled, like most crystalline materials, by a nucleation process, as can be verified by the linear relationship between $\ln t_g$ and $1/(\Delta\mu/kT)^2$ [46–48,51].

According to the 3D nucleation model, we should then have a linear relationship between $\ln(t_g)$ and $1/(\Delta\mu/kT)^2$ (cf. eq. (62) for the nucleation associated with a given f (including crystallographic mismatch nucleation). In Figure 5.34, the linear fits between $\ln(t_g)$ and $1/(\Delta\mu/kT)^2$ obtained for both systems in the absence and in the presence of EVACP indeed confirm the nucleation control in the early stage of fiber network formation.

Referring to Figure 5.30b, it can be seen that one of the key steps in building up the self-organized interconnecting fiber networks is via a crystallographic mismatch branching process at the growing tips of the nanofibers (cf. Figure 5.19a). The occurrence of mismatch nucleation, leading to the occurrence of new fiber branches (cf. Figure 5.18a), depends on the structural match between the substrate and the nucleating phase, on the concentration of additives (impurities), and on the supersaturation (cf. Figures 5.18 and 5.19) [74,75]. For the growth of L-DHL needles, the deposition of new layers on the existing surface of the growing tips requires a perfect structural match, which occurs at very low surface supersaturations [74,75] (Figure 5.19a).

The occurrence of CMB after introducing EVACP is actually attributed to an impurity-induced structural mismatch nucleation (cf. Figures 5.18a and 5.19a). Since EVACP is an agent that will selectively adsorb on certain surfaces of organic crystals [105], the adsorbed molecules will disrupt the structural match between the new layers and the tip surface of

the parent fibers. Such an adsorption as well as the hindering of the normal growth of the fibers can be examined from the slope of $G^* \sim t$ $(= \mathrm{d}G'/\mathrm{d}t)$ at $t \to t_g$ $(t > t_g)$.

The slope of $G^* \sim t$ $(= \mathrm{d}G'/\mathrm{d}t)$ at $t \to t_g$ $(t > t_g)$ (cf. Figure 5.33a) represents the growth rate of the L-DHL fibers. According to Einstein's relation [21], the specific viscosity η_{sp} can be related to the volume fraction of the L-DHL needles as $\eta_{sp} \approx F\phi$ with $\eta_{sp} = \eta^*/\eta_0 - 1$, *(F:* shape factor of the particle; η_0: viscosity of the solvent; the complex viscosity is $\eta^* = |G^*|/\omega)$ [21]. The growth of L-DHL fibers can be regarded as a lengthening of the fibers in their axial directions since the cross sectional area remains almost constant. It follows that

$$\frac{\mathrm{d}\eta_{sp}}{\mathrm{d}t} = \frac{1}{\eta_0}\frac{\mathrm{d}\eta^*}{\mathrm{d}t} = \frac{1}{\eta_0\omega}\frac{\mathrm{d}G^*}{\mathrm{d}t}, \qquad \text{and}$$

$$\frac{\mathrm{d}\eta_{sp}}{\mathrm{d}t} = \frac{F}{\eta_0}\frac{\mathrm{d}\varphi}{\mathrm{d}t} = \frac{F}{\eta_0}\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{N_{fib}v_{fib}}{V}\right) \approx \frac{FN_{fib}}{\eta_0 V}\frac{\mathrm{d}v_{fib}}{\mathrm{d}t} = \frac{FN_{fib}\phi}{\eta_0 V}\frac{\mathrm{d}L}{\mathrm{d}t},$$

where V is the total volume of the crystalline material at $t \to \infty$, N_{fib} is the number of fibers, v_{fib} is the average volume of the fibers and L denotes the length of the fibers. Combining the above two equations yields $\mathrm{d}G^*/\mathrm{d}t \propto \mathrm{d}L/\mathrm{d}t = R_g$ (R_g: the growth rate of fiber tips). This means that the slope of the straight lines in Figure 5.33a reflects the growth rate of the fibers.

Figure 5.33c shows the growth rate R_g of the L-DHL fiber tips as a function of supersaturation σ (10 wt% L-DHL in DIOP solutions without additive and with 0.05 wt% EVACP). The results can be summarized as follows:

(1) The growth rate of the L-DHL fibers increases drastically with the supersaturation at the beginning, then gradually levels off as supersaturation increases further.
(2) The introduction of EVACP causes a reduction in the growth rate of L-DHL fibers (only ½ in comparison with that in the solution without EVACP). Since EVACP is added by only a tiny amount, the mutual diffusivity may not change much. Therefore, the reduction in the growth rate is due to the adsorption of the fibers on the growth tips, slowing down the surface kinetics of the tip growth [51,61–64].

The structure mismatch caused by the adsorption of EVACP can be verified by the enhancement of f in heterogeneous nucleation experiments. The structural match between the newly created crystal layers and the tip surface can also be described by f (and f'). As already mentioned, the slope of $\ln(t_g)$ versus $1/(\Delta\mu/kT)^2$ plot corresponds to the specific f [46,47], In the case of a perfect structural match (such as crystal growth), one has f, $f' = 0$, while for a complete structural mismatch, one has f, $f' = 1$ [46,47]. If the adsorption of EVACP on any substrate (including the tips of the growing fibers) is capable of disrupting the structural match, one should observe a significant enhancement of f in the nucleation experiments [46,47]. Figure 5.34 shows that f indeed is raised by a factor of more than 183 after introducing EVACP.

Both the nucleation and the growth kinetics of the L-DHL fibers in DIOP confirm that the significant structural mismatch leading to CMB is caused by the adsorption of EVACP on the growth tips of the L-DHL fibers. This slows down the surface integration during the growth of the fibers, and causes interfacial structural mismatch nucleation and growth. Consequently, the (wide angle) crystallographic mismatch branching is promoted (Figure 5.35).

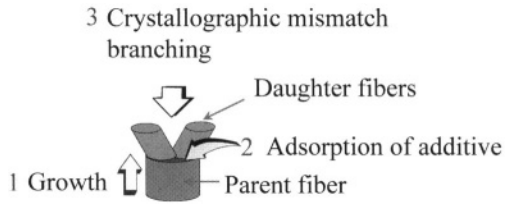


FIGURE 5.35. The adsorption of EVACP causes mismatch nucleation at the tip of the growing fiber, leading to crystallographic mismatch branching. The formation of interconnecting fiber networks via the process of primary nucleation–growth–branching–growth–branching–...

The average length $\langle \xi \rangle$ of the branches, which is proportional to the average mesh size in the fiber networks, was examined by SEM. The linear relation between $\ln[\langle \xi \rangle (\Delta\mu/kT)^{-1}]$ and $1/(\Delta\mu/kT)^2$ obtained for the aforementioned L-DHL/DIOP system with 0.03% EVACP (see Figure 5.25) confirms the above branching mechanism described by eq. (73).

The promotion of branching brought about by the addition of EVACP can be quantitatively understood as follows. According to eq. (73), a perfect structural match between the parent fibers and the newly created crystal layers ($m \sim 1$, $f \sim 0$) will give rise to an infinite $\langle \xi \rangle$, meaning that fiber growth will occur without any branching at all. To obtain a finite branching distance, a high degree of structural mismatch (a large non-zero f) is required. As shown by Figure 5.34, the addition of EVACP results in a drastic increase in f by almost a factor 183. This is the reason why CMB is promoted upon the addition of EVACP.

The addition of EVACP leads to much thinner L-DHL fibers (cf. Figures 5.32a and 5.32b). The thinning of the L-DHL fibers means that the ratio of the growth rate of the fibers to the growth rate of the tips increases when EVACP is added. This implies that EVACP will not only adsorb on the tips but also on the prism faces of the L-DHL fibers [51,74,75]. Note that CMB can also occur on the prism faces of the L-DHL fibers. Nevertheless, if the tip branch condition can be fulfilled, CMB will occur primarily at the tips of the fibers, since the tips point in the direction of rapid growth, leading to growth out of highly supersaturated bulk solutions [51,74,75].

5.4.4.3. Engineering of supramolecular functional materials As illustrated by Figures 5.30a and 5.30b, one aspect of the engineering of supramolecular functional materials is the architecture of self-organized 3D interconnecting nanofiber structures (Figure 5.30a); another aspect is to tune the micro/nano structure in a predictable way (Figure 5.30b). It has been shown in the previous sections that the variations in the micro/nano structure of functional materials will exert a direct impact on the rheological or other physical properties. This implies that it is possible to control the mesh size and the corresponding fiber network structure so as to obtain functional materials with certain desired micro/nano structure and rheological properties. In other words, the materials with the required properties can be obtained by altering the experimental conditions.

Figure 5.25b shows the dependence of $\langle \xi \rangle$ on the supersaturation, meaning that the micro/nano structure can be modified by changing the supersaturation according to eq. (73). A decrease in the mesh size of the 3D interconnecting networks of the materials will lead to an increase in G^*, G', etc. [15]. Therefore, changing the supersaturation should enable

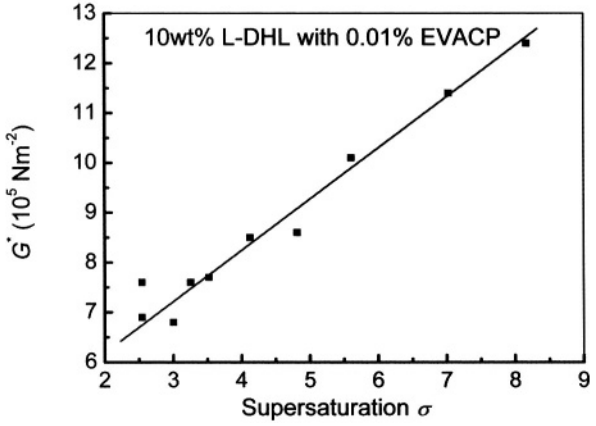


FIGURE 5.36. Dependence of G^* on supersaturation σ for the 10 wt% L-DHL in DIOP with 0.01% EVACP. Reproduced with permission from *J. Am. Chem. Soc.* **124**, 15055-15063 (2002). Copyright 2002 Am. Chem. Soc.

us to alter the rheological or other physical properties of these materials. Figure 5.36 shows the dependence of G^* on supersaturation for the system of 10 wt% L-DHL in DIOP with 0.01% EVACP. As expected, G^* increases with supersaturation (see Figure 5.36).

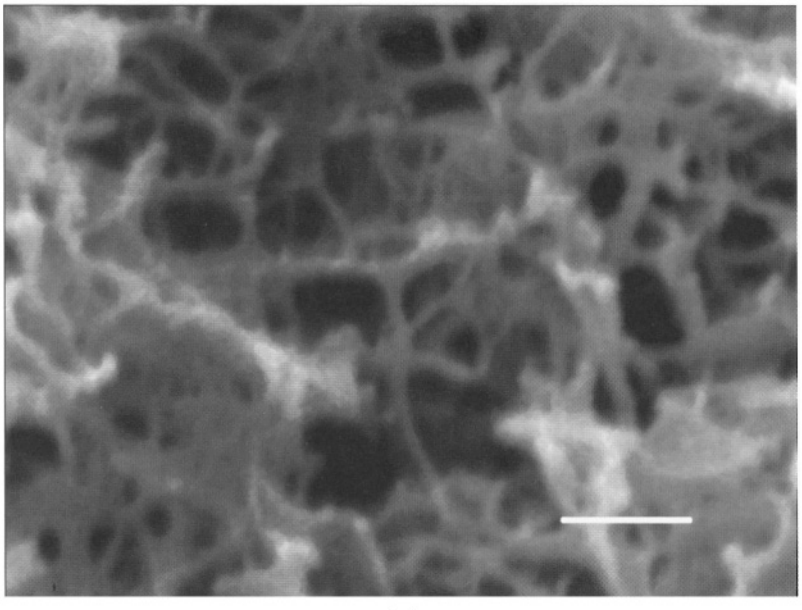
Under the given conditions, an increase of the concentration of the branching promoter (additive) will also cause a reduction of the mesh size of 3D interconnecting networks. Figures 5.37a and 5.37b show the 3D fiber network structures of L-DHL in 10 wt% L-DHL/DIOP gels with 0.01 wt% and 0.1 wt% EVACP, respectively. Raising the EVACP concentration C_{EVACP} causes a reduction of the mesh size of the networks. This in turn causes G^*, G', etc. (cf. Figure 5.35c) to increase.

Since the network can be created and controlled based on the above approaches, it should be possible to engineer directly the microstructure and macroscopic properties of the above materials in a controlled manner.

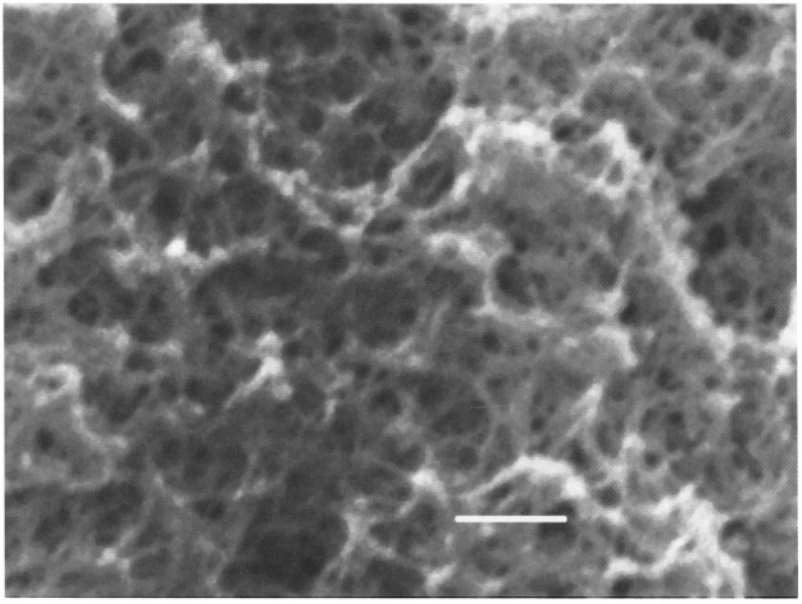
Note that to engineer supramolecular functional materials by additives, identifying an effective branching creator is one of the key steps [75]. The criteria for selecting and designing such molecules should take energetic and entropic factors into account. In order to promote the branching of the fibers, the branching promoters (or additives) should be adsorbed effectively on the growing tips in order to disrupt and hinder the normal growth of the fibers in the axial orientations. One of the key questions to be addressed is how to design or select branching promoters. In the following discussion, we will provide such guidelines from the molecular structural point of view, based on the results obtained from consistent-field theory calculations [60,63,65,66]. The branching promoter molecules should possess the following characteristics:

(1) Large molecules with relatively rigid basic structures. For molecules of similar types, larger molecules with somewhat rigid structures are easily adsorbed at interfaces [65]. This criterion is based on considerations of both energy and entropy.

Energy consideration. Given that there is a strong interaction between each monomer and the crystal surface, polymers or macromolecules composed of similar structural units have on the whole more interacting points with the crystal sur-



(a)



(b)

FIGURE 5.37. Effects of EVACP concentration on the micro/nano fiber structure and rheological properties of L-DHL. The increase of EVACP concentration C_{EVACP} gives rise to the reduction of the mesh size of the L-DHL network (cf. (a) and (b)) and the rise of G^*. (a) Fiber network of L-DHL obtained from 10 wt% L-DHL in DIOP with 0.01 wt% EVACP. Scale bar: 1 μm. (b) Fiber network of L-DHL obtained from 10 wt% L-DHL in DIOP with 0.1 wt% EVACP. Scale bar: 1 μm. (c) Dependence of G^* on C_{EVACP} (10 wt% L-DHL in DIOP). Reproduced with permission from *J. Am. Chem. Soc.* **124,** 15055–15063 (2002). Copyright 2002 Am. Chem. Soc.

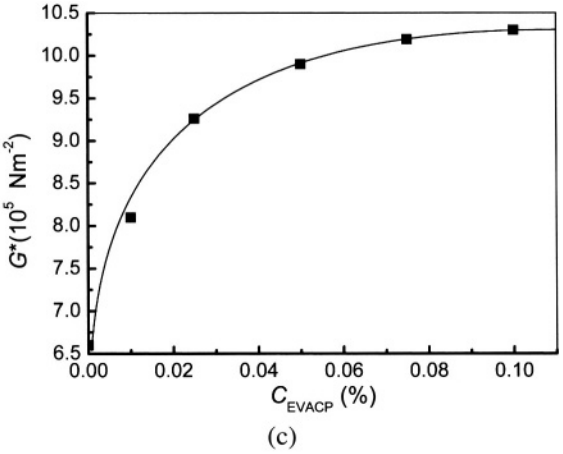


FIGURE 5.37. (Continued).

face than single monomers. Therefore, the desorption of polymers is much more difficult.

Entropy consideration. Our calculations [60,63,65,66] show that long and relatively rigid polymers have a much higher surface activity due to the entropy effect. Two contradictory entropic effects play a role in the adsorption of polymers or macromolecules. The first effect is the so-called *free particle entropic effect.* Normally, there is a solvation layer around larger molecules or polymers and the crystal surface (see Figure 5.38a), where solvent molecules are partially localized or bonded to the surface. If these additive molecules adsorb onto the crystal surface, these semi bonded solvent molecules will be liberated from the surfaces of the additive molecules and from the crystal surface (cf. Figure 5.38b), resulting in increased entropy. According to the definition of entropy [49], the increase of the free particle entropy (per molecule) is proportional to $k\ln(2N_{\text{surf}})$ (*k:* Boltzmann const; N_{surf}: the number of monomers or atoms on the surface of the macromolecule). Therefore, larger or longer molecules, such as polymers or macromolecules cause a large entropy increase when they become adsorbed onto the surface. This is why most effective tailor-made additives are either long polymers or macromolecules.

On the other hand, polymers or macromolecules have a very high conformational entropy, depending on the number of atoms or monomers included in the polymer or the macromolecule, and on the degree of freedom of the internal motion. This effect is known as the conformational entropic effect [49]. Molecules with the maximal degree of freedom of the internal motion (such as polymers with completely flexible chains), have a conformational entropy proportional to $k\ln(3N)$, whereas molecules with no freedom in the internal motion (such as polymers with completely rigid chains) have a conformational entropy of $k\ln 6$ ($k\ln 5$ for linear molecules) [49]. When additive molecules are adsorbed onto the crystal surface, the internal motion of the molecules becomes frozen to a large extent. This implies that when polymers or macromolecules are adsorbed onto the crystal surface, they lose an amount of conformational entropy from $k \ln 5$ (completely

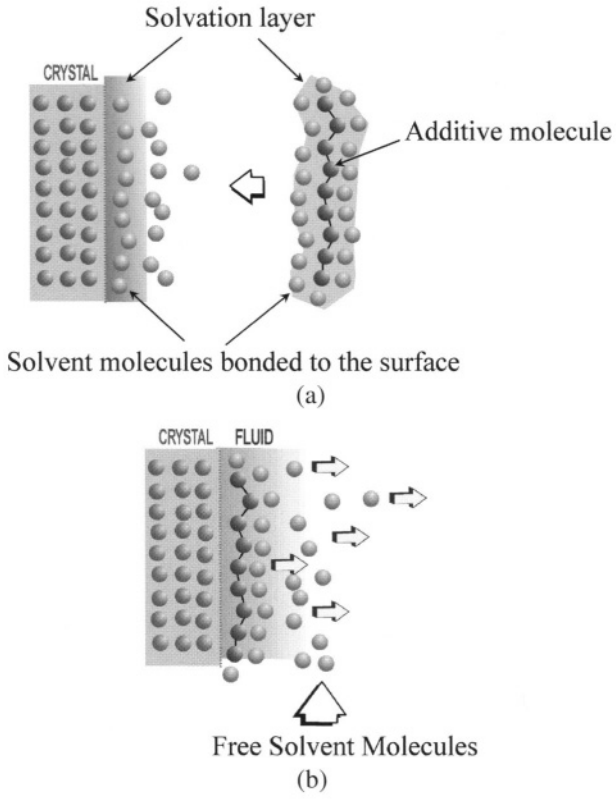


FIGURE 5.38. Illustration of the adsorption of additive molecules onto the crystal surface. (a) Solvent molecules are partially bonded to the surfaces of the additive molecule and of the crystal. (b) The adsorption of the additive molecule to the surface of the crystal will release semi-bonded solvent molecules, leading to an increase in the entropy. Reproduced with permission from *J. Am. Chem. Soc.* **124,** 15055–15063 (2002). Copyright 2002 Am. Chem. Soc.

rigid molecules) up to $k\ln(3N)$ (completely flexible molecules), depending on the rigidity of the molecules. Therefore, conformational entropy is opposed to the free particle entropy.

Combining these two contrasting effects together, one has for linear or flat molecules ($N_{\text{surf}} = N$) an increase in the entropy (per molecule) due to adsorption equalling $\sim k\ln(2/3)$ for completely flexible molecules, and $k\ln(2N/5)$ for completely rigid linear molecules. In other words, the adsorption of completely flexible molecules on the crystal surface leads to a loss of entropy due to the restriction of the internal molecular motion. This will consequently give rise to the depletion of the molecules on the surface.

On the other hand, the adsorption of completely rigid molecules will gain entropy since the loss of chain conformational entropy is very limited in this process. This will consequently result in a strong adsorption of the molecules. This trend is further enhanced as the number of monomers or atoms in the additive molecule increases. Normally, a semi-rigid chain will behave between the above two extremes. These results have been confirmed by our SCF calculations [60,63,65,66].

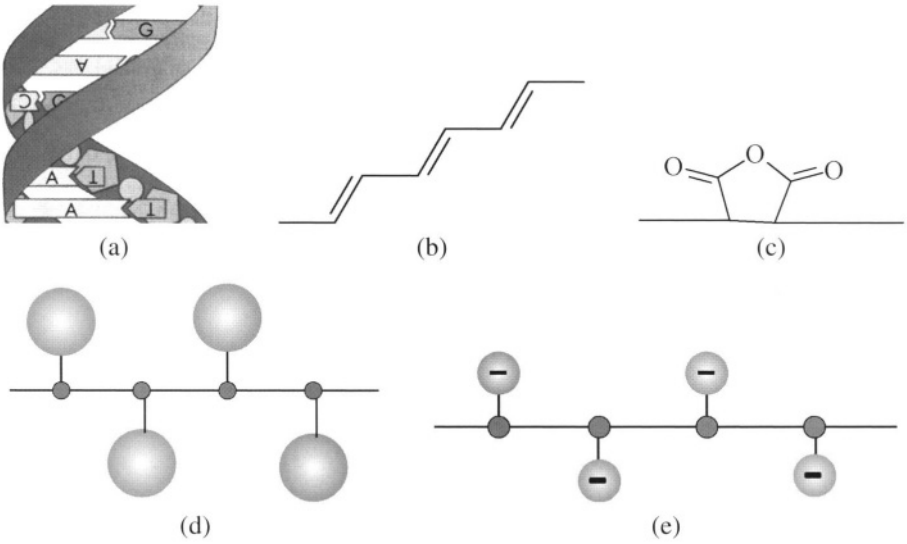


FIGURE 5.39. Rigidity of molecules due to (a) intra molecule hydrogen bond, (b) double covalent bond, (c) rigid ring structure, (d) neighboring bulky branch (hindered rotation), and (e) electrostatic repulsion (hindered rotation). Reproduced with permission from *J. Am. Chem. Soc.* **124,** 15055–15063 (2002). Copyright 2002 Am. Chem. Soc.

Note that the length of the polymers cannot be extended without limitation. Molecules with too long chains tend to fold and lose much chain conformational entropy in the adsorption. Normally, the relative rigidity of a molecule decreases with its size. For a given type of relatively rigid molecules, the increase of the entropy due to adsorption will reach its maximum at a certain N.

The rigidity of the additive molecules can be the result of a variety of reasons. In the following, we will summarize some typical examples of rigid molecules.

(i) *Intra molecular bonding.* Molecules will become relatively rigid if some intra molecular bonds, such as hydrogen bonds (Figure 5.39a), double or triple covalent bonds (Figure 5.39b) and rigid rings (Figure 5.39a), etc. occur in the backbone of additive molecules.

(ii) *Hindered rotation.* The folding of chain-like molecules is in many cases attributed to the rotation around single carbon-carbon bonds. Bulky functional groups attached to the neighboring units in polymers or macromolecules will cause them to avoid each other due to the steric repulsion (Figure 5.39d). In the case of a polyelectrolyte, groups with like charges attached to the backbone of additive molecules will avoid each other due to electrostatic repulsion (Figure 5.39e). These factors hinder chain rotation, leading to stiffer or more rigid chains.

Note that the additives (branching promoters) need not be linear polymers. Any type or shape of large molecule or macromolecule is suitable as long as the above criteria are satisfied.

(2) A stronger interaction between the additives and the crystal surface will lead to a stronger adsorption at the surface [60,63,65,66], Since a crystal surface is highly ordered and stiff, in order to maximize the interaction by matching the structure

of the crystal surface, it is desirable to have short and relatively flexible functional groups attached to the backbone of additive molecules, so that they can adjust their positions and thus optimize the interaction with the solid molecules at the crystal surface. Normally, some functional groups or branches attached to the backbones of additives are designed to have a structure similar to building units of the crystal or to interact strongly with the solid structure units at the crystal surface. In this way the interaction between the additives and the crystal surface can be enhanced.

(3) The adsorbed additives should interrupt the growth of the crystal layers along the crystal surface [60,63,65,66]. The repulsion with the growth units can originate from steric, electrostatic, polar/non-polar, or hydrophilic/hydrophobic forces, and can be achieved by attaching some functional groups to the backbone of the additives.

(4) Also the concentration of the additives plays a certain role. For a given system, the excess in the concentration may not enhance the adsorption, it sometimes can even cause a decrease in the adsorption [60,63,65,66]. This is attributed to the adsorption synergism by mixing additives with other structural units. Note that all of the above mentioned factors are correlated to each other. Therefore, specific cases should be considered from a number of different viewpoints.

The molecular structure of EVACP is as follows:

$$-(CH_2-CH_2)_x-(CH_2-CH(-O-C(=O)-CH_3))_y-$$

Obviously, this structure satisfies all of the structural requirements for branching promoters: $(C_4H_6O_2)_x$ in the molecules fulfill requirements (1) and (2) (cf. Figure 5.39d), while $(C_2H_4)_y$ satisfies requirement (3). For this reason EVACP serves as an effective branching promoter.

The molecular structure of polyvinyl chloride (PVC, MW: 100,000, from BDH, UK) is as follows:

$$-(CH_2-CHCl)_n-$$

Although PVC is a polymer with the polyethylene backbone (similar to EVACP), the molecules do not satisfy the rigidity and remaining criteria. Therefore, PVC is not effective in promoting the branching in L-DHL, and therefore it cannot serve as a branching promoter. This is verified by our experiments.

5.5. SUMMARY AND CONCLUSIONS

In this chapter, new kinetic models of micro/nano structure formation are presented within the framework of heterogeneous nucleation and the structure of the solid–fluid interface. The first part of this chapter deals with a generic mechanism, i.e. the mechanism of heterogeneous nucleation, based on thermodynamics and the relevant kinetic theories. Both the classic and the non-classical interfacial effects and the particle size effect are treated quantitatively in terms of the so-called interface correlation factor $f(m, R')$ and the relevant kink integration kinetics. The interface correlation factor $f(m, R')$ is a function of

the relative radius of curvature of the substrate R', and the interfacial match parameter m. Within a range of increasing values of the supersaturation, heterogeneous nucleation turns out to be a sequence of progressive processes, rather than just a single process, revealing a wide spectrum of heterogeneous characteristics described by $f(m, R')$ and $f''(m, R')$. At low supersaturations, nucleation is governed by a process dominated by an optimal structural match and a strong interaction between the substrate and the nucleating phase ($m \to 1$, $f(m, R') \to 0$); as the supersaturation increases, nucleation is controlled by a process dominated by a poor structural match ($m \to -1$, $f(m, R') \to 0$). Homogeneous nucleation is the upper limit of this spectrum. Due to the anisotropy of the crystalline phases, $f(m, R')$ assumes only discrete values that do not follow the continuous variation of the supersaturation.

When the radius of curvatrure of the subtrate is comparable with that of the critical nuclei, the nucleation kinetics will be affected directly by the curvatrure of the subtrate as the supetrsaturation changes. This curavture effect of the subtarte will vanish if the radius of curvature of the substrate is too large compared with that of the critical nuclei. On the other hand, the effect of the substrate will completely vanish if the radius of curvature of the substrate is much smaller than that of the critical nuclei.

The influence of impurities on nucleation can be summarized as follows: (i) the adsorption of impurity molecules at the interfaces of solid–liquid, solid–air, liquid–liquid, or liquid–air will modify the interfacial structural interaction/match between the crystallisation phase and the subtrate, and therefore it will change $f(m, R')$ and the nuclceation barrier (such as the epitaxial or counter-epitaxial effects); (ii) the adsorption of impurity molecules on the surfaces of the nuclei will reduce the solid–liquid interfacial energy, and therefore it will lower the nucleation barrier and enhance the nucleation rate (cf. eqs (34) and (52)); (iii) the adsorption of impurity molecules on the surface of the nuclei will block the kink integration of the growth units. The latter will directly enhance the kinetic energy barrier of the the kink (cf. eq. (66) and Figure 5.11), and consequently it will supress the nucleation rate. Conversely, if some specific additives (impurities) or particular substrates should induce a favored pre-ordering of the fluid molecules and facilitate the kink kinetics at the interface, nucleation will be promoted. Normally, the above controdictory effects (i.e. effects (ii) and (iii)) will co-exist. Nevertheless, in the case of nanophase nucleation, effect (iii) (a non-classical effect) may take place without coupling with effect (ii) (the classic effect), and therefore this occurrence could be very relevant for nanophase engineering.

The *supersaturation driven interfacial structural mismatch* is a new effect identified on the basis of the above model. With increasing supersaturation, the nucleation kinetics will induce random orientations of the growth fronts, unrelated to the orientation of the optimal structural match between the nucleating phase and the substrate. This implies that the epitaxial relationship will break down at high supersaturations.

If the surface of a growing crystal serves as the substrate for the mismatch nucleation and growth of new crystals of the same structure, an important phenomenon arising from the supersaturation driven interfacial structural mismatch is *the so-called crystallographic mismatch nucleation and growth:* new crystalline layers occurring on the surface of the growing crystal can adopt a mismatch orientation with respect to the crystallographic orientation of the parent crystal. In the case of needle-like crystals, this leads to the so-called *crystallographic mismatch branching.* Then the branching controls the formation of a typical structure having a spherulitic pattern, and other typical crystalline patterns.

The crystal networks created via *crystallographic mismatch branching* have the characteristics of a Cayley tree structure (a type of fractal structure). The structure of crystal networks can be then correlated to the rheological properties in terms of the fractal dimension.

Some supramolecular materials consisting of self-organized interconnecting fiber networks have a wide range of applications due to their novel in-use properties. As these novel in-use properties, in particular rheological properties of supramolecular functional materials, are to a large extent determined by the micro/nano structure of their micro/nano networks, creating and manipulating the self-organized networks provides an important method for producing new functional materials. Since the architecture of networks can be achieved by the branching of fibrous crystals, the engineering of supramolecular functional materials can then be realized by controlled branching.

We presented a completely new approach of creating a new supramolecular functional material using L-DHL by promoting the controlled branching of L-DHL needle-like crystals by means of a so-called "branching creator", EVACP, to generate self-organized interconnecting fibrous networks. The branching promotion is due to the adsorption of the branching creator molecules at the growing tips of the L-DHL needle-like crystals. The adsorption disturbs the normal growth of the needle crystals, leading to the (wide angle) crystallographic mismatch branching. In this regard, the effective branching promoters should strongly adsorb onto the tips and disrupt the regular growth of the growing fibers. As indicated above, another effective way of manipulating the micro/nano structure can also be achieved by changing the supersaturation (*supersaturation driven mismatch nucleation*).

ACKNOWLEDGEMENTS

The author would like to thank Dr. C. Strom for her valuable suggestions and proof reading of the manuscript. This research is supported by the Science and Engineering Research Council, Singapore (Project No: 022 101 0036).

REFERENCES

1. J.W. Mullin, *Crystallization* (Butterworth-Heinemann, Oxfort, 1997).
2. A.P. Alivisatos, Semiconductor clusters, nanocrystals, and quantum dots, *Science* **271**, 933–937 (1996).
3. C.M. Lieber, One-dimensional nanostructures: chemistry, physics & applications, *Solid State Communications,* **107,** 607–616 (1998).
4. X.Y. Liu, K. Tsukamoto and M. Sorai, New kinetics of $CaCO_3$ nucleation and microgravity effect, *Langmuir* **16**, 5499–5502 (2000).
5. D.W. Oxtoby, Phase transitions: catching crystals at birth, *Nature (London)* **406,** 464–465 (2000).
6. S.T. Yau and P.G. Vekilov, Quasi-planar nucleus structure in apoferritin crystallization. *Nature (London)* **406,** 494–497 (2000).
7. F.M. Ross, R.M. Tromp and M.C. Reuter, Transition states between pyramids and domes during Ge/Si island growth. *Science* **286,** 1931–1934 (1999).
8. G. Medeiros-Ribeiro, A.M. Bratkovski, T.I. Kamins, D.A.A. Ohlberg and R.S. Williams, Shape transition of germanium nanocrystals on a silicon (001) surface from pyramids to domes, *Science* **279**, 353–355 (1998).
9. U. Gasser, E.R. Weeks, A. Schofield, P.N. Pusey and D.A. Weitz, Real-space imaging of nucleation and growth in colloidal crystallization, *Science* **292**, 258–262 (2001).

10. B. Daudin, Self-assembled quantum dots systems: the case of GaN, in: *Advances in Crystal Growth Research,* Eds K. Sato, K. Nakajima and Y. Furukawa (Elsevier Science B.V., Amsterdam, 2001) pp. 233–251.
11. A.R. Ten Cate, *Oral Histology: Development, Structure and Function* (The C.V. Misby Company, St. Louis, 1985).
12. X.Y. Liu and S.W. Lim, Templating and supersaturation driven anti-templating: principles of biominerals architecture, *J. Am. Chem. Soc.* **125**, 888–995 (2003).
13. X.Y. Liu, Effect of microgravity on Ca mineral crystallization and implications for osteoporosis in space, *Appl. Phys. Lett.* **79**, 3539–3542 (2001); *J. Chem. Phys.* **115**, 9970–9974 (2001).
14. H.C. Anderson, Vesicles associated with calcification in the matrix of epiphyseal cartilage, *J. Cell Biology* **41**,59 (1969).
15. (a) M.J. Mio and J.S. Moore, Supramolecular materials: folded polymers as building blocks for adaptive organic materials, *MRS Bulletin* **25**, 36–41 (2000); (b) D. Derossi, K. Kajiwara, Y. Osada and A. Yamauchi, *Polymer Gels: Fundamentals and Biomedical Applications* (Plenum Press, New York, 1991); (c) J.M. Guenet, *Thermoreversible Gelation of Polymers and Biopolymers* (Academic Press, London, 1992); (d) R.J.P. Corriu and D. Leclercq, Recent developments of molecular chemistry for sol-gel processes, *Angew. Chem.* **108**, 1524–1540 (1996); (e) M.I. Reetz, Entrapment of biocatalysts in hydrophobic sol-gel materials for use in organic chemistry, *Adv. Materials* **9,** 943–954 (1997).
16. (a) P. Terech and R.G. Weiss, Low molecular mass gelators of organic liquids and the properties of their gels, *Chem. Rev.* **97**, 3133–3159 (1997); (b) U. Beginn, S. Keinath and M. Möller, New carbohydrate amphiphiles, 2. Gel formation and gel morphologies, *Macromol. Chem. Phys.* **199**, 2379–2348 (1998).
17. L. Brunsveld, B.J.B. Folmer and E.W. Meijer, Supramolecular polymers, *MRS Bulletin* **25**, 49–53 (2000).
18. J.H. van Esch and B.L. Feringa, New functional materials based on self-assembling organogels: from serendipity towards design, *Angew. Chem. Int. Ed.* **39,** 2263–2266 (2000).
19. T. Oya, T. Enoki, A.U. Grosberg, S. Masamune, T. Sakiyma, Y. Tekeoka, K. Tanaka, G. Wang, Y. Tilmaz, M.S. Feid, R. Dasari and T. Tanaka, Reversible molecular adsorption based on multiple-point interaction by shrinkable gels, *Science* **286**, 1543–1545 (1999).
20. P. Terech, V. Rodrigez, J.D. Barnes and G.D. Mckenna, Organogels and aerogels of racemic and chiral 12-hydroxyoctadecanoic acid, *Langmuir* **10**, 3406 (1994).
21. P.C. Hiemenz and R. Rajagopalan, *Principles of Colloid and Surface Chemistry* (Marcel Dekker, New York, 1997) pp. 145–192.
22. J.W. Gibbs, *Collected Works. Vol. I: Thermodynamics* (Longmans and Green, New York, 1928).
23. M. Votmer and A. Weber, *Z. Phys. Chem.* **119**, 277–289 (1926).
24. L. Farkas, Z. *Phys. Chem.* **125**, 236–242 (1927).
25. R. Kaischew and I. Stranski, Z. *Phys. Chem.* **B26**, 317–326 (1934).
26. R. Becker and W. Doering, *Ann. Phys.* **24,** 719–731 (1935).
27. J.B. Zeldovich, *Acta Physicochim. USSR* **18**, 1–12 (1943).
28. J.P. Hirth and G.M. Pound, *Condensation and Evaporation* (Pergamon, Oxford, 1963).
29. A.E. Nielsen, *Kinetics of Precipitation* (Pergamon, Oxford, 1964).
30. F.F. Abraham, *Homogeneous Nucleation Theory* (Academic, New York, 1974).
31. B. Lewis and J.C. Andersen, *Nucleation and Growth of Thin Films* (Academic, New York, 1978).
32. O. Söhnel and J. Garside, *Precipitation, Basic Principles and Industrial Application* (Butterworth-Heinemann, Oxford, 1992).
33. A.C. Zettlemoyer, *Nucleation* (Dekker, New York, 1969).
34. S. Toschev, Homogeneous nucleation, in: *Crystal Growth: An introduction,* Ed. P. Hamm (North-Holland, Amsterdam, 1973) pp. I–49.
35. I. Gutzow, Induced crystallization of glass-forming systems: a case of transient heterogeneous nucleation. Part 1, *Contemp. Phys.* **21**, 121-243 (1980).
36. V.P. Skripov, Homogeneous nucleation in melts and amorphous films, in: *Current Topics in Materials Science* **2**, Ed. E. Kaldis (North-Holland, Amsterdam, 1976), pp. 327–378.
37. S. Kotake and I.L. Glass, *Proig. Aerospace Sci.* **19**, 129–141 (1981).
38. D. Kashchiev, *Cryst. Res. Technol.* **19**, 1413–1423 (1984).
39. I. Gutzow, D. Kashchiev and I. Avramov, Nucleation and crystallization in glass-forming melts: old problems and new questions, *J. Non-Cryst. Solids* **73**, 477 (1985).
40. K.F. Kelton, Crystal nucleation in liquid and glasses, *Solid State Phys.* **45**, 75–83 (1991).

41. X.C. Zeng and D.W. Oxtoby, Gas-liquid nucleation in Lennard-Jones fluids, *J. Chem. Phys.* **94**, 4472–4478 (1991).
42. A. Laaksonen, V. Talanquer and D.W. Oxtoby, Nucleation: measurements, theory, and atmospheric applications, *Annu. Rev. Phys, Chem.* **46**, 489–524 (1995).
43. C.L. Weakliem and H. Reiss, Toward a molecular theory of vapor phase nucleation. IV. Rate theory using the modified liquid drop model, *J. Chem. Phys.* **101**, 2398–2406 (1994).
44. R. McGraw and A. Laaksonen, Scaling properties of the critical nucleus in classical and molecular-based theories of vapor-liquid nucleation, *Phys. Rev. Lett.* **76**, 2754–2757 (1996).
45. N.H. Fletcher, Size effect in heterogeneous nucleation, *J. Chem. Phys.* **29**, 572–581 (1958).
46. (a) X.Y. Liu, A new kinetic model for 3D heterogeneous nucleation, compared with experiments, *J. Chem. Phys.* **111**, 1628–1635 (1999); (b) X.Y. Liu, Heterogeneous nucleation or homogeneous nucleation, *J. Chem. Phys.* **112**, 9949–9955 (2000).
47. X.Y. Liu, Generic mechanism of heterogeneous nucleation and molecular interfacial effects, in: *Advances in Crystal Growth Research,* Eds K. Sato, K. Nakajima and Y. Furukawa (Elsevier Science B.V., Amsterdam, 2001) pp. 42–61.
48. Du Ning and X.Y. Liu, Controlled ice nucleation in microsized water droplet, *Appl. Phys. Lett.* **81**, 445–447 (2002).
49. R. Fowler and E.A. Giggenhein, *Statistical Thermodynamics* (Cambridge University, London, 1960).
50. B. Mutaftschiev, Nucleation theory, in: *Handbook on Cryst. Growth,* Ed. D.T.J. Hurle (North-Holland, Amsterdam, 1993)pp. 189–245.
51. A.A. Chernov, *Modern Crystallography III—Crystal Growth* (Springer-Verlag, Berlin, 1984).
52. D. Kashchiev, Solution of the non-steady state problem in nucleation kinetics, *Surface Sci.* **14**, 209–220 (1969).
53. G. Shi, J.H. Seinfeld and K. Okuyarna, Transient kinetics of nucleation, *Phys. Rev.* **A41**, 2101–2108 (1990).
54. D.T. Wu, The time lag in nucleation theory, *J. Chem. Phys.* **97,** 2644–2650 (1992).
55. V.A. Shneidman and M.C. Weinberg, Induction time in transient nucleation theory, *J. Chem. Phys.* **97,** 3621–3629(1992).
56. D. Kashchiev, Nucleation, in: *Sci. and Tech. of Crystal Growth,* Eds J.P. van der Eerden and O.S.L. Bruinsma (Kluwer Academic Publ., Dordrecht, 1995) pp. 53–66.
57. M. Gebhardt, Epitaxy, in: *Crystal Growth: an Introduction,* Ed. P. Hartman (North-Holland, Amsterdam, 1972) pp. 105–142.
58. (a) X.Y. Liu, Interfacial process of nucleation and molecular nucleation templator, *Appl. Phys. Lett.* 79, 39–41 (2001); (b) X.Y. Liu, Interfacial effect of molecules on nucleation kinetics, *J. Phys. Chem.* **B105,** 11550–11558(2001).
59. R.J. Davey, J. Garside, A.M. Hilton, D. McEwan and J.W. Morrison, Emulsion solidification of metachloronitrobenzene: purification and crystallization, *J. Crystal Growth* **166**, 971–975 (1996).
60. X.Y. Liu and P. Bennema, Models for the interface between crystal and ambient phase; calculation of actual bond energies at the interface, implications for growth kinetics and morphology, *Current Topics in Crystal Growth Research* **2,** 451–522 (1995).
61. X.Y. Liu, E.S. Boek, W.J. Briels and P. Bennema, Prediction of growth morphology of crystals based on interfacial structure analysis, *Nature* **374**, 342–345 (1995).
62. X.Y. Liu and P. Bennema, The relation between macroscopic quantities and the solid–fluid interfacial structure, *J. Chem. Phys.* **98**, 5863–5872 (1993).
63. X.Y. Liu, Modeling for effects of fluid phase on growth morphology of crystals, *Phys. Rev.* **B60**, 2810–2817 (1999).
64. X.Y. Liu and P. Bennema, The relation between macroscopic quantities and the solid–fluid interfacial structure, *J. Chem. Phys.* **98**, 5863–5872 (1993).
65. (a) X.Y. Liu, P. Bennema, C. Meijer and M. Couto, Ordering of chain-like molecules at the solid–fluid interface, *Chem. Phys. Lett.* **220**, 53–58 (1994); (b) X.Y. Liu, Self-consistent field calculation of structure and static properties of the solid–fluid interface: monomer systems, *Phys. Rev.* **E49**, 583–590 (1994).
66. M. da Silva Couto, X.Y. Liu, H. Meekes and P. Bennema, Scanning tunneling microscopy studies on odd and even n-paraffin molecules adsorbed on graphite, *J. Appl. Phys.* 75, 627–633 (1994).
67. T. Vicsek, *Fractal Growth Phenomena,* 2nd edn (World Scientific, Singapore, 1992).
68. A.L. Barabasi and H.E. Stanley, *Fractal Concepts in Surface Growth* (Cambridge University Press, Cambridge, 1995).

69. J.S. Langer, Dendrites, viscous fingers, and the theory of pattern formation, *Science* **243,** 1150–1156 (1989).
70. E. Ben Jacob and P. Garik, The formation of pattern in non-equilibrium growth, *Nature (London)* **343**, 523–530(1990).
71. X.Y. Liu and C.S. Strom, Self-epitaxial nucleation origin of fractal aggregation, *J. Chem. Phys.* **112,** 4408–4411 (2000).
72. C.S. Strom, X.Y. Liu and Mu Wang, Fractal pattern formation in gel-grown ammonium chloride: morphology and solution-induced reconstructive epitaxial nucleation on pseudo-flat surfaces, *J. Phys. Chem.* **B104,** 9638–9646 (2000).
73. (a) X.Y. Liu and P.D. Sawant, Mechanism of the formation of self-organized micro-structure in functional materials, *Adv. Materials* **14**, 421–426 (2002); (b) X.Y. Liu and P.D. Sawant, Formation kinetics of fractal nanofiber networks in supramolecular materials, *Appl, Phys. Lett.* **79**, 3518–3520 (2001).
74. X.Y. Liu, and P.D. Sawant, Micro/nanoengineering of self-organized three-dimensional fibrous structure of functional materials, *Angew. Chemie Int. Ed.* **41**, 3641–3645 (2002).
75. X.Y. Liu, P.D. Sawant, W.B. Tan, I.B.M. Noor, C. Pramesti and B.H. Chen, Creating new supramolecular materials by architecture of three-dimensional nano crystal fiber networks, *J. Am. Chem. Soc.* **124**, 15055–15063(2002).
76. (a) X.Y. Liu and P. Bennema, On the morphology of normal alkane crystals with triclinic structure: theory and observation, *J. Crystal Growth* **135**, 209–223 (1994); (b) X.Y. Liu and P. Bennema, On the morphology of normal alkane crystals with monoclinic structure: theory and observation, *J. Appl. Crystallography* **26,** 229–243(1993).
77. S.C. Huang and M.E. Glicksman, Fundamentals of dendritic solidification. II. Development of sidebranch structure, *Acta Metall.* **29**, 717–734 (1981).
78. J.S. Langer, Instabilities and pattern formation in crystal growth, *Rev. Mud. Phys.* **52**, 1–28 (1980).
79. D.A. Kessler, J. Koplik and H. Levine, Pattern selection in fingered growth phenomena, *Adv. Phys.* **37**, 255–339(1988).
80. E.A. Brener and V.I. Melnikov, Pattern selection in two-dimensional dendritic growth, *Adv. Phys.* **40**, 53–91 (1991).
81. E. Ben-Jocob, From snowflake formation to growth of bacterial colonies. Part I. Diffusive patterning in azoic systems, *Contemp. Phys.* **34**, 247–273 (1993).
82. J. Kepler, De *Nive Sexangula* (1611).
83. A. Cayley, *Phil. Maig.* **28**, 374–383 (1958).
84. N. Vandewalle and M. Ausloos, Construction and properties of fractal trees with tunable dimension: the interplay of geometry and physics, *Phys. Rev.* **E55**, 94–98 (1997).
85. X.Y. Liu and P.D. Sawant, Determination of the fractal characteristic of nanofiber-network formation in supramolecular materials, *Chem. Phys. Chem.* **3**, 374–377 (2002).
86. P. Bennema and G.H. Gilmer, Theory of crystal growth, in: *Crystal Growth: an Introduction,* Ed. P. Hartman (North Holland, Amsterdam, 1973) pp. 263–326.
87. X.Y. Liu, New understandings for two-dimensional nucleation, *Surface Review and Letters* **8**, 423–428 (2001).
88. (a) X.Y. Liu, K. Maiwa and K. Tsukamoto, Two-dimensional heterogeneous nucleation and the growth kinetics, *J. Chem. Phys.* **106,** 1870 (1997); (b) X.Y. Liu, E.P.G. van den Berg, A.R.A. Zauner and P. Bennema, Can foreign particle cause surface instability, *J. Phys. Chem.* **B104,** 11942–11949 (2000).
89. J.W. Cahn, W.B. Hillig and G.W. Sears, The molecular mechanism of solidification, *Acta Met.* **12**, 1421–1439 (1964).
90. T.A. Witten and L.M. Sander, Diffusion–limited aggregation, a kinetic critical phenomenon, *Phys. Rev. Lett.* **47**, 1400–1403(1981).
91. P.C. Hiemenz and R. Rajagopalan, *Principles of Colloid and Surface Chemistry* (Marcel Dekker, New York, 1997) pp. 145–192.
92. T. Vicsek, *Fractal Growth Phenomena,* 2nd edn (World Scientific, Singapore, 1992).
93. L. Barabasi and H.E. Stanley, *Fractal Concepts in Surface Growth* (Cambridge University Press, Cambridge, 1995).
94. M. Avrami, Kinetics of phase change. I: General theory, *J. Chem. Phys.* **7,** 1103 (1939).
95. A. Einstein, *Ann. Phys.* **19,** 289 (1906).
96. N. Vandewalle and M. Ausloos, Construction and properties of fractal trees with tunable dimension: the interplay of geometry and physics, *Phys. Rev.* **E55**, 94–98 (1997).

97. S. Havlin, R. Nossal and B. Trus, Cluster growth model for treelike structures, *Phys. Rev.* **A32,** 3829–3831 (1985).
98. M. Muthukumar and H.H. Winter, Fractal dimension of a crosslinking polymer at the gel point, *Macromolecules* **19**, 1284–1285(1986).
99. Klaas te Nijenhuis and H.H. Winter, Mechanical properties at the gel point of a crystallizing poly(vinyl chloride) solution, *Macromolecules* **22**, 411–414 (1989).
100. U. Beginn, S. Keinath and M. Möller, New carbohydrate amphiphiles, 2. Gel formation and gel morphologies, *Macromol. Chem. Phys.* **199**, 2379–2384 (1998).
101. W. Gu, L. Lu, G.B. Chapman and R.G. Weiss, Polymerized gels and reverse aerogels from methyl methacrylate or styrene and tetraoctadecylammonium bromide as gelator, *Chem. Commun.,* 543–544 (1997).
102. R.J. Hafkamp, B.P.A. Kokke, I.M. Danke, H.P.M. Geurts, A.E. Rowan, M.C. Feiters and R.J.M. Nolte, Organogel formation and molecular imprinting by functionalized gluconamides and their metal complexes, *Chem. Commun.,* 545–546 (1997).
103. J.R. Ilzhoefer and R.J. Spontak, Effect of polymer composition on the morphology of self-assembled dibenzylidene sorbitol, *Langmuir* **11**, 3288–3291 (1995).
104. W.H. Shih, W.Y. Shih, S.I. Kim, J. Liu and L. Aksay, Scaling behavior of the elastic properties of colloidal gels, *Phys. Rev.* **A42**, 4772–4779 (1990).
105. P. Bennema, X.Y. Liu, K. Lewtas, R.D. Tack, J.J.M. Rijpkema and K.J. Roberts, Morphology of orthorhombic long chain normal alkanes: theory and observation, *J. Cryst. Growth* **121**, 679 (1992).
106. C. Shi, Z. Huang, S. Kilic, J. Xu, R.M. Enick, E.J. Beckman, A.J. Carr, R.E. Melendez and A.D. Hamilton, The gelation of CO_2: a sustainable route to the creation of microcellular materials, *Science* **286**, 1540–1543 (1999).

This page intentionally left blank

6

Two centuries of morphology of crystals: integration of principles of mathematical crystallography, statistical mechanics of surface models and chemistry

P. Bennema, H. Meekes*
NSRIM Department of Solid State Chemistry, University of Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands

6.1. INTRODUCTION

This contribution is intended as an introduction in the subject morphology of crystals. This subject is already more than two centuries old. It will be shown, however, that recently quite a few modern developments took place, which are both from a purely scientific point of view and also from the point of view of all kinds of technical application, very interesting and important. Concerning practical applications we refer to industries growing single crystals and the numerous industries producing large quantities of tiny crystals of micrometer size and less. We can think of industries working in the fields of bulk chemicals, industries producing pigments, pharmaceutical industries etc. For all these industries knowledge of the shapes (habits) of the produced crystals in dependence of the growth conditions determined by the driving force for crystallization, solvent(s) or impurities is of crucial importance. It will be shown, here, that the logical derivation of theories of morphology of crystals is based on a logical integration of results of five theoretical and experimental well-founded disciplines. These are:

- classical mathematical crystallography and morphology,
- crystal structures resulting from X-ray or synchrotron radiation,
- statistical mechanical theories of surface roughening transitions,

*Corresponding author. E-mail address: hugom@sci.kun.nl

- models for various crystal growth mechanisms,
- models for the kinetics of the incorporation of growth units at the interface.

Classical crystallography and especially morphology of crystals are of essential importance for studying crystal morphology. A thorough introduction to these subjects seems to disappear from university curricula. Therefore, in Section 6.2 basic principles of classical crystallography leading to the concept of 3 dimensional translational invariant space, reciprocal space, etc. will be introduced. Special attention will be focused on the concept interplanar distance d_{hkl} between adjacent lattice planes, since this concept plays a key role in morphology studies. Starting from a quite different modern statistical mechanical point of view of critical surface transitions, it will be shown in Section 6.3 that the concept of thermal surface roughening plays a key role in investigations of crystal growth and morphology. In Section 6.4 ways to recognize and calculate relevant energies of overall bonds between growth units will be introduced leading to the concept of crystal graph and connected net. In addition, results from computer simulations of crystal growth processes concerning roughening will be discussed. Finally, in Section 6.5 an example of very recent research of crystal growth computer simulations of the crystal growth process taking place at crystal surfaces of a real crystal structure is presented.

Thus, in contrast to other survey papers on crystal morphology, special attention is paid to the introduction of two quite different concepts, which play a key role in modern morphological studies, namely *the interplanar distance* in Section 6.2 and *the concept of roughening transition* in Section 6.3. These concepts will be introduced starting from first principles. The allocated space for this contribution is not sufficient to treat all subjects, which are essential to understand crystal morphology from first principles, in detail. Therefore, quite some references to recent survey papers are included which, in turn, contain many references to experimental and theoretical papers.

6.2. INTRODUCTION TO CLASSICAL CRYSTALLOGRAPHY. THE CONCEPT OF INTERPLANAR DISTANCE d_{hkl}

6.2.1. Classical crystallography; law of constancy of angles and of rational indices

In the following subsection we will give a very short schematized historical introduction to two centuries of crystallography and crystal morphology.

Already in the 17^{th} century the law of constancy of angles was discovered by Steno and in the 18^{th} century by Romé de Lisle (see refs [1-3]). This law implies that although shapes of crystals of a certain compound are unique, angles between corresponding faces are within an accuracy of minutes the same. This law was discovered to hold for large mineral crystals and angles were measured with a simple contact goniometer.

In the beginning of the 19^{th} century Haüy discovered the law of rational indices. This historical event marked the beginning of the last two centuries of modern crystallography. This law describes the orientatation of crystallographic faces in reference to a coordinate system of the given crystal under investigation. It can be formulated as follows: three unit axes *A, B, C,* parallel to three edges of the crystal are selected and a plane, occurring at this crystal is selected as a unit plane. This plane is cutting pieces with a ratio $|a| : |b| : |c|$

from the axes *A, B, C* respectively. Now the law of rational indices implies: all similar crystallographic faces of crystals of the same (mineral) crystal cut pieces

$$p|a| : q|b| : r|c| \tag{1}$$

from the axes *A, B, C,* respectively, where *p, q, r* are integers. These integers are in practice chosen prime in reference to each other.

For mathematical reasons, which will be clarified in the following, reciprocal indices h,k,l, where $h = N/p$, $k = N/q$, $l = N/r$ are used. The integer N, is chosen in such a way that h, k, l are also relatively prime. In crystallography triples h,k,l are used to characterize the orientation of a crystallographic face and are written as *(hkl).* Using reciprocal indices *(hkl),* called Miller indices, the orientation of a crystallographic face is now given by

$$|a|/h : |b|/k : |c|/l, \tag{2}$$

analogously to eq. (1). Note that in case a face is parallel to for example the A axis the index p becomes $p = \infty$ and the reciprocal index $h = 0$. So a face *(hkl)* = (100) corresponds to a face, which is parallel to the B and C axes; see further Figure 6.1.

It is interesting to note that almost all crystals of inorganic and organic chemicals of simple compounds up to complex pharmaceutical, protein compounds and even viruses, varying in size from tens of nanometers or fractions of microns to meters, grown in nature, laboratories or factories, are bounded by flat faces. The orientation of almost all crystallographic faces, occurring on a large variety of crystals are fulfilling the law of rational indices with a high degree of accuracy. This law offers a possibility to characterize crystal faces unambiguously.

In passing we note that instead of three integers, two ratios *h/l, k/l* or in general *two* rational numbers (and unity) could have been used to characterize a crystal face. This explains the historically developed name: *"Law of Rational Indices"*.

We also note that in order to characterize crystals of a crystalline product it is necessary to understand which faces *(hkl)* occur on these crystals and how the relative sizes of faces *(hkl)* vary with the growth conditions. These faces determine the overall habit of the crystals of a product and hence the properties of this product.

6.2.2. 3D translational symmetry; crystallographic faces are parallel to netplanes

In the following we will study in more detail the symmetry, which almost all crystals have in common namely: translational symmetry in 3 dimensional (3D) space. Therefore, we take as a starting point three mutually independent vectors $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ which span the infinite set $\boldsymbol{T}$ of vectors $\boldsymbol{t}$:

$$\boldsymbol{t} = u\boldsymbol{a} + v\boldsymbol{b} + w\boldsymbol{c}, \tag{3}$$

where (u, v, w) triples of integers; $u, v, w \in \boldsymbol{Z}$. A set of points represented by eq. (3) is called a translational lattice. The infinite lattice, obviously, is invariant for a shift over a any vector given by eq. (3). The idea of a 3D lattice is illustrated in Figure 6.1.

Note that instead of taking a lattice of discrete points, represented by eq. (3) the whole translational invariant 3D space can be considered as a continuum represented by a 3D

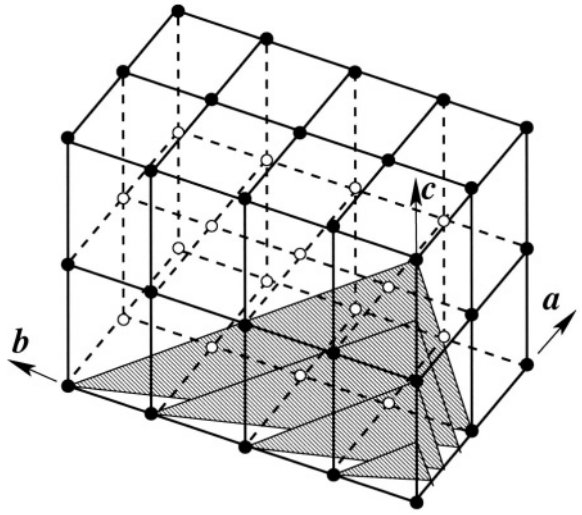


FIGURE 6.1. Small part of the lattice of a crystal showing $2 \times 4 \times 2$ elementary cells spanned on the basic vectors ***a, b, c.*** Each unit cell contains one atom or, in general, one growth unit. A few parallel lattice planes with the orientations *(hkl)* = (412) are drawn. It can be seen that the ***a*** axis is cut by adjacent lattice planes in $N|\boldsymbol{a}|/h$, the ***b*** axis in $N|\boldsymbol{b}|/k$ and the ***c*** axis in $N|\boldsymbol{c}|/l$, where N is integer. Note, that lattice planes need not intersect the crystallographic axes in lattice points.

function $F(x, y, z)$ with the same 3D translational symmetry, where x, y, z are fractional coordinates on the basis $\{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}\}$. Applying any translation out of the infinite set of translations $\boldsymbol{T}$ given by eq. (3) implies:

$$F(x\boldsymbol{a} + y\boldsymbol{b} + z\boldsymbol{c} + \boldsymbol{t}) = F(x\boldsymbol{a} + y\boldsymbol{b} + z\boldsymbol{c}). \tag{4}$$

This implies that after shifting the function of the right hand of eq. (4) over distances ***t*** the same function with the same environment is obtained.

In the following we will go back to the language of lattice points, in order to investigate planes going through a set of translational invariant lattice points. Looking at Figure 6.1, and assuming it to represent a small part of an infinite (or very large) lattice of a crystal, then a macroscopic crystal face with orientation *(hkl)* will be shown to be parallel to an infinite set of parallel lattice planes with an orientation *(hkl).* These lattice planes are sometimes called netplanes. As an example, according to eq. (2), for the face (412) the parallel lattice plane with the same Miller indices is cutting the axis ***a*** in $1/4\boldsymbol{a}$, the axis ***b*** in $1/1\boldsymbol{b}$ and the axis ***c*** in $1/2\boldsymbol{c}$. Equivalent lattice planes are found parallel to this lattice plane cutting the axes ***a, b, c*** at $N|\boldsymbol{a}|/4$, $N|\boldsymbol{b}|/1$, $N|\boldsymbol{c}|/2$, respectively, where N is any integer.

6.2.3. Reciprocal space

Before we will discuss the importance of the interplanar distance d_{hkl} for understanding the morphology of crystals, we will in the next sections introduce the mathematical concept of reciprocal space. This will allow us to derive a general expression for d_{hkl}.

To the set of the three microscopic vectors $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ we add a new set of vectors $\boldsymbol{a}^*$, $\boldsymbol{b}^*$, $\boldsymbol{c}^*$, spanning the unit cell of what will be called the reciprocal space, which is defined in the following way:

$$\boldsymbol{a}^* = \boldsymbol{b} \times \boldsymbol{c}/V, \qquad \boldsymbol{b}^* = \boldsymbol{c} \times \boldsymbol{a}/V, \qquad \boldsymbol{c}^* = a \times \boldsymbol{b}/V. \tag{5}$$

It can be seen from eq. (5) that for example $\boldsymbol{a}^*$ is a vector with a length of the vector product $\boldsymbol{b} \times \boldsymbol{c}$ divided by the volume of the elementary cell V:

$$V = \boldsymbol{a} \cdot (\boldsymbol{b} \times \boldsymbol{c}) = \boldsymbol{b} \cdot (\boldsymbol{c} \times \boldsymbol{a}) = \boldsymbol{c} \cdot (\boldsymbol{a} \times \boldsymbol{b}). \tag{6}$$

It can be seen from eq. (5) that the reciprocal vectors $\boldsymbol{a}^*$, $\boldsymbol{b}^*$, $\boldsymbol{c}^*$, have the dimension of reciprocal length.

It follows from eqs (5) and (6) that

$$\boldsymbol{a}^* \cdot \boldsymbol{a} = \boldsymbol{b}^* \cdot \boldsymbol{b} = \boldsymbol{c}^* \cdot \boldsymbol{c} = 1 \tag{7}$$

and

$$\boldsymbol{a}^* \cdot \boldsymbol{b} = \boldsymbol{a} \cdot \boldsymbol{b}^* = \boldsymbol{a}^* \cdot \boldsymbol{c} = \boldsymbol{a} \cdot \boldsymbol{c}^* = \boldsymbol{b}^* \cdot \boldsymbol{c} = \boldsymbol{b} \cdot \boldsymbol{c}^* = 0. \tag{8}$$

In the special case that the vectors $\boldsymbol{a,b,c}$ are mutually perpendicular the value for the volume V becomes: $|V| = |\boldsymbol{a}||\boldsymbol{b}||\boldsymbol{c}|$. Then $\boldsymbol{a}^*$, $\boldsymbol{b}^*$, and $\boldsymbol{c}^*$, are parallel to $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$, respectively and $|\boldsymbol{a}^*| = |\boldsymbol{a}|^{-1}$, $|\boldsymbol{b}^*| = |\boldsymbol{b}|^{-1}$, $|\boldsymbol{c}^*| = |\boldsymbol{c}|^{-1}$. This also demonstrates the "reciprocal relation" between the original lattice and the reciprocal lattice, *vice versa.* It also follows from eq. (8) and, of course, from the definition of reciprocal sapce in eq. (5) that, e.g., $\boldsymbol{a}$ is perpendicular to the plane that contains $\boldsymbol{b}^*$ and $\boldsymbol{c}^*$

6.2.4. 3D Fourier functions

Up to now, only crystal lattices consisting of points have been discussed. For real crystals these points have to be replaced by unit cells. To describe the contents of the unit cell, made up of atoms leading to an electron densitiy, making full use of the 3D lattice tranlational symmetry, 3D Fourier functions are appropriate.

In a one dimensional space a periodic function can be represented by the imaginary Fourier function

$$F(x) = \sum_h F_h \exp 2\pi i h x, \tag{9}$$

where, h is an integer. Here F_h is a fixed number called the Fourier component belonging to h. Each periodic function is fully characterized by its Fourier components F_h.

Equation (9) can be generalized to two, three and more dimensions and for a 3D space we write:

$$F(\boldsymbol{r}) = \sum_{\boldsymbol{k}} F_{\boldsymbol{k}} \exp 2\pi i \boldsymbol{k} \cdot \boldsymbol{r}, \tag{10}$$

where the vectors $\boldsymbol{r}$ are defined in normal (direct) space and the vectors $\boldsymbol{k}$ belong to reciprocal space; they are given by

$$\boldsymbol{r} = x\boldsymbol{a} + y\boldsymbol{b} + z\boldsymbol{c} \tag{11}$$

and

$$\boldsymbol{k} = h'\boldsymbol{a}^* + k'\boldsymbol{b}^* + l'\boldsymbol{c}^*. \tag{12}$$

Note that in eq. (11) (x, y, z) are real numbers as triples of fractional coordinates in direct space with respect to the basis $\{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}\}$. In eq. (12) the indices (h', k', l') are triples of integers with resepct to the reciprocal basis $\{\boldsymbol{a}^*, \boldsymbol{b}^*, \boldsymbol{c}^*,\}$. The fact that the latter triples are limited to integer values reflects the 3D lattice translational symmetry of the function $F(\boldsymbol{r})$, according to eq. (4). Contrary to the triples of integers *(hkl)* of Miller indices mentioned above, triples (h', k', l') need not be relatively prime.

If we take a point in direct space characterized by the vector $\boldsymbol{r}'$, given by:

$$\boldsymbol{r}' = (x+u)\boldsymbol{a} + (y+v)\boldsymbol{b} + (z+w)\boldsymbol{c} \tag{13}$$

in which $[u, v, w]$ are triples of integers, it can be seen using eqs (10), (7) and (8) that

$$F(\boldsymbol{r}') = \sum_{\boldsymbol{k}} F_{\boldsymbol{k}} \exp 2\pi i(\boldsymbol{k}\cdot\boldsymbol{r} + h'u + k'v + l'w) = F(\boldsymbol{r}). \tag{14}$$

Because $\exp 2\pi i(h'u + k'v + l'w) = \exp 2\pi i N = 1$, since N is an integer. So, also expressed as a Fourier transform the function $F(\boldsymbol{r})$ is periodic and translationally invariant for the infinite set of translations $\boldsymbol{T}$ of eq. (3).

6.2.5. *Interplanar thickness* $\boldsymbol{d}_{hkl}$

We define in reciprocal space a fixed vector $\boldsymbol{H}^*_{hkl}$ according to

$$\boldsymbol{H}^*_{hkl} = h\boldsymbol{a}^* + k\boldsymbol{b}^* + l\boldsymbol{c}^*. \tag{15}$$

Like for the Miller indices (see Section 6.2.1) h, k and l are chosen to be prime in reference to each other. Alternatively, we consider an infinite set of points in direct space defined by the vectors

$$\boldsymbol{H}_{uvw} = u\boldsymbol{a} + v\boldsymbol{b} + w\boldsymbol{c}, \tag{16}$$

with u, v, w, integer and which fulfil the relation

$$\boldsymbol{H}^*_{hkl}\cdot\boldsymbol{H}_{uvw} = \left(h\boldsymbol{a}^* + k\boldsymbol{b}^* + l\boldsymbol{c}^*\right)\cdot(u\boldsymbol{a} + v\boldsymbol{b} + w\boldsymbol{c}) = N, \tag{17}$$

for fixed N, thus, given by:

$$hu + kv + lw = N. \tag{18}$$

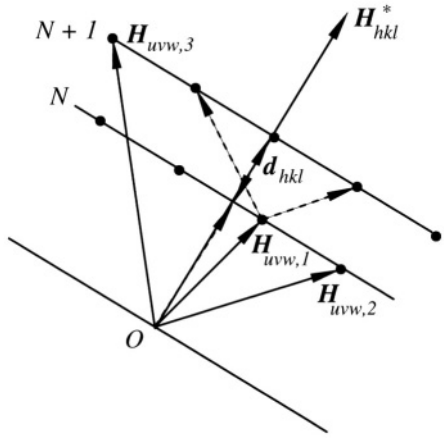


FIGURE 6.2. A reciprocal vector $\boldsymbol{H}^*_{hkl}$ is oriented perpendicular to a stack of parallel lattice planes labeled with an integer N with the orientation *(hkl).* One lattice plane consists of many lattice points, which correspond to the direct lattice vectors $\boldsymbol{H}_{uvw}$. The projection of these vectors $\boldsymbol{H}_{uvw}$ on the vector $\boldsymbol{H}^*_{hkl}$ coincide with the vector $N d_{hkl} \boldsymbol{H}^*_{hkl}/|\boldsymbol{H}^*_{hkl}|$, where N is the lattice plane $\boldsymbol{H}_{uvw}$ points to. The distance between the neighboring netplanes is the interplanar distance d_{hkl}. The dashed vectors indicate equivalent lattice points for neighboring lattice planes. The distance from the origin to the lattice plane labeled N is only schematic.

Equation (17) is the mathematical representation of a plane perpendicular to the vector $\boldsymbol{H}^*_{hkl}$ at a (dimensionless) distance N from the origin. If $\boldsymbol{H}_{uvw}$ were allowed to take on any value eq. (17) would describe all points of that plane. The fact that the values of $\boldsymbol{H}_{uvw}$ are limited to integer values for u, v, w limits eq. (17) to the lattice points of the translational lattice that lie in the plane perpendicular to $\boldsymbol{H}^*_{hkl}$ at a distance N from the origin. The geometrical interpretation of eq. (17), using a fixed integer N according to eq. (18) is illustrated in Figure 6.2. $\boldsymbol{H}^*_{hkl} \cdot \boldsymbol{H}_{uvw}$ is the projection of the vector $\boldsymbol{H}_{uvw}$ on $\boldsymbol{H}^*_{hkl}$. All vectors $\boldsymbol{H}_{uvw}$ pointing to a position in the plane at distance N are, thus, projected on the vector $\boldsymbol{H}^*_{hkl}$ $N/|\boldsymbol{H}^*_{hkl}|$ pointing in the direction of the normal of the plane.

Choosing another integer for N, say N', defines another parallel lattice plane at a distance N'. All lattice planes, thus defined, are parallel and at mutual unity distance. Therefore, the interplanar distance d_{hkl} between successive lattice planes perpendicular to the reciprocal lattice vetor $\boldsymbol{H}^*_{hkl}$ in terms of the measure of length in direct space is found by using eq. (17) with $N = 1$, resulting in

$$d_{hkl} = 1/|\boldsymbol{H}^*_{hkl}|. \tag{19}$$

So we arrive at two important theorems:

Theorem 1. An infinite set of parallel lattice planes is characterized by a reciprocal lattice vector $\boldsymbol{H}^*_{hkl}$, perpendicular to these lattice planes.

Theorem 2. The interplanar distance d_{hkl} between neighboring lattice planes of the set is the inverse of the length of the reciprocal lattice vector $\boldsymbol{H}^*_{hkl}$.

The interplanar distance represents the smallest distance over which the lattice plane is repeated. Note, that even the shortest of the corresponding crystallographic translational

symmetry vectors need not be parallel to $\boldsymbol{H}^*_{hkl}$. In fact, d_{hkl} is the projection of an infinite set of such crystallographic translational symmetry vectors on $\boldsymbol{H}^*_{hkl}$ as is illustrated in Figure 6.2.

6.2.6. *Meaning of Miller indices (hkl)*

We already discussed the relation between observed faces *(hkl)* and a large (infinite) number of parallel lattice planes in a preliminary way. In order to investigate the mathematical meaning of the indices *(hkl)* in more detail we again consider the lattice plane defined by eq. (17) for a certain integer N. We can find the intersection point of the line $\boldsymbol{r} = p\boldsymbol{a}$ along the basis vector a in direct space, and the plane, by evaluating $\boldsymbol{H}^*_{hkl} \cdot p\boldsymbol{a} = N$, where p is a real number. Using eqs (7) and (8) this leads to $hp = N$. For $N = 1$ this results in the conclusion that the first lattice plane cuts the $\boldsymbol{a}$ axis in $|\boldsymbol{a}|/h$. For th N-*th* lattice plane this amounts to $N|\boldsymbol{a}|/h$. Analogously, we find for the $\boldsymbol{b}$ and $\boldsymbol{c}$ axes as intersection points $N|\boldsymbol{b}|/k$ and $N|\boldsymbol{c}|/l$, respectively. So, we arrive at the following theorem as a mathematical formulation of eq. (2):

Theorem 3. A set of continuous lattice planes characterized by the reciprocal vector $\boldsymbol{H}^*_{hkl} = h\boldsymbol{a}^* + k\boldsymbol{b}^* + l\boldsymbol{c}^*$ cut the $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ axes, respectively, in pieces of size

$$\Delta x|\boldsymbol{a}| = |\boldsymbol{a}|/h, \qquad \Delta y|\boldsymbol{b}| = |\boldsymbol{b}|/k, \qquad \Delta z|\boldsymbol{c}| = |\boldsymbol{c}|/l. \tag{20}$$

This property was already illustrated for the face (412) in Figure 6.1. Note, that as was demonstrated in Figure 6.1 the intersection points need not be lattice points.

6.2.7. *Application to macroscopic morphology*

Let us as a thought experiment switch from a microscopic point of view of crystals to a macroscopic point of view. We than replace a stack of parallel lattice planes characterized by a reciprocal vector $\boldsymbol{H}^*_{hkl}$, which in turn is characterized by a triple of integers *(hkl)*, which are prime in reference to each other, by a single plane with Miller indices *(hkl)*. Since for a macroscopic description of the shapes of crystals, bounded by flat faces characterized by different Miller indices *(hkl)* first of all the mutual angles are essential parameters, we can place the macroscopic faces *(hkl)* in the origin of a coordinate system. Then, the directions of the faces *(hkl)* are characterized by the normals $\boldsymbol{H}^*_{hkl}$. In agreement with classical crystallographic conventions we will, therefore, in the following indicate a macroscopic crystal face of which the orientation is given by its normal $\boldsymbol{H}^*_{hkl} = h\boldsymbol{a}^* + k\boldsymbol{b}^* + l\boldsymbol{c}^*$ as *(hkl)*. Similarly, we will indicate a direction in the lattice in direct space $\boldsymbol{H}_{uvw} = u\boldsymbol{a} + v\boldsymbol{b} + w\boldsymbol{c}$ as $[uvw]$, with u, v, w integer. Since for the description of crystallographic directions only the ratios of u, v and w are relevant the triplet u, v, w is chosen as relatively prime.

If two macroscopic planes *(hkl)* cut each other in a line, this line is parallel to an edge of the corresponding macroscopic crystal. In the following we will derive relations, between the Miller indices *(hkl)*, defined in reciprocal space, of macroscopic faces and directions $[uvw]$, defined in direct space, describing edges and related directions, which are important for an adequate description of the macroscopic shapes of faceted crystals. These relations will be expressed in some further theorems. The proofs of these theorems are trivial.

Theorem 4. Two directions $[u_1v_1w_1]$ and $[u_2v_2w_2]$ which are not parallel span a lattice plane with a normal that can be found by evaluating the vector product of the two directions, according to

$$\boldsymbol{H}^*_{hkl} = \left(\boldsymbol{H}_{u_1v_1w_1} \times \boldsymbol{H}_{u_2v_2w_2}\right)/V. \tag{21}$$

Using eq. (5) we find as a result that the corresponding Miller indices are given by

$$(hkl) = \left((v_1w_2 - v_2w_1)(u_2w_1 - u_1w_2)(u_1v_2 - u_2v_1)\right). \tag{22}$$

Theorem 5. If a crystallographic direction *[uvw]* is an edge between two faces $(h_1k_1l_1)$ and $(h_2k_2l_2)$, then all faces *(hkl)* with orientations, which fulfill the relation $(hkl) = (\lambda_1h_1 + \lambda_2h_2, \lambda_1k_1 + \lambda_2k_2, \lambda_1l_1 + \lambda_2l_2)$, where λ_1 and λ_2 are arbitrary fractional numbers such that h, k and l are integer and mutually prime, are parallel with the same direction *[uvw].*

Theorem 5, in fact, defines all lattice planes that are parallel to the direction [*uvw*]; these are defined as the zone of *[uvw].* The special case of the face, for which $\lambda_1 = \lambda_2 = 1$, belonging to this zone was called by 19^{th} century crystallographers the "Law of Complication". Fom a modern point of view it can be considered as a trivial "law".

6.2.8. The morphological law of Bravais, Friedel, Donnay and Harker (BFDH)

6.2.8.1. Relation between morphological importance and interplanar distance In the period from about 1880 to 1937 the (in principle) purely mathematical law to describe and to predict the morphology of crystals was formulated first by Bravais and Friedel, taking the primitivity or non-primitivity of elementary cells into account and later by Donnay and Harker, taking the role of screw axes and glide planes of space groups into account. This law, which is nowadays known as the BFDH law and can be formulated as follows:

BFDH-law. The morphological importance of a form of faces *{hkl}* on the growth forms of any crystal is larger the larger the interplanar distance d_{hkl}.

The curly brackets are used to indicate the set of all faces that are equivalent to the face *(hkl)* under the (point)group symmetry of the crystal. This set is called the form *{hkl}.* The concept of morphological importance of faces *{hkl}* to be abreviated as $\boldsymbol{MI}_{hkl}$ is a, in practice not very precisely defined, statistical quantity for the combination of the relative size of the faces *{hkl}* and the frequency of occurrence of the faces *{hkl}.* These quantities run in principle parallel, because faces of the slowest growing forms *{hkl}* will dominate growth forms of faster growing faces. Note, that in order to use proper statistics the number of faces within a form *{hkl},* depending on the pointgroup symmetry of the crystal, should be taken into consideration.

The BFDH law can be expressed as follows

$$|d_{hkl,1}| > |d_{hkl,2}| \Rightarrow MI_{hkl,1} > MI_{hkl,2}, \tag{23}$$

or what comes to the same taking eq. (19) into account

$$|\boldsymbol{H}^*_{hkl,1}| < |H^*_{hkl,2}| \Rightarrow MI_{hkl,1} > MI_{hkl,2}. \tag{24}$$

Since it can be stated, statistically spoken, the higher the growth rate R_{hkl} of faces of a certain form *{hkl},* the lower its MI_{hkl}, eq. (24) can be replaced by

$$|\boldsymbol{H}^*_{hkl,1}| < |\boldsymbol{H}^*_{hkl,2}| \Rightarrow R_{hkl,1} < R_{hkl,2}. \tag{25}$$

6.2.8.2. Selection rules of space groups and morphology Up to now, we have focused the attention on translational symmetry. In order to derive all possibilities to fill 3D space in a symmetrical, translationally invariant way, rotational symmetry leading to *n*–fold rotation axes, mirror planes and centers of symmetry need to be taken into consideration. It is a tremendous job to derive logically all possible combinations of the symmetry elements, discussed above. It turns out that for 3D translationally invariant space there are 230 possibilities to fill this space. These 230 possibilities are called the 230 space groups. This name refers to group theory, a branch of mathematics, dealing with symmetry problems. For a derivation of these 230 space groups we refer for a treatment based on group theory, for example, to the book by T. Janssen [4] and for a derivation of these space groups in a geometric way to the book by Buerger [5]. Further, we refer for an introduction to the 230 space groups and a detailed description of all possible implications of these space groups to the International Tables for Crystallography [6]. It follows from mathematical crystallography that only mirror planes, centers of symmetry and 2, 3, 4 and 6 fold rotational axes are compatible with a 3D translationally invariant space. In addition, there can be glide mirror planes and screw axes which are combinations of non-primitive translations and mirror planes or rotation axes, respectively.

The symmetry of a given space group has a crucial influence on the shape of the elementary cell, spanned on the three basic vectors $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, making for the triclinic space groups, having the lowest symmetry, arbitrary angles α, β, γ, between $\boldsymbol{b}$ and $\boldsymbol{c}$, $\boldsymbol{c}$ and $\boldsymbol{a}$, and $\boldsymbol{a}$ and $\boldsymbol{b}$ respectively. In general, depending on the space group the shape of the elementary cell varies for the highest symmetries from the cube to, for the lowest symmetries, a triclinic elementary cell mentioned above. In Table 6.1 the different shapes of elementary cells are summarized. In this table $a = |\boldsymbol{a}|$, $b = |\boldsymbol{b}|$ and $c = |\boldsymbol{c}|$.

It follows from mathematical crystallography that in order to keep the highest possible symmetrical shape of the six elementary cells in the description of crystal structures of the six crystal classes, presented in Table 6.1 above, that it is for certain space groups necessary to add, apart from the basic translations vectors $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ one or more extra basic translations to these basic vectors of the elementary cell. In this way non-primitive cells are obtained that are called conventional unit cells. Looking at the last column of Table 6.1 it can be

TABLE 6.1.
The various crystal classes determining the shape of the conventional unit cell.

Crystal class	a, b, c	α, β, γ	Centered cells
Triclinic	$a \neq b \neq c$	α, β, γ, no special values	P
Monoclinic	$a \neq b \neq c$	$\beta = 90°, \alpha, \gamma$ no special values	P, (A, C)
Orthorhombic	$a \neq b \neq c$	$\alpha = \beta = \gamma = 90°$	P, (A, B, C), I, F
Tetragonal	$a = b \neq c$	$\alpha = \beta = \gamma = 90°$	P, I
Trigonal/hexagonal	$a = b \neq c$	$\alpha = \beta = 90°, \gamma = 60°$	P, R
Cubic	$a = b = c$	$\alpha = \beta = \gamma = 90°$	P, I , F

TABLE 6.2.
The morphological selecton rules due to the various centering translations. For the trigonal–hexagonal crystallographic system $\boldsymbol{d}$ represents the longest body digonal. n is the smallest integer for which h, k, l are mutually prime.

Centering	Non-primitive vectors	Selection rule
A	$\frac{1}{2}(\boldsymbol{b}+\boldsymbol{c})$	$k+l=2n$
B	$\frac{1}{2}(\boldsymbol{a}+\boldsymbol{c})$	$h+l=2n$
C	$\frac{1}{2}(\boldsymbol{a}+\boldsymbol{b})$	$h+k=2n$
I	$\frac{1}{2}(\boldsymbol{a}+\boldsymbol{b}+\boldsymbol{c})$	$h+k+l=2n$
R	$1/3\ \boldsymbol{d}, 2/3\boldsymbol{d}$	$h-k+l=3n$
F	$\frac{1}{2}(\boldsymbol{b}+\boldsymbol{c}), \frac{1}{2}(\boldsymbol{a}+\boldsymbol{c}), \frac{1}{2}(\boldsymbol{a}+\boldsymbol{b})$	h, k, l either all n or all $2n$

seen that that the types P, (A, B, C), I, F, R of cells occur corresponding to 14 types of lattice, playing their essential role in the 230 space groups. These lattices are called the 14 Bravais lattices. Here P refers to primitive, corresponding to the case, where no extra non-primitive translation vector is added to the elementary cell. A, B, C cells correspond to cells in which one extra non-primitive translation vector is added to the elementary, etc. These non-primitive translation vectors can be found in Table 6.2.

6.2.8.3. Non-primitive unit cells, screw axes, mirror glide planes and MI of forms {hkl}

The occurrence of centered non-primitive cells has important implications for the morphology of crystals. Due to the fact that extra translations occur for crystals having centered cells as compared to crystals having primitive cells the interplanar thickness d_{hkl}, representing the smallest distance over which the crystal structure is repeated, must for special triples *(hkl)* be divided by a factor n depending on the non-primitive translations, implying that the reciprocal vector $\boldsymbol{H}^*_{hkl} = h\boldsymbol{a}^* + k\boldsymbol{b}^* + l\boldsymbol{c}^*$ will be multiplied with n. Taking eqs (23)–(25) into account it then follows that the $\boldsymbol{MI}_{hkl}$ of such a face will be reduced. Because of this in morphological studies, often, a face (100), for example, is denoted as (200) in case $n = 2$, etc. The selection rules for conventional, i.e. non-primitive, cells are summarized in Table 6.2.

As an example we give the interpretation of Table 6.2 for the case of a C-centered cell. In case the sum of the Miller indices $h + k$ is even no correction is needed and n is equal to unity. However, if $h + k$ is odd $n = 2$.

In addition to these selection rules as a result of non-primitive cells also the presence of screw axes or glide planes that are perpendicular to a form *{hkl}* give rise to additional selection rules that make the interplanar distance smaller by an integer factor. These selection rules are summarized in Table 6.3. This table only gives the possibilities in settings that correspond to crystallographic conventions [6].

As an example to Table 6.3 consider a *d*-glide plane parallel to (100). The selection rule only applies to the faces perpendicular to this symmetry element, that is, the faces *(0kl)*. The Miller indices *hkl* have to be corrected such that $k + l$ is an integer multiple of 4 and *hkl* are mutually prime.

The selection rules described above are the same as the reflection conditions in (X-ray) diffraction theory with one difference, namely, that in diffraction n labels different reflections, while in crystal morphology n represents the smallest value for which h, k, l are still mutually prime.We refer for information on these selection rules to the International Tables for Crystallography [6] and to the books by Philips, Rousseau or Ham-

TABLE 6.3.
Selection rules due to glide mirror planes and screw axes parallel to $\boldsymbol{H}^*_{hkl}$. n is the smallest integer for which h, k, l are mutually prime; hkl refers to the orientations to which the selection rules apply.

Symmetry	Symbol	Non-primitive vector	hkl	Selection rule
Glide plane (001)	a	$\frac{1}{2}\boldsymbol{a}$	$hk0$	$h = 2n$
	b	$\frac{1}{2}\boldsymbol{b}$		$k = 2n$
	n	$\frac{1}{2}(\boldsymbol{a} + \boldsymbol{b})$		$h + k = 2n$
	d	$1/4(\boldsymbol{a} \pm \boldsymbol{b})$		$h + k = 4n$
Glide plane (100)	b	$\frac{1}{2}\boldsymbol{b}$	$0kl$	$k = 2n$
	c	$\frac{1}{2}\boldsymbol{c}$		$l = 2n$
	n	$\frac{1}{2}(\boldsymbol{b} + \boldsymbol{c})$		$k + l = 2n$
	d	$1/4(\boldsymbol{b} \pm \boldsymbol{c})$		$k + l = 4n$
Glide plane (010)	a	$\frac{1}{2}\boldsymbol{a}$	$h0l$	$h = 2n$
	c	$\frac{1}{2}\boldsymbol{c}$		$l = 2n$
	n	$\frac{1}{2}(\boldsymbol{a} + \boldsymbol{c})$		$h + l = 2n$
	d	$1/4(\pm\boldsymbol{a} + \boldsymbol{c})$		$h + l = 4n$
Glide plane (110)	c	$\frac{1}{2}\boldsymbol{c}$	hkl	$l = 2n$
	b	$\frac{1}{2}\boldsymbol{b}$		$h = 2n$
	n	$\frac{1}{2}(\boldsymbol{a} + \boldsymbol{b} + \boldsymbol{c})$		$h + k = 2n$
	d	$1/4(\boldsymbol{a} + \boldsymbol{b} \pm \boldsymbol{c})$		$2h + l = 4n$
Screw axis $\boldsymbol{c}$	$2_1, 4_2, 6_3$	$\frac{1}{2}\boldsymbol{c}$	$00l$	$l = 2n$
	$3_1, 3_2, 6_2, 6_4$	$1/3\boldsymbol{c}$		$l = 3n$
	$4_1, 4_2$	$1/4\boldsymbol{c}$		$l = 4n$
	$6_1, 6_5$	$1/6\boldsymbol{c}$		$l = 6n$
Screw axis $\boldsymbol{a}$	$2_1, 4_2$	$\frac{1}{2}\boldsymbol{a}$	$h00$	$h = 2n$
	$4_1, 4_3$	$1/4\boldsymbol{a}$		$h = 4n$
Screw axis $\boldsymbol{b}$	$2_1, 4_2$	$\frac{1}{2}\boldsymbol{b}$	$0k0$	$k = 2n$
	$4_1, 4_3$	$1/4\boldsymbol{b}$		$k = 4n$
Screw axis $\boldsymbol{a} + \boldsymbol{b}$	2_1	$\frac{1}{2}(\boldsymbol{a} + \boldsymbol{b})$	$hh0$	$h = 2n$

mond [7]. Garnets serve as an example. Garnets have as space group Ia3d. If a crystal with a corresponding space group without non-primitive translations, that is, Pm3m would show the faces (100), (110), (111), (210) and (211) according to the BFDH law described in the next section, the corresponding faces for the ganets would be (211), (110), (321), (100) and (210), respectively, as a result of the selection rules described above. For this class of compounds, as for many others, the implication of the BFDH theory is in agreement with the observed morphology of natural and synthetic crystals. For a survey see ref. [8].

In order to study relevant aspects of the macroscopic morphology of crystals, such as the shapes or habits of a three dimensional crystal, characterized by combinations of different faces *(hkl)* and zones *[uvw]* the stereographic projection is used. This projection allows to make 2D projections of sometimes complex 3D habits of crystals. Classical crystallographers used the stereographic projection extensively. This projection may still be an important tool to study from a modern point of view observed and predicted morphologies of crystals. The principles and application to crystallography of the stereographic projection can be found in the very interesting book by Donnay [9] and the book sby Philips and others [7].

6.2.9. Construction of growth forms

In order to predict or construct a (growth) form resulting from a given known crystal structure, the relative growth rates of the faces *{hkl}* need to be calculated. As a first approximation the relation between the rate of growth R_{hkl} of faces *{hkl}* given by the vectors $\boldsymbol{R}_{hkl}$ parallel to the reciprocal vectors $\boldsymbol{H}^*_{hkl}$ is usually assumed to be

$$\boldsymbol{R}_{hkl} = C\,\boldsymbol{H}^*_{hkl}, \tag{26}$$

where C is a constant, which is supposed to be the same for all faces. Equation (26) is often called the BFDH law, but, in fact, is a further simplification of the proportionality relation given by eq. (25). Then the enclosed figure formed by the faces erected perpendicular at the end points of the vectors $\boldsymbol{R}_{hkl}$ is believed to give a good representation of experimental growth forms and usually referred to as the BFDH morphology. There are various computer programs available that do the job.

Anticipating recent developments in crystal morphology prediction it is interesting to note that the BFDH aproach, presented so far is only based on purely geometrical concepts, resulting from the dimensions and shape of the elementary cell and the selection rules of the space group. Already from the beginning of the science of crystallography crystallographers had the following intuitive interpretation for the BFDH law; the thicker d_{hkl}, corresponding to a growth layer of a growing face, the more energy of chemical bonds between the growth units within the layer, the lower the growth rate R_{hkl} and the higher the Morphological Importance MI_{hkl}. The relevance of the bond energy between growth units for crystal morphology will be treated in the following sections.

In Section 6.3 a simple statistical mechanical surface model will presented, which explains why crystals grow, in principle, with flat faces. Next, it will be shown in Section 6.4 that by introducing the concept of overall chemical bonds between molecules of a given crystal structure and the crystallographic concept of faces *(hkl)* of which the layers have a thickness d_{hkl}, the morphology can be predicted.

6.3. STATISTICAL MECHANICAL ISING LATTICES APPLIED TO CRYSTAL SURFACES

6.3.1. Introduction

The aim of the previous section was to present an introduction to classical mathematical crystallography and the attention was at the end of that section focused to the concept of the interplanar distance d_{hkl}, which plays a key role in crystal morphology. In order to develop models which explain the experimental fact that almost all crystals grow with flat faces with a well-defined orientation *(hkl)* but, sometimes, with rounded non crystallographic faces, we will in this section introduce a statistical mechanical surface model based on Kossel (like) crystals. We will introduce Ising lattice models for the interface between the mother phase and a face *(hkl)* of the growing crystal. The mother phase can be vapour, melt or solution. It will be demonstrated that the introduction of interface Ising lattice models makes it possible to develop relevant statistical mechanical models for the interface of a crystal for, in principle, any mother phase.

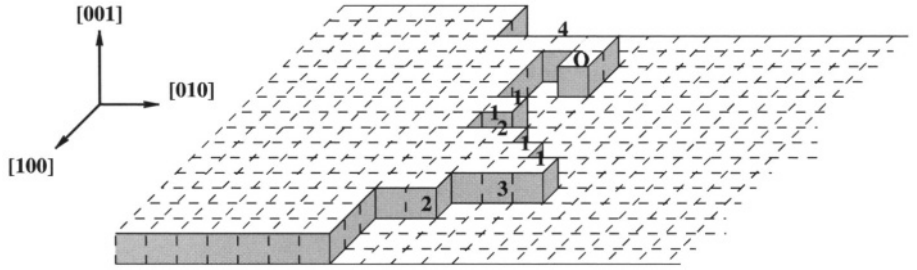


FIGURE 6.3. (001) Surface of an orthorhombic crystal as a generalization of a Kossel crystal with a step parallel to the [100] direction. The blocks represent a unit cell. The numbers label the heights of the columns perpendicular to the reference step parallel to the [100] direction. The SOS condition implies that the overhang indicated by 'O' is forbidden.

In Figure 6.3 a block model of a crystal and limiting surface is presented. The cells in such model are interpreted as the unit cells of a Kossel crystal that can be seen as a simple cubic crystal. Neighboring cells mutually interact via bonds between the centers of the cells, thus, representing the chemical bonds of such a crystal. These bonds are not drawn. Alternatively, the edges of such a block model can be interpreted as the bonds in which case the vertices of the blocks represent the growth units. Note, that there is only a single growth unit per unit cell for the simple cubic crystal. In the latter interpretation we call the model a crystal graph according to mathematical graph theory. Also note, that this reduction of the crystal to a crystal graph does not change the energetics of the crystal nor of the crystal growth process as long as we model the bond energies and the energy of the growth units in accordance with the actual growth process. Thus, we can make the model as realistic as we like by appropriate modelling of the relevant energy terms. Moreover, we can make the model more complicated by introducing different cell types and allowing for intra-cell bonds or defining inter-cell bonds that go beyond first neighbors. Such crystal graphs will de the subject of Sections 6.4 and 6.5.

In this Section 6.3 we will use statistical mechanics to derive information on the thermodynamic stability of the surfaces of a crystal, leading to a thermal roughening transition which is relevant for all crystal faces. For that, we will start with the model for an orthorhombic primitive cell with three different bonds along the crystallographic directions $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ with energies ϕ_a, ϕ_b and ϕ_c, respectively. The lengths of the crystallographic axes will be denoted as $a = |\boldsymbol{a}|$, etc. We will treat these bonds as broken bonds in case one of the neighboring growth units is absent. Often, we will, rather loosely but without loss of generality, call the growth unit a cell. In order to develop a statistical mechanical model for, for example, the (001) top face, this allows for the following model. We divide the whole space of crystal and the adjacent empty mother phase in a lattice of cells, having exactly the same shape as the cells in the crystal. For the time being we will consider the mother phase to be the vacuum as if the crystal is growing from a very dilute vapour. So the whole space of bulk crystal, and bulk vacuum and interface between crystal and vacuum becomes, as it were, one orthorhombic crystal. In order to describe the structure between crystal and vacuum we assume that the unit cells can have only one out of two properties; solid or vacuum. Such a model where cells can have only two properties, solid or vacuum, is called an Ising lattice in analogy with Ising models for spin up and spin down systems in magnetic systems.

6.3.2. Roughening transition and the solid on solid step model

In this section the Solid on Solid step model for the (001) surface of the orthorhombic cell model, presented in Figure 6.3 is discussed. We refer to the paper of Leamy, Gilmer and Jackson in which this model is treated extensively [10]. In Figure 6.3 a step on the (001) surface running, on the average, parallel to the ***a*** direction is seen meandering in perspective. Due to thermal fluctuations the step will not be straight and there will be columns of vacuum cells and columns of solid cells perpendicular to the straight reference step. These columns can be considered as the entropic noise of a straight step. We introduce the Solid On Solid (SOS) condition which is a simplification, implying that only those fluctuations are allowed which deepen or lengthen the columns in the negative or positive ***b*** direction. Thus, overhangs are forbidden. By introducing this approximation the problem of calculating the edge free energy of a step in an Ising model can be solved exactly. Thanks to the introduction of the SOS approximation each column can be treated separately. The SOS condition implies that the height of each column h varies in a discrete way with jumps of b or $-b$ from $-\infty b$ to ∞b. Then the partition function of one column Z_c in terms of the number h of broken bonds ϕ_a on one side of the column becomes

$$Z_c = \sum_{h=-\infty}^{\infty} \exp -|h|\phi_a/kT. \tag{27}$$

For a step with a length of $N\,a$, consisting of N columns the partion function for the step becomes, taking into consideration all possible independent fluctuations of the heights of the columns

$$Z_{\text{step}} = (Z_c)^N = \left(\sum_{h=-\infty}^{\infty} \exp -|h|\phi_a/kT \right)^N \tag{28}$$

which can be rearranged to

$$Z_{\text{step}} = \left[2\left(\sum_{h=0}^{\infty} \exp -h\phi_a/kT \right) - 1 \right]^N. \tag{29}$$

This becomes, taking the expression for an infinite geometrical series into account

$$\begin{aligned} Z_{\text{step}} &= \left[2/(1 - \exp -\phi_a/kT) - 1\right]^N \\ &= \left[(1 + \exp -\phi_a/kT)/(1 - \exp -\phi_a/kT)\right]^N. \end{aligned} \tag{30}$$

Multiplying numerator and denominator of eq. (30) with $\exp \phi_a/2kT$ gives

$$Z_{\text{step}} = \left[(\exp \phi_a/2kT + \exp -\phi_a/2kT)/(\exp \phi_a/2kT - \exp -\phi_a/2kT)\right]^N \tag{31}$$

or, alternatively,

$$Z_{\text{step}} = (\cosh \phi_a/2kT / \sinh \phi_a/2kT)^N = (\coth \phi_a/2kT)^N. \tag{32}$$

The edge free energy γ_x of a step parallel to the ***a*** direction of a SOS step, expressed in energy per unit length a, using $F = -kT \ln Z$, is therefore given by

$$\gamma_x = \phi_b/a - kT/a \ln \coth \phi_a/2kT, \tag{33}$$

where the first term on the right hand side represents the free energy of the reference (straight) step made up by the broken bonds ϕ_b. For a step consisting of N columns eq. (33) has to be multiplied by N. For a step in the ***b*** direction the edge free energy γ_y can be derived analogously to become

$$\gamma_y = \phi_a/b - kT/b \ln \coth \phi_b/2kT. \tag{34}$$

Within the framework of this Solid on Solid step model the roughening transition occurs if the edge free energy becomes zero because, then, the columns can be formed without cost of free energy. For the step in the ***a*** direction eq. (33) becomes

$$\coth \phi_a/2kT = \exp \phi_b/kT, \tag{35}$$

which in turn becomes after rewriting

$$(\exp \phi_a/2kT + \exp -\phi_a/2kT)/(\exp \phi_a/2kT - \exp -\phi_a/2kT) = \exp \phi_b/kT. \tag{36}$$

After multiplying the numerator and denominator by $\exp \phi_a/2kT$ of the left hand side of eq. (36) we get

$$(\exp \phi_a/kT + 1)/(\exp \phi_a/kT - 1) = \exp \phi_b/kT \tag{37}$$

which becomes after rewriting

$$\exp \phi_b/kT \exp \phi_a/kT - \exp \phi_b/kT - \exp \phi_a/kT - 1 = 0. \tag{38}$$

If ϕ_a and ϕ_b are known the crititical temperature T^c above which the edge free energy of the step becomes zero can de determined numerically from eq. (38). It can be seen that eq. (38) is symmetrical in ϕ_a and ϕ_b. If, therefore, only the ratio ϕ_b/ϕ_a is known the critical value of a dimensionless energy $(\phi/2kT)^c$ corresponding to a dimensionless roughening temperature $\theta^c = (2kT/\phi)^c$ can be calculated from eq. (38). In that case the convention is often used to express θ^c in the relative highest broken bond energy of ϕ_a and ϕ_b. The factor 2 in the definition of θ^c is purely conventional. If the condition $\phi_a = \phi_b = \phi$ applies, which is the case for the Kossel crystal, we get for eq. (38), after dividing by $2 \exp \phi/kT$, the expression

$$\sinh \phi/kT = 1, \tag{39}$$

which may be also written as

$$\sinh^2 \phi/kT = 1. \tag{40}$$

This happens to be a special case of one of the famous equations derived by Onsager [11]

$$\sinh \phi_x/kT \sinh \phi_y/kT = 1 \tag{41}$$

for the critical order–disorder phase transition of a 2D lattice consisting of first nearest interactions ϕ_x and ϕ_y. Contrary to the SOS step models discussed above leading to eq. (38), eq. (41) is an exact equation and has a general validity for all kinds of 2D lattice with first nearest neighbor interactions.

6.3.3. Meaning of order–disorder surface phase transitions

The critical dimensionless temperature $\theta^c = (2kT/\phi)^c$, which becomes a real critical temperature after substituting real bond energies ϕ_a and ϕ_b, is called the temperature of the order–disorder phase transition in a rectangular 2D lattice. We want to stress that the paper of Onsager [11], which was published in 1944 marked a real breakthrough, because Onsager, for the first time, could prove exactly, thanks to extensive mathematical derivations, that within a 2D Ising lattice a second order phase transition occurs. Onsager also showed how using an exact equation for the critical temperature such as eq. (41) the values of θ^c or T^c for a 2D lattice with two first nearest neigbor interactions could be calculated exactly, given the ratio of the two bond energies or the bond energies themselves respectively. Also for other simple trigonal or hexagonal 2D lattices exact equations from which θ^c or T^c could be calculated, were derived by Onsager [11].

Within the framework of theories of crystal surfaces and crystal growth this critical phase transition is called, nowadays, the critical roughening transition and the corresponding dimensionless roughening temperature are indicated as θ^R and the real roughening temperature as T^R.

In another pioneering paper on the fundamentals of statistical mechanical models of crystal surfaces, steps, 2D nucleation models and the resulting spiral growth theory Burton, Cabrera and Frank used the order disorder phase transition theory of Onsager as a basis for the development of theories of crystal surfaces and crystal growth [12]. They were able to simplify the very complex theory of Onsager. Moreover, the theory could be generalized to all kinds of complex 2D lattice, which correspond to the 2D interfaces of crystals or better their corresponding crystal gaphs. A formalism for this was developed by Rijpkema and Knops and for the first time succesfully applied to faces *(hkl)* of the complex garnet structure [13] and many other structures. We refer to the survey paper [14] and the references therein.

6.3.4. Roughening temperature for the SOS step model and onsager model

Looking at eq. (33) it can be seen that within the framework of the SOS step model the critical roughening phase transition can, to a certain approximation, be understood to occur if the broken bond step energy ϕ_b for the step along the a direction becomes equal to the step entropy times the temperature or when

$$\gamma_x = \phi_b/a - kT/aS_{\text{step}} = 0, \tag{42}$$

where S is the step entropy per unit length. An analogous expression can be formulated for a step along the $\boldsymbol{b}$ direction. In this approximation the step entropy is interpreted as

$$S_{\text{step}} = kT/a \ln \coth \phi_a/2kT. \tag{43}$$

For very strong bond energies as compared to the thermal energy $\phi_a/2kT \to \infty$ and $S_{\text{step}} \to 0$. This implies that the step becomes a straight step parallel to the $\boldsymbol{a}$ direction.

For a square lattice which implies that $a = b$ and $\phi_a = \phi_b = \phi$ we find from eq. (33)

$$\phi/kT = \ln\left[(\exp \phi/2kT + \exp -\phi/2kT)/(\exp \phi/2kT - \exp -\phi/2kT)\right] \tag{44}$$

which becomes after taking the exp of the equation and multiplying numerator and denominator and multiplying with $\exp \phi/2kT$ the quadratic equation

$$\exp 2\phi/kT - 2 \exp \phi/kT - 1 = 0. \tag{45}$$

The relevant positive root of this equation is $\exp \phi/kT = 1 + 2^{1/2}$, leading to

$$(\phi/kT)^{\text{c}} = \ln\left(1 + 2^{1/2}\right) = 0.8814. \tag{46}$$

The same solution is obtained from the exact equation of Onsager, eq. (41), applied to a square lattice with $\phi_x = \phi_y$.

6.3.5. Faceted rough and rounded-off faces

The habit or morphology of crystals plays a very important role in crystallization processes. Often crystals grow faceted with flat well-defined faces, but in some cases some of the faces appear as rounded-off roughened faces. In these cases the growth velocity of such faces is, in principle, much higher than the flat faces. Although the case of faceted crystals is mainly treated in this contribution, it turns out that the morphology of many crystals is to a large extent determined by the tendency of some of their faces to roughen easily. Closely connected to this tendency is the roughening temperature of a flat face above which they appear as rounded-off. This effect is known as thermal roughening. The growth velocity of these faces has already become high below their critical temperature. Also, as a result of its roughening temperature a flat face can become rough beyond a critical supersaturation, known as kinetic roughening, resulting in relatively high growth velocities. We refer to the survey paper [14], an experimental paper [15] and a recent paper of van Veenendaal *et al.* [16]. The implications of the concept of roughening transition to explain crystal growth phenomena together with crystal morphology are very profound. Rough faces have generally speaking higher growth rates than flat faces; these faces often do not appear on the crystal morphology.

6.3.6. The Burton–Cabrera–Frank step edge and growth spiral theory

Half a century ago, Burton, Cabrera and Frank [12] developed the spiral growth theory, after Frank [17] described the, originally, purely theoretical concept of a growth spiral. In

the now famous BCF paper [12], rightly called the "Bible of the science of crystal growth", surface models and step models, expressed in the statistical mechanical language of Ising models applied to the (001) face of a Kossel crystal, were logically integrated with the world of 2D nucleation theories and the world of (screw) dislocations. In the BCF theory, Onsager's theory [11] of the order-disorder phase transition in a 2D Ising model played a key role. As mentioned above Onsager's 2D order–disorder phase transition is, nowadays, interpreted as the roughening transition. This can be decribed in terms of the dimensionless critical temperature defined in Section 6.3.3

$$\theta < \theta^{c} \rightarrow \gamma > 0; \qquad \theta \geqslant \theta^{c} \rightarrow \gamma = 0. \tag{47}$$

The dimensionless critical emperature, θ^{c}, depends on the typical topology of the broken bond energies and bond structure of the surface *(hkl)* under consideration and its interaction with the mother phase; γ is the edge free energy of a step on that surface. The dimensionless temperatures are given by

$$\theta = 2kT/\phi, \tag{48}$$

where ϕ is a generalized bond energy, having the shape

$$\phi = \phi^{sf} - 1/2\left(\phi^{ss} + \phi^{ff}\right). \tag{49}$$

The superscript sf refers to the solid–fluid (in general, mother phase) bonds, the superscript ss to the solid–solid bonds within the crystal and the superscript ff to the fluid-fluid bonds with in the mother phase. Conventionally, θ and θ^{c} are expressed in reference to the strongest bond ϕ of the crystal. The reader is further referred to the survey papers written by Bennema [8] and Bennema and van der Eerden [14]. It is possible to calculate the (relative) order–disorder or Ising temperatures θ^{c} even for complex 2D Ising lattices consisting of (many) different cells [13,14].

It should be stressed that the concept of roughening expressed in eq. (47) (cf. eqs (33) and (34)) is the integrating principle of modern crystal growth theories. This concept explains that if on a crystal surface in at least one direction *[uvw]* the step edge free energy $\gamma_{[uvw]} = 0$ that crystal face grows as a microscopically rough and macroscopically rounded off or, in general, unstable face without any crystallographic orientation. The reason for this is that due to the zero edge free energy the formation of 2D nuclei on the surface will cost no energy and will, therefore, occur unbridledly. For temperatures at which there is no such condition, that is, for which in all directions *[uvw]* on the surface the step edge free energy $\gamma_{[uvw]} > 0$ the surface will grow as a flat face with a layer mechanism, i.e., a spiral growth or 2D nucleation mechanism. We refer here to a survey paper of Chernov and Nishinaga [18]. The authors show that in case of a spiral growth step mechanism differences in supersaturations between corners and centers (with spirals), which occur at crystal surfaces and are caused by volume diffusion, which, in principle, give non-faceted unstable growth forms, are suppressed. For most crystals, roughened facets are rarely observed, simply because they will grow out of the morphology leaving behind other flat orientations. Yet, the concept of roughening transition is of utmost importance for understanding the growth behaviour of all crystal faces and, therefore, the resulting crystal morphology.

6.3.7. *Computer simulations and recent experimental observations*

6.3.7.1. Computer simulations and roughening In Figure 6.4 the results of Monte Carlo computer simulations on a Kossel crystal by Gilmer and Bennema [19] and de Haan, *et al.* [20] are presented. In the figure the dimensionless growth rate is plotted against the dimensionless driving force for crystallization, $\beta = \Delta\mu/kT$, which is roughly equal to the relative supersaturation. Each curve corresponds to a value of $\alpha = 4\phi/kT$, which corresponds to a dimensionless reciprocal temperature. It can be seen that, the higher α or the lower the temperature, keeping the bond strength ϕ constant, the lower the growth rate. This is understandable because, the lower the temperature of the crystal surface, the less rough the surface will be on the atomic scale. Furthermore, the flatter the surface, the smaller is the fraction of growth units arriving at the surface that will actually stick. The most important finding of the simulations is the change from linear to non-linear (R versus β) curves. It is now well known that this change marks the roughening transition, which corresponds for the (001) surface to a value of $\alpha \approx 3.2$.

These results from computer simulations inspired theoreticians such Gilmer [21], Knops [22] and van der Eerden and Knops [23], Weeks and Gilmer [24] and van Beijeren [25] to develop a more sophisticated roughening transition theory for the relatively simple solid-on-solid (SOS) model of the (001) surface of the Ising–Kossel model. The SOS model implies in this case that, analogous to the SOS model for a 2D step model treated above, the approximation is introduced that the (001) surface consists of columns of cells perpendicular to the surface, with heights h which vary again in the same way as the SOS step model. Again, overhangs are forbidden. Van Beijeren [25] developed an alternative body-centered solid-on-solid (BCSOS) model. He proved that also for this model a

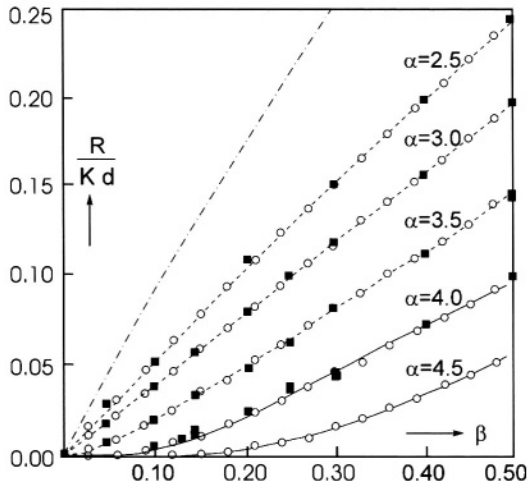


FIGURE 6.4. Results from Monte Carlo simulations for several values of the reciprocal temperature $\alpha = 4\phi/kT$. Plotted is the dimensionless growth rate R/Kd versus $\beta = \Delta\mu/kT$; K is the reciprocal reaction time unit and d the interplanar thickness. Full squares are results from a general computer [19], open circles are special-purpose computer results [20]; solid lines represent theoretical two-dimensional birth-and-spread nucleation curves; dashed lines empirical fitting curves. The uppermost line represents the maximal rate of growth (of a kinked or rough surface). This is the so-called Wilson–Frenkel law.

roughening transition of infinite order occurs. Moreover, he could calculate the roughening temperature for this model exactly.

The roughening temperature as determined for the (001) face of the SOS simple cubic or Kossel model from computer simulation data, is somewhat higher than the Onsager–BCF roughening temperature. The characters of both roughening phase transitions are also different. The SOS surface has a very smooth phase transition of infinite order while the Onsager 2D surface model shows a second-order phase transition. For the practice of studying crystal growth phenomena the subtle details of the precise character of the roughening transition may be not very important. It is, however, most important that different models do give roughening temperatures, which are close to each other.

It should be mentioned that recently new developments in the theory of critical phase transitions between surface phases in Ising surface models have been derived by Mazzeo *et al.* [26], As compared with the original BCSOS model of van Beieren, these authors discovered two new surface phases in the BCSOS-CsCl model, namely a disordered flat (DOF) phase and in the case of repulsive bonds a reconstructed phase. A transition from a normal flat surface phase to a DOF-like surface phase was to a high a extent, both experimentally and theoretically verified for the (011) face of naphthalene crystals growing from the vapour phase by Grimbergen *et al.* [27].

6.3.7.2. Implications of simulations for crystal growth mechanisms In the following we will focus the attention on the implications of computer simulations for understanding crystal growth kinetics. The non-linear curves in Figure 6.4, for values of α somewhat larger than 3.2, can be fitted very well with analytical 2D nucleation birth-and-spread curves. A birth-and-spread mechanism implies that 2D nuclei are formed by statistical fluctuations; these nuclei grow (spread) and form expanding terraces while on top of these terraces new nuclei are formed. Above the roughening temperature the edge free energy becomes zero so that the barrier for 2D nucleation vanishes. It can be seen from Figure 6.4 that below the roughening temperature at the lowest supersaturations the crystal surface in fact does not grow. There is a 'dead zone' for which the two-dimensional nucleation barrier is too high. The higher α, that is, the higher the bond energy or the lower the temperature, the larger is the dead zone.

Gilmer [21] showed that by introducing dislocations on surfaces which are below their roughening temperature that, in agreement with the BCF spiral growth theory, in the dead zones parabolic rate versus supersaturation curves developed.

Since the development of the BCF theory, surface Ising models have shown to be extremely fruitful to model crystal surfaces and crystal growth mechanisms. However, these models suffer from the *ad hoc* approximation that GU's at the surface have only two properties: they are in either a solid state or a fluid state. This latter state is a not well-defined average solute state of a mixture of solute and solvent molecules at the surface. To overcome these limitations of Ising models, simulations of crystal surfaces have been carried out using Molecular Dynamics (MD) techniques. Recently Huitema *et al.* [28] obtained from MD studies rate versus supersaturation curves for the octahedron and cube faces of a FCC crystal, consisting of spherical molecules with van der Waals first nearest neighbor bonds. The measured rate versus supersaturation curves for these MD simulations are very similar to the Monte Carlo curves as presented in Figure 6.4. This gives a justification for Monte Carlo simulation studies of Ising systems.

The BCF–Ising–spiral growth model is now, after half a century of theoretical and experimental research, directly as well as indirectly confirmed for a very large variety of crys-

tals growing from very different mother phases. During the last ten years both *in situ* and *ex situ* observations were made with advanced optical techniques and additional modern techniques, such as Atomic Force Microscopy (AFM). It is within this context interesting to mention just one example, namely the recent paper of Tsukamoto *et al.* [29]. Using highly advanced optical techniques the {001} surfaces of barium nitrate crystals, growing from aqueous solutions were examined. Growth spirals could be seen *in situ* on surfaces with dislocations and simultaneously growth rate versus supersaturation curves could be measured on surfaces with and without screw dislocations. The results are in perfect agreement with the BCF spiral growth theory and simulations. On dislocation free surfaces carried out under microgravity conditions, i.e., in space, a kind of birth-and-spread 2D nucleation mechanism could be observed *in situ.* In addition, a growth rate versus supersaturation curve was measured, which was in agreement with the theoretical 2D nucleation birth-and-spread model. This agrees in turn with results obtained from Monte Carlo simulations of Ising surfaces and the recent MD studies by Huitema *et al.* Tsukamoto *et al.* were so succesful because growth experiments were carried out under microgravity conditions. In microgravity, turbulence in the solution is absent; hence impurities present in the solution, cannot reach the crystal surface. So heterogeneous 2D nucleation is suppressed and homogeneous 2D nucleation can be studied.

6.4. CHEMICAL BONDS AND THE PREDICTION OF MORPHOLOGY. CRYSTAL GRAPH, CONNECTED NET AND HARTMAN–PERDOK THEORY

6.4.1. Introduction

So far, only crystallography and Ising interface models have been discussed emphasizing the importance of the two parameters interplanar distance and roughening transition temperature. To show the relevance of these parameters for any kind of crystal structure with the aim to predict the crystal morphology—in principle also in dependence of parameters like temperature or supersaturation—belonging to this structure we start in this Section from the crystal graph and give the procedure to derive the connected nets which are essential to understand the tendency of orientations to become flat. For the time being, a connected net can be seen as a net parallel to an orientation *(hkl)* having the interplanar thickness and consisting of uninterrupted chains of bonds of the crystal graph. For a review on crystal graphs and connected nets we refer to Grimbergen *et al.* [31].

6.4.2. The importance of the interplanar thickness for connected nets

In the following we will explore the importance of the interplanar thickness d_{hkl} for connected nets in some more detail. Assume that we have a crystal graph of a particular crystal structure and that we want to determine the connected nets corresponding to a certain crystallographic orientation *(hkl).* Let us assume that we can find a cut from the crystal graph parallel to the orientation *(hkl).* The boundary of this cut is in most cases not flat but wavy. Then, due to the crystallographic periodicity discussed in Section 6.2 after a repeat distance, being the interplanar distance d_{hkl}, exactly the same cut occurs. Note, that a surface in contact with its mother phase will be deformed as compared with a corresponding hypothetical cut in the bulk of the crystal. However if a crystal face with an

orientation *(hkl)* is growing below its roughening temperature as a flat face each time that a previous layer is covered with a new one exactly a same deformed layer will occur. Due to thermodynamic-statistical surface processes the same surface phases are formed on each newly formed layer. It is within this context important to note that the thermodynamic and statistical mechanical equilibrium processes to form a surface phase between crystal and mother phase is much faster than the process of a moving step.

6.4.3. Procedure to derive connected nets

6.4.3.1. Determination of growth units and bonds from the crystal structure After studying carefully a known crystal structure first of all the Growth Units (GU's) have to be determined. Growth units are molecules, complexes, ions, etc., occurring in the mother phase from which the crystal grows. Bonds within the growth units are not relevant for the crystal growth process. The attention is only focused on the overall bonds between growth units. These bonds are the bonds formed during the crystallization process. The difference between the energy of these bonds and the energy of the bonds in the mother phase determines the crystallization enthalpy. We consider bonds in the bulk of the solid and mother phase. These bonds may consist of many atom-atom pair potentials of atoms being part of separate growth units. Growth units are defined after considering the information of the complexes occurring in the mother phase and comparing these complexes with the crystal structure. Very often, growth units are not well known and then *ad hoc* hypotheses concerning growth units can be made. Also, alternative growth units can be chosen having mutually different bonds, which may lead to different morphologies. Comparing the alternative theoretical morphologies with the real one, conclusions concerning the structure of growth units in the mother phase may be drawn a posteriori.

When the growth units are defined, interaction energies in the bulk solid referenced to vacuum, ϕ_{ij}^{ss}, between growth units i and j have to be determined. In the past in the case of crystal structures of organic crystals like naphthalene, paraffin, fat-crystals, etc., overall bonds between molecules were calculated by adding atom pair potentials, occurring between molecules. For each of these pairs overall bond energies were calculated from reliable Lennard-Jones or Buckingham-like potentials, based on precise atomic distances which follow from the known crystal structure. Nowadays, software based on force fields is used to minimize the energy of crystal structures, with respect to the atomic positions and, if necessary, unit cell dimensions, by using Molecular Mechanics (MM). Overall bond energies between growth units can be calculated based on the, thus, optimized crystal structure for the force field used. In this way, for a certain crystal structure broken bond energies $(\phi_{ij}^{ss})_k$, where k stands for the first, second etc., are calculated as broken bond energy. These energies are broken bond energies referenced to vacuum. It can be cheched whether the sum of of the calculated bond energies per stoichiometric unit equals the measured sublimation enthalpy. In most cases crystals do not grow from a dilute vapor but from (concentrated) solutions. In that case the so-called proportionality approximation may be introduced, which implies that the ratio of bond energies referenced to vacuum and to a certain saturated mother phase are supposed to be the same. In order to obtain an approximation of the actual bond energies, the sum of the bonds energies may be put equal to the enthalpy of dissolution of the saturated solution from which the crystal grows.This is called the equivalent wetting principle [8]. Deviations from this principle may be identified and

further studied in detail; see the survey paper of Liu and Bennema [30]. Alternatively, the approach of eq. (49) can be used to determine effective overall bond energies.

6.4.3.2. Determination of the crystal graph Once the set of overall bonds between growth units are selected a crystal graph is defined. A crystal graph is defined as a 3D translationally symmetrical infinite set of points connected by the bonds between the growth units. In order to define the crystal graph of the crystal structure under consideration, the GU's are reduced to points, e.g. the centers of gravity of the GU's. The bonds of the crystal graph are chosen according to the methods outlined in the previous section. The set of bonds is limited by a maximum bond energy, which is usually set to the thermal energy kT or a fraction of it. This choice is disputable, but one can always enlarge the bond set to verify whether this has influence on the eventual morphology prediction. In most cases a limit of $\frac{1}{2}kT$ suffices. It is important to realize that for the crystal graph it is not necessary to maintain the point group symmetry; only the tranlational symmetry is essential.

6.4.3.3. Determination of connected nets From the thus-defined crystal graph the connected nets have to be determined. For complex crystal structures this can be a tedious job; see the paper of Rijpkema *et al.* [13] and the survey paper of Bennema and van der Eerden [14]. Nowadays, computer programs are used for this purpose, using software like FACELIFT [32] based on the graph theoretical method developed by Strom [33]. It is, moreover, possible to use software to visualize connected nets and also combinations of different connected nets for one orientation *(hkl).* We will, now, give a more precise definition of a connected net. A connected net *(hkl)* is a net of GU's connected by the bonds of the crystal graph parallel to the orientation *(hkl)* with a surface translational unit cell wich has a mesh area equal to the volume V of the crystallograhic unit cell divided by the thickness the interplanar distance d_{hkl}, corrected for the selection rules of the space group as treated in Section 6.2.8. Moreover, a connected net consists of at least two non-parallel uninterrupted chains of bonds of the crystal graph, each having an overall direction and periodicity of a vector [*uvw*] that have at least one growth unit in common. This latter condition, actually, makes the net connected. These chains are called Periodic Bond Chains (PBC's). This definition of a connected net implies that in all directions [*uvw*] coplanar with the connected net *(hkl)* all broken bond energies edge energies are larger than zero. This is, for example, obvious for the simple net of the (001) face of Figure 6.3. It can be seen that for this case because of the mere connected PBC's [100] and [010] the face with the orientation (001) has a single connnected net (001).

Very often connected nets may be quite complex and consist of many PBC's. Also for one orientation *(hkl)* many different connected nets may occur. Note, that the job aimed to understand the topology of a stack of alternative connected nets remains a tedious one. Therefore, recently a new tool was developed to derive the habits of crystals, based on Monte Carlo Simulation of the growth process using the crystal graph, also in dependence of the driving force for crystallization. We will come back to this in Section 6.5.

The essence of connected nets lies in their connectedness. Only nets that are connected according to the definition given in this Section show an order-disorder phase transition at a finite temperature. As soon as for a net at least one direction [*uvw*] occurs with an edge energy which is equal to zero, this net has a roughening transition temperature of 0 K. Non-connected nets occur especially for indices *(hkl)* that are large and for which d_{hkl} is, usually, small. Note, that here is relation between this latter rule of thumb and the BFDH law of Section 6.2.8; it can roughly be stated that the higher the indices h, k, l the thinner the

nets of the corresponding orientation and the larger the chance that within the interplanar thickness d_{hkl} no connected can be constructed.

6.4.3.4. Edge free energies in relation to connected nets Up to about 1998, in case more than one connected net could be identified for one orientation *(hkl)* the strongest connected net was considered as the dominant one and for this net the Ising temperature was calculated. Since that time Grimbergen *et al.* [31] discovered, realizing that the edge free energy of the weakest steps for an orientation *(hkl)* is both from the point of view of the roughening transition as well as from the point of view of a 2D nucleation process the crucial parameter, that quite often an orientation can have a much lower roughening temperature than expected on the basis of the bond strengths in the connected net. In such cases the edge free energy turns out be determined by differences in bond strengths rather than by the bond strengths of the connected net. In case of the presence of more than one connected net causing such low edge free energies for a certain orientation an alternative substitute connected can be defined. Such a substitute net mimics in certain directions [*uvw*] the relatively low edge free energy. This is demonstrated in Figure 6.5.

Figure 6.5a shows the crystal graph of a crystal structure that is only slightly more complicated than the Kossel crystal. The unit cell contains two growth units labeled *A* and *B* that are mutually connected by oblique bonds *p* and *q* with bond energies ϕ_p and ϕ_q, respectively. Along the *a* axes rows of either *A* or *B* occur that are connected via bonds *a*. The (001) face of this crystal is shown in Figure 6.5b projected along the direction [100]. Thus, only the bonds *p* and *q* can be seen in this figure. There are two connected nets for this orientation that are indicated as slices that are delimited by the dashed lines in

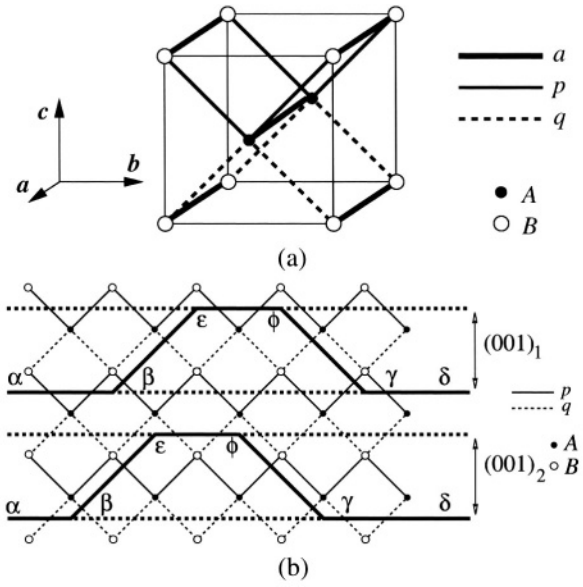


FIGURE 6.5. (a) Model crystal graph with two growth units *A* and *B* connected by oblique bonds *p* and *q* in the (100) plane. (b) The two connected nets for the (001) orientation $(001)_1$ and $(001)_2$ seen in the [100] projection, that is, along the *a*-bonds. The profiles of 2D nuclei are indicated as $\alpha\beta\varepsilon\phi\gamma\delta$ as compared to the surface $\alpha\beta\gamma\delta$ without a nucleus. The broken bond edge energy along the [100] direction of the nuclei is $2(\phi_q - \phi_p)$ for $(001)_1$ and $2(\phi_p - \phi_q)$ for $(001)_2$.

Figure 6.5b. For each of these connected nets a surface profile with a 2D nucleus on top of it is drawn by heavy solid lines. The profiles of these nuclei are chosen such that the edge energy along the projection direction of the nucleus is the smallest possible. If we assume that the energy ϕ_p of the p bond is smaller than ϕ_q of the q bond then the upper connected net $(001)_1$ in the figure will be the most stable one. We will now focus on this connected net; for the reverse case, an analogous reasoning can be found for the lower connected net $(001)_2$. The edge energy of the 2D nucleus along the [100] direction for $(001)_1$ is found by determining the difference in broken bond energy between the surface profile with the nucleus and the one without the nucleus, that is, the difference in broken bond energy for the profiles $\alpha\beta\varepsilon\varphi\gamma\delta$ and $\alpha\beta\gamma\delta$ for the upper connected net. For the 2D nucleus drawn this amounts to $(6\phi_p + 2\phi_q) - 8\phi_p = 2(\phi_q - \phi_p)$. Note, that this edge energy does not change when we enlarge the nucleus along the [010] direction; it is, therefore, independent of the number of GU's in the nucleus. The edge energy along the [010] direction, on the other hand, is $2\phi_a$ per GU. If we consider the case of a crystal structure for which the bond strenghths p and q are almost the same the edge energy $2(\phi_q - \phi_p)$ along the [100] direction can be much smaller than the bond energies ϕ_p and ϕ_q that make up the connected net. Thus, the roughening temperature for the (001) face will be much lower than expected on the basis of the bond energied making up the connected net.

In this case, which is comparable to the case of the (011) face of naphthalene [27], the two alternative connected nets can be substituted by a single alternative rectangular connected net with a bond a and a bond with energy $(\phi_q - \phi_p)$ also resulting in a relatively small roughening temperature. A very special case occurs when the bond strengths ϕ_p and ϕ_q are equal due to space group symmetry. Then the edge energy becomes zero and the face will have a roughening temperature of 0 K. Such a situation is called symmetry roughening. The orientations for which the space group can lead to symmetry roughening were studied by Meekes *et al.* [34].

Also, in case of orientations with no symmetry releated connected nets or even single connected nets the edge energy of 2D nuclei can be lower than expected on the basis of the bond strengths in the connected nets. We will come back to this issue in Section 6.4.

6.4.4. Hartman–Perdok theory

In the original Hartman–Perdok theory [35] crystallography played an essential role. The concept of roughening of crystal faces, however, was not part of that approach. In the Hartman–Perdok theory first the concept of PBC, defined as an uninterrupted chain of strong bonds, having the overall direction and periodicity [*uvw*] of the lattice of the crystal graph was introduced. From a modern point of view it can be said that the strong bonds are the bonds selected for the crystal graph. Then the, originally, introduced F, S and K faces are from a modern point of view defined in the following way. An F (Flat) face *(hkl)* is parallel to a connected net, which consists of at least two PBC's, which within the overall thickness d_{hkl} are connected to each other by bonds of the crystal graph. Thus, an F face has a finite roughening temperature. An S (Stepped) face is parallel to a non-connected net consisting of merely one PBC and is expected to be rough. The same holds for a K (Kinked) face which is parallel to a non-connected net consisting of no PBC.

6.5. MODERN MORPHOLOGY PREDICTIONS

In this section we dwell further on the situations for which the edge free energy of 2D nuclei is smaller than expected on the basis of bonds in the connected net as explained in a Section 6.4.3.4.

One of the drawbacks of conventional morphology prediction methods is that they do not take into account parameters like supersaturation or impurities, either in the form of the solvent, deliberately added impurities (tailor-made additives) or other impurities. The BFDH-method treated in Sections 6.2.8 and 6.2.9 is based on the assumption that the stability of a crystal orientation depends on the thickness of the corresponding growth slice as a measure for the binding energy of that slice, corrected for the selection rules described in Section 6.2.8. An alternative method that is used almost invariably, nowadays, is the attachment energy method. The atachment energy method determines the binding energy in the growth slice, E^{slice}, and uses the relation

$$E^{\text{cryst}} = E^{\text{slice}} + E^{\text{att}}, \tag{50}$$

where E^{cryst} is the crystallisation energy which is constant for a certain crystal structure and E^{att} is the attachment energy that is released when a growth slice is attached to the surface. All these energies are usually expressed per GU. The larger the slice energy the smaller the attachment energy and, therefore, the smaller the growth rate. The attachment energy method, in fact, is identical to the BFDH-method for isotropic crystal structures. For non-isotropic structures the method takes better account of the actual bonding energy in the growth slice. This method has been related to various growth mechanisms by Hartman and Bennema [36]. Both the BFDH and attachment energy methods are based on the growth slice only, thus, neglecting the effect of supersaturation, impurities and further details of the crystal structure. The effect of impurities has been studied in great detail by Lahav, Leiserowitz, Weissbuch and co-workers [37]. We will not treat impurities in this contribution.

The effect of supersaturation becomes eminent when one considers the analytical expressions for crystal growth mechanisms [38]. In all these expressions there are two essential parameters namely the driving force for crystallisation $\Delta\mu$ and the edge free energy γ of a 2D nucleus on the surface. The driving force which is directly related to the supersaturation, obviously, is determined by experimental conditions. The edge energy, on the other hand, depends on the crystallography of the crystal face under consideration. Moreover, the γ depends on the orientation of the steps that determine the 2D nucleus. If in any direction along the surface $\gamma = 0$ the crystal orientation will grow as a rough face and, therefore, generally will grow out of the crystal morphology as was discussed in Section 6.3.5. Therefore, a study of the edge (free) energy on a crystal face gives more direct information on the growth behaviour as compared to the attachment energy. In Section 6.4.3.4 an example was given where the edge free energy was much smaller than expected on the basis of the slice energy. This example was an inspiration for further studies on actual crystals for which the morphology could not be explained by the attachment energy morphology. One, not so obvious, example of naphthalene was already mentioned [27]. In the case of fat crystals the elongated plank-like habit was explained by carefully studying the edge free energies as a function of the length of the alkane chains of the fat molecules [39]. It turned out that the

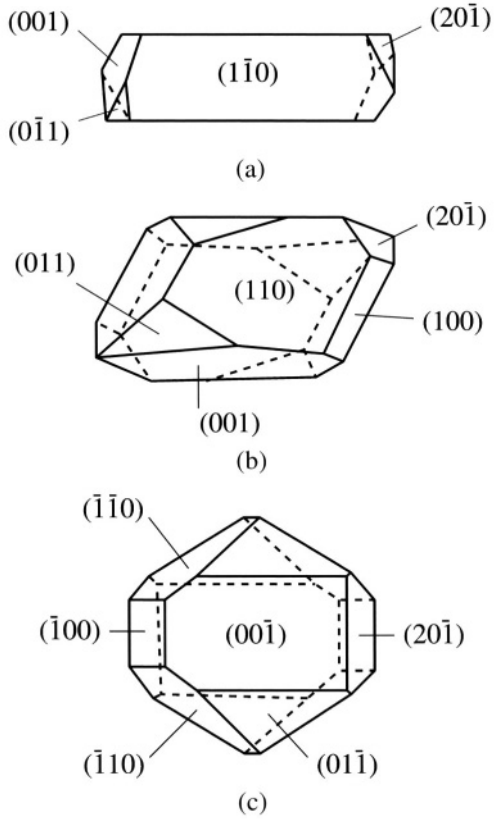


FIGURE 6.6. Three experimental morphologies of monoclinic paracetamol grown at (a) low, (b) medium and (c) high supersaturation in aqueous solution.

fast growing top faces of the crystals could be explained in great detail by the unexpectedly small edege energies of 2D nuclei on these faces. Moreover, by comparing three different fat crystal structures with diverse crystal habits the explanation for the diversity was found in the details of the step edges of the relevant faces, again, leading to unexpectedly small edge energies for the fast growing faces [40].

The complexity of crystal graphs and the corresponding surfaces led to the development of a Monte Carlo algorithm called MONTY based on the actual crystal graph which opens ways to study the growth behaviour of real crystal surfaces [41]. The formation of 2D nuclei and the propagation of spirals on crystal surfaces can, thus, be followed in detail. Growth rate versus supersaturation curves can be simulated and compared to analytical crystal growth expressions based on more simple models. Edge free energies of growth steps can be determined from the simulation results, leading to estimates for the roughening temperatures of the crystal faces. MONTY has been used to explain the growth behaviour and, especially, the morphology of various crystals ranging from ionic structures like Gibbsite, $Al(OH)_3$ [42] to molecular crystals like fat crystals [41,43] and paracetamol [44]. The latter case of paracetamol is illustrated in Figures 6.6 and 6.7.

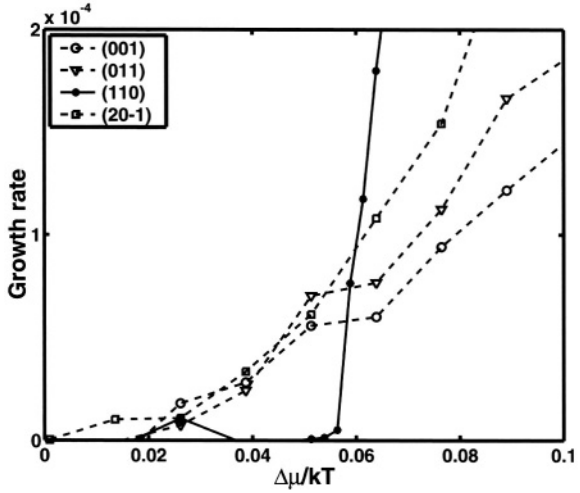


FIGURE 6.7. The growth rate versus supersaturation curves for four faces of monoclinic paracetamol resulting from the Monte Carlo simulations using MONTY.

Figure 6.6 shows three experimental morphologies of monoclinic paracetamol crystals [45]. The crystal habit changes with increasing supersaturation from needle like to a more block-like form. This supersaturation dependence cannot be predicted using the attachment energy method. Figure 6.7 gives the growth rate versus supersaturation curves determined by Monte Carlo simulations on the crystal graph of paracetamol using calculated bond energies. For these simulations it was assumed that the {110} faces grow via a 2D nucleation mechanism, while the other faces grow with a spiral growth mechanism. This assumption was based on the observations of Ristic *et al.* [45]. The {110} growth rate curve intersects the ones of the other faces leading to the supersaturation dependent morphology of Figure 6.6. For further details we refer to the paper [44].

6.6. CONCLUSION

In this contribution we have shown that the more than two centuries old science of crystal morphology is still very vivid. Conventional morphology predictions like the BFDH and attachment energy methods have been used for decades with reasonable success. One of the drawbacks of these methods is that they cannot explain the very often observed phenomenon of a supersaturation dependence of crystal morphology. Moreover, roughening effects which are essential to understand crystal growth behaviour are often neglected. Nevertheless, one of the essential parameters in conventional morphology predictions, the interplanar distance, is still considered to be essential. Therefore, the concept of interplanar distance is treated in detail in Section 6.2. Roughening of crystal faces, the subject of Section 6.3, is intimately related to the edge free energy of steps on crystal surfaces. During the last five years it was realised that the edge free energy of a crystal face is not always, and in fact quite often not, simply determined by the slice energy of the corresponding growth slice. Therefore, detailed studies of the edge free energies are essential to understand crystal growth behaviour. This is explained in Section 6.3. For that study the crystal

graph is a very useful tool as explained in Section 6.4. Real crystals, however, usually have rather complicated crystal graphs. Then, Monte Carlo simulations on real crystals as treated in Section 6.5 provide a valuable tool to study the growth behaviour. To understand the results of these simulations the crystal growth mechanism on real crystals has to be confronted with existing crystal growth theories which are almost always based on very simple crystal models. It will be a challenge for the future to generalise these theories to real crystal surfaces.

REFERENCES

1. J.G. Burke, *Origins of the Science of Crystals* (University of California Press, Berkeley and Los Angeles, 1966).
2. C.J. Schneer, Ed., *Crystal Forms and Structure* (Hutchinson and Ross, Dowden, 1977).
3. R. Hooykaas, *Geschiedenis der Natuurwetenschappen, A* (in Dutch) (Oosthoek Uitgeversmaatschappij N. V Utrecht, 1971) pp. 176–178 and 235–239.
4. T. Janssen, *Crystallographic Groups* (North Holland/American Elsevier, 1973).
5. M.J. Buerger, *Elementary Crystallography* (John Wiley & Sons. Inc., New York, London, 1963).
6. *International Tables for X-ray Crystallography,* Volume A, *Space-Group Symmetry,* Ed. Theo Hahn, fourth revised edition (Kluwer Academic Publishers, Dordrecht/Boston/London, 1996).
7. (a) F.C. Phillips, *An Introduction to Crystallography* (Longman, London, 1963); (b) J.-J. Rousseau, *Basic Crystallography* (John Wiley & Sons, Inc., New York, 1999); (c) C. Hammond, *The Basics of Crystallography and Diffraction* (Oxford University Press, 1997).
8. P. Bennema, in: *Handbook of Crystal Growth 1a Fundamentals, Thermodynamics and Kinetics,* Ed. D.T.J. Hurle (Elsevier, Amsterdam, 1993) pp. 477–581.
9. J.D.H. Donnay, *Spherical Trigonometry after the Cesàro Method* (Interscience Publishers, Inc. – New York, N.Y., 1945).
10. H.J. Leamy, G.H. Gilmer and K.A. Jackson, in: *Surface physics Materials,* Volume 1 (Academic Press, 1975).
11. L. Onsager, Crystal Statistics 1, *Phys. Rev.* **65**, 117–149 (1944).
12. W.K. Burton, N. Cabrera and F.C. Frank, *Phil. Trans. R. Soc. Lond. A.* **243**, 299–358 (1951).
13. J.J.M. Rijpkema, H.J.F. Knops, P. Bennema and J.P. van der Eerden, Determination of the Ising critical temperature of F-slices with an application to garnet, *J. Crystal Growth* **61**, 295–306 (1982).
14. P. Bennema and J.P. van der Eerden, in: *Morphology of Crystals,* Ed. Sunagawa (Terra Scientific Publishing Company (TERRAPUB), Tokyo, 1987) pp. 1–75.
15. L.A.M.J. Jetten, H.J. Human, P. Bennema and J.P. van der Eerden, On the observation of the roughening transition of organic crystals, growing from solution, *J. Crystal Growth* **68**, 503–516 (1984).
16. E. van Veenendaal, P.J.C.M. van Hoof, J. van Suchtelen, W.J.P. van Enckevort and P. Bennema, *Surface Science* **417,** 121 (1998).
17. F.C. Frank, *Disc. Farad. Soc.* **5**, 48–54 (1949).
18. A.A. Chernov and T. Nishinaga in: *Morphology of Crystals,* Ed. Sunagawa (Terra Scientific Publishing Company (TERRAPUB), Tokyo, 1987) pp. 207–267.
19. G.H. Gilmer and P. Bennema, *J. Appl. Physics* 43, 1347–1360 (1972).
20. S.W.H. de Haan, V.J.A. Meeuwsen, B.P. Veltman, P. Bennema, C. van Leeuwen and G.H. Gilmer, *J. Crystal Growth* **24**, 491–494 (1974).
21. G.H. Gilmer, Growth on imperfect crystal faces, *J. Crystal Growth* **35**, 15 28 (1976).
22. H.J.F. Knops, *Phys. Rev. Lett.* **39**, 766–769 (1977).
23. J.P. van der Eerden and H.J.F. Knops, *Phys. Lett.* **38**, 993–996 (1977).
24. J.D. Weeks and G.H. Gilmer, *Adv. Chem. Physics* **4**, 157–228 (1979).
25. H. van Beijeren, *Physics Lett.* **38**, 993–996 (1977).
26. G. Mazzeo, E. Carlon and H. van Beijeren, *Phys. Rev. Lett.* **74**, 5258–5365 (1995).
27. R.F.P. Grimbergen, H. Meekes, P. Bennema, H.J.F. Knops and M. den Nijs, *Phys. Rev. B* **58**, 5258–5265 (1998).
28. H.E.A. Huitema, B. van Hengstum and J.P. van der Eerden, *J. Chem. Phys.* **111**, 10248–10260 (1999).

29. K. Tsukamoto, E. Yokoyama, S. Maruyama, K. Maiwa, K. Shimizu, R.F. Sekerka, T.S. Morita and S. Yoda, *J. Jap. Soc. Microgravity Applic.* **1**, 6–11 (1998).
30. X.Y. Liu and P. Bennema, in: *Current Topics in Crystal Growth Research* **2**, 451–534 (1995).
31. R.F.P. Grimbergen, H. Meekes, P. Bennema, C.S. Strom and L.J.P. Vogels, On the prediction of crystal morphology. {I}. The Hartman–Perdok theory revisited, *Acta Cryst. A* **54**, 491–500 (1998).
32. S.X.M. Boerrigter, R.F.P. Grimbergen and H. Meekes, FACELIFT-2.50, a program for connected net analysis, Dept. of Solid State Chemistry, University of Nijmegen; e-mail: hugom@sci.kun.nl (2001).
33. C.S. Strom, *Z. Kritatlographie* **153**, 99–113 (1980); *ibid.* C.S. Strom, *Z. Kritallographie* **154**, 31–43 (1981); *ibid.* C.S. Strom, *Z. Kritallographie* **172**, 11–24 (1985).
34. H. Meekes, P. Bennema and R.F.P. Grimbergen, On the prediction of crystal morphology. {II}. Symmetry roughening of pairs of connected nets, *Acta Cryst. A* **54**, 501–510 (1998).
35. P. Hartman and W.G. Perdok, *Acta Crystallographica* **8**, 49–52 (1955); *ibid.* P. Hartman and W.G. Perdok, *Acta Crystallographica* **8**, 521–524 (1955); *ibid.* P. Hartman and W.G. Perdok, *Acta Crystallographica* **8**, 525–529 (1955).
36. P. Hartman and P. Bennema, The attachment energy as a habit controlling factor: I theoretical considerations, *J. Crystal Growth* **49**, 145–156 (1980).
37. I. Weissbuch, R. Popovitz-Biro, M. Lahav and L. Leiserowitz, *Acta Cryst. A* **51**, 115–148 (1995).
38. J.P. van der Eerden in: *Handbook of Crystal Growth 1a Fundamentals, Thermodynamics and Kinetics,* Ed. D.T.J. Hurle (Elsevier, Amsterdam, 1993) p. 317.
39. F.F.A. Hollander, S.X.M. Boerrigter, C.J. van de Streek, R.F.P. Grimbergen, H. Meekes and P. Bennema, The implication of the topology of connected nets on the morphology of β-monoacid triacylglycerol crystals, *J. Phys. Chem. B* **103**, 8301–8309 (1999).
40. F.F.A. Hollander, S.X.M. Boerrigter, J. van de Streek, P. Bennema, H. Meekes, J. Yano and K. Sato, Comparison of the morphology of β-$n.n.n$, β'-$n.n{+}2.n$ and β'-$n.n.n$-2 monoacid triacylglycerol crystals, *J. Phys. Chem. B* **107**, 5680–5689 (2003).
41. S.X.M. Boerrigter, G.P.H. Josten, J. van de Streek, F.F.A. Hollander, J. Los, P. Bennema and H. Meekes, MONTY: Monte Carlo crystal growth on any crystal structure in any crystallographic orientation: application to fats, *J. Phys. Chem. B,* submitted.
42. S.X.M. Boerrigter, F.F.A. Hollander, J. van de Streek, P. Bennema and H. Meekes, Explanation for the needle morphology of crystals applied to a β' triacylglycerol, *Crystal Growth & Design* **2**, 51–54 (2002).
43. C. Sweegers, S.X.M. Boerrigter, R.F.P. Grimbergen, H. Meekes, S. Fleming, I.D.K. Hiralal and A. Rijkeboer, Morphology prediction of gibbsite crystals—an explanation for the lozenge-shaped growth morphology, *J. Phys. Chem. B* **106**, 1004–1012 (2002).
44. S.X.M. Boerrigter, H.M. Cuppen, R.I. Ristic, J.N. Sherwood, P. Bennema and H. Meekes, Explanation for the supersaturation dependent morphology of monoclinic paracetamol, *Crystal Growth & Design* **2**, 357–361 (2002).
45. R.I. Ristic, S. Finnie, D.B. Sheen and J.N. Sherwood, Macro- and micromorphology of monoclinic paracetamol grown from pure aqueous solution, *J. Phys. Chem. B* **105**, 9057–9066 (2001).

This page intentionally left blank

7

Crystal dissolution kinetics studied by vertical scanning interferometry and Monte Carlo simulations: a brief review and outlook

Andreas Lüttge*
Department of Earth Science – MS126, Rice University, P.O. Box 1892, Houston, Texas 77251-1892, USA

Abstract: Various interferometric techniques have been used successfully to study crystal growth kinetics of relatively fast growing organic crystals. More recently, vertical scanning interferometry (VSI) has been applied to investigate dissolution and growth processes of slowly dissolving rock-forming minerals.

In this chapter, we present an integrated approach of analytical and computational methods to study mineral dissolution kinetics. VSI is a non-destructive, fast data acquisition method complementary to atomic force microscopy (AFM). It provides a large field of view and allows to quantify mineral (solid) surface topographies with up to 1 millimeter height difference. In addition, VSI is capable of precisely quantifying the processes that occur at the solid-fluid interface, i.e., dissolution, corrosion and growth processes. The use of a hybrid reactor allows the parallel investigation of mineral powders and single crystals at identical experimental conditions. These characteristics make VSI a centerpiece in our approach to link macroscopic with microscopic results.

At present, processes at the molecular scale can not be studied directly in mineral dissolution and growth reactions. Therefore, computational techniques like *ab initio,* density functional theory (DFT), and Monte Carlo simulations that treat crystal dissolution and growth as a many-body problem are important tools to gain insight in the reaction kinetics. Interferometry techniques provide the possibility to test and link the computational results to macroscopic observations. AFM can provide additional high-resolution information that complements the approach.

In our discussion, we first highlight some key questions of mineral dissolution, and focus particularly on the problem of quantification of surface area, which is important to normalize bulk rates determined from solution chemistry. Another topic is the dependence of the dissolution rate on the distance from equilibrium (ΔG). Next, we explain the VSI technique in some detail and, then, we present brief reviews of recent experimental and theoretical results of studies that used VSI and Monte Carlo techniques. Finally, we discuss our inte-

grated experimental and computational approach towards a fundamental understanding of crystal dissolution based on the available results and show some developments for future use in this important field of research.

7.1. INTRODUCTION

Crystal dissolution and growth processes occur at the molecular, i.e., the nanometer scale but they influence or control processes that occur at much larger scales, even at the global scale. The composition of earth's soils, waters, and atmosphere is largely determined by the submicroscopic reactions between minerals and aqueous solutions. And more recently, we have begun to understand the key role that (micro)organisms can play in these fluid–solid interactions (e.g., Banfield and Nealson [1]). Equally important, our climate is significantly affected by the same processes if we consider for example some mechanisms of the global carbon cycle and questions of carbon sequestration. Scientific advice for many environmentally crucial policy decisions with respect to soil protection, groundwater pollution, waste management, and carbon sequestration requires knowledge of reaction rates that is sufficiently accurate to permit extrapolation of reaction paths to up to 1000 years in the future. Therefore, a fundamental understanding of crystal dissolution is increasingly important and justifies efforts to understand and quantify the fluid–solid interactions occurring at the nanoscale.

During the last few decades, advanced techniques like electron microscopy, atomic-force microscopy (AFM), and various interferometry techniques including vertical scanning interferometry (VSI) have been developed and successfully applied to image the crystal surface and to quantify the processes at the fluid–solid interface. These techniques provide the ability to observe the crystal surface directly. In contrast, other experimental and analytical techniques seek to measure reaction rates indirectly, i.e., from changes in solution chemistry. This approach is often easier because the result of a microscopic process can be quantified at a larger scale over longer periods of time. In this case, rates are integrated over the entire mineral surface. Thus, these experiments cannot resolve the surface distribution of the integrated rate. By increasing the ratio of surface area to reactor volume, very slow rates are measurable that are characteristic for many minerals at earth surface conditions.

Independently of the scale at which the rates are obtained, they need to be extrapolated, compared, or applied to an often largely different scale (see discussion below). Therefore, a major goal of all experimental dissolution studies is the application of laboratory data to natural or technically engineered systems. However, transfer of laboratory data to natural systems is usually difficult and often unsuccessful (e.g., Velbel [2,31; White [4]). We suggest here that this scaling problem can be tackled eventually by an integrated approach based on VSI and AFM that link molecular modeling techniques like *ab initio* calculations in combination with density functional theory (DFT), and Monte Carlo simulations with macroscale analytical techniques. Although, we do not intend to imply that this approach will be the final solution of the longstanding problem, experimental investigations in the laboratory in combination with increasingly sophisticated computer simulations are pow-

*E-mail address: aluttge@rice.edu

erful tools to measure reaction rates and to improve our fundamental understanding of the dissolution process of crystalline matter.

Ab initio calculations, DFT (e.g., Lasaga [5,6], Cygan and Kubicki [7] and citations therein) and Monte Carlo techniques (e.g., Van der Eerden *et al.* [8]; Gilmer [9–11]; Blum and Lasaga [12]; Lasaga and Lüttge [13–15]) are valuable tools if computer-based predictions of the behavior of kinetically controlled crystal/solution systems are required. Such computer models need accurate input data if they are to simulate a "real" system. Therefore, laboratory measurements are needed to produce the input data. Numerous experimental studies have provided data on the dissolution behavior of many minerals over a wide temperature range. The result of this effort is a large pool of experimental data too extensive to be listed here in detail, and we refer instead to the following list of review articles: White and Brantley [16], Kerrick [17], Morse and Arvidson [18] [and refs therein].

Before we focus on the experimental and analytical strategy in more detail, it is important to discuss some key problems in the quantification of crystal dissolution. A comparison of the literature shows that the theory of crystal growth has been successfully developed over the last fifty years and is now well advanced. In contrast, the theory of crystal dissolution has not been developed in a comparable way (e.g., Lasaga [6]; Lasaga and Lüttge [13,15]). A quick answer to this disparity might be that crystal dissolution theory is in fact not necessary because crystal dissolution can be understood as negative crystal growth. However, it can be demonstrated that this assumption is incorrect. We will show in the subsequent discussion that crystal dissolution is a complex problem that can not be explained and quantified simply by changing the sign in a crystal growth rate equation.

7.2. MINERAL DISSOLUTION RATES AND SURFACE AREA

Any quantitative treatment involving fluid flow and reactions of aqueous solutions with crystal surfaces needs to extract a crystal dissolution and/or growth rate law from experimental data, and this treatment must include the possibility of approaching crystal-fluid equilibrium. While many computer models routinely input such "rate laws" (e.g., Lichtner *et al.* [19]), the experimental studies of mineral dissolution reactions have not provided definitive results in many cases (e.g., Lüttge *et al.* [20]) except for conditions that are quite far from equilibrium at the so-called "dissolution plateau", where the dissolution rate of a crystal becomes independent of the difference in Gibbs free energy (ΔG) (for further information see, e.g., Lasaga [6]). However, extrapolation of laboratory data to either natural conditions or to repository time scales necessitate a description of the kinetic changes expected as conditions approach near-equilibrium conditions. As an additional challenge, we need to understand and quantify not only the interactions between aqueous solutions and crystalline (or glass) surfaces but also the participation of microbes in this complex interplay (e.g., Banfield and Nealson [1], and citations therein; Conrad *et al.* [21]; Lüttge *et al.* [22]).

A review of the literature reveals that the broad discussion has identified three key problems. These are: (i) the quantification of dissolution rate as a function of distance from equilibrium; (ii) the definition and quantification of the "correct" surface area with which to normalize the bulk rates measured experimentally or in the field (e.g., Hochella [23]; Brantley *et al.* [24]; Brantley and Mellott [25]; Jeschke and Dreybrodt [26]; Lüttge [27];

and (iii) how to link experimental observations at atomic or near-atomic scales to bulk dissolution rates. More generally, these three problems indicate that our understanding of the dissolution process of crystals is not sufficiently advanced to provide a complete theoretical description.

Experimentally measured bulk rates have led to a formulation of the dissolution rate based on the formation of surface molecular complexes by surface reaction with solution species (e.g., Brady and Walther [28]; Stumm and Wieland [29]; Stumm and Wollast [30]; Van Capellen *et al.* [31]; Gautier *et al.* [32]; Oelkers *et al.* [33]; Lasaga [5]; Casey and Ludwig [34]; Dove [35]; Blum and Stillings [36]; Oelkers [37]. This approach highlights the chemical bonding properties of mineral surfaces as well as the role of adsorption-desorption steps on the crystal dissolution rate. The important role of surface complexes such as $\equiv Si–OH$, $\equiv Si–OH_2^+$, $\equiv Al–OH$, $\equiv Al–OH_2^+$, $\equiv Si–O^-$ in alumino-silicates has been recognized. The equilibrium distribution of surface complexes has also been used to predict the pH dependence of the rate. The molecular approach, therefore, has emphasized the role of *surface speciation* on mineral surfaces as a function of solution composition and temperature. The model then tries to identify particular surface complexes whose activities limit the overall dissolution rate (e.g., Blum and Lasaga [38]; Oelkers *et al.* [33]; Oelkers [37]).

As long as the dissolution reaction is surface-controlled, we need to normalize bulk reaction rates by the surface area involved in the dissolution process to obtain the rate constant, k, in units of $[\mathrm{mol\,m^{-2}\,s^{-1}}]$. An outstanding question that has received considerable attention in the literature is how to "correctly" quantify the surface area involved in crystal-fluid interaction (e.g., Hochella and Banfield [23]; Gautier *et al.* [39]; Jeschke and Dreybrodt [26]). Theoretically, there are three different possibilities that could be used by researchers: (1) geometric, (2) total or BET surface area, and (3) the "reactive" surface area. However, our discussion will show that practically only BET surface area is a robust measurement. Geometric surface area is merely an abstraction or estimate, and reactive surface area is problematic and more of a conceptual tool than a directly measurable quantity. Geometric surface area is usually calculated from the crystal dimensions and ignores surface roughness (compare e.g., Anbeek [40]). This calculation may also involve assumptions of an ideal geometric body, such as a cube, rhombohedron or sphere. If the geometric surface area is calculated for a mineral powder, it will depend on the way the average grain diameter of the grain population is computed. Therefore, a value for geometric surface area derived from a model tends to underestimate the true value.

The term "total surface area" is often used synonymously with BET surface area, which is measured by gas-adsorption techniques developed in 1938 by Brunauer, Edward, and Teller (Brunauer *et al.* [41]). The apparatus measures the adsorption of an inert gas on the sample surface. A key assumption is that the gas adsorbs in a monomolecular layer on the sample surface. In practice this technique works well for a relatively large surface area, i.e., 1 m^2/g to 1000 m^2/g. Smaller surface area values are difficult to measure with sufficient precision. However, measurements are usually successful down to about 0.1 m^2/g by using krypton. However, the extent to which the BET surface area measurement reflects the total surface area is a function of both the gas used and the structure of the solid surface. Unless the crystal surface is atomically flat, i.e., there is no surface roughness, the total/BET surface area is always larger than the geometric surface area. Consequently, the calculated reaction rate constants in $[\mathrm{mol\,m^{-2}\,s^{-1}}]$ are slower than they would be by normalization with geometric surface area. This fact becomes an important issue if the difference between

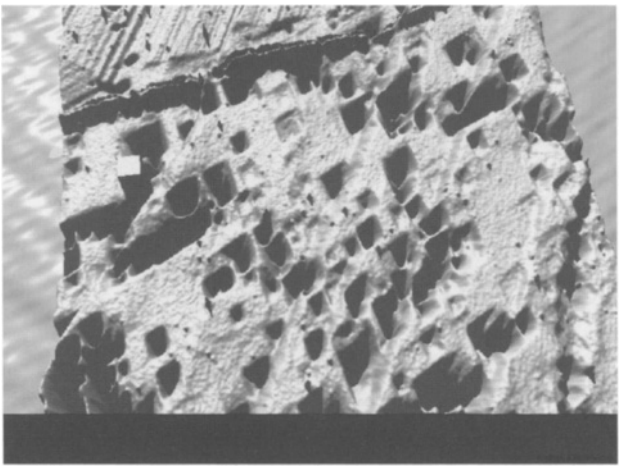


FIGURE 7.1. 3-D false color image of a VSI height map; shown is a calcite cleavage surface (163 μm × 124 μm) after dissolution at 25°C at pH 8.8 for 90 minutes. The z-scale is about 250 nm from top to bottom. The image is produced by rendering VSI height data with Bryce4 software.

geometric and total surface area is significant with respect to the problem under investigation. It is likely also that geometric and total surface area will increasingly deviate during the course of a dissolution process. Model calculations show that the difference between different types of surface area measurements may change significantly during a dissolution reaction due to the development of etch pits and other dissolution features (Lüttge [27]).

In general, the surface area (however described) will change over the course of a reaction. The extent to which the surface is self-modifying reflects the balance between two competing processes: (1) a particle's tendency to minimize its total surface energy and exposure to the solution by reducing its area to volume ratio, and (2) the mechanistic restriction of possible paths by which the bulk free energy can be minimized, through material transfer to solution through propagation of kinks, steps, and etch pits, which tend to increase the surface area.

The picture may be strongly modified and complicated if we take into account that the crystal surface is usually not homogeneously reactive, and thus certain areas of the crystal surface react faster than other areas. This has been observed by scanning electron microscopy (SEM), AFM, and VSI studies, that often show the formation of etch pits and canyon-like structures at crystal surfaces during the dissolution process (e.g., Figure 7.1). These features are often linked to some micro-damage within the crystal lattice (e.g., microcracks) or lattice dislocations (e.g., screw dislocation, line defects or point dislocations) that form high energy sites at the crystal surface. The observation that some parts of the surface are significantly more reactive while others remain more passive has led to the idea of reactive surface area (e.g., Hochella [42]; Hochella and Banfield [23]; Lasaga [6]). Unfortunately, this term is merely of conceptual importance because it is currently impossible to measure reactive surface area. However, the discussion has motivated a reappraisal of the approach we take to quantify surface area, and it has emphasized the fact that the (reactive) surface area of a crystal does change during the dissolution process. In this context, the development towards an often assumed quasi steady-state surface may be questionable.

As already mentioned above, crystal dissolution rate measurements often suffer from a scaling problem. Bulk dissolution rates are measured from changes in solution chemistry using mineral powders in flow-through reactors, batch reactors, or rotating disk experiments. Measurements on the atomic, i.e., nanometer or angstrom scale are typically obtained by AFM studies. These investigations can provide rates of step movement on the crystal surface. Unfortunately, it is not always easy to compare and link rates obtained from step movement as observed with the AFM with bulk dissolution rates. A key observation provided by vertical scanning interferometry (see below) is that the variation in rates distributed over the surface is potentially large. This variation most likely is related to the concept of reactive surface area. However, what controls the distribution of rates is the distribution of surface energy. Therefore, if one could understand the relationship between rate and surface energy then we would have the means to quantify reactive surface area. This means that our understanding of reactive surface area depends on the determination of the distribution of sites on a crystal surface. Consequently, reactive surface area is not just a measurement of surface area in units of cm^2 or m^2. In addition, we need to understand the interaction of these surface sites with the solution. While it is not quite possible to measure distribution and development of sites across a crystal surface in the laboratory, Monte Carlo simulations that treat mineral dissolution as a many-body problem can do that precisely (e.g., Lasaga and Lüttge [15], and see discussion below).

In the experimental practice, bulk rates are usually composed of an average of various step rates across a topographically complex crystal surface, and the relationship between e.g., step density and step velocity is not well understood. It is also likely that the dissolution of mineral powder grains is governed by different mechanisms than the dissolution of relatively large single crystals (e.g., Arvidson *et al.* [43]). With these considerations in mind, it is hardly surprising that significant variations in dissolution rates are reported from single crystal and powder experiments (e.g., Arvidson *et al.* [44]).

The quantification of crystal dissolution can be even more complicated if we consider the formation of secondary, i.e., alteration phases (e.g., kaolinite and gibbsite) or the formation of so-called "leached" layers. These effects may become so strong that in some cases, the reaction control may even shift from a surface controlled process to transport controlled dissolution, where the transport through the layer of alteration products determines the rate of the overall dissolution reaction. However, for many geologically relevant processes, it may be sufficient to use geometric or total surface area for the normalization of bulk rates but there are other problems, e.g., the prediction of groundwater pollution or nuclear waste management that require significantly better constraints of mineral or glass dissolution rates.

It is beyond the scope of this chapter to provide a comprehensive review of the experimental and theoretical progress towards an overall understanding of crystal dissolution. Also, it is not our goal here to cover the ongoing complex discussion in depth. Instead, we will address what approaches and developments are necessary to advance our quantitative understanding of crystal dissolution and how interferometry can contribute towards this goal. Although we recognize that there are a variety of opinions and possible methods, we think that direct observations and measurements of the crystal surface during the dissolution process, in combination with computer models and simulations, hold great potential in this area. AFM has tackled the task of direct observations at high resolution successfully, and a growing number of experimental studies has created a valuable pool of data, e.g.,

Bosbach and Rammensee [45]; Dove and Platt [46]; Eggleston [47]; Higgins *et al.* [48]; Liang *et al.* [49]; Teng *et al.* [50]; Jordan and Rammensee [51]; Lea *et al.* [52]. However, it is still unclear in many cases how the detailed high-resolution results produced by AFM (e.g., precise rates of step movement) can be used to calculate overall dissolution rates of entire crystals. In addition, AFM's small field of view is both a handicap and a strength of the technique. A detailed comparison and comprehensive discussion of AFM and VSI techniques and their applicability to crystal dissolution and growth processes have been presented recently by Eggleston and Lüttge [53].

Interferometry has been successfully applied to a wide range of problems in crystal growth and dissolution kinetics. In addition to VSI (described below), related techniques include laser Michelson interferometry, real-time phase shifting interferometry, and in situ differential phase-shifting interferometry. Development is sufficiently advanced and permits the deployment of miniaturized instruments in exotic environments: e.g., crystal growth and protein crystallization under microgravity conditions (e.g., Kim *et al.* [54]). Several research groups of, e.g., Tsukamoto, Vekilow, and Lal have successfully applied a variety of interferometry techniques to study organic and inorganic crystal growth kinetics (e.g., Vekilov *et al.* [55–57]; Vekilov and Kuznetsov [58]; Vekilov and Rosenberger [59]; Vekilov and Alexander [60]; Booth *et al.* [61]; Gliko *et al*. [62]; Kim *et al.* [63]; Tsukamoto [64,65]; Onuma *et al.* [66,67]; Maiwa *et al.* [68]; Yokoyama *et al.* [69]; Li and Tsukamoto [70].

In contrast, the application of interferometry to study crystal *dissolution* has been relatively limited. One exception is the experimental study by Vekilov *et al.* [71] who investigated the dissolution morphology and kinetics of the (101) ADP face using X-ray topography and laser Michelson interferometry. They observed the movement of step patterns to be symmetrical with respect to crystal growth and dissolution, and determined the limiting undersaturation for the appearance of a dislocation etch pit. This important datum allowed them in turn to estimate the Burger's vector/effective free surface energy ratio for one of the dislocation sources. Experimental results of this quality play a significant role in our attempt to improve our fundamental understanding of the dissolution process of crystalline matter. In this context, our understanding of the dependence of the dissolution rate on ΔG, i.e., the difference in Gibbs free energy and consequently the saturation state of the solution, is critical (see discussion below).

At first glance it is surprising that analytical techniques based on interferometry have not been used more extensively in growth and dissolution studies of minerals. The main reason is that most rock-forming minerals, particularly the group of (alumino)silicates dissolve extremely slowly under Earth's surface conditions. The following examples illustrate the experimental challenge to measurement of dissolution rates in the laboratory via direct, i.e., in situ real time experiments. The mean lifetime of a cubic millimeter-sized quartz grain would be 34,000,000 years in an aqueous solution at 25°C and pH 5 far away from equilibrium (Lasaga [6] p. 109; Rimstidt and Barnes [72]). The mean lifetime of an aluminosilicate, e.g., a feldspar grain (albite, $Na[AlSi_3O_8]$) under similar conditions would be 575,000 years (Lasaga [6] p. 109; Burch *et al* [73]; Knauss and Wolery [74]; Chou and Wollast [75]). Albite ($Na[AlSi_3O_8]$) is an endmember of the solid solution of plagioclase feldspars, one of the most abundant minerals in the earth's crust. For a long time, it seemed therefore not feasible to use interferometric techniques for the measurements of mineral dissolution rates.

Slow dissolution rates pose unique problems for any analytical or experimental technique. These problems can be accommodated through various strategies that come at the expense of additional complexity, assumptions, and technical requirements.

As discussed above, one possibility is to conduct experiments with fine-grained mineral powders in flow-through cells, batch reactors, or column reactors. In experiments that utilize flow-through or column reactors the dissolution rate is determined from the difference between the composition of the input solution and the effluent solution. Flow rates, cell geometry and other experimental parameters (e.g., pH, temperature, saturation state of the solution) are usually quite well known and controlled. Beside the problem of correctly normalizing bulk rates by the surface area of the reactant, an additional problem is often caused by the use of extremely fine-grained powders in stirred reactors. Here, we have to deal with issues beyond that of surface area alone: "excess" surface free energy, enhanced solubility, boundary layer transport, characteristic etch pit density, etc. need to be considered because they may influence measured dissolution rates significantly.

As an alternative to the analytical methods above, one can conduct direct observations and measurements by VSI and AFM if (1) it is feasible to increase the dissolution rate by an increase of temperature and/or a change to more extreme pH conditions (assuming the rate *has* a strong pH dependence). Note that it may be necessary in some cases to raise the temperature even up to hydrothermal conditions. (2) Measurements can be conducted ex situ. Even a combination of the two options may be desirable in some circumstances. There are pros and cons for each attempt. An increase in temperature is usually possible, but it often causes significant technical difficulties, particularly for all types of in situ observations. It is mandatory to pressurize the cell, which must also contain a window to allow the observation of the sample surface. If a VSI is used, the reference path of the interferometer needs the addition of a (flexible) compensator to compensate for all changes in the refractive index of the sample pass. The data measured under these conditions would be high-temperature dissolution rates, and it is important that the extrapolation of such data to low temperature environments involves important assumptions regarding invariance of mechanism, activation energy, and solution components. In general, such an extrapolation is difficult.

The final approach sacrifices the in situ strategy and seeks to measure surface-normal retreat (and advance) rates ex situ. This may be not a disadvantage at all; at least in cases of slow dissolution rates it is often preferable to perform measurements ex situ over days, weeks or even months (see discussion below).

MacInnis and Brantley [76,77] have conducted pioneering work using an ex situ approach. They used Mirau interferometric objectives to study the dissolution of the geologically important mineral calcite. They directly addressed the relationship of dissolution rate and surface area discussed above. In rotating disk experiments they found the rate of calcite dissolution, measured by normalization of changes in fluid composition (pH) to interferometrically-measured surface area, to agree favorably with published rates for calcite powders, and related this rate to the calculated volume removed by etch pits. They observed a relatively modest rate increase in strained (dislocation-rich) versus unstrained samples, consistent with a bimodal model for etch pit generation. From the same experimental data set, MacInnis and Brantley [77] also presented detailed observations of the nucleation, evolution and interaction of etch pits, and related these processes to a maintenance of a steady state rate.

Our approach builds on this work and the pioneering work of Maiwa *et al.* [78], Ohmoto *et al.* [79], Onuma *et al.* [80–82], Tsukamoto *et al.* [83], MacInnis and Brantley [76,77], and MacInnis *et al.* [84]. We utilize a vertical scanning interferometer in our experiments as a complementary technique to AFM and mineral powder methods that have been so successfully applied to quantify dissolution rates of numerous minerals, e.g., carbonates, (alumino-)silicates, and barite (e.g., Lüttge *et al.* [20,85]; Arvidson *et al.* [43,44]; Beig and Lüttge [86]). We are using VSI to measure surface-normal advance (or retreat) of single crystal faces relative to a fixed (unreacted) reference surface. Measurements are made before, during, and after reaction for fixed time intervals at constant fluid composition in a temperature-controlled, flow-through cell system. VSI makes it possible to quantify intrinsic reaction rates, k, without the need to independently measure the surface area of the solid (see discussion). Another advantage of VSI is its large field of view, making it possible to measure absolute rates (see below). Each scan provides up to 100,000 height data that can be converted into rates. This large number of data allows a vigorous statistical treatment that can be directly related to bulk dissolution or growth rates. Therefore, VSI can be seen as the "missing link" between the large pool of nanoscale data produced with atomic, scanning or tunneling force microscopy and the results obtained from bulk solution chemistry.

7.3. THE TECHNIQUE

The present generation of vertical scanning interferometers, like the MicroXAM MP8 (ADE-Phase Shift, Tucson), provides high (angstrom- to nanometer-scale) vertical resolution, and a lateral resolution of about 500 × 500 nanometer (with a Nikon 50× Mirau objective). The lateral resolution of the system depends on the wavelength of the light source, e.g., 550 nm (green light) and the aperture of the Mirau objective. The MicroXAM MP8 has a field of view of up to $845 \times 630\ \mu m^2$ (10× Nikon Mirau objective). This range is large compared to the range provided typically by an AFM. The high vertical resolution of VSI is achieved by moving the sample surface topography through the focal plane of the Mirau objective. Only those parts of the surface which are exactly in the focal plane of the objective produce an interferogram. That means that the CCD camera "sees" only a partial interferogram at each scan step unless the sample surface is perfectly leveled and atomically flat. The vertical movement of the sample is achieved by a piezoceramic (PZT).

The VSI is capable of operating in two different modes: (a) using a narrow band of green light; (b) using white light. White light allows for larger vertical scans, i.e., up to 100 nanometer with better than two nanometer precision or even up to one millimeter with ten microns precision. The use of green light (e.g., $\lambda \sim 550$ nm) reduces the vertical scan height to a maximum of four microns, but improves the precision by an order of magnitude, and provides angstrom scale vertical resolution. The two microns per second vertical scan rate allows the acquisition of data orders of magnitude faster than any other surface mapping technique can provide. As an example, a surface area of almost one mm^2 and up to 20 microns of surface relief can be quantified in less than ten seconds.

Figures 7.2–7.4 show a photograph of the MicroXAM MP8 and schematic sketches of some of the principles of white light VSI, and the flow-through cell design. A more detailed description of the technique can be found in Lüttge *et al.* [50]. Commercially

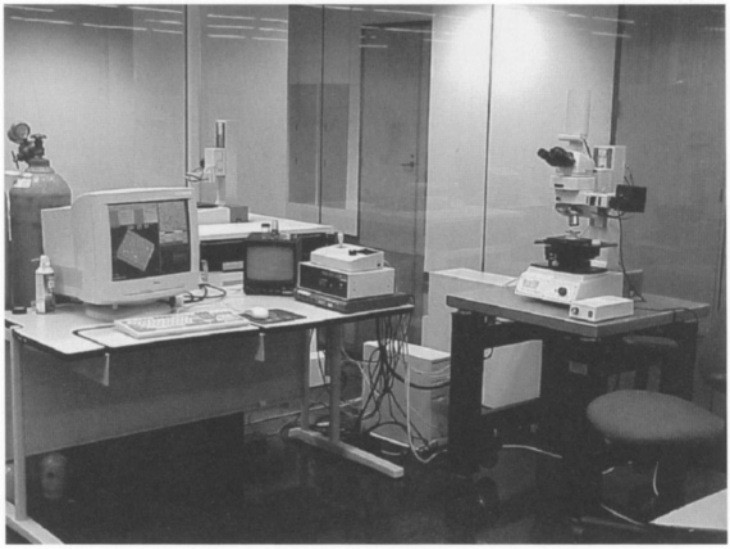


FIGURE 7.2. ADE-Phase Shift's MicroXAM MP8 positioned on an anti-vibration air table and equipped with a l0× and a 50× Mirau objective (Nikon).

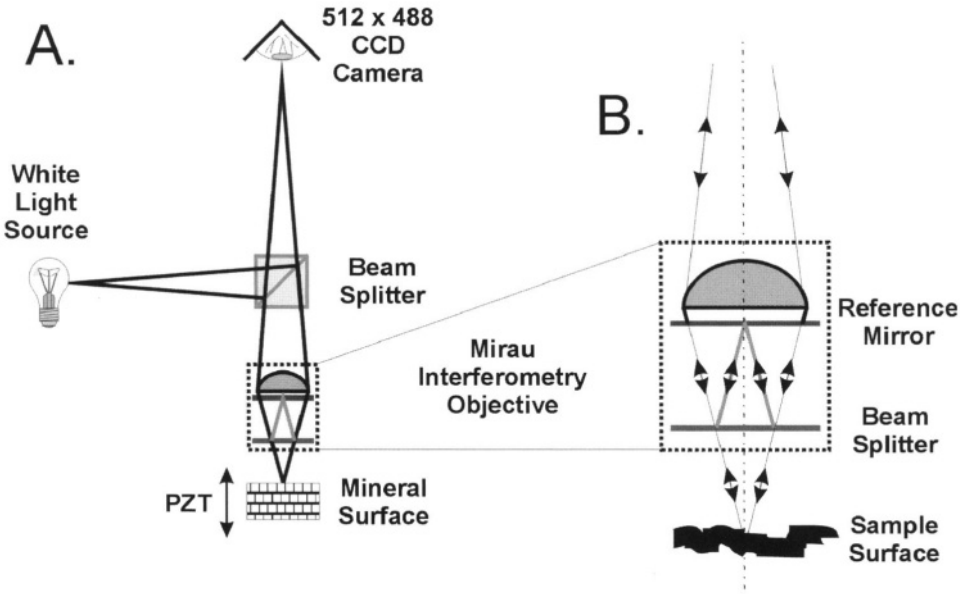


FIGURE 7.3. (A) Sketch of a double-beam Mirau interferometer with CCD camera. (B) Detail of (A) showing the sample and the reference light paths in the Mirau objective as well as between the objective and the sample surface. [Adapted from Lüttge *et al.* [85], Figure 1].

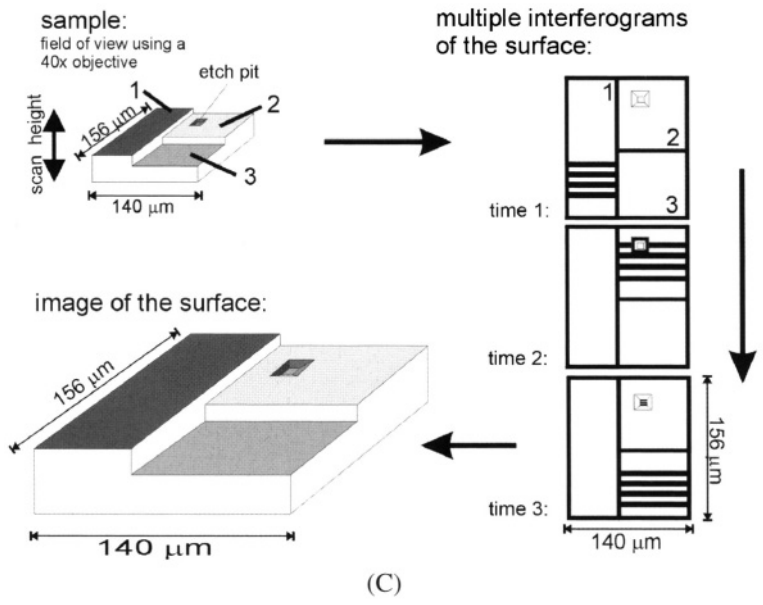


FIGURE 7.3. (Continued). (C) The sketch illustrates how VSI works in principle. Data are acquired by moving the sample surface through the focal plain of the Mirau objective. Only the parts of the surface that are in focus at a certain time produce a (partial) interferogram. When the scan is complete, a computer program (Mapvue, ADE-Phase Shift) is used to analyze the set of interference data and to calculate the actual height map of the sample surface. [Adapted from Lüttge *et al.* [85], Figure 2].

available state-of-the-art software packages [MAPVUE, ADE-Phase Shift (Tucson) and self-developed software programs, Lüttge *et al.* [20,85]] in combination with an automated sample positioning stage allow it to "stitch" several scans together. The task is accomplished by overlapping two scans by 10%. That means that in case of two scans that contain 100,000 pixels (i.e., height data) each, 10,000 pixels per scan are identical (within the error of the measurement) and can be used to match the two scans. In this way, large data sets can be produce that cover most or all of the entire reactant surface area. As described previously, the ability to create reference surfaces (Figure 7.5) within the field of view provides absolute height measurements. As a result, VSI can precisely quantify both the absolute topography of a crystal surface and critical changes in this topography due to ongoing dissolution, corrosion, or growth processes. VSI measurements of surface-normal height changes can then be converted directly into intrinsic dissolution or growth rates and averaged over the entire solid surface. This leads to a rate constant that is independent of any external measurement of the surface area, e.g., BET surface area. Note, that the lateral resolution of the VSI system defines the "footprint" of a pixel (e.g., 500 × 500 nm), and that each height measurement, therefore, represents an average height across the pixel.

In this way, the MicroXam system quantifies surface topography in great detail (e.g., Figure 7.6). Time-lapse observations quantify the evolution of surface topography (Figure 7.7a–d). As said above, rates measured with VSI are usually surface-normal velocities (e.g., Lasaga [5]; Lüttge *et al.* [20,85]). The change in height, h, of a given point (i, j) on

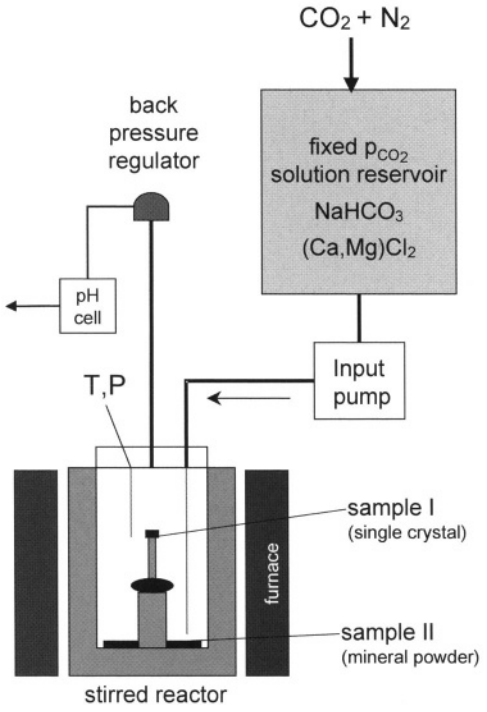


FIGURE 7.4. Schematic diagram of mixed single crystal-powder reactor and associated fluid delivery system used in combination with VSI. Reactant solution is delivered from a gas-sparged (fixed fugacity) reservoir by a high-pressure metering pump to the fluid cell. The cell consists of magnetically stirred, titanium alloy volume (~100 ml nominal capacity) that contains both a mineral powder and mounted single crystal(s). Cell pressure is maintained by gas-loaded back-pressure regulator (200 bars maximum); temperature is held constant by a ceramic furnace and microprocessor feedback controller (~350°C maximum). Solution pH is measured after exit from back-pressure regulator prior to atmospheric exchange and sampling.

the *(hkl)* crystal surface or cleavage plane during the time interval from t_1 to t_2 is

$$\frac{\Delta h_{i,j}}{\Delta t} = \frac{(h_{i,j})_2 - (h_{i,j})_1}{t_2 - t_1}, \\ = v_{i,j(hkl)}, \tag{1}$$

where $v_{i,j(hkl)}$ is the velocity of dissolution or growth of every *i,j* coordinates at the *(hkl)* surface in units of $[\mathrm{m\,s^{-1}}]$. The rate constant, $k_{i,j}$, is then computed by dividing by the molar volume $(\bar{V})$ of the material dissolved or grown,

$$k_{i,j} = v_{(hkl)} \times \bar{V}^{-1}. \tag{2}$$

Rates computed in this way have units of flux $[\mathrm{mol\,m^{-2}\,s^{-1}}]$. Now, we have to sum over all local rate constants, $k_{i,j}$, and take the average,

$$k_{(hkl)} = \sum \frac{k_{i,j(hkl)}}{N_{i,j}}, \tag{3}$$

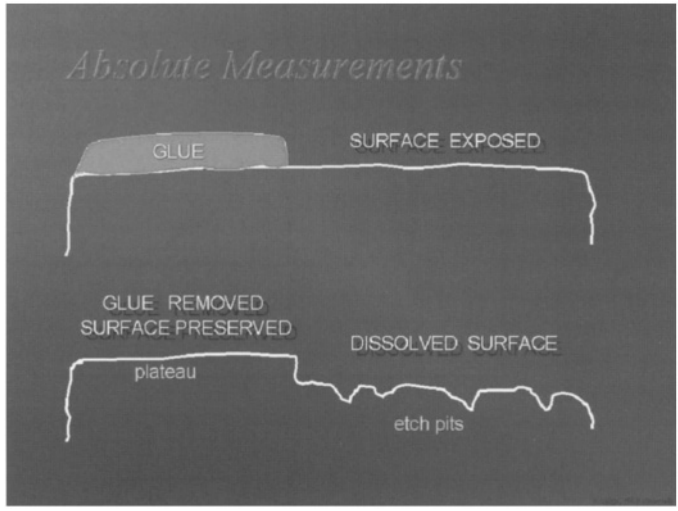


FIGURE 7.5. The sketch demonstrates the effect of an internal reference surface. Micron-scale sized spots of the pristine surface are covered with an inert cement or a "cap" (see Arvidson *et al.* [43]). These spots are protected against the dissolution, corrosion or growth process occurring at the rest of the sample surface. After some time, the mask(s) can be removed and any part of the sample surface can be measured and compared to the reference spot(s). This procedure allows absolute height measurements.

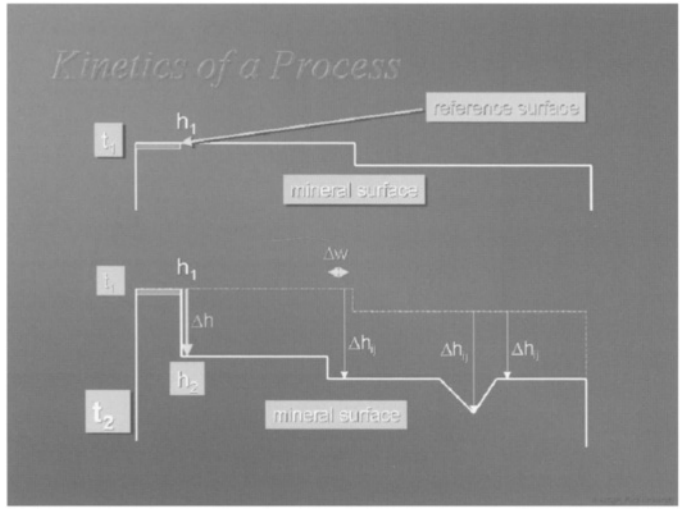


FIGURE 7.6. The diagram explains how height differences are measured that can be then converted into dissolution or growth rates. (Top) An initial sample surface with a fixed height reference at t_1. (Bottom) The same surface at time t_2. The entire surface has now retreated but with different rates—with the exception of the reference spot. A subtraction of the height data at t_1 minus the height data at t_2 produces a height-difference file that can be converted into surface-normal retreat velocities and into rate(s) (see text for a more detailed description).

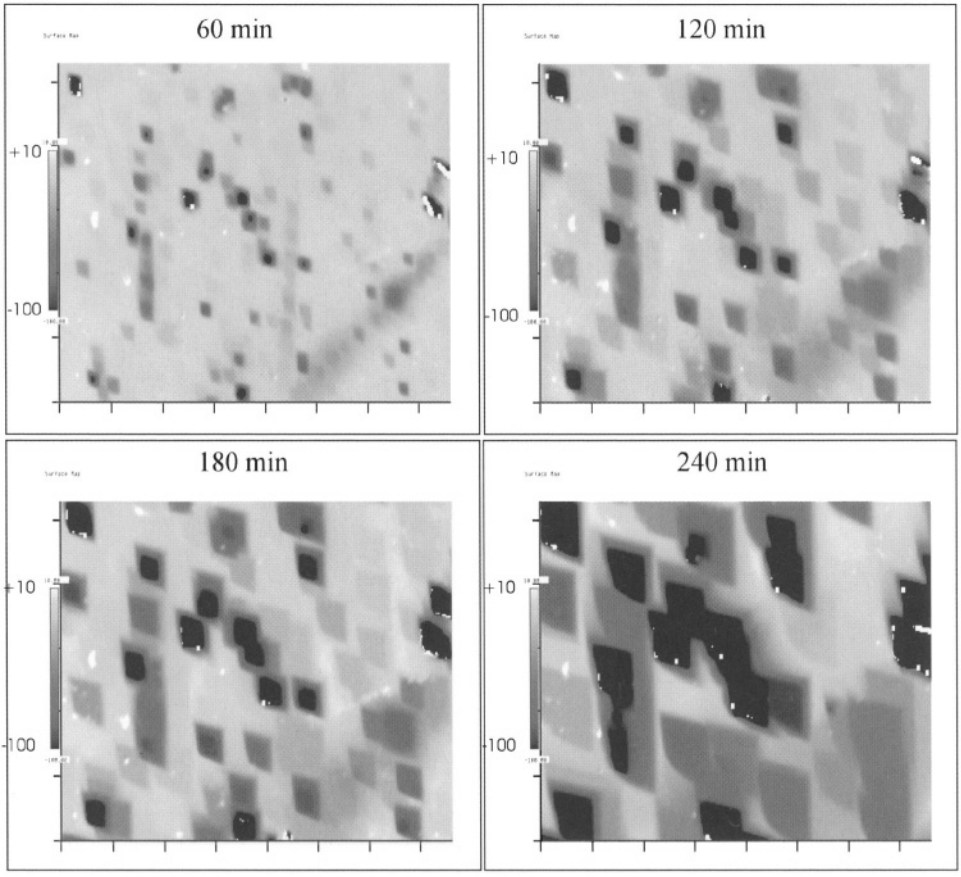


FIGURE 7.7. VSI topography of a calcite cleavage surface (about 75 μm × 75 μm) after (A) 60 min, (B) 120 min, (C) 180 min, and (D) 240 min dissolved in an aqueous solution at 25°C and pH 8.8. Orientation of etch pits is clearly crystallographically-controlled, with progressive etch pit growth giving rise to extensive coalescence and cannibalization with sufficient reaction progress. [Adapted from Arvidson *et al.* [44], Figure 4].

to obtain a rate constant $k_{(hkl)}$ for the entire surface. N_{ij} is the number of all pixels, *i,j*, measured in a VSI scan; as said earlier, one scan provides typically up to 100,000 data (pixels). From this treatment, we see that the rate constant is itself not a "true" constant but depends on the number and distribution of sites across the crystal surface. This approach leads directly into a statistical treatment of reaction rates on the crystal surface and a revision of our understanding of the so-called reactive surface area (e.g., Lüttge [87]).

The reference surface within the field of view of the Mirau-interferometer objective allows it to quantify absolute height differences. A longstanding problem was to establish such a reference surface directly on the mineral surface. Lüttge *et al.* [85] have solved the problem by masking one or several portions of the mineral surface with an inert adhesive (Figure 7.5). Different types of adhesives should be used depending on the reactant, temperature and pH range of the experiments. The adhesive can be removed during and/or post-reaction to expose an unreacted reference plane (e.g., Figure 7.8). By measuring heights relative to the preserved plane, the computed rates are absolute, and can be compared with

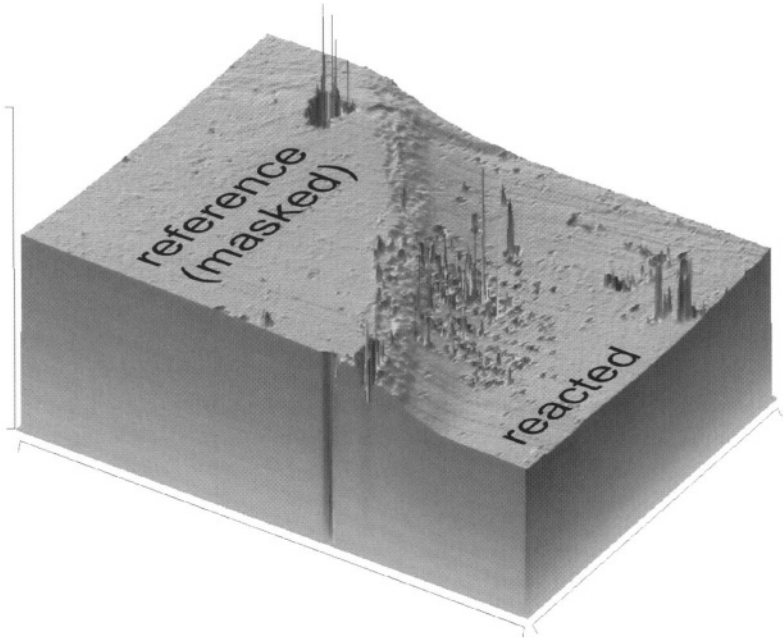


FIGURE 7.8. The VSI image demonstrates how a "cap" protects a part of the sample surface (anorthite cleavage surface, treated for 7 hours at pH 8.56, 34 bars, and 200°C). The protected and, therefore, unreacted surface serves as a height reference after the "cap" has been removed. The surface-normal retreat is $\sim 1\ \mu$m adjacent to the unreacted masked plateau. [Adapted from Figure 1 of Arvidson *et al.* [43]; see also Figure 7.10, this paper].

those obtained for example from conventional powder experiments. Note that as explained above, unlike powder experiments, in which reaction rates must be normalized to *total* surface area, rates measured by VSI do not require and are independent of bulk surface area determinations (compare also discussion in Lüttge *et al.* [20,85].

7.4. REVIEW OF SELECTED RESULTS OF RECENT VSI-STUDIES

In the following, we will discuss the problems identified above in the light of recent experimental and theoretical results. First, let us focus on two studies conducted to investigate (alumino) silicate dissolution kinetics.

7.4.1. Anorthite ($Ca[Si_2Al_2O_8]$) dissolution kinetics by Lüttge et al. [85]

This paper is the first publication of a series of studies on dissolution kinetics measured by VSI. The paper provides a detailed description of the experimental/analytical technique. Lüttge *et al.* used VSI in combination with a white-light source to quantify the dissolution of a (010)-cleavage surface of a near-end-member calcium feldspar (An_{98}). A flow-through cell was used with an aqueous solution at pH 3 (perchloric acid) at 25°C and a fast flow rate of ten ml per minute and a cell volume of about 2.5 cm^3. These parameters set far-from-equilibrium conditions for the dissolution experiment, and thus the dissolution rate is independent of ΔG. The dissolution rates were calculated from surface-normal retreat rates measured after 48, 84, 120, and 168 h run duration. All measurement were taken ex

situ: the experiment was stopped, the sample removed from the reactor and mounted on the interferometer's stage, and analyses made of the surface topography. Although molecular-scale alteration of mineral surfaces due to their reactivity in air has been well documented (e.g., Stipp *et al.* [88,89]), this alteration is beyond the detection limit of VSI, and thus will not change the outcome of the measurement. The rates measured by VSI are large compared to the limits and precision of its vertical resolution (~2 nm), typically several hundred nanometers per run interval. In this study, absolute height changes were obtained by masking micron-sized areas of the pristine crystal surface, and thus protecting them from dissolution. The individual masks were later removed during the experiment, and the preserved pristine areas of the anorthite surface served then as internal reference heights.

The development of fifteen different areas at the (010)-cleavage surface were studied by VSI. All scans showed the formation of two different types of etch pits, i.e., numerous shallow pits and more rare but very deep ones. The corresponding dissolution rates were 2.7×10^{-12} and 6.6×10^{-12} [mol/cm^2/s], respectively. In addition, Figures 7.9a, 7.9b

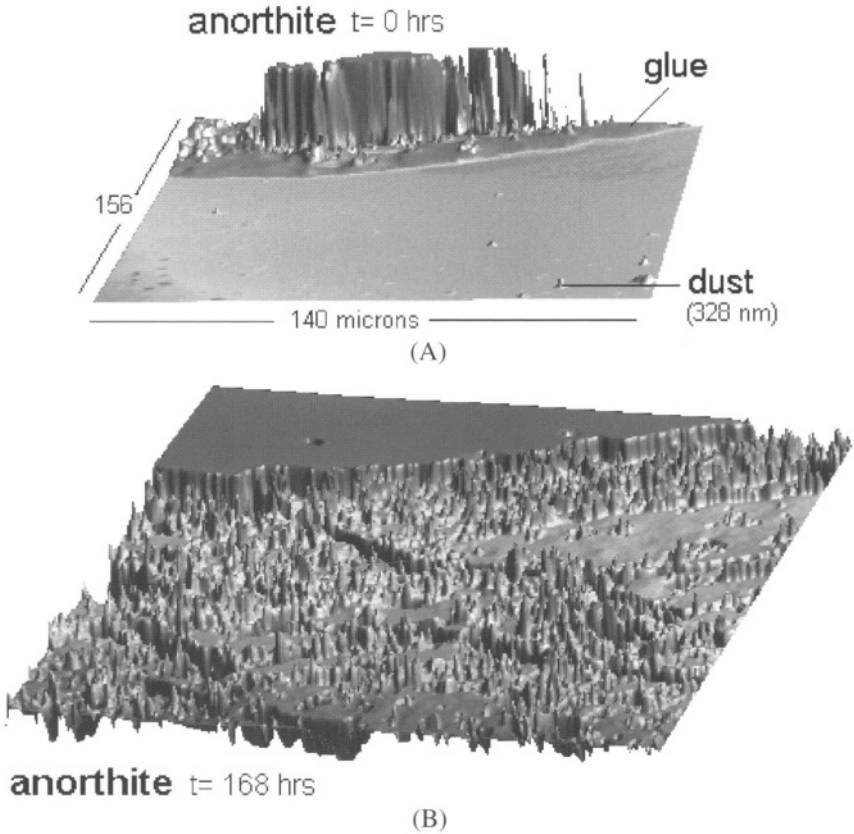


FIGURE 7.9. It is shown the pristine cleavage surface of an anorthite crystal (A) (156 μm $\times$ 140 μm measured with a 40 × Mirau objective). A part of the surface (red) is protected with a mask made from silicon rubber. (B) The same anorthite surface as in (A) but after 168 hours of dissolution at 25°C (pH 3) with the mask removed. It can be seen that the unprotected part of the surface has been significantly roughened by etch pit formation and has retreated, forming a "cliff".

[modified from Figure 4 in Lüttge *et al.* [85]] show the formation of a "cliff" after the protecting mask had been removed. This "cliff" represents the global surface-normal retreat. However, Lüttge *et al.* did not recognize the significance of this observation although they incorporated this global retreat into the measurement of the overall bulk dissolution rate (5.7×10^{-13} [mol/cm^2/s]. The importance of the "cliff"-formation was later recognized by Lasaga and Lüttge [13] (see discussion below).

7.4.2. Albite ($Na[AlSi_3O_8]$) dissolution kinetics by Arvidson et al. [43]

Arvidson *et al.* have developed an experimental method for studying and quantifying the dissolution of both single crystals and mineral powders at identical conditions in the same flow-through cell. The method employs a titanium reactor that combines the capabilities of two existing experimental designs: single-crystal and mixed flow powder reactors. The new reactor allows *ex situ* VSI measurement at elevated temperatures of one or more single crystals (one masked) in the same volume of fluid as a stirred mineral powder (Figure 7.10). The elastomeric masking compound is stable up to 250°C, allowing the mainte-

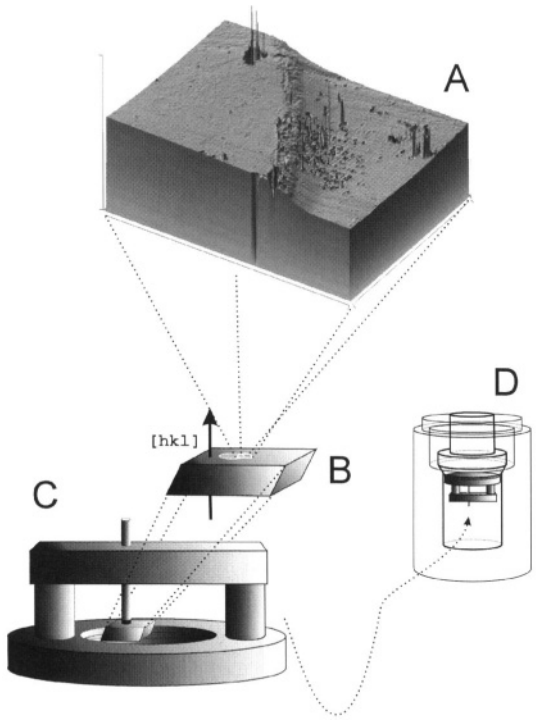


FIGURE 7.10. Scaled figure showing component relationships of hybrid single crystal/powder reactor. (A) 3-D rendering of VSI data collected from single crystal anorthite cleavage surface (010) showing 1000+ nm of surface-normal dissolution adjacent unreacted (masked) plateau. Masked single crystal (B) is attached to sample mount (C), which is in turn contained within the reactor volume together with crushed mineral powder (D). VSI data were collected at pH 8.56, 200°C, 34 bars. [Figure adapted from Arvidson *et al.* [43], with some aspects omitted for clarity. Although the data shown in the figure are for anorthite, the albite dissolution experiments described in the text use the identical setup].

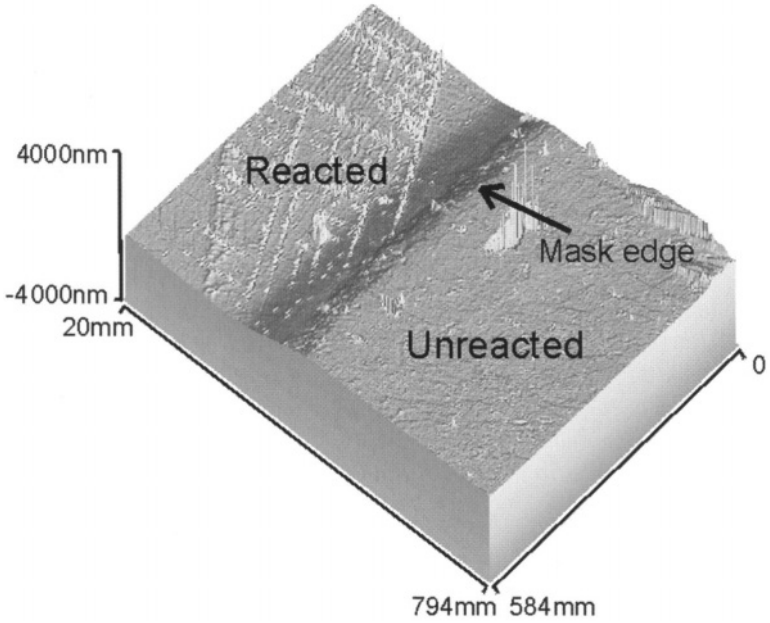


FIGURE 7.11. 3-D topographic image of an anorthite single-crystal after 7 hours of dissolution at pH 9 and 200°C. A portion of the surface was masked to protect it from reaction and provide an absolute height reference surface. The height difference between the reacted and unreacted surfaces is approximately 1 μm.

nance of a reference surface over temperatures typical of burial environments. This feature also permits VSI study of the kinetics of materials whose reaction rates are otherwise too slow to be studied at room temperature over run times feasible in the laboratory. Mounting different single crystal cleavage faces allows the variability in rate and mechanism to be explored as a function of crystallographic orientation during the same run. The ability to simultaneously react single crystals and powdered mineral samples under the same experimental conditions will also help to resolve questions regarding the comparability of dissolution rates derived by different techniques.

The authors used albite (001 and 010 cleavage faces + powder, $d \leqslant 20$ μm) from Amelia Courthouse, Virginia for a first case study. Key results from this work include the observation that etch pit distributions and rates determined by VSI are significantly higher ($45–110 \times 10^{-10}$ mol/m^2/s) on a powder versus large ($d \sim 0.6$ cm) single crystals (4.5×10^{-10} and 6.8×10^{-10} mol/m^2/s, (010) and (001) faces, respectively). In contrast, the rate determined from changes in solution chemistry is 7.0×10^{-10} mol/m^2/s, suggesting a "reactive" surface area of less than 16% of the total BET surface area. These results imply considerable complexity in the relationship between absolute and "bulk" reaction rates, and confirm earlier suggestions that rates may be a nonlinear function of grain diameter (e.g., Holdren and Speyer [90]).

7.4.3. Calcite ($CaCO_3$) dissolution kinetics by Arvidson et al. [44]

Current studies of calcite dissolution rates at pH ~ 9 using VSI (Arvidson *et al.* [44,91]) have produced the following central results:

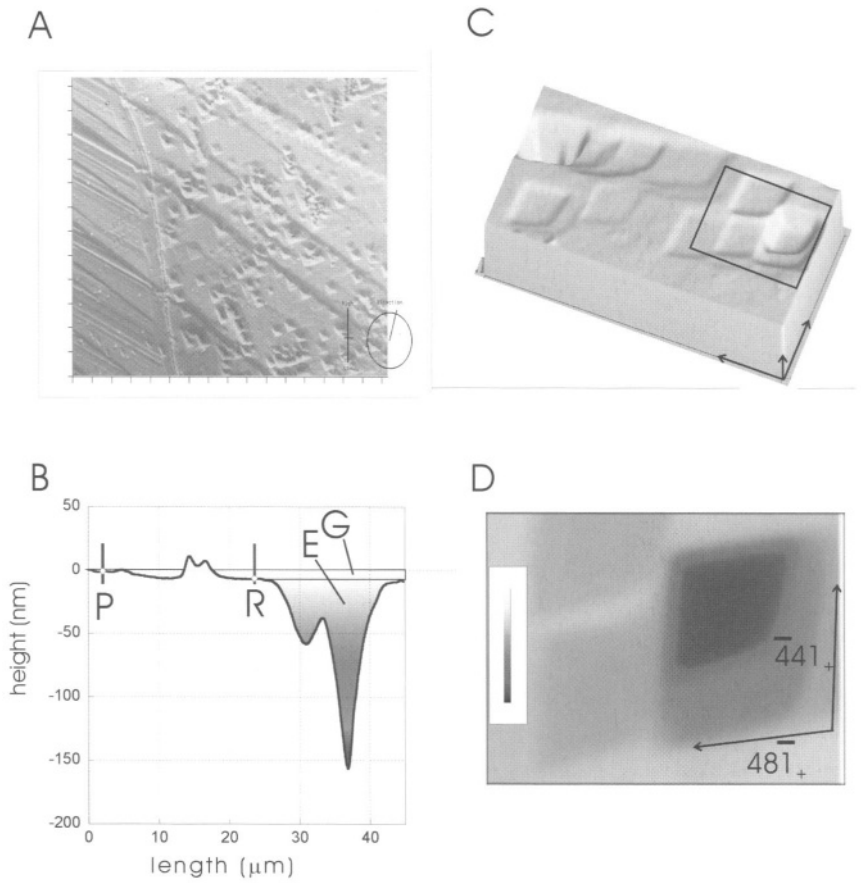


FIGURE 7.12. Calcite dissolution at pH ~ 9, 25°C. (A) Cleavage surface of calcite after 3 hours total dissolution, showing boundary between unreacted surface (left) and reacted area (right) with etch pit development. (B) Profile across this boundary, showing that the total change in surface topography as the sum of a slow "global" dissolution ('G') and local rapid removal at etch pits ('E'). Adhesive mask residue is present only between 'P' and 'R'. Line separating 'G' and 'E' is parallel to original cleavage surface. (C) 3-dimensional height map showing shallow, coalescing etch pits. $[48\bar{1}]_+$ and $[\bar{4}41]_+$ steps show "rounding", *x, y* scale bars are 10 μm, $z = 100$ nm. (D) Gradient height map of insert from (C) showing coalescent etch pits. Height bar spans 88 nm. [Adapted from Figures 3A and 3B of Arvidson *et al.* [44]].

(1) Calcite dissolution is dominated by the removal of material at etch pits, with a measured rate of $10^{-11.0}$ [mol/cm^2/s] ($P_{CO_2}10^{-3.41}$ atm, pH 8.8, 25°C). This rate is consistent with measured step velocities in published AFM work (Lea *et al.* [52]), assuming a step density $\sim 9\ \mu m^{-1}$.
(2) In areas outside of major etch pit development, the removal process is limited by a slower, "global" rate of $10^{-11.68}$ [mol/cm^2/s], although the ultimate coalescence of etch pits renders this rate measurable only at early run times (Figures 7.12a, 7.12b).
(3) The rate of surface-normal retreat in areas of etch pit development is 5–100 slower than rates derived from experiments using mineral powders or single crystal rotating disks (Figure 7.13).

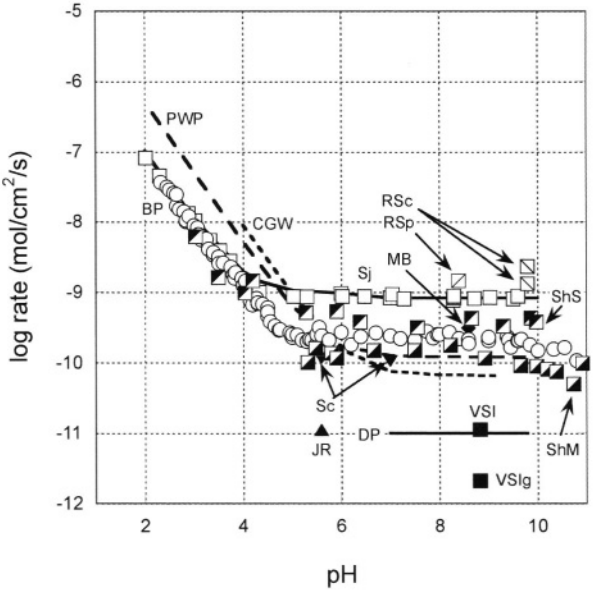


FIGURE 7.13. Variation in published calcite dissolution rates far-from-equilibrium as a function of pH. Note lower VSI rates (solid square 'VSI' and 'VSIg' relate to total dissolution rate and "global" rate, respectively (see also Figure 7.12 this paper). [Adapted from Figure 1 of Arvidson *et al.* [44]].

Although not fully resolved, the most likely reason for the differences between VSI and powder data is the higher reactivity and accelerated dissolution at grain edges or boundaries, along which cleavage step density is maximized (Figure 7.14). Because the ratio of edge (boundary) length to surface area in a free grain varies as the reciprocal of grain diameter, this creates the possibility of a nonlinear relationship between grain size and specific dissolution rate. As particle size increases, the contributions of material from edge reactions relative to etch-pit-controlled wasting on terraces becomes negligible, and the "bulk" rate of very large particles may approach the surface-normal-retreat rate measured by VSI.

7.4.4. Dolomite ($CaMg(CO_3)_2$) dissolution kinetics by Lüttge et al. [20]

The authors have studied the dissolution rates and mechanism of three different cleavage faces of a dolomite crystal from Navarra, Spain using the same vertical scanning interferometer techniques previously used in the anorthite study (Lüttge *et al.* [85], see above). On the three sample surfaces, Lüttge *et al.* monitored a total of 37 different regions (each about 124 × 156 microns in size) as a function of time during dissolution at 25°C and pH 3. The dissolution process produced shallow etch pits with widths reaching up to 20 microns during eight hours of dissolution (Figures 7.15, 7.16). They reported that the depth development of etch pits as a function of time was remarkably similar for all pits on a given crystal surface.

An increase of dissolution rate from near zero to 4×10^{-11} [mol cm^{-2} s^{-1}] over a run duration of five hours was calculated based on etch pit distribution and volume as a function of time. The authors found that the time variation was different for each of the

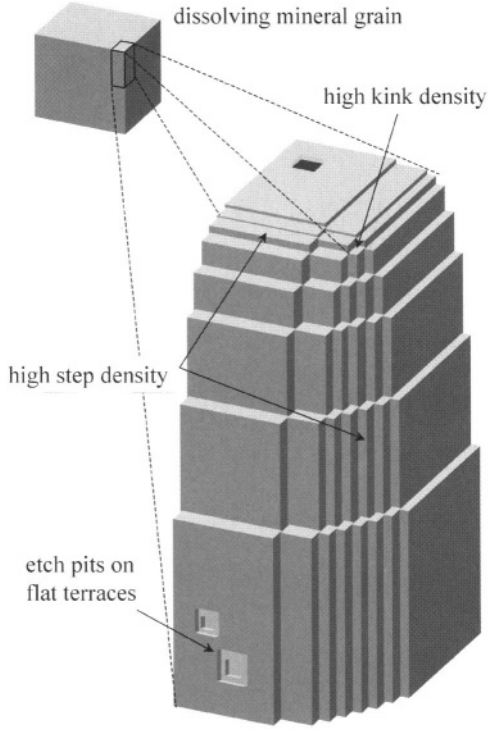
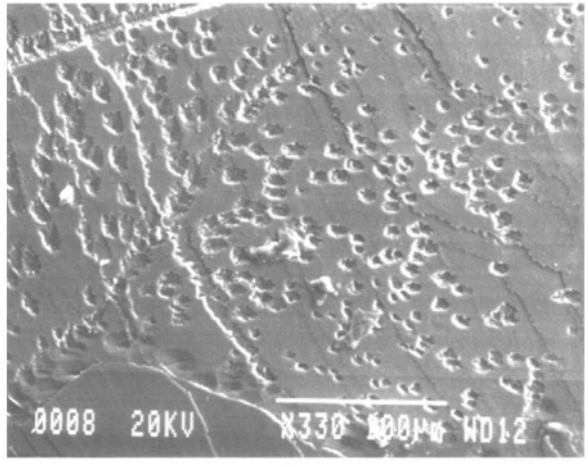
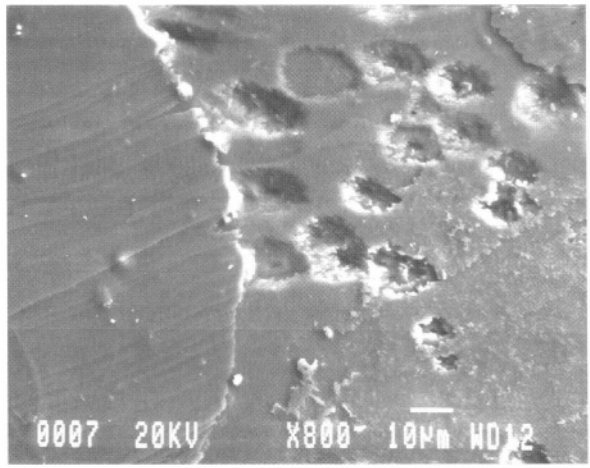


FIGURE 7.14. Schematic illustration of the possible distribution of steps and kinks on a mineral particle, illustrating higher "cleavage" or macro-step and kink densities at crystal face boundaries than flat terraces. [Adapted from Figure 7 of Arvidson *et al.* [44]].

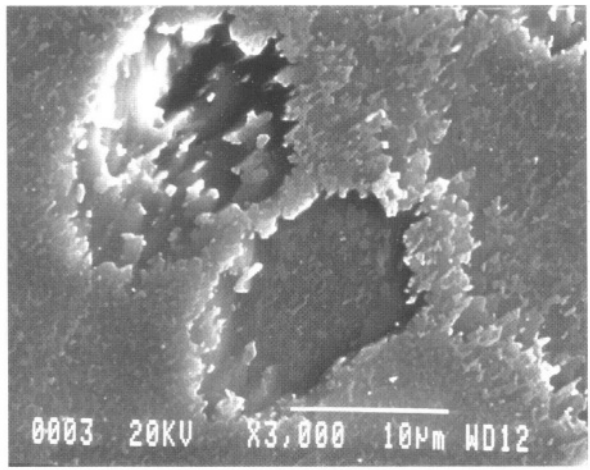


(A)

FIGURE 7.15. (A)–(C) SEM photomicrographs of a dolomite cleavage surface in different magnifications. Relatively shallow etch pits have formed during the dissolution process at pH 3 and 25°C. In comparison, these etch pits appear much deeper and more cone-shaped in the VSI measurements (compare e.g., Figure 7.16). This effect is due to the large exaggeration of the z-scale in VSI measurements. [Adapted from Lüttge *et al.* [20], Figure 2].



(B)



(C)

FIGURE 7.15. (Continued).

three cleavage surfaces studied (Figure 7.17). In addition, the absolute dissolution rates of different parts of the dolomite crystal surface were computed by using the internal reference technique described above. The different surfaces yielded an "average" rate of 1.08×10^{-11} [mol cm^{-2} s^{-1}] with a standard deviation of 0.3×10^{-11} [mol cm^{-2} s^{-1}] based on about 60 analyses. The mean absolute rate of the dolomite surface was about ten times slower than the rate calculated from etch pit dissolution alone. However, earlier published batch rate data that used BET surface areas yielded rates that were at least 30 to 60 times faster than their directly measured mean dissolution rate for the same pH and temperature.

Lüttge *et al.* also present a conceptual model for mineral dissolution that has been inferred from the surface topography obtained by the interferometry investigations (Fig-

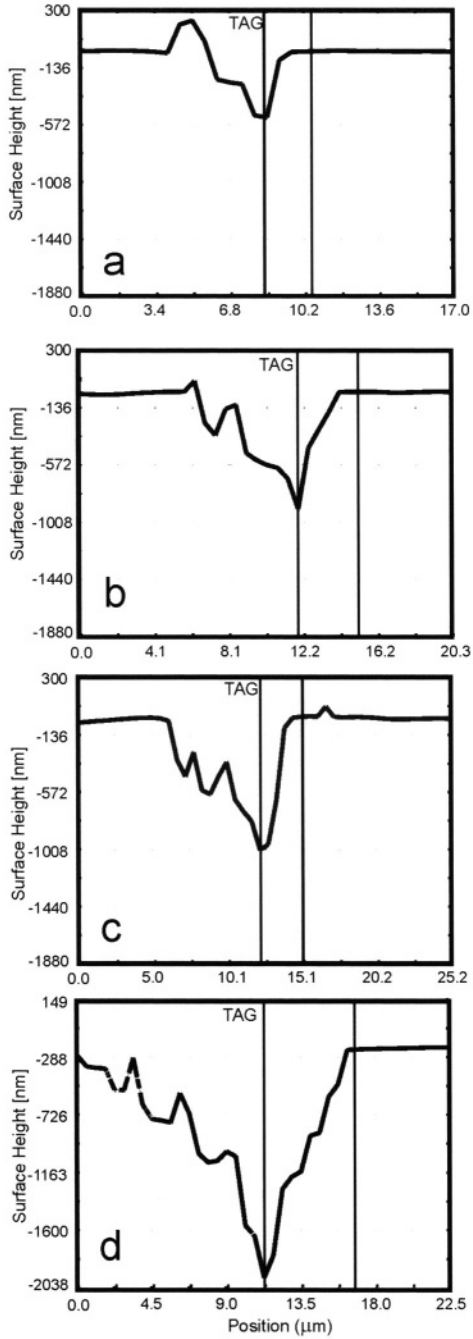


FIGURE 7.16. Series of four line-profiles that show the development of a single etch pit in a rhombohedral dolomite cleavage surface after: (A) 30 min, (B) after 60 min, (C) after 90 min, and (D) after 180 min of dissolution in hydrochloric acid solution (pH 3) at 25°C and 1 bar. [Adapted from Lüttge *et al.* [20], Figure 7].

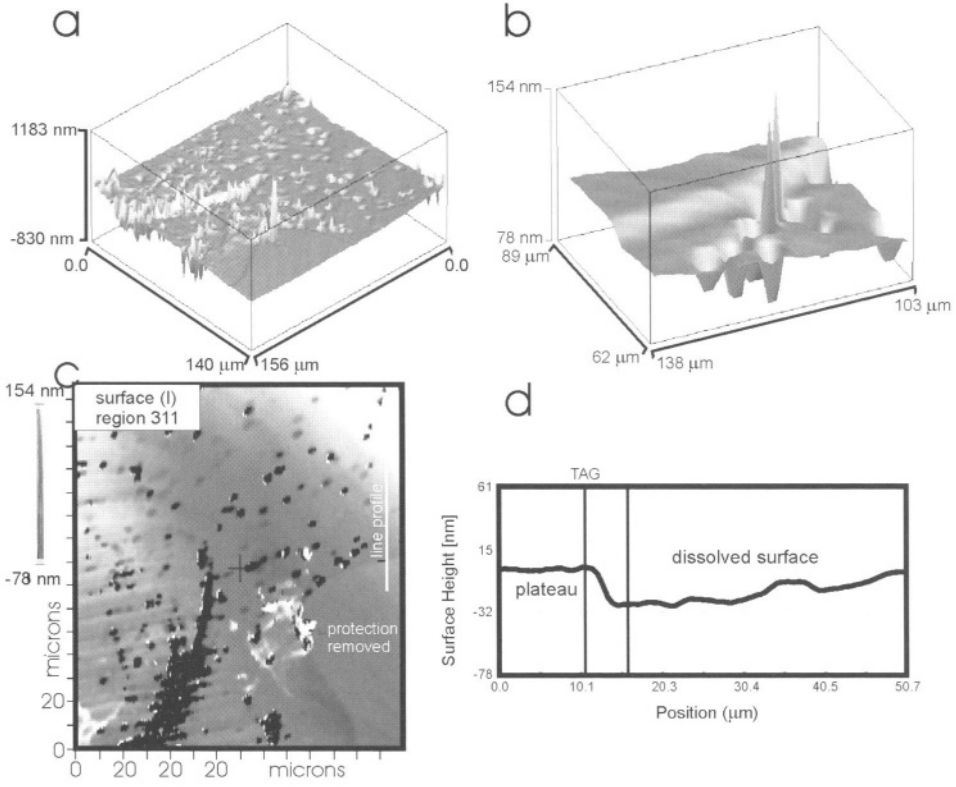
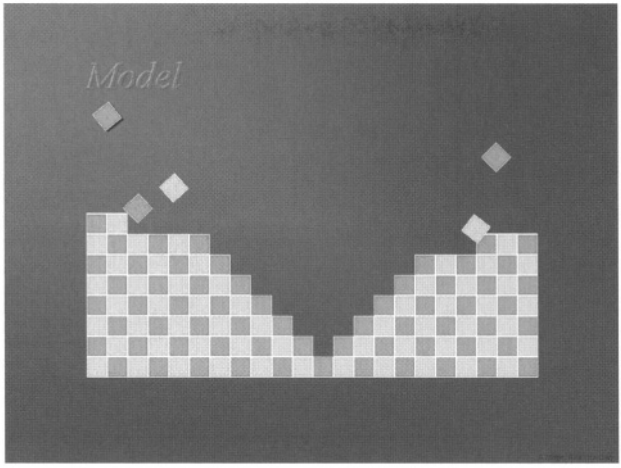


FIGURE 7.17. (A) Three-dimensional view of a dolomite cleavage surface (164 μm × 124 μm) after reacting for 60 minutes with a hydrochloric acid solution (pH 3, 25°C, 1 bar), and after a mask made from silicon rubber has been removed. Notice the small plateau used for reference; (B): blow-up of the reactive-non-reactive interface shown in (A); (C): contour plot of the same region after reacting 60 minutes; (D): line profile along the white line indicated in (C). [Adapted from Lüttge *et al.* [20], Figure 14].

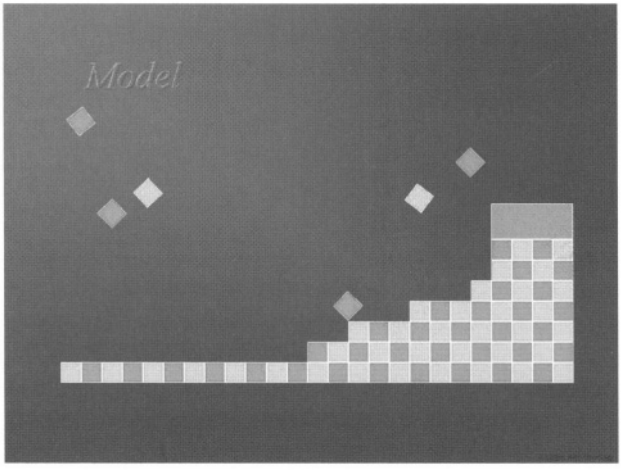
ure 7.18). This model is derived in detail by Lasaga and Lüttge [13,14]. In their model, mineral dissolution is not dominated by etch pit formation itself but rather by extensive trains of steps that originate at the outskirts of each etch pit. These stepwaves control the overall dissolution process as well as its dependence on temperature and saturation state.

7.4.5. Barite ($BaSO_4$) dissolution kinetics by Fewless and Lüttge [92]

Barite is a commercially important mineral in the oil and gas industry, both as a component of drilling fluid, and because of its tendency to form scales during production. There is also recent recognition of the role of barite in biogeochemical cycles (e.g., Dickens *et al.* [93]). Barite ($BaSO_4$) has long been used in studies of growth kinetics; initial nucleation theory results were produced using barite (Nielsen [94]). Barite was chosen because of its high insolubility ($K_{sp} = 10^{-9.98}$; Blount [95]), the lack of a significant gas phase, its structural and compositional properties as a simple, binary solid, and the ease with which it can be grown at room temperature. This latter property eliminates the need for mechanical



(A)



(B)

FIGURE 7.18. The sketch demonstrates how etch pits can generate steps (A). (B) shows that the lateral movement of steps across the crystal surface can form a "cliff" or plateau if a part of the surface is protected against dissolution (or corrosion).

grinding to produce reactant powders. Cleavage on barite's (001) surface is perfect and can yield an atomically flat surface, making it ideal for AFM and VSI studies (see also Becker *et al.* [96]). Recent work is described below.

Fewless and Lüttge studied barite dissolution by using VSI to examine the development of the (001) cleavage surface directly, at 25°C in both pure water and water + EDTA. In experiments similar in design to those described above for carbonates the authors quantified the surface-normal retreat as a function of time and fluid composition. All experiments were conducted at "far-from-equilibrium" conditions. SEM and VSI studies showed the development of different dissolution features beside a global surface-normal retreat, e.g., etch pits and canyon-like structures (Figure 7.19). The etch pits are attributed to screw disloca-

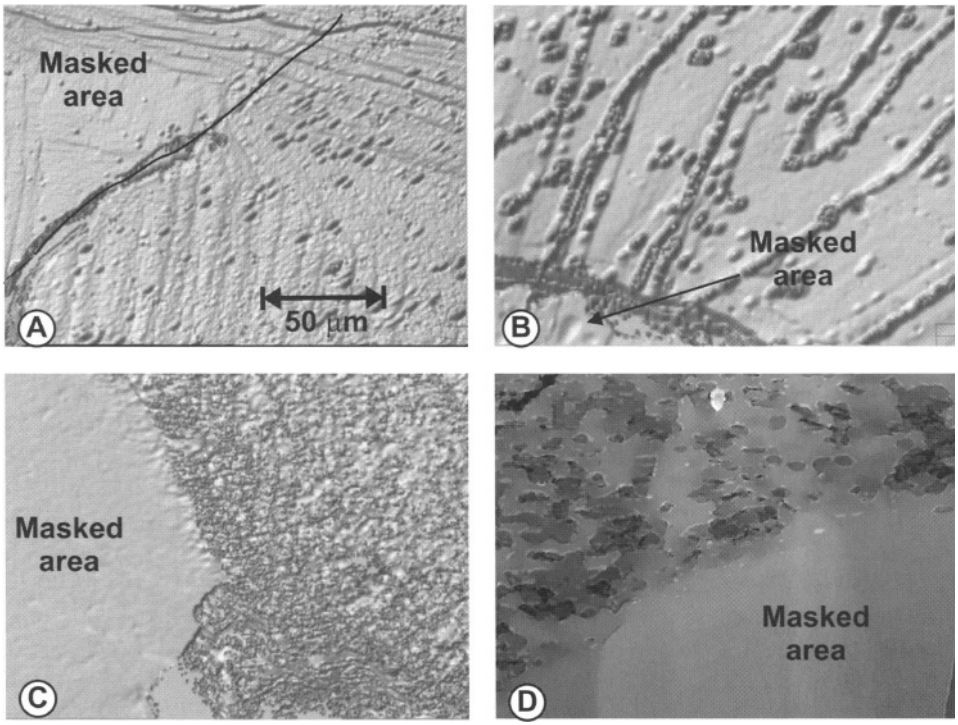


FIGURE 7.19. Various VSI images of a (001) cleavage surface of barite taken after 10 hours of dissolution. Masks were present during dissolution to preserve a pristine reference surface and were removed prior to data collection. (A) Scattered etch pits showing lineation in a different direction then the cleavage steps. (B) Coalescing etch pits following cleavage steps to form trenches. (C) Many etch pits in the process of coalescence. (D) Larger etch pits beginning to coalesce.

tions while the "canyons" are attributed to line defects. This result suggests that different dissolution processes contribute to the overall dissolution rate. The authors reported a heterogeneous distribution of etch pits across the surface, with some areas containing only a few scattered pits, while other areas were totally covered with pits. Etch pit morphology evolved during the course of the dissolution. Initially, pits were circular and cone shaped but the diameter and depth of the pits increased and developed crystallographically defined faceted walls. Others, however, maintained their conical shape. Statistical evaluations of the VSI results show that the two different types of etch pits occur in zones. Figure 7.19 shows the evolution of a typical (001) barite surface during dissolution in pure water.

Barite (001) surfaces dissolve at a reported rate of 5×10^{-8} [mol m^{-2} s^{-1}] $\pm 2 \times 10^{-8}$ [mol m^{-2} s^{-1}], in pure water (18 MΩ cm). This rate is slightly faster then reported AFM rates of 4×10^{-8} [mol m^{-2} s^{-1}] which were taken at 5°C higher temperatures, i.e., ~30°C (Dove and Platt [46]). However, the AFM rate was calculated based on step movement measurements and may be altered due to the heterogeneity of the dissolution process. In the presence of EDTA, which is used in several technical applications, dissolution rates of the (001) barite surface was expectedly faster.

7.5. SUMMARY AND DISCUSSION

The selection of VSI studies reviewed above has produced a number of observations and reaction rates that are specific for the minerals investigated. Beyond these details there are results that can be generalized and therefore lead to more fundamental insight. The most important result is that the direct (VSI) studies of the dissolving surface topography not only show the formation and development of numerous etch pits but also another quite general phenomenon: after some reaction time that varies depending on the mineral type, crystallographic orientation, and experimental conditions, the surface spots that had been masked and, therefore, protected against the dissolution attack always formed plateaus. This observation indicates that the entire mineral surface is retreating significantly during the dissolution process. Lasaga and Lüttge [13] termed this overall surface normal retreat "global dissolution" (compare also Lüttge *et al.* [20,85]; Arvidson *el al.* [44]). Note that the flat, often smooth parts of the crystal surface between the developing etch pits were previously thought to be passive, i.e., nonreactive, or at least much less reactive than the pitted surface. The VSI studies have modified this picture, suggesting that the dissolution process could be quantified by measuring the height of the plateau as a function of time. The VSI studies conducted so far indicate that the global dissolution is fairly uniform and systematic over the entire crystal surface. However, we have recognized that the dissolution process can be quite heterogeneous at a small scale, and the picture may change over a long reaction time (Arvidson and Lüttge [97]). An example that the dissolution process is often quite uniform and systematic are measurements of the surface retreat for 60 different regions of three different cleavage surfaces in dolomite (see above). All produced the same drop of surface topography to within 25% (Lüttge *et al.* [20]). Another important result is that, in most cases, neither the local development of etch pits nor their ultimate coalescence can account for the dissolution of the entire surface. Therefore, a universal process must govern the bulk crystal dissolution.

Lasaga and Lüttge [13,14] showed that VSI results on dolomite dissolution (Lüttge *et al.* [20]) can be interpreted by looking for effects generated by the etch pits beyond the impact of the localized formation of "dissolution holes". This model can now be extended and generalized to interpret the similar results of other mineral dissolution studies (see above). The importance of etch pits stems not only from the localized formation of a "hole", i.e., simple mass removal but in their ability to generate a continual sequence of steps. Direct observations by VSI show that the outskirts of etch pits consist of locations where the crystal surface first drops significantly from the pre-etch pit level. At the outskirts of the etch pit, therefore, one always finds steps. This consideration is in agreement with AFM observations, e.g., by Jordan and Rammensee [51]. The steps, however, are not static but can move into the rest of the mineral surface. In this way, the moving step "peels" away the top molecular layer. After this step moves far enough from the center of the dislocation-generated etch pit, another step can then be generated at the outskirts of the etch pit. The continuous movements of these steps into the crystal surface produces a train of steps, which Lasaga and Lüttge [13] termed "dissolution stepwaves".

The mechanism described above leads directly to a model of crystal dissolution in which the overall dissolution rate no longer depends on the mass removed by localized dissolution at etch pits, but is eventually controlled by the generation of step waves at the outskirts of the pits. Dissolution rate thus depends on the rates of step movements and

the step density. In this model, etch pit walls represent maximum step densities, while flat terraces represent a minimum in step density. Therefore, both parameters, step density and rate of step movement can serve as basic descriptors of crystal dissolution rates, and step spacing must in turn represent a combined functional dependence on kink nucleation rate and step velocity. This dependence can be better understood if we remember that kink sites must be generated to "move" the step. For example, a perfect step will be stable as long as no atom is released from the step. It takes a higher activation energy to remove the first molecule. The removal of the first molecule generates two kink sites. Now, the molecules in the kink positions can be removed with less activation energy, and since kink sites are self-reproducing, molecule after molecule can be removed with the same low activation energy. As a result, the step will move one molecular row forward. Based on these considerations, the coupling between kink sites and step movement, and as shown in Lasaga and Lüttge [13,14], crystal dissolution can consequently be treated and understood as a many-body problem.

7.6. COMPUTER SIMULATIONS

Building on the increasing power of computers and based on the assumption that the dissolution kinetics of crystals are not necessarily governed by a single rate-controlling elementary reaction but can be treated as a many-body problem, Lasaga and Lüttge [13-15,98] have developed a sequence of computer simulations with increasing complexity. These simulations range from a very simple model (Lasaga and Lüttge [13,14]) that treats the crystal structure as being composed of molecular "blocks", to models of AB and A_3B molecular structures (Lasaga and Lüttge [15]) to "real" mineral structures like that of anorthite ($Ca[Al_2Si_2O_8]$) and albite ($Na[AlSi_3O_8]$) feldspars or phyllosilicates (e.g., Lasaga and Lüttge [98]). All these models make extensive use of Monte Carlo techniques as used earlier by, e.g., Blum and Lasaga [12]). The reaction mechanism is generated by a combination of elementary reactions, i.e., bond breaking and bond forming, as well as some basic reactions, i.e., the dissolution of surface units (units = molecules), adsorption and incorporation of solution units, and surface mobility of surface units. In addition, the crystal structure is incorporated to calculate all the interactions of neighboring atoms as well as possible defects (e.g., screw or point dislocations). This approach is different from approaches based on either precursor complexes or adsorption models (compare e.g., Gautier *et al.* [32]; Oelkers *et al.* [33] and Oelkers [37]).

Monte Carlo techniques treat crystal dissolution or growth as a stochastic process. Rates of microscopic processes, e.g., the bond breaking/forming processes during dissolution are translated into probabilities (Van der Erden *et al.* [8], Gilmer [9–11]). Also, the interaction between the solution and the solid can be simulated by arrival rates (probabilities) of molecules from the solution at the crystal surface and vice versa for the departure rates (probabilities). Therefore, it is no problem to adjust the model for variations, e.g., in the saturation state of the solution. We can illustrate the procedure for the elementary reaction of bond-breaking/forming with a simple example from Lasaga and Lüttge [15]. Let us assume that the total number of bonds between two A molecules (i.e., A–A bonds) on the dissolving surface is N_{AA}, and that the total number of broken A–A bonds on the same

surface is N_{AA}^{broken}. If steady state for these processes is assumed, then we can write

$$\frac{N_{AA}^{broken}}{N_{AA}} = \frac{k_-}{k_+}, \tag{4}$$

where k_- and k_+ are the rate constants for the breaking and the formation of an A–A bond. If there are a few bonds broken at any one time, then $N_{AA} \sim N_{AA}^{tot} \sim$ constant. As a result, the probability that any one given A–A bond is broken is given by

$$P_{AA} = \frac{k_-}{k_+}. \tag{5}$$

If we ignore all other complexities in the dissolution process, e.g., arrival of ligands and other units, surface diffusion, etc., then, the rate of dissolution of units with n bonds would depend on the number of A units with n bonds on the surface, N_n^A. It would depend also on the probability that all n bonds are found broken at the same time (although the actual breakdown of each bond occurs as a separate event)

$$\text{dissolution rate} = (P_{AA})^n N_n^A. \tag{6a}$$

This last equation can be rewritten as

$$\text{rate} = \left(\frac{k_-}{k_+}\right)^n N_n^A. \tag{6b}$$

The comprehensive approach that was used in Lasaga and Lüttge [15,98] includes other complexities in the dissolution process, e.g., arrival of ligands and other units, surface diffusion, etc. This example nicely illustrates the potential of this general approach. The authors were able to analyze fundamental concepts such as activation energy, surface free energy, the solubility product, inhibition and catalysis, and the dependence on the saturation state of the solution. The crystal-based reaction mechanism used in the simulation has lead to a unified explanation of many experimentally observed water-rock features. In addition, it has produced a series of unexpected but essential modifications of kinetic results that had not been completely understood before.

The ability to model the processes at the molecular level is of particular interest because, as highlighted before, there are currently no analytical techniques available that allow observation of these processes in the laboratory. Therefore, it is even more important to improve our capabilities to simulate the processes at the molecular level. The increasing speed of computer processors and the ability of parallel processing will allow us to treat to large systems with millions of molecules in the near future. If successful, this work will help to produce valuable insight into crystal dissolution kinetics and will provide the tools to predict dissolution behavior that can then be tested by experiments.

Currently, results are based on relatively small solid models of 80 by 80 or 200 by 200 molecules per layer. We use mirror boundary conditions which create a quasi endless crystal surface. However, the first results are promising. As an example, Figure 7.20 is generated by a computer simulation (Lasaga and Lüttge [13]) and shows a typical scenario of the stepwave model introduced above. In an idealized case, on a freshly cleaved

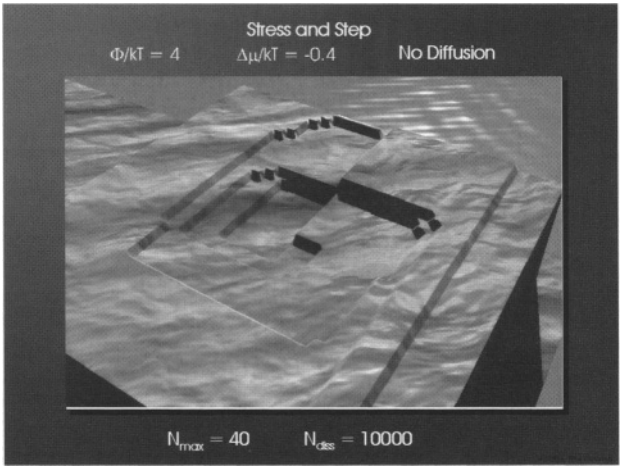


FIGURE 7.20. Monte Carlo simulation for the formation of an etch pit at the site of a screw dislocation. It can be seen that the pit generates steps that move outwards and lower the surrounding crystal surface. The smallest height of a step is one molecular unit. Steps that show a larger height, i.e.. a multiple of a molecular unit have "bunched" and do not move any more.

unreacted crystal surface, the surface is initially (atomically) Hat and free of etch pits. In crystal dissolution studies, experimentalists often attempt to create such initial conditions. However, we know from AFM, VSI, and SEM observations that in reality actual mineral surfaces are much more complex. When the crystal surface is exposed to the solution, etch pits will develop if the saturation state is far enough removed from equilibrium (see discussion in Lasaga and Lüttge [13–15]). During this stage of the dissolution process, etch pits are surrounded by the remaining flat cleavage surface. Without an absolute reference present, the observation of this kind of topography would suggest that the dissolution is all coming from the etch pits themselves. However, throughout the dissolution process, stepwaves from each of the etch pits have been lowered the entire mineral surface (Figures 7.18, 7.21). It is important to note that different from "real" waves, e.g., electromagnetic waves, these stepwaves can not constructively or destructively interfere with each other. As a result, the actual surface topography would look like that in Figure 7.22, exhibiting a uniform surface drop if a reference surface were available (as in the interferometry investigations).

Natural crystals always contain dislocations, and consequently there are many dislocations outcropping on a cleavage surface. These dislocations produce many etch pits during dissolution (e.g., Berner and Schott [99]). As explained above, the general formulation of the rate constant is based on the velocity and separation of the steps generated by each etch pit. If the mineral surface had only one etch pit site, the steps would move out like waves on the surface (Figure 7.23). The dissolution, i.e., the surface normal retreat at any point on the surface would depend on the passage of each wave. If we think of each step as having a step height of Δh, then after the passage of the first wave, the surface will move down by Δh. After the second wave passes our spot, the surface will move down another Δh.

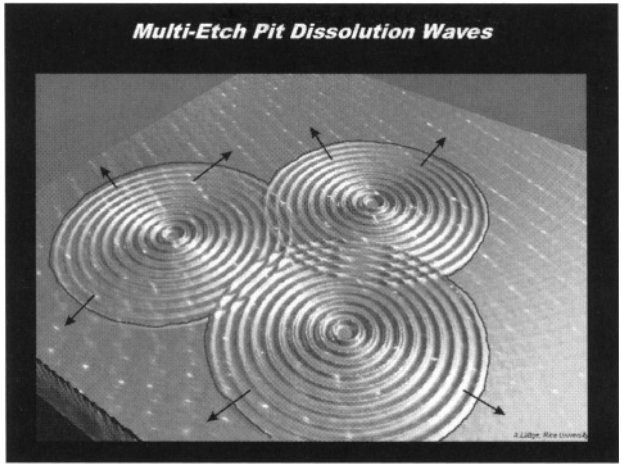


FIGURE 7.21. Computer-generated sketch shows "stepwaves" moving outward from their centers (etch pits). The trains of steps do not truly "interfere" with each other, they eliminate each other. Therefore, the front of steps generated by multiple etch pits is not significantly larger than the front generated by just one pit. This observation can explain why the dissolution rate of a crystal does not depend strongly on its etch pit density (compare also Lasaga and Lüttge [13,14]).

Based on this model, Lasaga and Lüttge [13,14] have derived the following fundamental rate law:

$$\text{rate} = v_\infty(1 - e^{\Delta G/kT})\tanh\left(\frac{1}{2x_s f(r_{\text{pit}})}\right) f(r_{\text{pit}}), \tag{7}$$

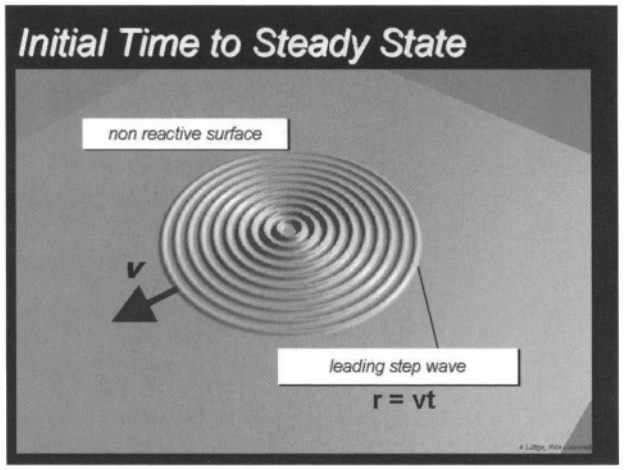
where the key function $f(r_{\text{pit}})$ is:

$$f(r_{\text{pit}}) \equiv \left(1 - \frac{1 - e^{-\frac{\sigma v}{rkT}} e^{\frac{u(r)v}{kT}}}{1 - e^{\frac{\Delta G}{kT}}}\right), \tag{8}$$

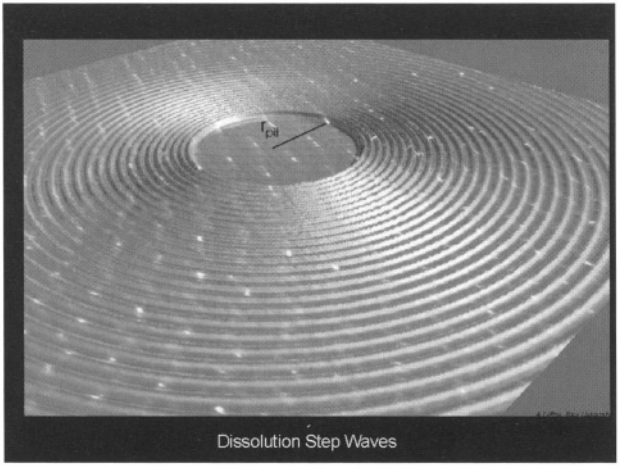
where ΔG is the difference in Gibbs free energy, k is Boltzmann's constant, T is absolute temperature, r_{pit} is location of the (step) velocity minimum (Å), x_s is a surface diffusion distance expressed in molecular units, σ is the surface free energy, and v is the molecular volume.

This expression formally links the velocity of a step itself with the free energy available for the dissolution reaction, and predicts the overall dissolution rate for crystals based on the expansion of etch-pit induced stepwaves. This equation can be understood as the counter part to the Burton–Cabrera–Frank (BCF) equation for dislocation-step-controlled crystal growth (Burton *et al.* [100]).

Calculations based on this rate law can help to explain a number of long standing problems and observations, e.g., the fact that dissolution rates do not vary within the same range dislocation densities on crystal surfaces do. The rate law can also be used to fit the ΔG dependence measured in a number of flow-through experiments using mineral powders (e.g., Figure 24, Lasaga and Lüttge [13–15]). These theoretical considerations have in turn provided the basis for Monte Carlo computer simulations described above.



(A)



(B)

FIGURE 7.22. Computer-generated sketch shows "stepwaves" moving outward from a center of a single etch pit. (A) As long as the front of the stepwave has not reached the edge of the crystal, steady state is not reached. (B) Steady state has been established after the front of the steps has reached the outskirts of the crystal face, and step spacing and step movement are constant.

7.7. FUTURE WORK

Future work will focus on two general approaches: (1) further development of experimental/analytical techniques like VSI and AFM that allow the direct observation of the processes at the crystal surface, and (2) more powerful Monte Carlo based computer simulations that treat crystal dissolution and growth as many-body problems. As mentioned above, Lüttge and Lasaga [98] have shown recently that it is possible to simulate "real" crystal structures of feldspars, e.g., albite and anorthite, or phyllo-silicates, e.g., phlogopite or muscovite, with large numbers of atoms. It will be important to expand the size of

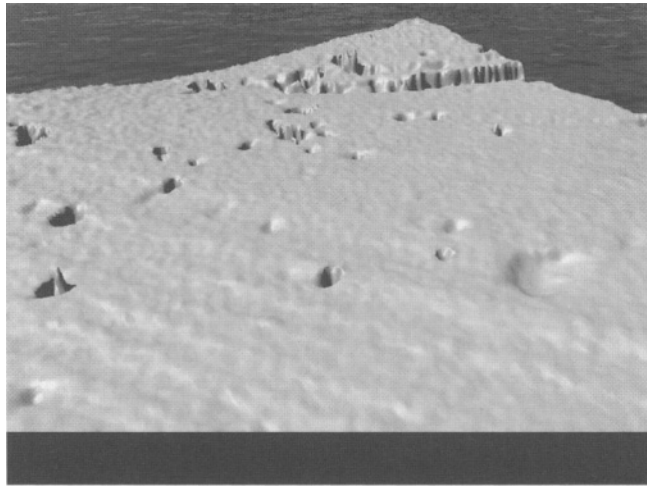


FIGURE 7.23. VSI/Bryce4 image of a dolomite cleavage surface dissolved for 1 hour at pH 3 and 25°C. The false color image shows that the lateral movement of pit generated steps has formed a "cliff" where the surface was formerly protected by a mask. [Modified from Figure 1A of Lasaga and Lüttge [13]].

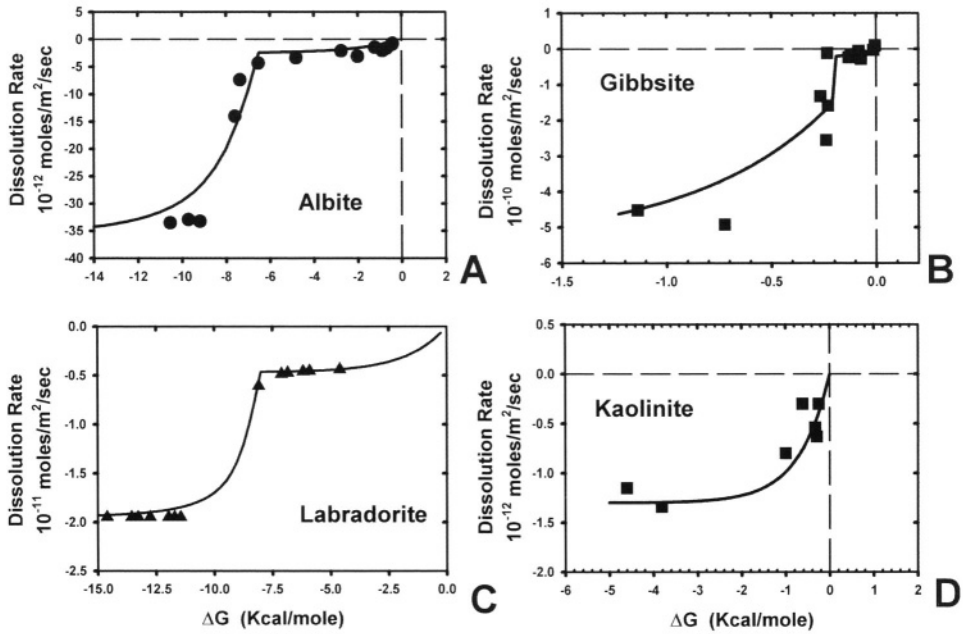


FIGURE 7.24. Comparison of the full dissolution theory (Lasaga and Lüttge [2,3]) with four published experimental data sets generated by Burch *et al.* [73] (albite); Taylor *et al.* [107] (labradorit); Nagy and Lasaga [108] (gibbsite); and Nagy *et al.* [109] (kaolinite).

such Monte Carlo simulations to a size representing a significant part of a crystal surface. Parallel processing and improvements in optimization will allow the modeling of systems composed of many millions of molecules. Systems of this size can be explored in terms of distinct surface sites, step–step, and step–dislocation interactions. Computer simulations of this size and quality are the prerequisit for a direct comparison of high-resolution laboratory observations at the mineral surface.

Such a combined approach will allow the investigation of crystal dissolution and growth kinetics from a quantum mechanical to the macroscopic scale. Of course, modeling holds the most promise for those processes that are beyond the reach of laboratory experiments but, in addition, model predictions based on molecular simulations should be tested experimentally to provide a robust link between results from the nano-, micro-, and macroscale.

We recognize that earlier applications of *ab initio* and DFT have suffert from the fact that calculations always had to be conducted for molecule clusters in the gas phase (e.g., Xiao and Lasaga [101,102]). However, we are now able to model, e.g., feldspar dissolution in a much more realistic manner. Recently, Xiao and Lüttge [103] have presented a simulation where solute molecules are surrounded by solvation shells and solvent water is modeled as a dielectric continuum (compare also Kubicki [104] and Cygan and Kubicki [7]). The combination of such "realistic" *ab initio* and DFT based calculations with Monte Carlo simulations will allow the prediction of crystal dissolution in laboratory experiments. These predictions can then be tested if analytical tools like AFM and VSI are available that allow the direct observation of the crystal surface and the quantification of the dissolution process.

At this point, it is important to note that currently computer simulations based on Monte Carlo techniques allow us to investigate and visualize kinetic processes at the molecular scale, while, at least in the forseeable future, neither VSI nor AFM will be capable of accomplishing this task, even if AFM is already capable of providing "true" atomic, i.e., angstrom-scale resolution. With respect to VSI, it is our goal to modify and complement the VSI hardware and software in a way that it will provide extensive *in situ* capabilities in the forseeable future (Lüttge [105]).

Developments are already underway that will significantly improve the lateral resolution of VSI (Lüttge [105], Sawyer *et al.* [106]). As discussed above, the current VSI MicroXAM system acquires images with lateral pixel spacing of ~500 nm by using a 50× Mirau objective. There are other methods like AFM for acquiring higher resolution images of surfaces but none have the non-destructive character of VSI. In addition, VSI features significantly faster data acquisition and provides a field of view orders of magnitude larger than high-resolution AFM. The non-destructiveness is highly desirable for imaging surfaces involving living microorganisms.

We are optimistic that our new approach will help to improve the lateral resolution of VSI to better than 100 nanometers, i.e., 0.5 to 1 order of magnitude more resolution than currently available, while the vertical resolution should remain better than 1 nanometer. This exciting development is based on the use of current VSI methods but exploiting the addition of a high-precision (sub-pixel) sample positioning stage. This technology allows now the positioning of samples with a precision at the single digit nanometer scale. We utilize the piezzo-positioning technique to acquire multiple overlapping images of the sample surface, offset by sub-pixel spacing offsets. The acquired images, which have the

original resolution of the standard MicroXAM interferometer, are then assembled into a single image with finer pixel spacing. The resulting image is equivalent to a blurred and defocussed higher resolutionimage. The image is restored to a higher resolution by digital image restauration (deblurring). Experiments with synthetic height maps with perfect data (i.e., full floating point resolution in three dimensions, and known point spread function) have already demostrated the great potential of this approach (Sawyer *et al.* [106]). However, we recognize that the VSI data are not perfect. The basic images from the VSI have very good, but limited, vertical resolution. Because digital deblurring methods involve taking numerical derivatives of the surface height of the sample, they are very sensitive to noise and rounding errors. We support that this will ultimately define the limit of the improvement of the lateral resolution. Therefore, we will apply this strategy seeking progressively higher resolution until the method fails to yield additional improvement. This has the potential to make VSI a true nano-scale analytical technique and would have a broad range of applications in a variety of fields. VSI could become the first non-destructive instrument that could quantify surface topography and reaction rates with "true" nanoscale-resolution.

ACKNOWLEDGEMENTS

I thank especially R.S. Arvidson for many interesting discussions and for substantial help with editing the manuscript. I appreciate the critical and helpful reviews of the manuscript by G.R. Dickens, C.-T. Lee, and D.S. Sawyer. Special thanks also to A.C. Lasaga, E.W. Bolton, D.M. Rye, P.G. Conrad, K.H. Nealson, J.P. Icenhower, J.E. Amonette, K. Tsukamoto and my graduate students U. Winkler, M.S. Beig, T.A. Fewless, K.J. Davis, L. Zhang, and M.D. Vinson for their collaboration, help, and many thoughtful discussions.

I gratefully acknowledge support for the research studies from the US Department of Energy (DOE) (grant #DE-FG07-01R63295), DOE/Batelle (grant #400635A9E), the US Jet Propulsion Laboratories (JPL) (grant #RFP 400635-A9E), the US National Science Foundation (NSF) (grant #EAR-0125667, EAR-0236761), and subcontracts from University of Southern California (USC) (grant #DE-FG03-02ER63331 and DE-FG03-02ER63427), and last not least generous support from ADE-Phase Shift, EXXONMOBIL, and Halliburton.

REFERENCES

1. J.F. Banfield and K.H. Nealson, *Geomicrobiology: Interactions Between Microbes and Minerals. Reviews in Mineralogy & Geochemistry* **35** (Mineralogical Society of America, Washington, DC, 1997).
2. M.A. Velbel, Effect of chemical affinity on feldspar hydrolysis rates in two natural weathering systems, *Chem. Geol.* **78**, 245–253 (1989).
3. M.A. Velbel, Influence of temperature and mineral surface characteristics on feldspar weathering rates in natural and artificial systems, *Water Resour. Res.* **26**, 3049 (1990).
4. A.F. White, in: *Chemical Weathering Rates of Silicate Minerals. Reviews in Mineralogy,* **31**, Eds A.F. White and S.L. Brantley (Mineralogical Society of America, Washington, DC, 1995) pp. 407–461.
5. A.C. Lasaga, in: *Chemical Weathering Rates of Silicate Minerals. Reviews in Mineralogy* **31**, Eds A.F. White and S.L. Brantley (Mineralogical Society of America, Washington, DC, 1995) pp. 23–81.
6. A.C. Lasaga, *Kinetic Theory and Applications in Earth Sciences* (Princeton Press, Princeton, 1998).
7. R.T. Cygan and J.D. Kubicki (Eds), *Molecular Modeling Theory: Applications in the Geosciences. Reviews in Mineralogy & Geochemistry* **42** (Mineralogical Society of America, Washington, DC, 2001).

8. J.P. Van der Eerden, P. Bennema and T.A. Cheiepanova, Survey of Monte Carlo simulations of crystal surfaces and crystal growth, *Progr. Crys. Growth Character* **1**, 219–251 (1978).
9. G.H. Gilmer, Growth on imperfect crystal faces. I. Monte Carlo growth rates, *J. Crys. Growth* **35**, 15–28 (1976).
10. G.H. Gilmer, Computer simulations of crystal growth, *J. Crys. Growth* **42**, 3–10 (1977).
11. G.H. Gilmer, Computer models of crystal growth, *Sience* **208**, 355–363 (1980).
12. A.E. Blum and A.C. Lasaga, in: *Aquatic Surface Chemistry; Chemical Processes at the Particle–Water Interface,* Ed. W. Stumm (*Environm. Sci. & Techn. Series,* J. Wiley, New York. 1987) pp. 255–292.
13. A.C. Lasaga and A. Lüttge, Variation of crystal dissolution rate based on a dissolution stepwave model, *Science* **291**, 2400–2404 (2001).
14. A.C. Lasaga and A. Lüttge, A model for crystal dissolution, *Eur. J. Min.* **15**, 603–615 (2003).
15. A.C. Lasaga and A. Lüttge. A fundamental approach to mineral dissolution kinetics, *Am. Min.* **89** (2004, in revision).
16. A.F. White and S.L. Brantley (Eds), *Chemical Weathering Rales of Silicate Minerals. Reviews in Mineralogy* **31** (Mineralogical Society of America, Washington, DC, 1995).
17. D.M. Kerrick (Ed.), *Contact Metamorphism. Reviews in Mineralogy* **26** (Mineralogical Society of America. Washington, DC, 1991).
18. J.W. Morse and R.S. Arvidson, The dissolution kinetics of major sedimentary carbonate minerals, *Earth Sci. Revs.* **58**, 51–84(2002).
19. P.C. Lichtner, C.I. Steefel and E.H. Oelkers (Eds.), *Reactive Transport in Porous Media. Reviews in Mineralogy* **34** (Mineralogical Society of America, Washington, DC, 1996).
20. A. Lüttge, U. Winkler and A.C. Lasaga, Dolomite dissolution kinetics studied with vertical scanning white light interferometry, *Geochim. Cosmochim. Acta* **67**, 1099–1116 (2003).
21. P.G. Conrad, A. Lüttge and K.H. Nealson, Surface recognition and attachment of Shewanella oneidensis: Initial results and environmental implications, Department of Energy's Natural and Accelerated Bioremediation Research Program (DOE-NABIR) PI Workshop, Warrenton, Virginia (March 18–20, 2002).
22. A. Lüttge and P.G. Conrad, Bacterial control of calcite dissolution kinetics, *Appl. and Environml. Microbiol.* **70** (2004, in press).
23. M.F. Hochella, Jr. and J.F. Banfield, in: *Chemical Weathering Rates of Silicate Minerals. Reviews in Mineralogy* **31**, Eds A.F. White and S.L. Brantley (Mineralogical Society of America, Washington, DC, 1995) pp. 353–406.
24. S.L. Brantley, A.F. White and M.E. Hodson, Surface area of primary silicate minerals, in: *Growth, Dissolution and Pattern Formation in Geosystems,* Eds B. Jamtveit and P. Meakin (Kluwer Academic Publishers. Dordrecht, 1999) pp. 291–326.
25. S.L. Brantley and N.P. Mellott, Surface area and porosity of primary silicate minerals. *Am. Min.* **85**, 1767–1783(2000).
26. A.A. Jeschke and W. Dreybrodt, Dissolution rates of minerals and their relation to surface morphology, *Geochim. Cosmochim. Acta* **66**, 3055–3062 (2002).
27. A. Lüttge, Model calculations of BET, geometric, and reactive surface area during a dolomite dissolution reaction, *Geochim. Cosmochim. Acta* (2004, subm.).
28. P.V. Brady and J.V. Walther, Controls on silicate dissolution rates in neutral and basic pH solutions at 25 degrees C, *Geochim. Cosmochim. Acta* **53**, 2823–2830 (1989).
29. W. Stumm and E. Wieland, in: *Aquatic Chemical Kinetics* (J. Wiley, New York, 1990) pp. 367–400.
30. W. Stumm and Wollast, Coordination chemistry of weathering; kinetics of the surface-controlled dissolution of oxide minerals, *Revs. in Geophys.* **28**, 53–69 (1990).
31. P. Van Cappellen, L. Chariet, W. Stumm and P. Wersin, A surface complexation model of the carbonate mineral–solution interface, *Geochim. Cosmochim. Acta* **57**, 3505–3518 (1993).
32. J.-M. Gautier, E.H. Oelkers and J. Schott, Experimental study of K-feldspar dissolution rates as a function of chemical affinity at 150 degrees C and pH 9, *Geochim. Cosmochim. Acta* **58**, 4549–4560 (1994).
33. E.H. Oelkers, J. Schott and J.L. Devidal, The effect of aluminum, pH, and chemical affinity on the rates of aluminosilicate dissolution reactions, *Geochim. Cosmochim. Acta* **58**, 2011–2024 (1994).
34. W.H. Casey and C. Ludwig, in: *Chemical Weathering Rates of Silicate Minerals. Reviews in Mineralogy* **31**, Eds A.F. White and S.L. Brantley (Mineralogical Society of America, Washington, DC, 1995) pp. 87–114.
35. P.M. Dove, in: *Chemical Weathering Rates of Silicate Minerals. Reviews in Mineralogy* **31**, Eds A.F. White and S.L. Brantley (Mineralogical Society of America, Washington. DC, 1995) pp. 235–282.

36. A.E. Blum and L.L. Stillings, in: *Chemical Weathering Rates of Silicate Minerals. Reviews in Mineralogy* **31**, Eds A.F. White and S.L. Brantley (Mineralogical Society of America, Washington, DC, 1995) pp. 291–346.
37. E.H. Oelkers, General kinetic description of multioxide silicate mineral and glass dissolution, *Geochim. Cosmochim. Ada* **65**, 3703–3720 (2002).
38. A.E. Blum and A.C. Lasaga, The role of surface speciation in the dissolution of albite, *Geochim. Cosmochim. Acta* **55**, 2193–2201 (1991).
39. J.-M. Gautier, E.H. Oelkers and J. Schott, Are quartz dissolution rates proportional to B.E.T. surface areas'? *Geochim. Cosmochim. Acta* **65**, 1059–1070 (2001).
40. C. Anbeek, Surface roughness of minerals and implications for dissolution studies, *Geochim. Cosmochim. Acta* **56**, 1461–1469, (1992).
41. S. Brunauer, P.H. Emmet and E. Teller, Adsorption of gases in multimolecular layers, *J. Amer. Chem. Soc.* **60**, 309–319(1938).
42. M.F. Hochella, Jr., in: *Mineral Surfaces, Min. Soc. Series* **5**, Eds D.J. Vaughan and R.A.D. Pattrick (Chapman & Hall, London, 1995) pp. 17–60.
43. R.S. Arvidson, M.S. Beig and A. Lüttge, A new method for comparison of single crystal and powder dissolution rates by vertical scanning interferometry, *Am. Min.* **89**, 51–56 (2004).
44. R.S. Arvidson, I.E. Ertan, J.E. Amonette and A. Lüttge, Variation in calcite dissolution rates: a fundamental problem? *Geochim. Cosmochim. Acta* **67**, 1623–1634 (2003).
45. D. Bosbach and W. Rammensee, In situ investigation of growth and dissolution on the (010) surface of gypsum by scanning force microscopy, *Geochim. Cosmochim. Acta* **58**, 843–849 (1994).
46. P.M. Dove and F.M. Platt, Compatible real-time reaction rates for in situ imaging of mineral–water interactions using scanning force microscopy, *Chem. Geol.* **127**, 331–338 (1996).
47. C.M. Eggleston, High-resolution scanning probe microscopy; tip–surface interaction, artifacts, and applications in mineralogy and geochemistry, *Clay Min. Soc. Workshop Lec.* **7**, 3–90 (1994).
48. S.R. Higgins, C.M. Eggleston, K.G. Knauss and C.O. Boro, A hydrothermal atomic force microscope for imaging in aqueous solution up to 150°C, *Rev. Sci. Instrum.* **69**, 2994–2998 (1998).
49. Y. Liang, D.R. Baer, J.M. McCoy, J.E. Amonette and J.P. LaFemina, Dissolution kinetics at the calcite–water interface, *Geochim. Cosmochim. Acta* **60**, 4883–1887 (1996).
50. H.H. Teng, P. Fenter, L. Cheng and N.C. Sturchio, Resolving orthoclase dissolution processes with atomic force microscopy and X-ray reflectivity, *Geochim. Cosmochim. Acta* **65**, 3459–3474 (2001).
51. G. Jordan and W. Rammensee, Dissolution rates of calcite $(10\ \bar{1}4)$ obtained by scanning force microscopy: microtopography-based dissolution kinetics on surface with anisotropic step velocities, *Geochim. Cosmochim. Acta* **62**, 941–947 (1998).
52. A.S. Lea, J.E. Amonette, D.R. Baer, Y. Liang and N.G. Colton, Microscopic efects of carbonate, manganese, and strontium ions on calcite dissolution, *Geochim. Cosmochim. Acta* **65**, 369–379 (2001).
53. C.M. Eggleston and A. Lüttge, A comparison of atomic force microscopy (AFM) and vertical scanning interferometry (VSI), *Geochem. Trans.* (2004, in prep.).
54. Y.K. Kim, J.D. Trolinger, B.R. Reddy and R.B. Lal, Refractive index variation measurements of fluids using common-path Zeeman interferometry, *J. Mod. Optics* **45**, 1637 (1998).
55. P.G. Vekilov, Yu.G. Kuznetsov and A.A. Chernov, Interstep interaction in solution growth; (101) ADP face, *J. Crys. Growth* **121**, 643–655 (1992).
56. P.G. Vekilov, M. Ataka and T. Katsura, Laser Michelson interferometry investigation of protein crystal growth, *J. Crys. Growth* **130**, 317–320 (1993).
57. P.G. Vekilov, L.A. Monaco and F. Rosenberger, High-resolution interferometric technique for in-situ studies of crystal growth morphologies and kinetics, *J. Crys. Growth* **146**, 289–296 (1995).
58. P.G. Vekilov and Yu.G. Kuznetsov, Growth kinetics irregularities due to changed dislocation source activity; (101) ADP face, *J. Crystal Growth* **119**, 248–260 (1992).
59. P.G. Vekilov and F. Rosenberger, Dependence of lysozyme growth kinetics on step sources an impurities, *J. Crys. Growth* **158**, 540–551 (1996).
60. P.G. Vekilov and J.I.D. Alexander, Dynamics of layer growth in protein crystallization, *Chem. Rev.* **100**, 2061–2089 (2000).
61. N.A. Booth, A.A. Chernov and P.G. Vekilov, Characteristic lengthscales of step bunching in KDP crystal growth: in-situ differential phase-shifting interferometry study, *J. Crys. Growth* **237–239**, 1818–1824 (2002).
62. O. Gliko, N.A. Booth, E. Rosenbach and P.G. Vekilov, Phase-shifting interferometry for the study of the step dynamics during crystallization of proteins, *Ciys. Growth and Design* **2**, 381–385 (2002).

63. Y. Kim, B.R. Reddy, T.G. George and R.B. Lal, Optical heterodyne interferometry technique for solution crystal growth rate measurement. *Opt. Eng.* **37**, 616 (1998).
64. K. Tsukamoto, In situ observation of mono-molecular growth steps on crystals growing in aqueous solution, *J. Crys. Growth* **61,** 199–209 (1983).
65. K. Tsukamoto, In situ observation of crystal growth from solution, *Faraday Discussion* **95,** 183–189 (1994).
66. K. Onuma, K. Tsukamoto and I. Sunagawa, Measurements of surface supersaturations around a growing K-Alum crystal in aqueous solution, *J. Crys. Growth* **98,** 377–383 (1989).
67. K. Onuma, K. Tsukamoto and I. Sunagawa, Growth kinetics of K-alum crystals in relation to the surface supersaturations, *J. Crys. Growth* **100,** 125–132 (1990).
68. K. Maiwa, K. Tsukamoto and I. Sunagawa, Activities of spiral growth hillocks on the (111) faces of barium nitrate crystals growing in an aqueous solution, *J. Crys. Growth* **102,** 43–53 (1990).
69. E. Yokoyama, K. Okada and K. Tsukamoto, Precise visualization of the non-steady stage concentration field around a growing crystal in solution, *Forma* **13,** 363–374 (1998).
70. C. Li and K. Tsukamoto, The direct interference intensity phase analyzing technique for in situ Michelson interference and its application in studying of the fluctuation of crystal growth rates, *J. Crys. Growth* **223**, 336–342 (2001).
71. P.G. Vekilov, Yu.G. Kuznetzov and A.A. Chernov, Dissolution morphology and kinetics of (101) ADP face: Mild etching of possible surface defects, *J. Crys. Growth* **102,** 706–716 (1990).
72. J.D. Rimstidt and H.L. Barnes, The kinetics of silica–water reactions, *Geochim. Cosmochim. Acta* **44,** 1683–1699 (1980).
73. T.E. Burch, K.L. Nagy and A.C. Lasaga, Free energy dependence of albite dissolution kinetics at 80 degrees C and pH 8.8, *Chem. Geol.* **105,** 137–162 (1993).
74. K.G. Knauss and T.J. Wolery, Dependence of albite dissolution kinetcs on pH and time at 70°C, *Geochim. Cosmochim. Acta* **50**, 2481–2498 (1986).
75. L. Chou and R. Wollast, Steady-state kinetics and dissolution mechanisms of albite, *Amer. J. Sci.* **285**, 963–993 (1985).
76. I.N. MacInnis and S.L. Brantley, The role of dislocations and surface morphology in calcite dissolution, *Geochim. Cosmochim. Acta* **56**, 1113–1126 (1992).
77. I.N. MacInnis and S.L. Brantley, Development of etch pit size distributions on dissolving minerals, *Chem. Geol.* **105**, 31–49 (1993).
78. K. Maiwa, K. Tsukamoto and I. Sunagawa, Activities of spiral growth hillocks on the (111) faces of barium nitrate crystals growing in an aqueous solution, *J. Crys. Growth* **102**, 43–53 (1990).
79. H. Ohmoto, K. Hayashi, K. Onuma, K. Tsukamoto, A. Kitakaze, Y. Nakano and Y. Yamamoto, Solubility and reaction kinetics of solution-solid reactions determined by in situ observations, *Nature* **351**, 634–636 (1991).
80. K. Onuma, K. Tsukamoto and I. Sunagawa, Dissolution kinetics of K-alum crystals as judged from the measurements of surface undersaturations, *J. Crys. Growth* **110,** 724–732 (1991).
81. K. Onuma, K. Tsukamoto and S. Nakadate, Application of real time phase shift interferometer to the measurement of concentration field, *J. Crys. Growth* **129**, 706–718 (1993).
82. K. Onuma, T. Karneyama and K. Tsukamoto, In situ study of surface phenomena by real time phase shift interferometry, *J. Crys. Growth* **137,** 610–622 (1994).
83. K. Tsukamoto, K. Onuma and I.N. MacInnis, Application of real time phase shift interferometry to the slightly dissolved crystals in water, *Proc. 6th Topical Meeting on Crystal Growth Mechanism,* Awara, Japan (1993) pp. 227–231.
84. I.N. MacInnis, K. Onuma and K. Tsukamoto, In situ study of the dissolution kinetics of calcite using phase shift interferometry, *Proc. 6th Topical Meeting on Crystal Growth Mechanism,* Awara, Japan (1993) pp. 233–238.
85. A. Lüttge, E.W. Bolton and A.C. Lasaga. in: *Biogeochemical Cycles and Their Evolution Over Geologic Time. A special triple issue as a tribute to the career of Robert A. Berner,* Eds D. Canfield and B. Boudreau (*American Journal of Science* **299**, Yale University, New Haven) pp. 652–678.
86. M.S. Beig and A. Lüttge, Anorthite–albite dissolution kinetics. A comparison of VSI results of single crystal and conventional mineral powder samples (2004, in prep.).
87. A. Lüttge, Heterogeneity of mineral dissolution processes—a statistical approach using vertical scanning white light interferometry, (abstr.) *EOS* **81** supp., 231 (2000).

88. S.L.S. Stipp, W. Gutmannsbauer and T. Lehmann, The dynamic nature of calcite surfaces in air, *Am. Min.* **81,** 1–8 (1996).
89. S.L.S. Stipp, J. Konnerup-Madsen, K. Franzreb, A. Kulik and H.J. Mathieu, Spontaneous movement of ions through calcite at standard temperature and pressure, *Nature* **396,** 356–359 (1998).
90. G.R. Holdren, Jr. and P.M. Speyer, Reaction rate surface area relationships during the early stages of weathering—I. Initial observations, *Geochim. Cosmochim. Acta* **49**, 675–681 (1985).
91. R.S. Arvidson, M. Collier, K.J. Davis, M.D. Vinson, J.E. Amonette and A. Lüttge, The role of magnesium in calcite dissolution, *Geochim. e Cosmochim. Acta* (2004, subm.)
92. T.A. Fewless and A. Lüttge, Barite dissolution rates measured with different methods, *12th Annual VM Goldschmidt Conference,* Davos, Switzerland (August 18–23, 2002).
93. G.R. Dickens, T.A. Fewless, E. Thomas and T. Bralower, Excess barite accumulation during the Paleocene/Eocene thermal maximum: massive input of dissolved barium from the seafloor gas hydrate reservoirs, *Geological Society of America Special Volume on Paleogene Climates* (2002).
94. A.E. Nielsen, The kinetics of electrolyte precipitation, *J. Coll. Sci.* **10,** 576–586 (1955).
95. C.W. Blount, Barite solubilites nad thermodynamic quantities up to 300°C and 1400 bars, *Am. Min.* **62**, 942–957 (1977).
96. U. Becker, P. Risthaus, D. Bosbach and A. Putnis, Selective attachment of monovalent background electrolyte ions and growth inhibitors to polar steps on sulfates as studied by molecular simulations and AFM observations, *Molec. Simul.* **28**, 607–632 (2002).
97. R.S. Arvidson and A. Lüttge, VSI data from long-term calcite dissolution experiments (unpub.).
98. A. Lüttge and A.C. Lasaga, Feldspar dissolution kinetics—A fundamental approach using Monte Carlo simulations, *Chem. Geol.* (2004, in revision).
99. R.A. Berner and J. Schott, Mechanism of pyroxene and amphibole weathering; II. Observations of soil grains, *Am. J. Sci.* 282, 1214–1231 (1982).
100. W.K. Burton, N. Cabrera and F.C. Frank, The growth of crystals and the equilibrium structure of their surfaces, *Phil. Trans. Royal Soc. London* **A243,** 299–358 (1951).
101. Y. Xiao and A.C. Lasaga, Ab initio quantum mechanical studies of the kinetics and mechanisms of silicate dissolution; $H^+(H_3O^+)$ catalysis, *Geochim. Cosmochim Acta* **58,** 5379–5400 (1994).
102. Y. Xiao and A.C. Lasaga, Ab initio quantum mechanical studies of the kinetics and mechanisms of silicate dissolution: OH^- catalysis, *Geochim. Cosmochim. Acta* **60,** 2283–2295 (1996).
103. Y. Xiao and A. Lüttge, Solvated ab initio and density functional theory (DFT) modeling of mineral-fluid surface reactions: towards a better understanding of alumino-silicate dissolution mechanisms, GSA 2002 Fall meeting, Denver, CO (October 27–30, 2002).
104. J.D. Kubicki, in: *Molecular Modeling Theory: Applications in the Geosciences. Reviews in Mineralogy & Geochemistry* **42**, Eds R.T. Cygan and J.D. Kubicki (Mineralogical Society of America, Washington, DC, 2001) pp. 459–483.
105. A. Lüttge, Vertical scanning interferometry: super-resolution and in situ capabilities for the studies of gas-hydrates, *Energy Exploration & Exploitation* **21**, 329–332 (2003).
106. D.S. Sawyer, A. Lüttge and P.G. Conrad, Towards a high resolution vertical scanning interferometer: applications for fluid–solid–microbe interactions, *GSA 2002 Fall meeting,* Denver, CO (October 27–30, 2002).
107. A. Taylor, J.D. Blum and A.C. Lasaga, The dependence of labradorite dissolution and Sr isotope release rates on solution saturation state, *Geochim. et Cosmochim. Acta* **64,** 2389–2400 (2000).
108. K.L. Nagy and A.C. Lasaga, Dissolution and precipitation kinetics of gibbsite at 80 degrees C and pH 3; the dependence on solution saturation state, *Geochim. et Cosmochim. Acta* **56,** 3093–3111 (1992).
109. K.L. Nagy, A.E. Blum and A.C. Lasaga, Dissolution and precipitation kinetics of kaolinite at 80 degrees C and pH 3; the dependence on solution saturation state, *Am. J. Sci.* **291**, 649–686 (1991).

This page intentionally left blank

8

Machining with chemistry: controlling nanoscale surface structure with anisotropic etching

Melissa A. Hines*
Department of Chemistry, Cornell University, Ithaca, NY 14853-1301, USA

8.1. INTRODUCTION

For more than three decades, the size of commercially available micromachined devices has decreased exponentially. In the microprocessor industry, these advances are codified by "Moore's Law," which states that the density of transistors on a microprocessor doubles every 18 months [1,2]. As widely acknowledged, this rapid evolution has had enormous scientific, technological and cultural benefits. What is perhaps less well recognized is that the capital costs associated with these advances have also increased exponentially. Some have termed this scaling "Moore's Second Law" [3]. With the cost of new semiconductor fabrication lines now in the billions of dollars [4], there is increasing impetus to find new methods for nanoscale machining.

In contrast to current nanofabrication techniques, chemistry is an inherently atomic-scale process. Can chemical etching be used to control nanoscale structure? Chemical machining is by no means a new idea. Almost 200 years ago, mineralogists used anisotropic etchants—etchants that selectively attack specific crystallographic planes—to reveal the inherent symmetry of crystals [5], In the 1950s, scientists showed that macroscopic single crystal spheres could be selectively etched to form cubes [6], dodecahedra [7] and other geometric shapes [8]. More recently, aqueous bases, such as KOH and TMAH (tetramethylammonium hydroxide), have been used to machine self-aligning, ultraflat Si(111) surfaces for applications ranging from precision nozzles for ink-jet printers [9], to textured surfaces for solar cells [10], to nanoscale V-grooves for novel microelectronic devices [11]. Advances in scanning probe microscopy have shown that this chemical precision extends to the *atomic scale.*

*E-mail address: Melissa.Hines@cornell.edu

Nevertheless, there are many challenges in turning this concept into a workable reality. Very little is known about the connection between etchant composition, chemical reactivity and nanoscale morphology. Theories of chemical reactivity are based on local, atomic-scale considerations, such as steric hindrance, relative electronegativity and structural rigidity, which have an effective range of no more than a few nanometers. Nanoscale structures, on the other hand, develop over the course of many such reactions—tens or hundreds of monolayers may need to be removed to reach the final, steady-state morphology. Is the final structure with its potentially long length scales, a simple consequence of the cumulative effect of many short-scale events, or do more subtle perturbations, such as strain, contamination or even random fluctuations, come to dominate the final morphology?

In the following, I will present atomic-scale studies of three types of nanoscale surface structures that can be produced by chemical etching and explain the chemical processes that govern the production of ultraflat surfaces, nanoscale textured surfaces, and nanoperiodic surfaces. All of these studies involve aqueous silicon etchants, which are of interest because of their widespread use in the microelectronics and MEMS (microelectromechanical systems) industries; however, the same principles will apply to the etching of many materials. One of the biggest problems facing the etching community is the rational development of new etchants with controlled properties. In the final section, I will present a new technique for the development of anisotropic etchants, which provides further insight into the atomic-scale processes that govern etching.

8.2. PRODUCTION OF EXTREMELY FLAT SURFACES

For reasons that are poorly understood, the mechanical, chemical, and electronic properties of nanoscale structures are often strongly influenced by their surface and interface morphologies. For example, atomic-scale roughness of the Si/SiO_2 interface in a MOSFET gate can reduce the mobility of electrons in the channel by a factor of four [12]. More recently, energy losses in 50–250 nm silicon resonators have been shown to be proportional to their surface-to-volume ratio [13], suggesting that surface effects, such as surface chemistry or morphology, play an important role in mechanical energy dissipation at the nanoscale. Atomically flat surfaces are also needed for many scientific applications, such as neutral atom focussing [14].

Can chemical etching produce the atomically flat surfaces needed for these applications? Intuitively, a chemical etchant that selectively etches defect sites, such as steps and kinks, much faster than terrace sites would be expected to produce a flat surface. But what are the limitations to this perfection? Would an infinitely anisotropic etchant, an etchant that only etches defects, produce an atomically perfect surface? Relatively flat surfaces can also be produced by thermal annealing, a purely thermodynamic process. Can chemical etching ever "beat thermodynamics" and produce surfaces flatter than the equilibrium limit?

The production of a macroscopically perfect surface, one with absolutely no steps or defects, that is tens or hundreds of centimeters in diameter is likely impossible, as macroscopic cutting tools are limited to an accuracy of $\approx 0.1°$. Because of this, all macroscopic surfaces are actually *vicinal surfaces* with an average atomic step spacing of ≈ 100 nm or less. Given this, the flattest possible macroscopic surface must have regularly spaced vicinal steps, as illustrated by Figure 8.1a. A surface with unequally spaced steps, such as

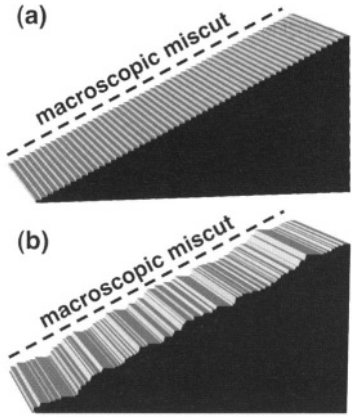


FIGURE 8.1. The long-range roughness of a vicinal surface is controlled by the step distribution. (a) A perfect surface, and (b) a rough surface produced by one-dimensional step flow etching.

the one pictured in Figure 8.1b, would have pronounced long-range roughness even if the individual terraces were perfectly flat. A "perfect" etchant must therefore meet two criteria: it must produce perfect, unpitted terraces as well as evenly spaced atomic steps. The second criterion seems especially challenging, as it would seem to require very long-range chemical interactions between etching steps.

In the following, I will show that the repeated action of simple, site-specific chemical reactions can lead to very long-range structural effects. In some cases, chemical etching can be used to create surfaces that are very flat (even on long length scales), whereas other etchants are inherently unstable, and lead to bumpy surfaces that are covered with etch hillocks. Interestingly, infinitely anisotropic etchants, which only etch defect sites, *do not* produce the flattest possible surfaces. In fact, a low density of etch pits induces a phenomenon known as *dynamic step–step repulsion* which suppresses the development of long-range roughness.

8.2.1. Step-flow etching and long-range roughness

The simplest and most commonly invoked type of etching is *step-flow etching.* In this mechanism, step edges are assumed to be infinitely more reactive than other sites on the surface; however, etching steps are not allowed to undercut one another. Since the etchant selectively attacks step edges, the etching steps appear to retreat or flow across the surface. Etch pits cannot be nucleated, so step flow etching must lead to perfectly flat surfaces on short length scales. But what about longer length scales, where the roughness is controlled by the spacing of atomic steps? In this case, the dimensionality of the surface becomes important.

The surfaces in Figure 8.1 illustrate the one-dimensional limit, where there are only two types of surface sites, steps and terraces. In this dimension, step-flow etching continuously roughens the surface, but it does not lead to step bunching (e.g. the predominant production of double- and triple-steps). In fact, the rough morphology in Figure 8.1b was produced by prolonged step-flow etching of the perfect surface in Figure 8.1a. In one dimension, prolonged step-flow etching leads to a *random distribution* of steps.

One common measure of surface roughness is W, the rms roughness, which is defined to be

$$W = \sqrt{\langle (z(x) - \bar{z}(x))^2 \rangle}, \tag{1}$$

where $z(x)$ is the height of the surface at position x, $\bar{z}$ is the best fit line to $z(x)$ and the brackets denote an average over the entire surface. In one dimension, step-flow etching leads to a divergent rms surface roughness that scales as $D^{1/3}$, where D is the number of monolayers etched. Although commonly used, the rms surface roughness is a relatively uninformative measure. First, W contains no information about the distribution of the roughness. For example, it cannot distinguish between rippled (one dimensionally rough) and bumpy (two dimensionally rough) surfaces. Second, for a given surface morphology, W is not uniquely defined. Instead, it is highly dependent on both the resolution of the surface probe (i.e. the density of z measurements) as well as the lateral extent of the averaging. Because of this, the rms roughness measured from a large image is almost always larger than the roughness measured from a smaller image of the same surface.

If atomically resolved images are available, terrace width distributions (TWDs) are a more informative measure of surface roughness. Obviously, the TWD of the flattest possible macroscopic surface would be a delta function centered at L, the average distance between steps. Since step-flow etching of a 1D surface leads to a random distribution of steps, the normalized terrace width distribution $P_{1D}(w)$ is simply an exponential,

$$P_{1D}(w) = e^{-w/L}, \tag{2}$$

where w is the terrace width.

The situation is more complex in two dimensions, as there are at least two different types of step flow etching. In this section, I will consider the simplest and most commonly invoked model of step flow etching, which allows for the etching of only two different surface sites: steps and kinks. In the next section, I will discuss the surprising role of etch pits, which are the result of a finite terrace etch rate. Finally, I will discuss a more complicated type of step flow, which is kinetically unstable. In all of these cases, undercutting is explicitly forbidden—a reasonable approximation for most experimental systems.

In two dimensions, two kinks are produced when a single atom is etched from an otherwise perfect step. Once formed, subsequent etching of a kink site removes that kink site and generates a new kink at a neighboring site. When repeated many times, sequential kink etching effectively leads to the "unzipping" of a single row of atoms from the step. In the (unphysical) limit where the kink etch rate is infinitely faster than the step etch rate, this two-dimensional model reduces to the previously discussed one-dimensional model, and the TWD is exponential.

The effects of site-specific etching on two-dimensional surface morphology can be studied quantitatively using an atomistic kinetic Monte Carlo (KMC) simulation of surface etching. All of the KMC simulations described here were performed with a simulator developed by Flidr *et al.* [15] which models the full structure of the Si(111) surface, including the interlayer stacking pattern, and all possible reactive sites within the solid-on-solid approximation. Toroidal bounds are used to simulate an infinite surface. The simulations typically start with a perfect miscut surface. Before the simulation begins, a site-specific etch rate

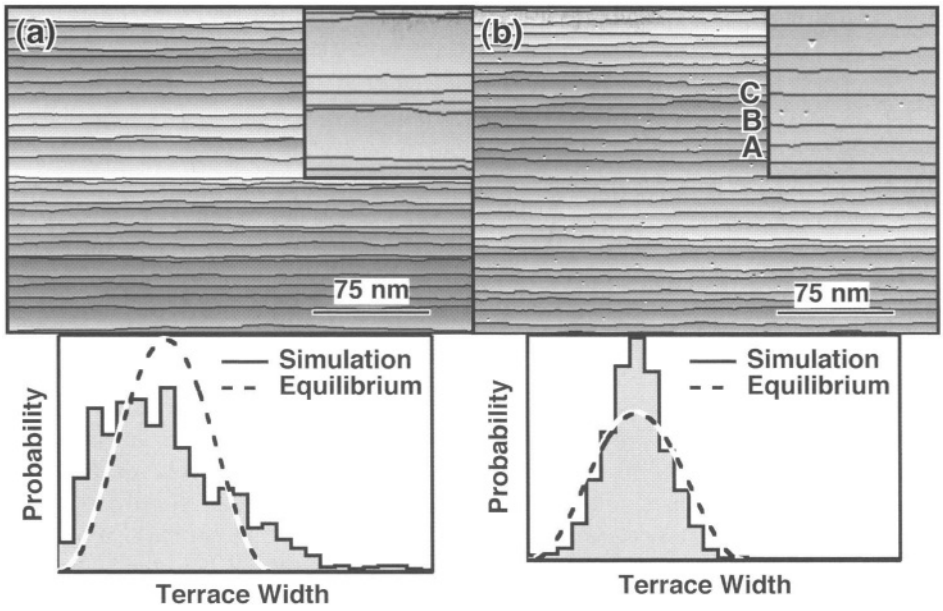


FIGURE 8.2. (a) Pure step-flow etching of two-dimensional surfaces leads to modest step–step repulsion and a terrace width distribution similar to that of an equilibrium surface. (The discrepancies between the simulated distribution and the equilibrium distribution are due to the small size of the simulation, as discussed in ref. [17].) (b) A low rate of terrace etching, which produces a low density of etch pits, increases the long-range smoothness of the surface and narrows the terrace width distribution.

is assigned to each type of surface site (e.g., kink, step, terrace). During the simulation, single atoms are randomly removed from the lattice subject to their assigned site-specific etch rates. Diffusion of surface atoms and redeposition of etch products was explicitly forbidden. Both of these approximations are excellent for the aqueous etching of silicon.

When only step and kink etching are allowed, step flow etching results, and the morphology of the etching step is solely determined by the ratio of the step etch rate (i.e. the kink nucleation rate) to the kink etch rate (i.e. the rate of unzipping) [16]. When the kink etch rate is much faster than the step etch rate, nearly straight steps are produced, as illustrated by the simulated morphology in Figure 8.2a. In contrast, when the two etch rates are comparable, a relatively rough step morphology is produced (not shown).

Two-dimensional step flow etching leads to much smoother surfaces than one-dimensional etching [17], as illustrated by the terrace width distributions in Figure 8.2a. Regardless of the step morphology, the two-dimensional model leads to a near-Gaussian TWD. This distribution implies that the etching steps are repelling one another, even though there are no repulsive forces in this purely kinetic model.

This weak step–step repulsion is a kinetic effect that is due to the high energetic barrier to undercutting. (In the KMC model, undercutting is explicitly forbidden, which implies an infinite barrier.) When a statistical fluctuation causes two steps to collide, only a short region of the two steps are affected. During the collision, the colliding length of the trailing step must stop etching to avoid undercutting the leading step. Regions of the trailing step that are adjacent to the collision continue to etch. As a result, most collisions will slightly

reduce the curvature (i.e. kink density) and the total length of the trailing step in the region close to the collision. These morphological changes have no effect on the leading step, but induce a transient decrease in the trailing step etch rate. In effect, step collisions introduce a small negative feedback into the step–step distribution. In the absence of collisions, all steps propagate (i.e. etch) at the same average rate. When two steps collide, the trailing step transiently slows, allowing the leading step to pull away.

As a result of this negative feedback, step-kink etching leads to surfaces with relatively smooth morphologies on both short and long length scales. Interestingly, the long-range morphology produced by step-kink etching is nearly identical to that produced by simple thermal annealing. At high temperatures, steps fluctuate due to the continual detachment and reattachment of atoms from the step edge, and the step morphology is determined by the relative energies of step and kink sites [18]. Since the energetic cost of creating an overhang is high, colliding steps do not cross. These avoided crossings reduce the total number of allowed step configurations (in comparison to an isolated step) and increase the step free energy. Because of this, there is an apparent *entropic repulsion* between steps—thermodynamics favors a well ordered series of steps. In the absence of long-range strain fields, the equilibrium morphology has a near-Gaussian TWD [19] which is essentially identical to that produced by step-kink etching.

Of course, the similarity between the TWDs of etched and annealed surfaces does not imply that etching produces an equilibrium surface. On the contrary, the etching simulation is far from equilibrium, because redeposition of surface atoms is explicitly forbidden.

8.2.2. Beating thermodynamics: the surprising role of etch pits

The step-kink model of step flow etching presented in the previous section is overly simplistic, as most etchants also attack terrace sites with a finite probability. Even though the terrace site etch rate is often orders of magnitude slower than the rate of kink and step etching, the density of terrace sites is typically orders of magnitude higher than the density of defect sites. As a result, terrace site etching cannot be ignored. In fact, a low density of etch pits can have a profound affect on the long-range roughness of the etched surface, as this section will show.

When a terrace site is etched on an otherwise flat terrace, an etch pit is nucleated. To a first approximation, this etch pit can be modeled as three or four unit-length steps (depending on the symmetry of the lattice). Once nucleated, subsequent step etching will cause the etch pit to grow in diameter. Etch pits cannot grow indefinitely. On a vicinal surface, their lifetime is usually limited by step-flow etching. When a step collides with an etch pit on the terrace above the step, the pit first attaches to the etching step, creating an indentation in the step, and is then “erased” by subsequent step etching. Because of this, the average etch pit size and density is determined by kinetic competition between etch pit nucleation, which is governed by the terrace etch rate, and etch pit annihilation, which is governed by the rate of step flow etching and thus by the relative rates of step and kink etching.

The effects of this kinetic competition are illustrated by the three steady state etch morphologies in Figure 8.3. All three of these simulated morphologies were generated using identical site-specific rates of etching. The only difference between these surfaces is the surface miscut (i.e. the average step density). When the step density is high, the

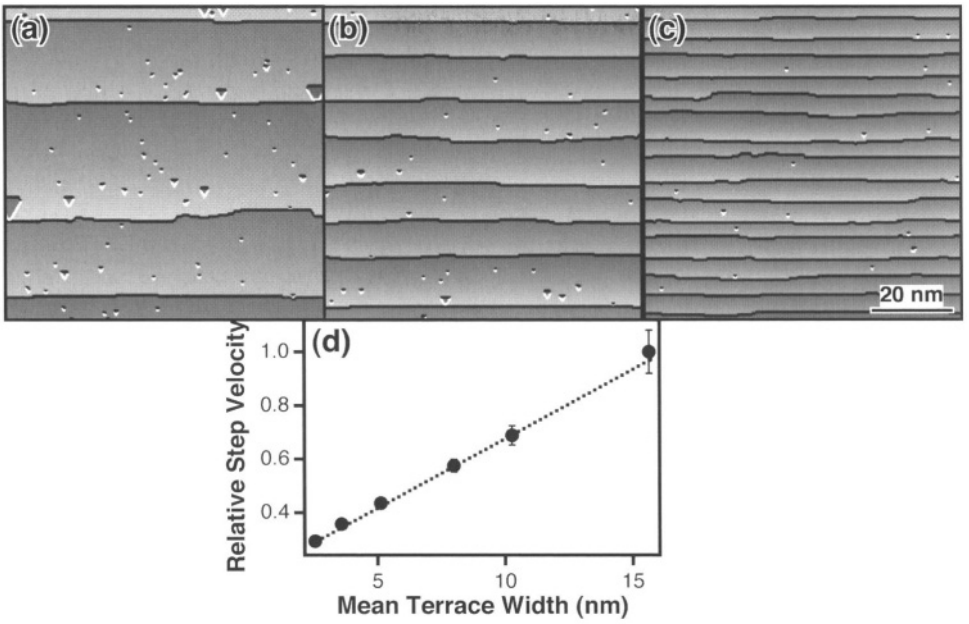


FIGURE 8.3. KMC simulations of terrace-step-kink etching using identical site-specific kinetics but varying step density. (a–c) The density of etch pits is controlled by the surface miscut, and (d) the step-density dependent etch pit density leads to a step velocity that is proportional to the mean terrace width.

pit annihilation rate is also high. As a result, Figure 8.3c has a relatively low density of small etch pits. In contrast, a surface with a relatively low step density, such as the one in Figure 8.3a, develops a higher density of pits, and these pits grow to a larger size before they are annihilated.

Etch pits have an indirect, but sometimes profound, effect on step flow etching. For simplicity, consider the case where kink etching is much faster than step etching. As illustrated in Figure 8.3, rapid kink etching leads to the production of nearly straight steps, because step unzipping is much faster than kink nucleation. In the absence of etch pits, every step will flow with the same average velocity, independent of the step density. This velocity will be determined by the rate-limiting reaction—kink nucleation. The situation changes when etch pits are present. When a step collides with an etch pit, new kinks are also created. Under these kinetics, the rate of step flow etching, and thus the step velocity, is determined by both the kink nucleation rate and the etch pit density. As a result, *the step velocity is step-density dependent.* Wide terraces lead to fast etching steps, while narrow terraces lead to slow etching steps, as shown by average velocities in Figure 8.3d.

As a result of this step-density dependent step velocity, etch pits can *suppress* long-range roughness, as shown by the two morphologies in Figure 8.2. Both of these surfaces have the same miscut, and both were etched with the same site-specific rates of kink and step etching. These surfaces differ only in the assumed rate of terrace etching: the surface in Figure 8.2a is pit-free, whereas the surface in Figure 8.2b has a low density of tiny etch pits. Even by eye, these two surfaces have dramatically different terrace width distributions. Many step–step collisions are visible on the pit free surface, and the terrace width distri-

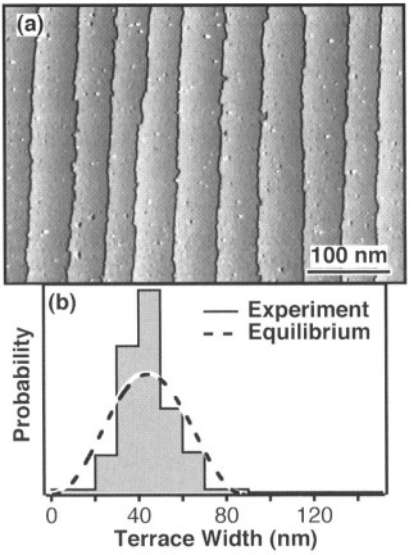


FIGURE 8.4. Steady state morphology of Si(111) miscut by 0.3° towards the $\langle 11\bar{2} \rangle$ direction etched in stirred $NH_4F(aq)$. (a) STM image showing equally distributed single atomic layer steps. (b) Distribution of terrace widths based on 168 data points. For comparison, the (calculated) equilibrium distribution is also shown.

bution is comparable to that of an equilibrium surface, as expected from the discussion in Section 8.2.1. In contrast, the pitted surface has a more uniform step distribution, and the TWD is significantly narrower than the equilibrium distribution.

This pit-induced narrowing of the TWD is due to a second form of dynamic step–step repulsion that is mediated by the step-density dependent step velocity. For concreteness, consider the three marked steps in Figure 8.2b. Because of statistical fluctuations in the step etch rate, step *A* is too close to the step preceding it (step *B*), while step *B* is somewhat too far behind step *C*. Since pit density is determined in part by the terrace width, the narrow *A*–*B* terrace is pit free, while the wide *B*–*C* terrace has a number of small pits. Because of this anisotropic pit distribution, the laggard *B* step will soon collide with the pits, increase its etch rate, and begin to catch up to *C*. In contrast, step *A* will not accelerate, as there are no pits in its path. As a result, the pit-induced, step-density dependent step velocity leads to negative feedback in the TWD.

Although both forms of dynamic step–step repulsion were originally observed in atomistic simulations [17], kinetic terrace width narrowing was later experimentally measured in the $NH_4F(aq)$ etching of Si(111) [23], as shown by the data in Figure 8.4. To a first approximation, $NH_4F(aq)$ is a prototypical terrace-step-kink etchant towards vicinal Si(111) surfaces (when the miscut is towards the $\langle 11\bar{2} \rangle$ direction.) By comparing Figure 8.4a to kinetic Monte Carlo simulations, the kink/step etch ratio is estimated to be ≈ 50, whereas the kink/terrace etch rate ratio is $\approx 10^6$ [16]. This pronounced anisotropy leads to surfaces of exceptionally low short- and long-range roughness. On an atomic scale, these NH_4F-etched surfaces are characterized by large defect-free regions ($\approx 10^4$ Å^2) separated by small pits. The remarkable regularity of the etched surface on longer length scales is confirmed by

the terrace width distribution in Figure 8.4b. For comparison, the calculated TWD for a (unstrained) surface in thermal equilibrium is also shown.

These data clearly show that *chemical etching can produce smoother surfaces than simple annealing,* at least on long length scales. In contrast to annealing, the chemical etch is carried out at room temperature, where it will have no effect on the overall thermal budget. The repeated action of simple site-specific surface chemistry, which is sensitive to surface structure over no more than ≈ 1 nm, leads to surprisingly long-range correlations in surface morphology. This observation also suggests that there is no one "perfect etchant," as the etchant needs to be tuned to the surface miscut. High miscut surfaces will need a higher pit nucleation rate than low miscut surfaces.

8.3. THE PRODUCTION OF TEXTURED SURFACES: ETCH HILLOCKS

In the previous section, I showed that simple site-specific chemistry can produce surfaces of exceptional long-range smoothness, but other etchants are known to spontaneously produce etch hillocks (i.e. bumps) with length scales ranging from nanometers to microns [24]. What controls the production of these features? And their length scale? Can their size be controlled chemically? Their shape? Their formation? From a scientific standpoint, these are probably the most perplexing and most challenging questions facing the etching (and growth!) community. Unfortunately, there are currently no good general answers to any of these questions, and atomic-scale studies are just beginning to explain these phenomena.

The formation of macroscopic etch features is not just an intellectual curiosity—it is a major technological concern as well. For example, aqueous silicon etchants are widely used in the production of MEMS, such as the accelerometers used to trigger automobile airbags. In these systems, macroscopic surface roughness is often thought to be detrimental. Interestingly, there are some cases where the controlled production of rough surfaces may be beneficial. For example, photovoltaic manufacturers sometimes boost the light-trapping properties of solar cells by texturing the surface with chemically generated microscopic pyramids [10]. Since asperities decrease the effective contact area, rough surfaces may also be less susceptible to stiction, a major problem in MEMS [25].

In the following, I will first provide an overview of the leading theory of hillock formation and point out some of the limitations of this theory. I will then give two examples where atomic-scale studies of etching have yielded further insight into the hillock formation process. In the first example, atomic-scale step bunching, the presumed precursor to hillock formation, is shown to be very sensitive to solution homogeneity, suggesting that bunching can be driven by either the local build-up of etch products or the depletion of reactants. In the second, and much more unexpected, example, I will show that some site-specific chemical reactions are *inherently unstable* and lead to the production of two- and three-dimensional etch hillocks.

8.3.1. Kinematic wave theory

The first, and by far the most commonly invoked, theory of hillock formation is a one dimensional model of step flow etching that was proposed essentially simultaneously

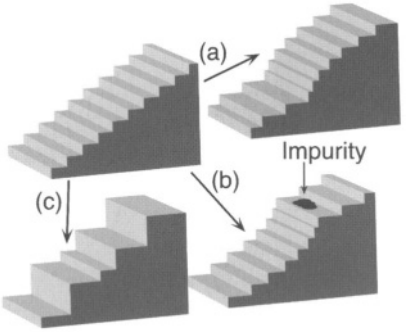


FIGURE 8.5. The three different types of step bunching predicted by one-dimensional models: (a) concave, which may be formed during local etchant depletion, (b) convex, which may be induced by contamination, and (c) facetted.

by Frank [26], who described growth hillocks, and by Cabrera and Vermilyea [27], who described etch hillocks. (Technically, a one-dimensional model describes macrosteps, not hillocks.) Both of these theories were based on *kinematic waves* [28], a mathematical construct developed to model flow in one-dimensional fluids (e.g., rivers).

In kinematic wave theory, etching is only allowed at step sites, and the etch rate (or velocity) v_{step} of an individual step is assumed to be only a function of the *local* step density, ρ_{step}:

$$v_{\text{step}} = f(\rho_{\text{step}}). \tag{3}$$

If the steps etch independently (i.e. no step–step interactions), all steps will move with the same velocity, independent of local step density and miscut. Under these kinetics, kinematic wave theory predicts that hillocks will be unstable.

The situation is very different if the step velocity is step-density dependent. For example, consider a case of very rapid step-flow etching in which the etchant is somewhat depleted in the vicinity of each etching step. This depletion would be enhanced in regions of locally high step density. Not surprisingly, this situation leads to step bunching in much the same way that my grandmother, a notoriously slow and timid driver, causes traffic jams. When a random fluctuation brings two steps close to one another, enhanced depletion causes the pair to travel more slowly. This slowing allows the step behind the pair to catch up, which leads to slower etching of the step trio, and so on. Mathematically, this scenario occurs when

$$\partial v_{\text{step}}/\partial \rho_{\text{step}} < 0. \tag{4}$$

Kinematic wave theory shows that a function of this general form will lead to the production of *concave step bunches* as shown in Figure 8.5a.

A second type of kinematic wave is also possible. For example, consider an etching surface that is exposed to a continuous flux of contaminants. During etching, new surface sites are continuously exposed by step flow. If the flux of contaminants is uniform, the probability of contaminating a specific site on the surface increases with its exposure time (i.e. the

time since etching of the previous step uncovered this site). As a result, the contaminants will be anisotropically distributed across the surface. Sites that are at the upper edge of a wide terrace will have a high probability of being contaminated, while those at the upper edge of a narrow terrace will have a small contamination probability. If contaminated sites etch more slowly that clean sites, a reasonable scenario, narrow terraces, which have a low probability of contamination, will etch more quickly than wide terraces. Mathematically, this relationship leads to

$$\partial v_{\text{step}}/\partial \rho_{\text{step}} > 0. \tag{5}$$

This situation is similarly unstable, and kinematic wave theory predicts the formation of *convex* hillocks as sketched in Figure 8.5b.

In summary, kinematic wave theory predicts the formation of kinetic instabilities—macrosteps or etch hillocks—during step flow etching in any system where the etch rate of individual steps is step-density dependent. At least in one-dimensional systems, *no other criteria must be met.* Conversely, systems where the steps etch with a constant rate, independent of step density, should be stable to macrostep formation, at least in one dimension.

In the almost 45 years since Frank, Cabrera and Vermilyea produced the first model of step bunching/macrostep formation, many other models have been advanced to account for these phenomena. Almost all theories of step bunching or macrostep formation require some type of step density dependent step velocity. This dependence can be introduced by many different physical phenomena, such as elastic strain [29–31], anisotropic diffusion [32] (e.g., Ehrlich–Schwoebel barrier [33–35], electromigration [36–40]), contamination [41–44] and the buildup of reaction products [45]. Although these mechanisms differ in their details, they all have the same mathematical underpinning—*a step-density dependent etch/growth rate.*

Only one type of step bunching does not require step-density dependent etch rates. In a number of systems, macrostep formation has been explained experimentally [46,47] and theoretically [48,49] by the existence of a thermodynamically stable facet near the original surface orientation as illustrated in Figure 8.5c. For example, vicinal Si(111) surfaces form large (7 × 7)-reconstructed, (111)-facets separated by step bunches at temperatures near the (7 × 7)–(1 × 1) phase transition [46]. This type of facet-driven macrostep formation leads to very straight macrosteps that bound well-defined facets.

Kinematic wave theory, as originally formulated, has two major limitations. First, the step density dependence is symmetric; kinematic wave theory does not distinguish between the upstairs and downstairs directions. Because of this, kinematic wave theory cannot explain the pit-induced long-range *smoothing* of surfaces described in the Section 8.2.2, even though this phenomenon is also driven by a step-density dependent step velocity. Second, kinematic wave theory does not include effects due to the two-dimensional, extended nature of surface steps. As shown in Section 8.2.1, this increased dimensionality leads to weak step–step repulsion that is similar to equilibrium entropic repulsion. In spite of these limitations, the simplicity of the model is compelling.

8.3.2. Atomic-scale step bunching and solution inhomogeneity

One of the major predictions of kinematic wave theory is that etch hillocks may be formed by localized inhomogeneities at the etchant/surface interface. These inhomo-

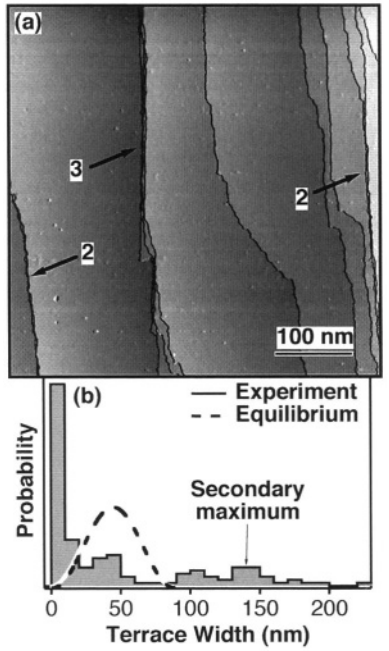


FIGURE 8.6. Steady state morphology of Si(111) miscut by 0.3° towards the $\langle 11\bar{2} \rangle$ direction etched in *unstirred* NH_4F(aq). (a) STM image showing step bunches. The number of steps in each bunch is indicated. (b) Distribution of terrace widths based on 146 data points. The equilibrium distribution is shown for comparison.

geneities can be introduced by the depletion of the etchant or by the build-up of reaction products. To the extent that a two-dimensional surface can be modeled as a one-dimensional array of steps, kinematic wave theory predicts the formation of macrosteps with a characteristic concave or convex shape.

Solution homogeneity has a profound effect on the nanoscale morphology of vicinal Si(111) surfaces etched in aqueous NH_4F solutions. When the solution is homogenized by stirring, this etchant produces surfaces with very evenly spaced steps and a narrow terrace width distribution, as discussed in Section 8.2.2 and shown in Figure 8.4 [23]. Surprisingly, surfaces etched in an *unstirred* etchant of identical composition have a dramatically different steady state morphology as shown by the STM micrograph in Figure 8.6a. This surface is dominated by *step bunches.* The 0.35° miscut of the surface dictates an average step–step spacing of 50 nm; however, both double and triple steps are seen in the STM image. These bunches cannot be explained by facet formation, as the bunches frequently curve. Step bunching also has a profound effect on the TWD, as shown by Figure 8.6b. Instead of the narrow, near-Gaussian distribution characteristic of stirred solutions, unstirred solutions produce a bimodal distribution that cannot be explained by site-specific chemistry in a homogeneous etchant.

The marked dependence of etch morphology on stirring strongly suggests that solution homogeneity plays an important role in determining the final morphology. The observed terrace width distribution can be qualitatively explained by a simple one-dimensional model of step-flow etching that includes contributions from both dynamic step–step re-

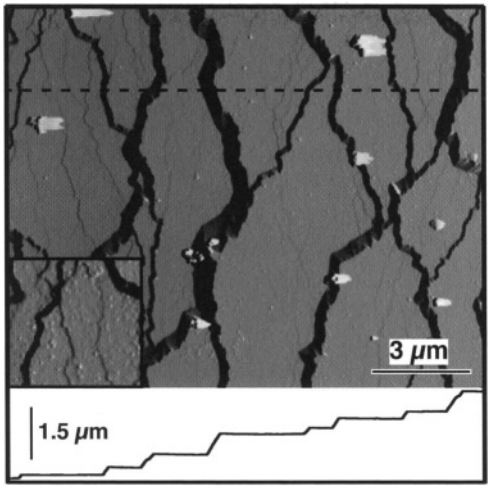


FIGURE 8.7. AFM image of macrosteps formed on 7°-miscut vicinal Si(111) surfaces etched in 70°C, 50% (w/v) KOH for 45 min. The dashed line indicates the position of the bottom cross-section. The inset is a magnified region showing the existence of atomic height steps between the macrosteps.

pulsion and step–step attractions due to solution inhomogeneity [23]. The importance of solution-phase inhomogeneities on a 50 nm length scale is surprising, given the rapidity of diffusion in solution. At some length scale, diffusion should be fast enough to remove these inhomogeneities. Interestingly, surfaces with much more closely spaced steps do not bunch, even in unstirred solutions. In fact, surfaces with an average step–step spacing of 5 nm displayed the near-Gaussian terrace width distribution expected of a homogenous solution [23].

In summary, the formation of step bunches during NH_4F etching of Si(111) surfaces is qualitatively consistent with the predictions of kinematic wave theory. This system is not ideal for further study, though, as NH_4F is a very slow etchant. Uncontrolled contamination would be unavoidable during the very long etch times needed for the production of large macrosteps.

In recent experiments [50], macrostep formation was observed in one of the most common silicon etchants, aqueous KOH; however, these macrosteps are in apparent violation of kinematic wave theory! Vicinal Si(111) surfaces etched in KOH develop pronounced macrosteps (small miscuts) and hillocks (large miscuts); however, kinetic measurements are consistent with a constant, step density independent step etch rate. Why then do macrosteps and hillocks form?

Figure 8.7 shows images of *huge macrosteps* that spontaneously formed on 7°-miscut vicinal Si(111) surfaces during a 30 min, 70°C KOH etch. On these surfaces, a uniform array of atomic steps would have a step–step spacing of 2.6 nm. In contrast, the macrosteps have a characteristic spacing of *a few microns* and heights ranging from ≈20–500 nm (65–1600 atomic steps!). On an atomic scale, the macrosteps must be regions of high step density—step bunches. Perhaps more interestingly, the AFM images suggest that the regions between macrosteps are almost, but not quite, devoid of steps. Can controlled KOH

etching lead to the production of large step-free regions on miscut wafers? If so, this chemistry may lead to the first production of large-scale, atomically perfect surfaces.

In kinematic wave theory, hillock production is predicated on the existence of step-flow etching with a step-density-dependent step velocity. As described in more detail in Section 8.5, measurements of the orientation-dependent etch rate of vicinal Si(111) surfaces strongly suggest that KOH is a step-flow etchant and that v_{step} is profoundly *independent* of step density over a wide range of densities [51]. In fact, surfaces with miscuts ranging from $\approx 2°$ to 20°, corresponding to average step densities ranging from $(9\ \text{nm})^{-1}$ to $(0.9\ \text{nm})^{-1}$, all etch with the same apparent step velocity. According to kinematic wave theory, KOH etched surfaces *should not* develop macrosteps, but they clearly do.

In summary, kinematic wave theory has been very successful in providing qualitative insights into the formation of macrosteps and etch hillocks; however, atomic-scale investigations of etching are beginning to point out the limitations of this theory.

8.3.3. Kinetically unstable chemistry

The previous sections have considered the implications of the simplest and most generalizable atomistic model of etching: terrace-step-kink etching. Although this model completely describes the etching of simple cubic lattices, real crystals have more complicated geometries. Because of this, many surfaces have more exotic reaction sites that may complicate the kinetics and morphology of an etching surface. Since most theoretical studies of etching predate atomistic probes of surface structure, the effects of these species have rarely been considered. In the following, I will show that the production of one such species—a so-called point site—can profoundly change the etched surface morphology and lead to the production of two- or three-dimensional etch hillocks. I will also show that an atomistic examination of two-dimensional hillock formation leads to new insights into the criteria for hillock formation. Since unwanted hillocks can be a major technological problem, a deeper understanding of hillock formation and stability are needed.

8.3.3.1. Two-dimensional etch hillocks When Si(111) surfaces miscut toward the $\langle \bar{1}\bar{1}2 \rangle$ direction are etched in NH_4F solutions, a characteristic *pseudo-facetted* morphology is observed, as illustrated by Figure 8.8a. This morphology, with its characteristic shark's tooth shaped steps, stands in stark contrast to the straight step morphology of NH_4F-etched surfaces miscut towards the $\langle 11\bar{2} \rangle$ direction which is shown in Figure 8.4. After etching, the $\langle \bar{1}\bar{1}2 \rangle$ steps form relatively straight *step segments* that point in three different directions—the original direction and ±60° to the original step. Since the Si(111) surface is three-fold symmetric, the ±60° step segments, which account for approximately 2/3 of the total step length, correspond to regions of $\langle 11\bar{2} \rangle$ oriented step. All of the etched step segments are relatively straight, which indicates that NH_4F attacks kink sites much more rapidly than either the $\langle 11\bar{2} \rangle$ or $\langle \bar{1}\bar{1}2 \rangle$ steps. In the following, we will show that these characteristic shark's-tooth-shaped features are not facets; they are kinetic structures, two-dimensional etch hillocks, that propagate across the etching surface [52,53].

In some sense, etch hillocks are a product of the three-fold symmetry of the (111) surface. Like many silicon etchants, NH_4F produces hydrogen-terminated surfaces [541. As shown by detailed spectroscopic investigations [55,56] and *ab initio* cluster calculations [57], steps miscut toward the $\langle \bar{1}\bar{1}2 \rangle$ direction etch to form the highly strained vertical dihydride moiety sketched in Figure 8.9a. Hillocks are nucleated when two opposing kinks

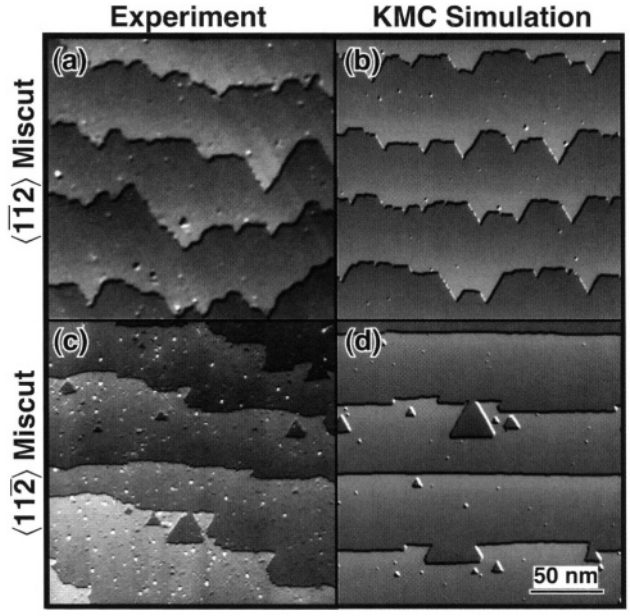


FIGURE 8.8. A comparison of the (a, c) experimental and (b, d) simulated morphologies of Si(111) surfaces miscut by 0.3° towards the (a, b) $\langle \bar{1}\bar{1}2 \rangle$ and (c, d) $\langle 11\bar{2} \rangle$ directions and etched in stirred $NH_4F(aq)$. The KMC simulations are the best fit to experiment. In these, the relative site specific rates of etching are kink:point:vertical dihydride step:monohydride step:terrace $= 1 : 0.10 : 10^{-2} : 5 \times 10^{-4} : 10^{-7}$.

on a $\langle \bar{1}\bar{1}2 \rangle$ step collide as illustrated by Figure 8.9b. Since the three-fold symmetry of the surface cants the kink sites by 30° from the step normal, colliding kinks do not annihilate one another. Instead, they produce a new surface site known as a *point site.* Although similar to the $\langle \bar{1}\bar{1}2 \rangle$ step site, the point is a horizontal, unstrained dihydride. Because of this difference in structure, the vertical and horizontal dihydride structures have different reactivities and thus different etch rates.

Once a hillock is formed, it can either grow or decay, as illustrated by Figure 8.10. Hillock growth occurs when a $\langle \bar{1}\bar{1}2 \rangle$ step site (i.e. a vertical dihydride site) is etched. Sequential etching of the two kinks formed by this process leads to rapid "unzipping" of an entire row of atoms. When an etching kink collides with an existing hillock, the process terminates, and the hillock grows by one atomic unit. The sides of the hillock are $\langle 11\bar{2} \rangle$ steps, which have the very stable monohydride termination as shown in Figure 8.9c. In contrast, hillock decay can be initiated by either of two events. If one of the $\langle 11\bar{2} \rangle$ step sites on the hillock face is etched, two kinks will be nucleated. Sequential kink etching will lead to the removal of one row of atoms, and the decay of the hillock by one atomic unit. A similar process occurs if the point site is etched. Although not obvious from the sketch in Figure 8.9c, the etching of a point site also leads to the formation of two kink sites and the removal of an entire row of atoms from one face of the hillock [53].

In steady state, the rate of hillock growth, k_{grow}, must exactly balance the rate of hillock decay, k_{decay}. This criterion determines, in part, the average hillock size. In the limit where

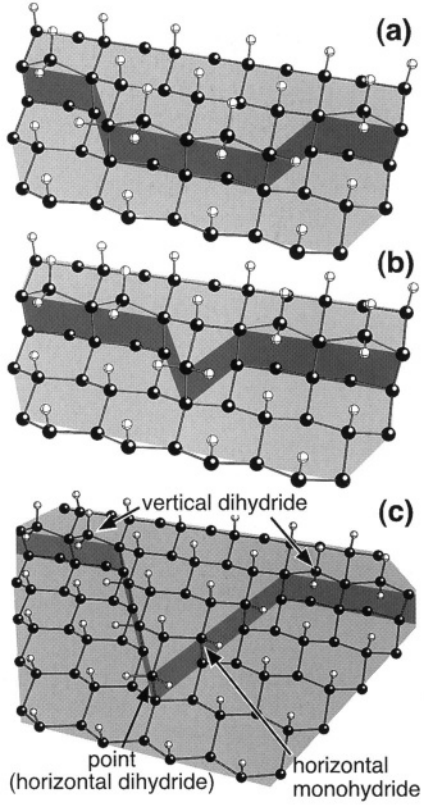


FIGURE 8.9. The formation of hillocks during aqueous etching. The dark balls represent Si atoms, while the light balls represent H atoms, (a) Two opposing kinks on a $\langle 1\bar{1}2\rangle$ step, (b) the newly nucleated hillock, (c) a well-developed hillock showing horizontal monohydride-terminated, $\langle 11\bar{2}\rangle$ faces.

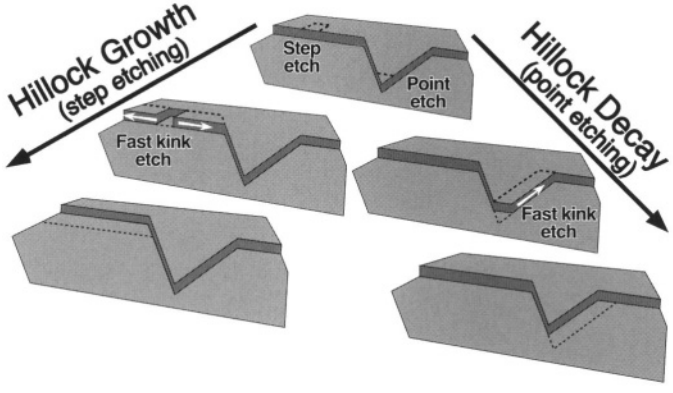


FIGURE 8.10. Hillock size is determined by a balance between growth, which is initiated by the etching of a vertical dihydride step, and decay, which is initiated by point or horizontal monohydride (not shown) step etching.

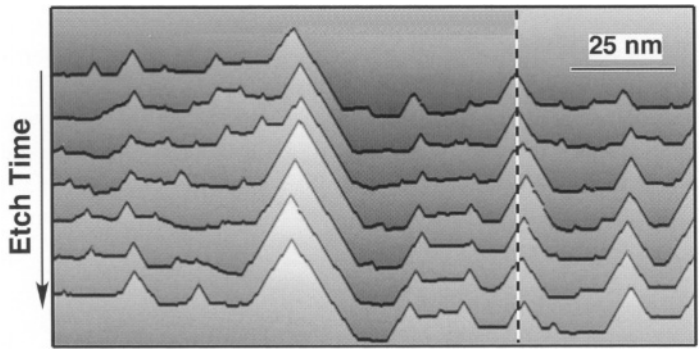


FIGURE 8.11. The time-dependent morphology of an etching $\langle \bar{1}\bar{1}2 \rangle$ step showing the propagation of two-dimensional hillocks. The apex of each hillock performs a random walk in the direction perpendicular to propagation, as illustrated by the dashed line.

kink etching is very fast [a good approximation for $NH_4F/Si(111)$ etching], the average rate of hillock growth, k_{grow}, will be given by

$$k_{grow} = \rho_{vDi} k_{vDi}, \tag{6}$$

where ρ_{vDi} and k_{vDi} are the average density and etch rate of $\langle \bar{1}\bar{1}2 \rangle$ step sites (i.e. vertical dihydrides), respectively. Similarly, the average rate of hillock decay, k_{decay}, will be given by

$$k_{decay} = \rho_{pnt} k_{pnt} + \rho_{mono} k_{mono}, \tag{7}$$

where the subscript pnt refers to point sites, and the subscript mono refers to $\langle 11\bar{2} \rangle$ step sites (i.e. monohydride steps). Since the site-specific etch rates (k_{vDi}, k_{mono}, and k_{pnt}) are determined by the chemical etchant, one more constraint is needed to completely specify the hillock size and density. This constraint is provided by the kink etch rate. Since hillocks are nucleated by the collision of two *opposing* kinks on a $\langle \bar{1}\bar{1}2 \rangle$ step segment, the average distance between hillocks (i.e. the average length of a straight $\langle \bar{1}\bar{1}2 \rangle$ step segment) is determined by the relative rates of kink nucleation (k_{vDi}) and kink propagation (k_{kink}).

The balance between growth and decay events implies that hillocks are not static features; they are self-propagating kinetic features that evolve with the etching surface through sequential growth and decay events. Their kinetic nature is illustrated in Figure 8.11, where a kinetic Monte Carlo simulation of the temporal evolution of a single etching step is displayed. In this sequence, large hillocks are seen to propagate across the surface with minor morphological fluctuations, while small hillocks are both nucleated or annihilated. As they propagate, the apices of the two-dimensional hillocks undergo a random walk in the direction parallel to the step edge. A QuickTime™ movie of this process is available online [58].

This simple example illustrates an important criterion for hillock formation that has often been overlooked. *The stability of a hillock depends crucially on the stability of its apex.* In other words, the mere existence of a slower etching step (or face in three dimensions) is not a sufficient criterion for hillock formation. In this example, hillock formation depends

crucially on the rate of point site etching. If this rate is too high, hillocks will be etched away as soon as they are nucleated, the etch kinetics will revert to simple terrace-step-kink etching, and the etched steps will be relatively straight.

The correlation between etch rate and hillock formation was first discussed by Batterman [7] in the context of three-dimensional hillock formation on germanium in an H_2O_2/HF etchant (Superoxol) and later expanded upon by Jaccodine [59]. Batterman pointed out that a pre-existing hillock with faces of orientation $\hat{h}$ will only be stable to further etching on a surface of orientation $\hat{s}$ if

$$R_h < R_s \cos\theta_{hs}, \tag{8}$$

where R_h and R_s are the etch rates of surfaces $\hat{h}$ and $\hat{s}$, respectively, and θ_{hs} is the angle between surfaces $\hat{h}$ and $\hat{s}$. Although necessary, this criterion should not be viewed as sufficient for hillock formation. As will be discussed further in Section 8.5, three-dimensional hillock formation during chemical etching is ubiquitous, especially on highly miscut surfaces. At the present time, there are *no predictive theories* of hillock formation, hillock stability, or hillock shape. Although hillock formation is often attributed to contamination, strikingly similar morphologies have been reported by laboratories separated by decades and by continents. Either contamination is surprisingly uniform or more subtle processes are at work.

At this point, the reader must surely be wondering about the supporting evidence for this detailed mechanism of two-dimensional hillock formation. Anisotropic etching literally writes a record of the reactivity of the etchant into the etched surface morphology [15]. As described in some detail in a recent review [16], the site-specific rates of chemical etching can therefore be *quantitatively measured* using a combination of experimental determinations of the steady state etch morphology, which are measured with a scanning tunneling microscope (STM), and atomistic, kinetic Monte Carlo simulations of the underlying chemical reactions. This process is enabled, in part, by previous spectroscopic investigations that determined the detailed structures of many defect sites on etched, H-terminated Si(111) surfaces [55–57].

Figure 8.8 shows representative experimental and simulated images of both principal miscuts of Si(111) etched in stirred, air-saturated 40% $NH_4F(aq)$ [53]. Both simulations were performed with the *same set of site-specific etch rates*; the only difference between the simulations is the orientation of the surface miscut. As expected, surfaces miscut towards the $\langle\bar{1}\bar{1}2\rangle$ direction form two-dimensional etch hillocks in both experiment and simulation. The good agreement between experiment and simulation in terms of the size, shape and density of the various etch features confirms the validity of this approach. Interestingly, even though all of the surface sites are silicon hydrides, the measured site-specific reactivities span a huge range. For example, the strained and distorted kink site is 10^7 times more reactive than the terrace site! The correlation between the structure and reactivity of the sites can also be used to infer the reaction mechanism [16]. This technique has also been used to study the site-specific reactivity of etch additives, such as isopropanol [60], and impurities, such as dissolved O_2 [61].

In summary, real surfaces may have reactive sites that fall outside of the simple terrace-step-kink classification, and these reactive sites may induce the formation of new etch structures. There are only two prerequisites for the formation of two-dimensional hillocks:

the existence of fast- and slow-etching steps, and the existence of a site that can form a relatively stable hillock apex. As such, two-dimensional etch hillocks may be relatively common.

8.3.3.2. Three-dimensional etch hillocks Three-dimensional etch hillocks have attracted significantly more attention than two-dimensional hillocks for obvious reasons. As discussed in previous sections, hillock formation is often attributed to heterogeneities in the etch system, which might be caused by contaminants or by localized depletion of the etchant. But can an etchant be *inherently unstable* to three-dimensional hillock formation? In this section, I will show that under certain circumstance, two-dimensional etch hillocks can spontaneously self-organize into three-dimensional hillocks [53]. In addition to creating three-dimensional "bumps," the hillocks also significantly increase the long-range roughness of the surface.

On an atomic scale, three-dimensional hillock formation depends critically on the kinetics of step collisions. In Section 8.2, simple terrace-step-kink etching was shown to lead to two forms of dynamic step–step repulsion that are mediated by step collisions and etch pits, respectively. In the absence of additional complications, terrace-step-kink etching is stable to hillock formation. To investigate the role of two-dimensional hillocks on step collisions, kinetic Monte Carlo etch simulations were performed on surfaces with variable surface miscut. The site-specific etch rates were chosen to mimic $NH_4F(aq)/Si(111)$ etching; however, terrace etching was specifically forbidden to minimize the effects of pit-induced step repulsion. Under the chosen kinetics, isolated steps developed two-dimensional hillocks that were $\approx$50 nm in height.

On surfaces where the average terrace width was smaller than the steady state hillock size, step-hillock collisions led to the production of three-dimensional hillocks, as shown by the time-lapse images in Figure 8.12. In this simulation, the surface steps had an average spacing of 3.5 nm; however, the etch kinetics produced 50 nm high two-dimensional hillocks on *isolated* steps. After only 10 ML of etching (Figure 8.12a), the etched steps developed characteristic, $\approx$5 nm tall hillocks. Since the hillocks were taller than the mean terrace width, step collisions forced the hillocks to partially align with one another, as shown by the inset in Figure 8.12a. As the etching continued, this process of hillock growth and self-alignment continued (Figure 8.12b–d). As a result, the etched surface developed three-dimensional hillocks which were composed of propagating trains of aligned two-dimensional-hillocks, as shown by the rendered image of the final surface (Figure 8.12e).

Even though these three-dimensional hillocks were relatively small ($\approx$50 nm), their presence had *profound effects* on the long-range roughness of the surface. Figure 8.13 shows the effects of hillock formation on steady-state terrace width distributions (TWD) extracted from three simulations of etching. All three of these simulations had very similar site-specific kinetics; the only parameter changed in the simulations was the point site etch rate. When the rate of point site etching was high, the surfaces underwent simple terrace-step-kink etching and were stable against hillock formation. As expected, these simulations led to the near-Gaussian TWD characteristic of terrace-step-kink etching (Figure 8.13a). The long-range morphology of these surfaces should resemble that of an annealed (i.e. equilibrium) surface. In contrast, when the rate of point site etching was low (Figure 8.13c), stable two- and three-dimensional hillocks formed on the surface. Under these kinetics, the TWD was well fit by P_{1D}, the exponential function characteristic of one-dimensional step etching [eq. (2)]! In other words, two-dimensional hillocks *suppressed* the dynamic step-step repulsion characteristic of step collisions in the terrace-step-kink model. Experimental

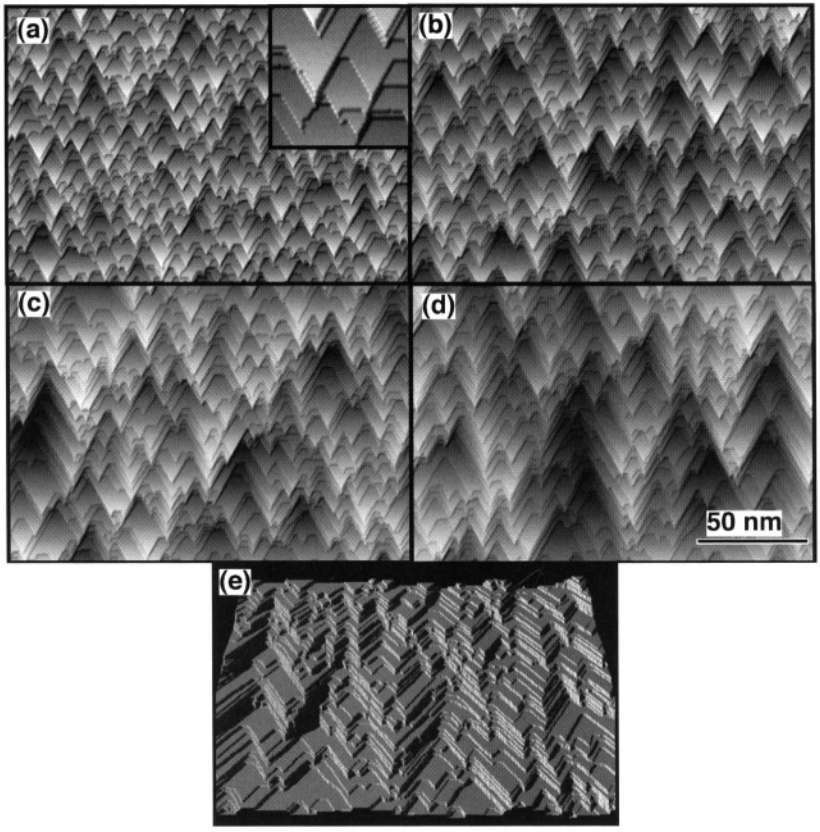


FIGURE 8.12. KMC simulation of the development of three-dimensional hillocks on a Si(111) surface miscut by 5° towards the $\langle \bar{1}\bar{1}2 \rangle$ direction, (a) After 10 ML of etching. The inset is an expanded, 200 × 200 Å view, (b–d) After 30 ML, 100 ML and 300 ML of etching, respectively; (e) a rendered view of the surface in *(d)*.

measurements of the TWD of highly miscut, NH_4F-etched Si(111) surfaces miscut in the $\langle \bar{1}\bar{1}2 \rangle$ direction confirm this prediction [62].

As discussed in Section 8.2, an exponential TWD is characteristic of a surface that *continually roughens with etching.* This behavior is inconsistent with the simple terrace-step-kink model that has dominated discussions of etching for the past decades and shows that some etchants may be inherently unstable to hillock formation. Nevertheless, this model is highly specific and not easily generalized. Phenomenological studies of chemical etching have uncovered a wide variety of hillock shapes and sizes [24]. Because of this, I expect that many hillock formation mechanisms are waiting to be discovered.

8.4. THE PRODUCTION OF NANOPERIODIC SURFACES

Today, the smallest lateral feature on a microprocessor in commercial production is ≈90 nm. In contrast, the distance between binding sites on an antibody is 10 nm—our smallest machines are 9 times larger than Nature's. Although impressive strides are being

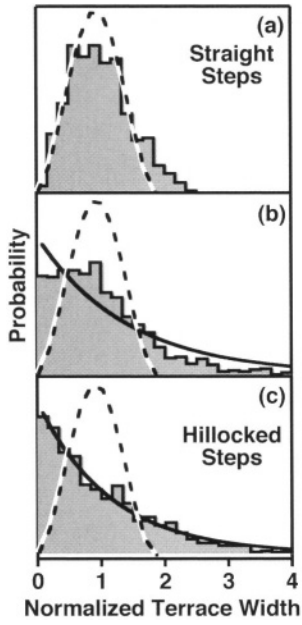


FIGURE 8.13. Normalized terrace width distributions for etched Si(111) surfaces miscut by 2.25° towards the $\langle \bar{1}\bar{1}2 \rangle$ direction. The kinetics were chosen to mimic NH_4F etching, except terrace etching was specifically disallowed, (a) When the rate of point site etching is high, hillocks are unstable, and straight steps are formed. In this case, the TWD mimics the equilibrium distribution. (b) An intermediate rate of point site etching leads to small hillocks and an intermediate TWD. (c) When the rate of point site etching is small, large hillocks form. In this case, the TWD is exponential, indicating the complete absence of dynamic step–step repulsion.

made in silicon processing, inexpensive, parallel nanofabrication at the 10 nm length scale will likely remain beyond the grasp of conventional nanofabrication for the foreseeable future. Under favorable conditions, electron-beam lithography can produce patterns at this length scale; however, the serial nature of this technique (and the complexity of the equipment) limit its widespread use. Scanned-probe-based nanofabrication techniques, such as direct atom manipulation [63] and dip pen lithography [64], are similarly hindered by their relatively slow scan rate.

To circumvent these limitations, many groups are developing techniques based on self-assembly—a so-called bottom-up approach. Although ultimately more limited in their capabilities, these techniques are attractive for both their cost and their simplicity. In the following, I describe an approach to the controlled fabrication of nanoperiodic surfaces that uses the inherent spacing between atoms to control the lateral dimensions of the structure and controlled chemical etching to develop the structure [65]. In principle, this technique can be used to prepare periodic features with a continuously variable spacing of $\approx$5–100 nm in any crystalline material.

This technique is based on the Moiré pattern that is generated whenever two misoriented, periodic structures are superimposed. For example, when two square arrays are misoriented by an angle θ, a square superstructure is formed, as sketched in Figure 8.14. Interestingly, the size of the superstructure is dependent on the misorientation angle. Large

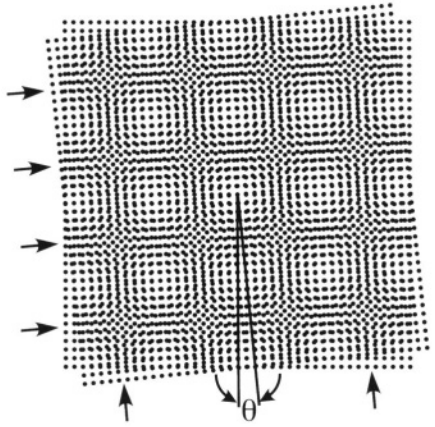


FIGURE 8.14. A pure twist boundary produces orthogonal screw dislocations, indicated by arrows, as illustrated by this Moiré representation of the bonding of two Si(100) surfaces.

misorientations lead to small periodicities, while small misorientations lead to large periodicities.

In this technique, a single crystal surface, such as Si(100), is used as the initial periodic pattern. When two very smooth and very clean silicon surfaces are brought into close contact at room temperature, a surprisingly strong van der Waals bond is formed between the two surfaces. In fact, a well-bonded surface is stable for weeks and often survives mechanical shocks. After annealing, strong chemical bonds are formed between the surfaces, and the initial Moiré pattern is transformed into an array of *screw dislocations.* In a pure twist boundary, the spacing between the (undissociated) screw dislocations, d, is related to the misorientation angle by Frank's rule,

$$d = \frac{|\mathbf{b}|}{2\sin(\theta/2)}, \tag{9}$$

where **b** is the Burgers vector of the dislocation. By varying the twist angle—an easily adjusted macroscopic parameter—between 0.2° and 10°, dislocation spacings ranging from 100–2 nm can be generated at a silicon twist boundary.

This *interfacial* pattern can be readily imaged using transmission electron microscopy (TEM). For example, Figure 8.15a shows the dislocation structure at the interface of a twist-bonded bicrystal fabricated from two Si(100) wafers. As expected, a square array of orthogonal screw dislocations with a mean spacing of 37.5 nm, corresponding to a twist angle of 0.6°, is visible. In addition, an array of meandering lines (indicated by arrows) is also visible. These lines are *edge dislocations* formed by the slight difference in miscut between the two surfaces which introduces a tilt component to the boundary.

The production of periodic interfacial structures at grain boundaries has been known for decades [66]. The challenge lies in bringing this interfacial structure to the surface. Fortunately, materials scientists developed a number of dislocation etchants in the '50s and '60s to aid in the development of high performance crystal growth technology for the then-developing semiconductor industry [67]. Although the mechanism of dislocation etching

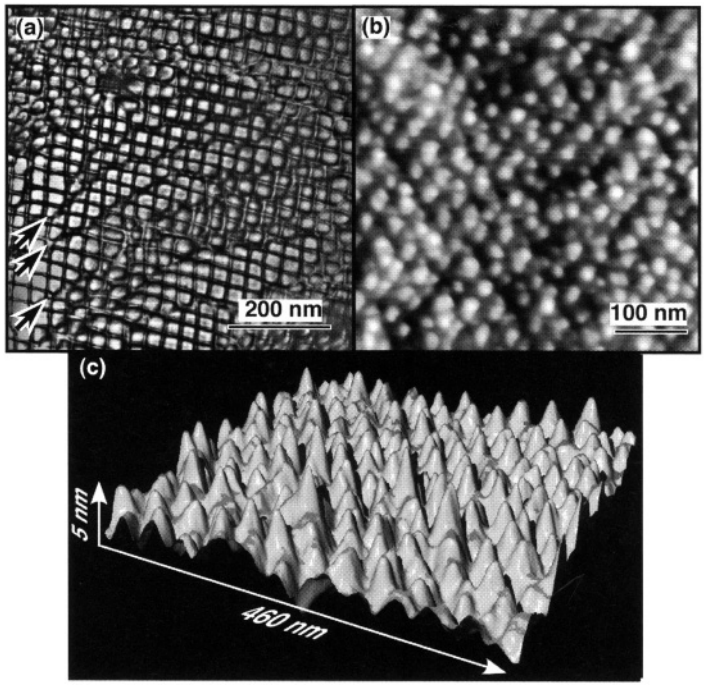


FIGURE 8.15. Experimental images obtained from samples cut from the same bicrystal displaying 38 nm periodicity. (a) TEM image of tilt-twist boundary showing orthogonal screw dislocations and meandering edge dislocations (indicated by arrows), (b) AFM image of etched surface, (c) Rendered representation of AFM image.

is poorly understood, selective dislocation etching is usually attributed to local etch rate enhancement by the strain energy in the vicinity of the dislocation [24].

To bring the interfacial structure to the surface, the upper crystal was first thinned to a thickness of ≈50 nm, then a dilute form of a common dislocation etchant, the Sirtl etch [68], was used to selectively attack the dislocations [65]. After etching, the entire surface was covered by an array of 5 nm tall, periodic structures, as shown by the AFM images in Figures 8.15b and 8.15c. When the initial interfacial structure (measured with TEM) and the final surface structure (measured with AFM) were compared using samples cut from the same bicrystal, the measured periodicities were in excellent agreement (37.5 nm vs. 38.6 nm, respectively).

Many applications for nanotextured surfaces can be envisioned. For example, these surfaces could be used as templates for the ordering of quantum dots or biological molecules. Additionally, the bonding of two dissimilar materials, such as silicon and germanium, will lead to misfit dislocations, and the spacing of these dislocations can also be varied with misorientation angle [69]. The controlled etching of hetero-bicrystals would lead to the formation of periodic heterostructures with controlled spacing.

From the perspective of etching chemistry, these bicrystals also present a unique opportunity to study the fundamental mechanisms of dislocation etching. Calculations by Bourret [70] have shown that the strain energy surrounding interfacial screw dislocations produced by twist-bonding is approximately two orders of magnitude smaller than the

strain energy surrounding edge dislocations produced by tilt-bonding. This difference has a simple origin. Since the strain fields of neighboring screw dislocations tend to cancel each other, buried screw dislocations produce little elastic strain at the surface. If preferential dislocation etching were due to stored elastic energy—as often assumed in the literature [24]—Bourret's calculations suggest that etching of a buried tilt-twist boundary would produce meandering trenches from the edge dislocations. A two-dimensional array should not be produced.

The AFM images in Figure 8.15 are inconsistent with this strain-mediated mechanism. The absence of meandering lines in the image, which would be produced by rapid edge dislocation etching, leads us to conclude that screw dislocations etch at least as fast as edge dislocations. Thus, the Sirtl etch is sensitive to the structure and bonding of the dislocation cores, not just the elastic energy.

8.5. THE SEARCH FOR NEW ETCHANTS: ORIENTATION-RESOLVED ETCHING

Etching is often considered more an art than a science, because the formulation of new anisotropic etchants is based upon trial-and-error much more than upon chemical insight. The problem, at least in part, is that anisotropic etchants are by their very nature highly defect selective. This selectivity has hindered chemical investigations. For example, the NH_4F etchant discussed in Section 8.2 selectively etches kink sites, which are present in ppm surface concentrations, while leaving the majority species, the terrace site, essentially unetched. (The kink site is 7 orders of magnitude more reactive than the terrace.) Direct investigation of the chemistry of the kink site is impossible, though, as surface spectroscopies can only probe species present in concentrations exceeding $\approx 1\%$ of a monolayer. Although indirect methods based upon STM analysis have been developed [16], these techniques are very slow and poorly suited to the evaluation and comparison of many etchants.

With these problems in mind, we developed a new technique for the rapid assay of surface defect reactivity that is based on orientation-dependent changes in the density of surface defects [71]. For example, consider the search for a new etchant to produce atomically smooth surfaces. As discussed in Section 8.2, the best etchant for this purpose would be a step-flow etchant. In other words, the new etchant must attack surface steps much more rapidly than perfect terraces. If every step etches independently, the macroscopic etch rate, R, should be proportional to both the density of surface steps, which is controlled by the surface miscut, and by the velocity of each step as it flows across the surface, v_{step}. For perfect step-flow etching of vicinal steps, simple geometry predicts that

$$R(\theta) = v_{step} \sin|\theta|, \tag{10}$$

where θ is the miscut angle measured from the nearest close-packed plane. In other words, a perfect step-flow etchant should have a *characteristic orientation-dependent etch rate.* The close-packed surface should be essentially unreactive [i.e. $R(0) = 0$], while vicinal surfaces should etch with a rate proportional to their miscut.

8.5.1. Assaying etchant anisotropy with micromachined test patterns

Our assay is based on the geometric amplification inherent to the etching of a wedge. When the sides of a wedge of included angle α are etched by a distance Δt, the length of the wedge is reduced by a distance

$$\Delta l = \Delta t \cot(\alpha/2) = 115\Delta t \quad \text{for } \alpha = 1^\circ. \tag{11}$$

A 1° wedge, which is easily produced using standard microfabrication techniques, produces a more than 100-fold amplification! In the case of basic silicon etchants, where etch rates are typically tens of microns per hour (at ≈70°C), the retraction rate of 1° wedges is a few millimeters per hour—a distance easily seen by eye and measured with an optical microscope.

The initial pattern consists of 180 1°-wide silicon wedges arranged in an evenly spaced, circular array as diagrammed in Figure 8.16a. The entire pattern is 18 mm in diameter, and the individual wedges are ≈130 μm tall. The wedges are 140 μm in width at their widest and taper to a 2 μm wide apex. Since the initial structure is fabricated from a single crystal silicon wafer, each wedge is bounded by a different set of vertical planes. The wafer orientation is important, as the surfaces measured by this technique are all *perpendicular* to the starting wafer. For example, if a Si(110) wafer is used, the pattern will contain wedges

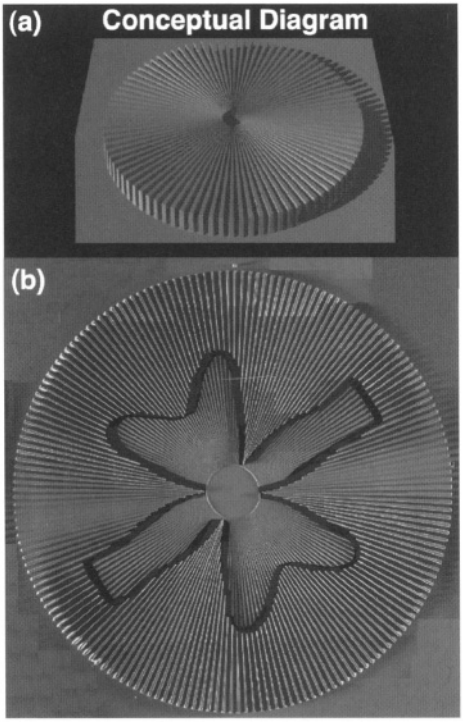


FIGURE 8.16. The micromachined test pattern. (a) Conceptual diagram of the initial pattern. (b) Composite optical micrograph of etched pattern. The pattern is 18 mm in diameter, and each wedge is 1° wide.

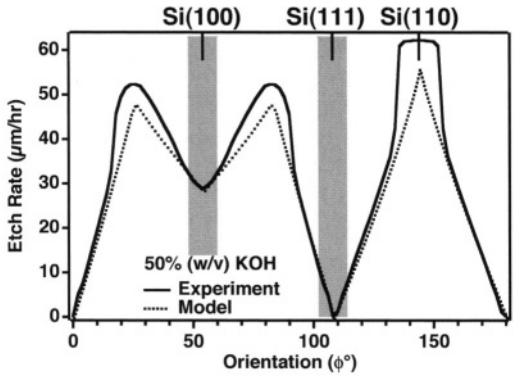


FIGURE 8.17. Measured orientation dependent etch rates for 70°C, 50% (w/v) KOH (solid line) and best fit to the step-flow model described by eq. (12) (dotted line). Data in the shaded regions were used to extract model parameters. The orientations of the close-packed planes are indicated at top.

representing the close-packed Si(100), Si(111) and Si(110) surfaces as well as their vicinal surfaces.*

When the test pattern is immersed in an etchant, a characteristic flower pattern develops, such as the one shown in Figure 8.16b. This flower pattern is a direct measure of the site-specific etch rates. After etching, the lengths of the etched wedges are quantified by optical microscopy and image analysis, and these data are then analyzed to obtain orientation-dependent macroscopic etch rates. For example, the solid line in Figure 8.17 represents the measured, orientation-dependent etch rates of a 70°C, 50% w/v KOH(aq) solution, where ϕ represents the orientation of the etching face with respect to one of the (1 1 1) surfaces. Local etch rate minima are observed at both of the close-packed faces, Si(111) and Si(100). Interestingly, an etch rate *maximum* is observed at the (110) orientation, another relatively close-packed plane.

Since the Si(100) surface etches with a finite etch rate, as evidenced by the shallow minimum in Figure 8.17, the simple step flow model described by eq. (10) must be modified to include the possibility of terrace etching. If the terraces etch with a velocity v_{terr} (measured in the direction perpendicular to the terrace), simple geometry shows that the macroscopic rate of etching should be approximately

$$R(\theta) = v_{\text{step}} \sin|\theta| + v_{\text{terr}} \cos\theta. \tag{12}$$

In eq. (12), the first term accounts for the contributions from step-flow etching, while the second term accounts for the rate of terrace etching. Of course, surfaces on the micromachined array have a number of different types of vicinal steps, and each type of step is expected to have a different etch rate. For example, Si(1 1 1) surfaces miscut towards the (110) direction (i.e. $\phi > 108°$ in Figure 8.17) are expected to have monohydride-terminated steps, while the surfaces miscut towards the (100) direction (i.e. $\phi < 108°$) should have dihydride-terminated steps.

*To avoid confusion, all geometrically equivalent surfaces are referred to by their "principle" name. For example, wedges with $(\bar{1}11)$, $(1\bar{1}1)$, $(11\bar{1})$ and $(\bar{1}1\bar{1})$ sidewalls are formed on Si(110) surfaces; however, all of these surfaces are referred to as "(111) faces."

This very simple step flow model explains the orientation-dependent rates of KOH/Si etching surprisingly well. For example, the dotted lines in Figure 8.17 represent the best fit to the simple model described by eq. (12). The fit is particularly good near the Si(100) and Si(111) minima. Interestingly, there is no evidence of direct step–step interactions or step-density-dependent step etching on vicinal Si(111) surfaces within $\approx 20°$ of the close-packed plane.

This good fit strongly suggests that 50% w/v KOH is a step-flow etchant of vicinal Si(111) surfaces. Indeed, morphological investigations are in good agreement with this prediction, as shown by the AFM image in Figure 8.7. Nevertheless, many question remain. Since these kinetics also suggest that the vicinal steps etch with a step-density independent rate, vicinal Si(111) surfaces should be stable to macrostep and hillock formation, which is in obvious disagreement with the morphology in Figure 8.7 and others [51].

8.5.2. Factors affecting etchant anisotropy

As discussed earlier, the goal of chemically controlling surface morphology relies on an ability to chemically control etchant anisotropy. But what factors affect etchant anisotropy? At least in the case of basic silicon etchants, such as KOH and tetramethyl ammonium hydroxide (TMAH), etchant anisotropy is surprisingly sensitive to a wide range of chemical parameters, including temperature, etchant concentration and etchant composition. More importantly, these dependencies are complex and not easily analyzed using concepts from homogenous chemical kinetics.

This complexity is exemplified by the concentration-dependent anisotropies for the KOH etching of silicon, which are displayed in Figure 8.18. Interestingly, all orientations show a maximum etch rate at intermediate KOH concentrations As originally suggested by Palik and coworkers [72], this behavior is consistent with a mechanism in which both OH^-

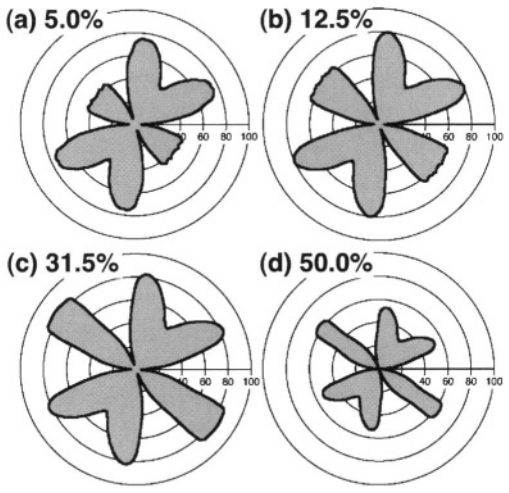


FIGURE 8.18. Concentration dependent etch rates of silicon in a variety of 70°C KOH solutions. The etch rates are in μm/h, and the horizontal axis denotes the Si(111) face at $\phi = 0°$. All graphs have the same scaling.

and H_2O are reactants. Since this reaction evolves $H_2(g)$, Palik and coworkers suggested a net reaction of the form

$$Si + 2H_2O + 2OH^- \rightarrow 2H_2 + Si(OH)_2O_2^{2-}. \tag{13}$$

In homogeneous chemical kinetics, the rate of a chemical reaction is usually well fit by an empirical rate law that scales as the concentration of the reactants raised to small powers (e.g., 0.5, 1, 2). The scaling exponents in turn often reflect the molecularity of the rate limiting step. Thus, the concentration-dependent rate of KOH etching is expected to have the functional form

$$R(\phi) = k[OH^-]^a[H_2O]^b, \tag{14}$$

where *k, a,* and *b* are constants that may be orientation dependent.

Even when additional complexities are taken into account (e.g., solvation effects, activities), the concentration dependent anisotropies in Figure 8.18 cannot be credibly fit to an empirical rate law of the form suggested in eq. (14). This breakdown of classical chemical kinetics is likely due to the multisite nature of chemical etching. As the etch parameters are changed, the steady-state morphology of the etching surface also changes. In the case of KOH etching, the morphological changes are often dramatic and observable in an electron microscope. On an atomic scale, these morphological changes imply that the etch site densities (e.g., the densities of steps and kinks) are also concentration dependent. If every type of site *i* on the surface etches with a well defined empirical rate law, the macroscopic etch rate will be determined by

$$R(\phi) = \sum_{i=1}^{N} \rho_i\left([OH^-], [H_2O]\right) k_i [OH^-]^{a_i} [H_2O]^{b_i}, \tag{15}$$

where ρ_i is the concentration-dependent density of site *i*. Even in the case where all sites etch with the same empirical rate law (i.e. the same *k, a,* and *b),* the concentration dependent site densities preclude the reduction of eq. (15) to a macroscopic empirical rate law [e.g., eq. (14)]. In general, eq. (15) implies that macroscopic etch rate may have a complicated concentration dependence. More importantly, the molecularity of the chemical mechanism *may not be simply related* to the concentration dependence of the macroscopic etch rate.

The anisotropy of KOH etching is also surprisingly sensitive to etchant temperature. Like the concentration dependence, the temperature dependence defies simple explanation, as the macroscopic etch rates display non-Arrhenius behavior. This behavior is similarly attributed to temperature dependent changes in etch morphology and site density.

In spite of a pressing need for an understanding of the chemical mechanisms that govern etchant anisotropy, these preliminary investigations suggest that simple answers will not be readily forthcoming for many etchants. Indeed, a detailed study of an exceedingly simple etch system, KOH/silicon, raised more questions than it answered. In particular, the richness of the orientation-, concentration- and temperature-dependent etch morphologies was surprising. This complexity is likely a reflection of the multisite nature of chemical etching. Nevertheless, this simple screening method may be very useful in testing a wide range of potential etchants for step-flow activity.

8.6. CONCLUSIONS AND OUTLOOK

Although some have characterized etching chemistry as more of an art than a science, recent advances in scanned probe microscopies have enabled detailed, atomic-scale studies of the mechanisms that govern the development of nanoscale features during chemical etching. The simplest type of chemical etching, which only involves the etching of terrace, step and kink sites, is now well understood, and highly anisotropic terrace-step-kink etchants can be used to produce ultra-flat surfaces. Interestingly, a low terrace site etch rate can lead to smoother long-range morphologies than are possible with a step-kink etchant. Since very few crystalline materials have a simple cubic lattice, simple terrace-step-kink etching may only be applicable in a small subset of cases. More exotic reaction sites, such as the point site found on Si(111) surfaces, may complicate etching. In some cases, these exotic species can lead to inherently unstable chemistry and the production of two- or three-dimensional etch hillocks—nanoscale structures with well-controlled geometry and size. Chemistries of this type are useful for the production of nanotextured surfaces. Similarly, nanoperiodic morphologies can be generated through the selective etching of patterned dislocations formed at intentional grain boundaries.

The major challenge facing etching chemistry today is the rational design of new etchants for the production of specific nanoscale features. To this end, a micromachined test pattern that can rapidly screen etchants for step-flow etching has been developed. Nevertheless, there is room for much more improvement in this area.

ACKNOWLEDGMENTS

The work described in this chapter was supported by the National Science Foundation, the Beckman Young Investigator Program, the Petroleum Research Foundation, and Philip Morris USA. MAH is a Cottrell Scholar of the Research Corporation and has been privileged to collaborate with a number of talented students on this work, including Dr. Jaroslav Flidr, Dr. Yi-Chiau Huang, Dr. Theresa A. Newton, Dr. Rikard A. Wind, Simon P. Garcia, Hailing Bao, Helen Jones, and Michael J. Little. Special thanks goes to the Cornell Center for Materials Research and the Cornell Nanofabrication Facility for their support and for the use of their excellent facilities, without which this research would not have been possible.

REFERENCES

1. G.E. Moore, Cramming more components onto integrated circuits, *Electronics* **38**(8), 114–117 (1965).
2. G.E. Moore, Lithography and the future of Moore's Law, *SPIE Proc.* **2440,** 2–17 (1995).
3. R.R. Schaller, Moore's law: past, present, and future, *IEEE Spectrum* **34,** 52–59 (1997).
4. J. Birnbaum and R.S. Williams, Physics and the information revolution, *Physics Today* **53**(1), 38–42 (2000).
5. C.V. Schreibers, *Beitrag z. Geschichte u. Kenntnis d. meteorischen Steine u. Eisenmassen* (Wien, 1820).
6. H.C. Gatos and M.C. Lavine, Chemical behaviour of semiconductors: etching characteristics, in: *Progress in Semiconductors,* Vol. 9, Eds A.F. Gibson and R.E. Burgess (Temple Press, London, 1965) pp. 3–45.
7. B.W. Batterman, Hillocks, pits, and etch rate in germanium crystals, *J. Appl. Phys.* **28**(11), 1236–1241 (1957).
8. R.C. Ellis, Jr., Etching of single crystal germanium spheres, *J. Appl. Phys.* **25**(12), 1497–1499 (1954).

9. E. Bassous and E.F. Baran, The fabricaton of high precision nozzles by the anisotropic etching of (100) silicon, *J. Electrochem. Soc.* **125**(8), 1321–1327 (1978).
10. P. Campbell and M.A. Green, Light trapping properties of pyramidally textured surfaces, *J. Appl. Phys.* **82**, 243–249 (1987).
11. J. Knoch, J. Appenzeller, B. Lengeler, R. Martel, P. Solomon, Ph. Avouris, Ch. Dieker, Y. Lu, K.L. Wang, J. Scholvin and J.A. del Alamo, Technology for the fabrication of ultrashort channel metal-oxide-semiconductor field-effect transistors, *J. Vac. Sci. Technol. A* **19**(4), 1737–1741 (2001).
12. T. Ohmi, K. Kotani, A. Teramoto and M. Miyashita, Dependence of electron channel mobility on $Si–SiO_2$ interface microroughness, *IEEE Electron. Dev. Lett.* **12**(12), 652–654 (1991).
13. D.W. Carr, S. Evoy, L. Sekaric, J.M. Parpia and H.G. Craighead, Measurement of mechanical resonance and losses in nanometer scale silicon wires, *Appl. Phys. Lett.* **75**(7), 920–922 (1999).
14. B. Holst and W. Allison, An atom focussing mirror, *Nature* **390**, 244 (1997).
15. J. Flidr, Y.-C. Huang, T.A. Newton and M.A. Hines, Extracting site-specific reaction rates from steady state surface morphologies: kinetic Monte Carlo simulations of aqueous Si(111) etching, *J. Chem. Phys.* **108**(13), 5542–5553 (1998).
16. M.A. Hines, The picture tells the story: using surface morphology to probe chemical etching reactions, *Intl. Rev. of Phys. Chem.* **20**(4), 645–672 (2001).
17. Y.-C. Huang, J. Flidr, T.A. Newton and M.A. Hines, Dynamic repulsion of surface steps during step flow etching: controlling surface roughness with chemistry, *J. Chem. Phys.* **109**(12), 5025–5035 (1998).
18. B.S. Swartzentruber, Y.-W. Mo, R. Kariotis, M.G. Lagally and M.B. Webb, Direct determination of step and kink energies on vicinal Si(001), *Phys. Rev. Lett.* **65**(15), 1913–1916 (1990).
19. E.E. Gruber and W.W. Mullins, On the theory of anisotropy of crystalline surface tension, *J. Phys. Chem. Solids* **28**(5), 875–887 (1967).
20. N.C. Bartelt, T.L. Einstein and E.D. Williams, The influence of step–step interactions on step wandering, *Surf. Sci.* **240**(1–3), L591–L598 (1990).
21. B. Joós, T.L. Einstein and N.C. Bartelt, Distribution of terrace widths on a vicinal surface within the one-dimensional free fermion model, *Phys. Rev. B* **43**(10), 8153–8162 (1991).
22. T.L. Einstein and O. Pierre-Louis, Implications of random-matrix theory lor terrace-width distributions on vicinal surfaces: improved approximations and exact results, *Surf. Sci.* **424**(1), L299–L308 (1999).
23. Y.-C. Huang, J. Flidr, T.A. Newton and M.A. Hines, The effects of dynamic step–step repulsion and auto-catalysis on the morphology of Si(111), *Phys. Rev. Lett.* **80**, 4462 (1998).
24. K. Sangwal, *Etching of Crystals: Theory, Experiment and Applications* (North-Holland, Amsterdam, 1987).
25. R. Maboudian, Surface processes in MEMS technology, *Surf. Sci. Reports* **30**(6–8), 207–270 (1998).
26. F.C. Frank, On the kinematic theory of crystal growth and dissolution processes, in: *Growth and Perfection of Crystals,* Eds R.H. Doremus, B.W. Roberts and D. Turnbull (Wiley, New York, 1958), pp. 411–419.
27. N. Cabrera and D.A. Vermilyea, The growth of crystals from solution, in: *Growth and Perfection of Crystals,* Eds R.H. Doremus, B.W. Roberts and D. Turnbull (Wiley, New York, 1958), pp. 393–404.
28. M.J. Lighthill and G.B. Whitham, On kinematic waves: I. the movement of floods in long rivers, *Proc. Royal Soc. London A* **229**, 281–316 (1955).
29. C. Duport, P. Noziéres and J. Villain, New instability in molecular beam epitaxy, *Phys. Rev. Lett.* **74**(1), 134–137 (1995).
30. R.V. Kukta and K. Bhattacharya, A three-dimensional model of step flow mediated crystal growth under the combined influences of stress and diffusion, *Thin Solid Films* **357**(1), 35–39 (1999).
31. V.B. Shenoy, S. Zhang and W.F. Saam, Step-bunching transitions on vicinal surfaces with attractive step interactions, *Surf. Sci.* **467**(1–3), 58–84 (2000).
32. M.I. Larsson, Surface diffusion mechanism for step bunching, *Phys. Rev. B* **56**(23), 15157–15166 (1997).
33. G.S. Bales and A. Zangwill, Morphological instability of a terrace edge during step-flow growth, *Phys. Rev. B* **41**(9), 5500–5508(1990).
34. M. Vladimirova, A. Pimpinelli and A. Videcoq, A new model of morphological instabilities during epitaxial growth: from step bunching to mounds formation, *J. Crystal Growth* **220**(4), 631–636 (2000).
35. A. Videcoq, A. Pimpinelli and M. Vladimorova, Kinetic Monte Carlo study of the terrace width distribution during step bunching in homoepitaxial growth, *Appl. Surf. Sci.* **177**(4), 213–220 (2001).
36. A.V. Latyshev, A.L. Aseev, A.B. Krasilnikov and S.I. Stenin, Transformations on clean Si(111) stepped surface during sublimation, *Surf. Sci.* **213**(1), 157–169 (1989).

37. D. Kandel and J.D. Weeks, Simultaneous bunching and debunching of surface steps: theory and relation to experiments, *Phys. Rev. Lett.* **74**(18), 3632–3635 (1995).
38. E.D. Williams, E. Fu, Y.-N. Yang, D. Kandel and J.D. Weeks, Measurement of the anisotropy ratio during current-induced step bunching, *Surf. Sci.* **336**(1–2), L746–L752 (1995).
39. E.S. Fu, D.-J. Liu, M.D. Johnson, J.D. Weeks and E.D. Williams, The effective charge in surface electromigration, *Surf. Sci.* **385**(2–3), 259–269 (1997).
40. D.-J. Liu, J.D. Weeks and D. Kandel, Current-induced step bending instability on vicinal surfaces, *Phys. Rev. Lett.* **81**(13), 2743–2746 (1998).
41. J.P.v.d. Eerden and H. Müller-Krumbhaar, Dynamic coarsening of crystal surfaces by formation of macrosteps, *Phys. Rev. Lett.* **57**(19), 2431–2433 (1986).
42. J.P.v.d. Eerden and H. Müller–Krumbhaar, Dynamics of macrostep formation on crystals: recursive differential equations, *Physica Scripta* **40**(3), 337–339 (1989).
43. D. Kandel and J.D. Weeks, Step bunching as a chotic pattern formation process, *Phys. Rev. Lett.* **69**(26), 3758–3761 (1992).
44. D. Kandel and J.D. Weeks, Theory of impurity-induced step bunching, *Phys. Rev. B* **49**(8), 5554–5564 (1994).
45. A.J. Nijdam, E. van Veenendaal, H.M. Cuppen, J. van Suchtelen, M.L. Reed, J.G.E. Gardeniers, W.J.P. van Enckevort, E. Vlieg and M. Elwenspoek, Formation and stabilization of pyramidal etch hillocks on silicon (100) in anisotropic etchants: experiments and Monte Carlo simulations, *J. Appl. Phys.* **89**(7), 4113–4123 (2001).
46. R.J. Phaneuf, N.C. Bartelt, E.D. Williams, W. Swiech and E. Bauer, Low-energy electron-microscopy investigations of orientational phase separation on vicinal Si(111) surfaces, *Phys. Rev. Lett.* **67**(21), 2986–2989 (1991).
47. H. Hibino and T. Ogino, Transient step bunching on a vicinal Si(111) surface, *Phys. Rev. Lett.* **72**(5), 657–660 (1994).
48. H.-C. Jeong and J.D. Weeks, Two-dimensional dynamical model for step bunching and pattern formation induced by surface reconstruction, *Phys. Rev. B* **57**(7), 3939–3948 (1998).
49. S.V. Khare, T.L. Einstein and N.C. Bartelt, Dynamics of step doubling: simulations for a simple model and comparison with experiment, *Surf. Sci.* **339**(3), 353–362 (1995).
50. S.P. Garcia, H. Bao, R.A. Wind and M.A. Hines (unpublished results).
51. R.A. Wind, H. Jones, M.J. Little and M.A. Hines, Orientation-resolved chemical kinetics: using microfabrication to unravel the complicated chemistry of KOH/Si etching, *J. Phys. Chem. B* **106**(7), 1557–1569 (2002).
52. J. Flidr, Y.-C. Huang, T.A. Newton and M.A. Hines, The formation of etch hillocks during step-flow etching of Si(111), *Chem. Phys. Lett.* **302**(1–2), 85–90 (1999).
53. J. Flidr, Y.-C. Huang and M.A. Hines, An atomistic mechanism for the production of two- and three-dimensional etch hillocks on Si(111) surfaces, *J. Chem. Phys.* **111**(15), 6970–6981 (1999).
54. G.S. Higashi, Y.J. Chabal, G.W. Trucks and K. Raghavachari, Ideal hydrogen termination of the Si(111) surface, *Appl. Phys. Lett.* **56**(7), 656–658 (1990).
55. P. Jakob and Y.J. Chabal, Chemical etching of vicinal Si(111): dependence of the surface structure and the hydrogen termination on the pH of the etching solutions, *J. Chem. Phys.* **95**(4), 2897–2909 (1991).
56. M.A. Hines, Y.J. Chabal, T.D. Harris and A.L. Harris, Measuring the structure of etched silicon surfaces with Raman spectroscopy, *J. Chem. Phys.* **101**(9), 8055–8072 (1994).
57. K. Raghavachari, P. Jakob and Y.J. Chabal, Step relaxation and surface stress at H-terminated vicinal Si(111), *Chem. Phys. Lett.* **206**(1–4), 156–160 (1993).
58. Hines group web page, Department of Chemistry, Cornell University; (http://www.chem.cornell.edu/mah11/Hillock.html).
59. R.J. Jaccodine, Use of modified free energy theorems to predict equilibrium growing and etching shapes, *J. Appl. Phys.* **33**(8), 2643–2647 (1962).
60. T.A. Newton, Y.-C. Huang, L.A. Lepak and M.A. Hines, The site-specific reactivity of isopropanol in aqueous silicon etching: controlling morphology with surface chemistry, *J. Chem. Phys.* **111**(20), 9125–9128 (1999).
61. S.P. Garcia, H. Bao, M. Manimaran and M.A. Hines, Measuring the site-specific reactivity of impurities: the pronounced effect of dissolved oxygen on silicon etching, *J. Phys. Chem. B* **106**, 8258 (2002).
62. Y.C. Huang, Evolution of silicon surface morphologies during aqueous fluorine etching (Ph.D. Dissertation, Cornell University, 2000).

63. H.C. Manoharan, C.P. Lutz, and D.M. Eigler, Quantum mirages formed by coherent projection of electronic structure, *Nature* **403**(6769), 512–515 (2000).
64. R.D. Piner, J. Zhu, F. Xu, S. Hong and C.A. Mirkin, "Dip-pen" nanolithography, *Science* **283**(5402), 661–663 (1999).
65. R.A. Wind, M.J. Murtagh, F. Mei, Y. Wang and M.A. Hines, Fabrication of nanoperiodic surface structures by controlled etching of dislocations in bicrystals, *Appl. Phys. Lett.* **78**(15), 2205–2207 (2001).
66. C.B. Carter, H. Föll, D.G. Ast and S.L. Sass, Electron diffraction and microscopy studies of the structure of grain boundaries in silicon, *Phil. Mag. A* **43**(2), 441–167 (1981).
67. S. Amelinckx, *The Direct Observation of Dislocations* (Academic Press, New York, 1964).
68. E. Sirtl and A. Adler, Chromsäure-Flußsäure als spezifisches System zur Ätzengrubenentwicklung auf Silizium, *Z. Metalkunde* **52**(8), 529–531 (1961).
69. F.-S. Shieu and S.L. Sass, Experimental and theoretical studies of the dislocation structure of NiO-Pt interfaces, *Acta Metall. Mater.* **38**(9), 1653–1667 (1990).
70. A. Bourret, How to control the self-organization of nanoparticles by bonded thin layers, *Surf. Sci.* **432**(1–2), 37–53 (1999).
71. R.A. Wind and M.A. Hines, Macroscopic etch anisotropies and microscopic reaction mechanisms: A micromachined structure for the rapid assay of etchant anisotropy, *Surf. Sci.* **460**(1–3), 21–38 (2000).
72. E.D. Palik, V.M. Bermudez and O.J. Glembocki, Ellipsometric study of orientation-dependent etching of silicon in aqueous KOH, *J. Electrochem. Soc.* **132**(4), 871–84 (1985).

Index to Volume I

Index to Volume II