Materials Chemistry

Materials Chemistry

by

Bradley D. Fahlman
Central Michigan University,
Mount Pleasant, MI, USA

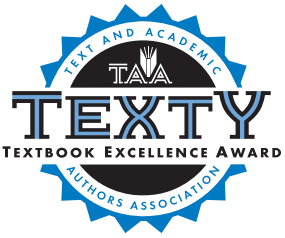
Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN 978-1-4020-6119-6 (HB)
ISBN 978-1-4020-6120-2 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com



RECOGNIZING CURRENT TEXTBOOKS AND LEARNING MATERIALS FOR EXCELLENCE

This textbook was awarded a "Textbook Excellence Award" from the Text and Academic Authors Association (TAA), a nonprofit professional membership association that serves textbook and academic materials authors (www.TAAonline.net). The "Textbook Excellence Award" recognizes works for their excellence in the areas of content, presentation, appeal, and teachability. These awards are given each year in several academic disciplines for both elementary and high school works and college titles. TAA's panel of judges are all noted textbook authors in these disciplines.

Printed on acid-free paper

First published 2007
Reprinted 2008

TABLE OF CONTENTS

Preface

Though most colleges and universities now have courses and degree programs related to materials chemistry, there is a need for a textbook that addresses inorganic-, organic-, and nano-based materials from a structure *vs.* property treatment. As I quickly discovered, trying to fill this void represented a daunting task of providing a suitable breadth and depth coverage of the rapidly evolving field of materials – all in a concise format. The material contained herein is most appropriate for junior/senior undergraduate students, as well as first-year graduate students in chemistry, physics, and engineering fields. In addition, this textbook will be extremely useful for researchers in industry as an initial source to learn about materials/techniques, with references provided for more detailed investigation.

After providing a historical perspective for the field of materials in the first chapter, the first concentration of the textbook focuses on solid-state chemistry. Though there are many popular textbooks that deal with this topic, my approach contains some unique perspectives. In addition to colored illustrations of archetypical crystalline unit cells, digital photos of models are also provided to add clarity to their structures. Further, a large section on amorphous solids including sol–gel techniques and cementitious materials is provided – largely left out of most solid-state textbooks. The next chapter on metals contains a thorough treatment of traditional and powder metallurgical techniques, with a focus on the complex phase behavior exhibited by the Fe–C system and steels. However, also left out of most metallurgy textbooks, I cover topics such as corrosion inhibition, magnetism, hydrogen storage, and the structure/properties of other metallic classes, such as the coinage metals and other alloys, such as those exhibiting shape-memory properties.

The next chapter deals with semiconducting materials, which consists of a discussion of band theory and semiconductor physics. Unique among other texts, I also describe in great detail the evolution of the transistor, with a discussion of current limitations and solutions currently being investigated by researchers in the field. Also described in this chapter is IC fabrication, including vapor deposition techniques, photolithography, and ion implantation. The current trends in applications such as LEDs/OLEDs, thermoelectric devices, and photovoltaics (including emerging technologies such as dye-sensitized solar cells) are also provided in this chapter.

Polymers and organic-based "soft" materials represent one of the largest materials classes; however, these materials are often left out of solid-state textbooks. Herein, I cover the chemistry of the five classes of polymers, from simple chains to complex branched dendritic architectures. Again, the approach is unique in that it covers traditional mechanisms and structure/property relationships for polymers, in addition to advanced topics such as homogeneous catalysis, polymer additives, and self-healing polymers. A detailed discussion of "molecular magnets" is also provided in this chapter, due to their relatively mild syntheses and "soft" properties relative to traditional inorganic-based magnets.

I devoted a significant effort in the next chapter to nanomaterials, due to their increasing popularity and relevance for current/future applications. In addition to structure/property descriptions and applications, essential topics such as nomenclature, synthetic techniques, and mechanistic theories are described in detail. The last chapter is also of paramount importance for the materials community – characterization. From electron microscopy to surface analysis techniques, and everything in between, this chapter provides a thorough description of modern techniques used to characterize materials. A flowchart is provided at the end of the chapter that will assist the materials scientist in choosing the most suitable technique(s) to characterize a particular material.

At the end of each chapter, a section entitled "Important Materials Applications" is provided, along with open-ended questions and detailed references/bibliography. Appendices are also provided that contain an interesting timeline of major materials developments, the complete Feynman speech "There's Plenty of Room at the Bottom," and a preliminary collection of materials-related laboratory modules. These additions were provided to promote student engagement through effective student–instructor interactions. Though I attempted to hit all of the "high notes" in the materials world, an obvious omission would be a detailed discussion of biomaterials – of increasing importance throughout the world. Applications such as biomimetics and drug delivery are presented in this edition; however, a detailed discussion of this topic was beyond the scope of providing a *concise* first edition – a separate chapter on this topic will appear in future editions.

The realization of a major milestone such as the completion of a textbook would not have been possible without the influence of many people in my life. First and foremost, I wish to thank my precious wife Diyonn for her patience and support during the many months of seclusion, as I crouched behind the laptop monitor. I am eternally grateful to my parents Frank and Pearl for their continuing support and godly wisdom, to whom I attribute all of my many blessings and successes. I must also acknowledge my Ph.D. advisor Andy Barron (Rice University) for his guidance and advice. I truly have never met anyone with as much drive and excitement for both materials research and teaching (as well as professional autoracing!). I thank him for being such an effective role model for the challenging and rewarding life of academia.

Of course, this textbook is the compilation of input from a number of my professional colleagues. I wish to thank Profs. Anja Mueller, Bob Howell (at CMU), Richard Finke (Colorado State University), and Jean-Claude Thomassian (at Georgia

Southern University) for their input and suggestions regarding various sections of this textbook. I am also very appreciative for the input provided by the first students to have access to evolving versions of the textbook – Nick Bedford, Jesse Thompson, Brandon Rohde, Brian Smith, Dan Denomme, Jason MacDonald, Laura Slusher, David Moyses, Michael Todd, Leontios Nezeritis, and Megan McCallum. The administration at CMU has also been a constant source of support. In particular, I wish to thank President Michael Rao, Dean Bob Kohrman, chemistry Chairman David Ash, and my faculty/staff colleagues from the chemistry department. I cannot express in words what their support has meant to me.

The close proximity to Midland, MI, a leading center of the chemical industry, has also provided much inspiration for this textbook and my research projects. In particular, Petar Dvornic (Michigan Molecular Institute) is a constant source of inspiration and also provided much feedback for this textbook. I also gratefully acknowledge Don Tomalia (Dendritic Nanotechnologies, Inc. at CMU) for his support since my arrival at CMU in 2002. Thanks for putting CMU on the map in the dendrimer field! I would also like to acknowledge the funding agencies of the Dreyfus Foundation, Research Corporation, and the Army Research Laboratory who provided support for my research interests during my first years at CMU.

Last, but certainly not least, I thank every reader of this book, and solicit your comments to my email fahlm1b@cmich.edu. Please let me know what you think of this edition; I will earnestly try to incorporate your suggestions to strengthen future editions.

Bradley D. Fahlman, Ph.D.
Mount Pleasant, Michigan
March 2007

CHAPTER 1

WHAT IS *MATERIALS CHEMISTRY*?

Life in the 21st century is ever dependent on an unlimited variety of advanced materials. In our consumptive world, it is easy to take for granted the macro-, micro-, and nanoscopic building blocks that comprise any item ever produced. We are spoiled by the technology that adds convenience to our lives, such as microwave ovens, laptop computers, digital cell phones, and improved modes of transportation. However, we rarely take time to think about and appreciate the materials that constitute these modern engineering feats.

The term *material* may be broadly defined as any solid-state component or device that may be used to address a current or future societal need. For instance, simple building materials such as nails, wood, coatings, *etc*. address our need of shelter. Other more intangible materials such as nanodevices may not yet be widely proven for particular applications, but will be essential for the future needs of our civilization. Although the above definition includes solid nanostructural building blocks that assemble to form larger materials, it excludes complex liquid compounds such as crude oil, which may be more properly considered a *precursor* for materials.

A general description of the various types of materials is illustrated in Figure 1.1. Although this indicates sharp distinctions between various classes, there is often ambiguity regarding the proper taxonomy for a specific material. For example, a *thin film* is defined as having a film thickness of less than 1 μm; however, if the thickness drops to below 100 nm, the dimensions may be more accurately classified within the nanoscale regime.[1] Likewise, liquid crystals are best described as having properties intermediate between amorphous and crystalline phases, and *hybrid composite* materials involve both inorganic and organic components.

The broadly defined discipline of *materials chemistry* is focused on understanding the relationships between the arrangement of atoms, ions, or molecules comprising a material, and its overall bulk structural/physical properties. By this designation, common disciplines such as polymer, solid-state, and surface chemistry would all be placed within the scope of materials chemistry. This broad field consists of studying the structures/properties of existing materials, synthesizing and characterizing new materials, and using advanced computational techniques to predict structures and properties of materials that have not yet been realized.

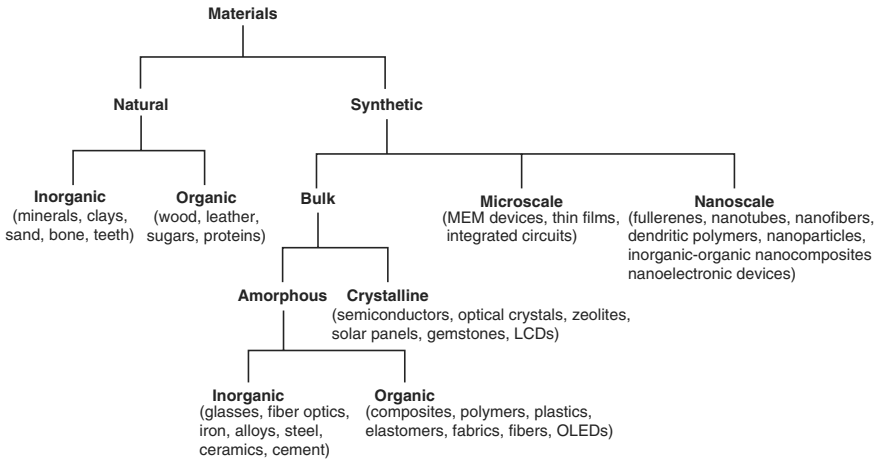


Figure 1.1. Classification scheme for the various types of materials.

1.1. HISTORICAL PERSPECTIVES

Although the study of materials chemistry is a relatively new entry in both undergraduate and graduate curricula, it has always been an important part of chemistry. An interesting timeline of materials developments from Prehistoric times to the present may be found in Appendix A. By most accounts, Neolithic man (10,000–300 B.C.) was the first to realize that certain materials such as limestone, wood, shells, and clay were most easily shaped into materials used as utensils, tools, and weaponry. Applications for metallic materials date back to the Chalcolithic Age (4,000–1,500 B.C.), where copper was used for a variety of ornamental, functional, and protective applications. This civilization was the first to realize fundamental properties of metals, such as malleability and thermal conductivity. More importantly, Chalcolithic man was the first to practice *top-down* materials synthesis (see later), as they developed techniques to extract copper from oxide ores such as malachite, for subsequent use in various applications.

Metal alloys were first used in the Bronze Age (1,400 B.C.–0 B.C.), where serendipity led to the discovery that doping copper with other compounds drastically altered the physical properties of the material. Artifacts from the Middle East dating back to 3,000 B.C. are found to consist of arsenic-doped copper, due to the wide availability of lautite and domeykite ores, which are rich in both arsenic and copper. However, due to arsenic-related casualties, these alloys were quickly replaced with tin–copper alloys (bronze) that were widely used due to a lower melting point, higher hardness, and lower brittleness relative to their arsenic forerunner.

The Iron Age (1,000 B.C.–1,950 A.D.) first brought about applications for iron-based materials. Since the earth's crust contains significantly more iron than copper (Table 1.1), it is not surprising that bronze was eventually abandoned for materials applications. An iron silicate material, known today as *wrought iron*, was

Table 1.1. Natural Abundance of Elements in the Earth's Crust[a]

Oxygen	46.1%
Silicon	28.2%
Aluminum	8.2%
Iron	5.6%
Calcium	4.2%
Sodium	2.4%
Magnesium	2.3%
Potassium	2.1%
Titanium	0.57%
Hydrogen	0.14%
Copper	0.005%
Total	99.8%

[a] Data taken from Reference [2].

accidentally discovered as a by-product from copper processing. However, this material was softer than bronze, so it was not used extensively until the discovery of steel by the Hittites in 1,400 B.C. The incorporation of this steel technology throughout other parts of the world was likely an artifact of the war-related emigration of the Hittites from the Middle East in 1,200 B.C. The Chinese built upon the existing iron-making technology, by introducing methods to create iron alloys that enabled the molding of iron into desired shapes (*i.e.*, cast iron production). Many other empirical developments were practiced in this time period through other parts of the world; however, it must be stated that it was only in the 18th and 19th century A.D. that scientists began to understand why these diverse procedures were effective.

Figure 1.2 presents the major developmental efforts related to materials science, showing the approximate year that each area was first investigated. Each of these areas is still of current interest, including the design of improved ceramics and glasses, originally discovered by the earliest civilizations. Although building and structural materials such as ceramics, glasses, and asphalt have not dramatically changed since their invention, the world of electronics has undergone rapid changes. Many new architectures for advanced material design are surely yet undiscovered, as scientists are now attempting to mimic the profound structural order existing in living creatures and plant life, which is evident as one delves into their microscopic regimes.

As society moves onto newer technologies, existing materials become obsolete, or their concepts are converted to new applications. A prime example of this is related to phonographs that were commonplace in the early to mid-1900s. However, with the invention of magnetic tape by Marvin Camras in 1947, there was a sharp drop in record usage due to the preferred tape format. The invention of compact disk technology in 1982 has driven the last nail in the coffin of records, which may now only be found in antique shops and garage sales. The needles that were essential to play records no longer have marketability for this application, but have inspired another application at the micro-and nanoscale regime: *atomic force microscopy*, more generally referred to as *scanning probe microscopy* (SPM, see Chapter 7). This materials characterization technique uses a tip, analogous to the record needle that

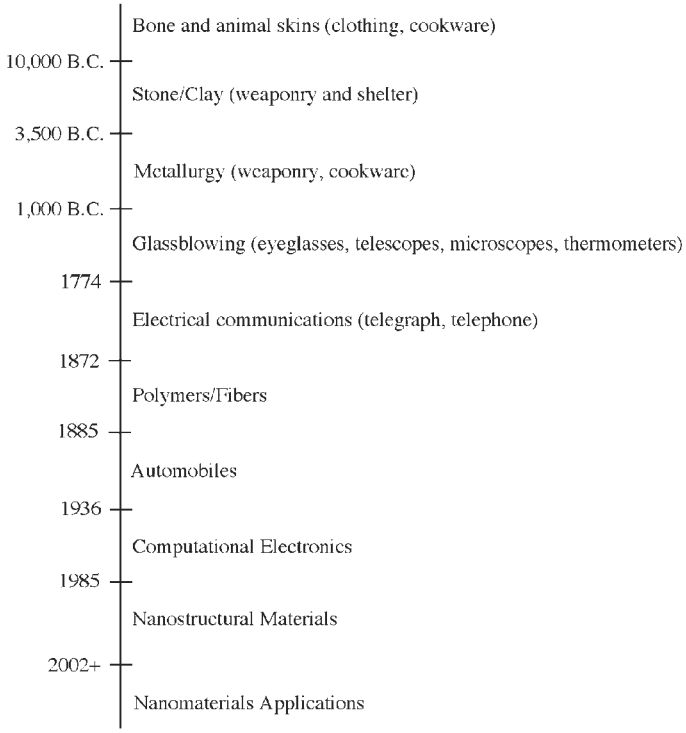


Figure 1.2. Timeline of major developmental efforts related to materials science.

was once used in phonographs, to create images of the surface topology of a sample – even including the controlled placement of individual atoms (Figure 1.3)! Hence, even though the needs and desires of society are constantly changing, the antiquated materials that are being replaced may still be of benefit toward the design of new materials and technology.

The early world of materials discovery consisted solely of empirical observations, without an understanding of the relationship between material structure and properties. Each civilization had specific needs (*e.g.*, materials for shelter, clothing, warfare), and adapted whatever materials were available at the time to address these desires. Although this suitably addressed whatever issues were of societal concern at the time, such a trial-and-error manner of materials design resulted in slow growth.

Interestingly, until the 19th century, the practice of chemistry was viewed as a religion, being derived from alchemical roots that focused on a spiritual quest to make sense of the universe.[3] The alchemists searched for a number of intriguing discoveries including the keys to immortality, a "philosopher's stone" to transform base matter into higher matter, methods to synthesize gold (or transform any other metal into gold), and magic potions to cure diseases. However noble these pursuits were, they remained unaccomplished due to the lack of an underlying chemical theory to guide their experimentation. In addition, their trial-and-error methodology

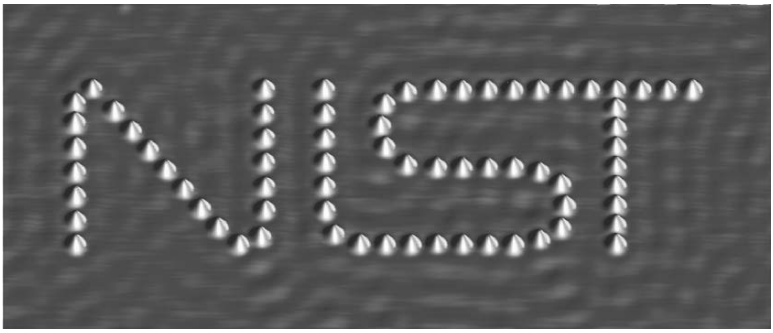


Figure 1.3. A 40-nm wide logo for NIST (National Institute of Standards and Technology), made by the manipulation of Co atoms on a Cu(111) surface. The ripples in the background are due to electrons in the fluid-like layer at the copper surface, bouncing off the cobalt atoms – much like the patterns produced when pebbles are dropped in a pond. Image provided courtesy of J. A. Stroscio and R. J. Celotta (NIST, Gaithersburg, MD).

involved only qualitative characterization, and it was extremely difficult to control the reaction conditions, making it virtually impossible to repeat the exact procedure a number of times.

As a result, from 1,000 B.C. to 1,700 A.D., only a few new substances were discovered which later turned out to be elements such as copper, iron, and mercury. Although this foundation resulted in the development of many experimental techniques of modern chemistry, it is not hard to see that true progress toward new material design may only be accomplished through foresight, based on an intimate understanding of specific relationships between the structure and property of a material. However, as you will see throughout this text, even with such knowledge, many important materials discoveries have been made by accident – the result of an unplanned occurrence during a carefully designed synthesis of an unrelated compound!

1.2. CONSIDERATIONS IN THE DESIGN OF NEW MATERIALS

The development of new materials is governed by the current societal need and availability of resources. However, the adoption of a material depends primarily on its cost, which is even observed by changes in the chemical makeup of currencies through the years. Coins currently comprise worthless ferrous alloys rather than high concentrations of metals such as gold, silver, copper, and nickel that comprised early coins. When a new technology or material is introduced, there is almost always a high price associated with its adoption. For example, consider the cost of computers and plasma televisions when they first became available – worth tens of thousands of dollars!

The market price of a device is governed by the costs of its subunits. Shortly after the invention of germanium-based transistors in the late 1940s, the price of

an individual transistor was approximately US $8–10. However, as germanium was substituted with silicon, and fabrication techniques were improved, the price of these materials has exponentially decreased to its current price of one-millionth of a penny! This has allowed for an unprecedented growth in computational expediency, without a concomitant increase in overall price.

There are two rationales for the synthesis of materials – "top-down" and "bottom-up"; Figure 1.4 illustrates examples of materials synthesized from both approaches. Whereas the transformation of complex natural products into desirable materials occurs primarily via a top-down approach (*e.g.*, gemstones from naturally occurring mineral deposits, etching features on silicon wafers for chip production), the majority of synthetic materials are produced using the bottom-up approach. This latter technique is the easiest to visualize, and is even practiced by children who assemble individual LEGO^{TM} building blocks into more complex architectures. Indeed, the relatively new field of nanotechnology has drastically changed the conception of bottom-up processes, from the historical approach of combining/molding bulk precursor compounds, to the self-assembly of individual atoms and molecules. This capability of being able to manipulate the design of materials from the atomic level will provide an unprecedented control over resultant properties. This will open up possibilities for an unlimited number of future applications, including faster electronic devices, efficient drug-delivery agents, and "green" energy alternatives such as hydrogen-based and fuel cell technologies.

The recent discovery of self-repairing/autonomic healing structural materials is an example of the next generation of "smart materials." Analogous to the way our bodies are created to heal themselves, these materials are designed to undergo spontaneous physical change, with little or no human intervention. Imagine a world where cracks in buildings repair themselves, or automobile bodies actually appear in showroom condition shortly following an accident. Within the next few decades, these materials could be applied to eliminate defective parts on an assembly line, and could even find use in structures that are at present impractical or impossible to repair, such as integrated circuits or implanted medical devices. This is the exciting world that lies ahead of us – as we learn more about how to reproducibly design materials with specific properties from simple atomic/molecular subunits, the applications will only be limited by our imaginations!

1.3. DESIGN OF NEW MATERIALS THROUGH A "CRITICAL THINKING" APPROACH

Although it is essential to use critical thinking to logically solve problems, this method of reasoning is not being taught in most baccalaureate and postbaccalaureate curricula. Unfortunately, the curricular pattern is focused on memorization and standardized-exam preparation. Further, with such a strong influence of television, movies, and the Internet on today's society, the practice of invoking a deliberate flowchart of thought is not generally applicable.

From the MD that must properly diagnose an illness, to the lawyer that must properly follow logic to defend his/her client, critical thinking skills are a necessity for

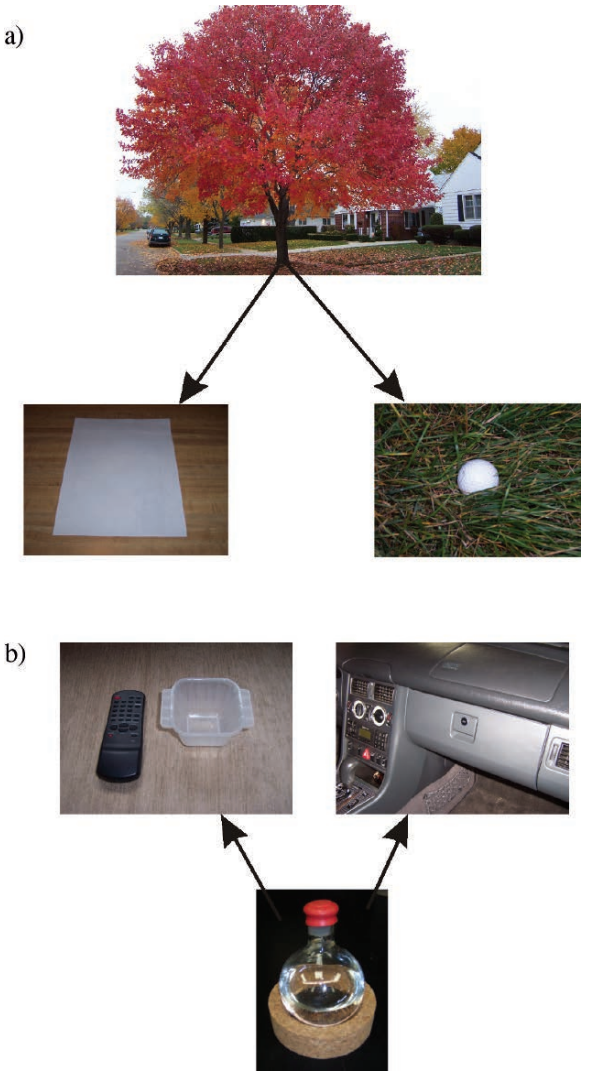


Figure 1.4. Illustrations for the "top-down" and "bottom-up" approach to materials synthesis. (a) The top-down route is often used to transform naturally occurring products into useful materials. Representations shown above include the conversion of wood into paper products, as well as certain golf ball covers.[4] (b) The bottom-up route of materials synthesis is most prevalent. The representation shown above is the fabrication of plastics and vinyl found in common household products and automotive interiors, through polymerization processes starting from simple monomeric compounds (see Chapter 5).

any career path. These skills are also very applicable for the design of new materials – the topic of this textbook. Figure 1.5 illustrates one example of a critical thinking flowchart that could be applied for the design of a new material. Although there are many possibilities for such development, the following are essential components of

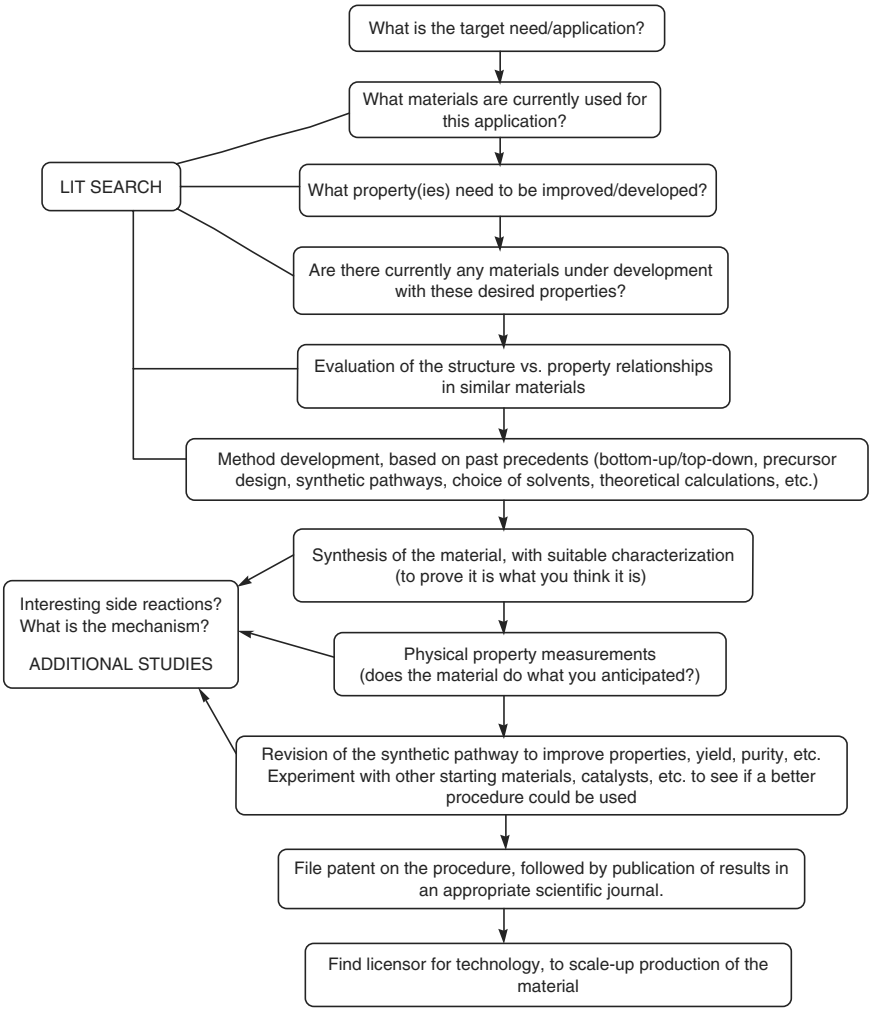


Figure 1.5. An example of a critical thinking scheme for the design of a new material.

any new development:

(i) Define the societal need, and what type of material is being sought. That is, determine the desired properties of the new material.

(ii) Perform a comprehensive literature survey to determine what materials are currently being used. This must be done in order for the new product to successfully compete in the consumer/industrial market. It is essential to search both scientific (*e.g.*, http://www.pubs.acs.org – for all journals published by the American Chemical Society) and patent literature (*e.g.*, http://www.delphion.com), so that extensive research efforts are not wasted by reinventing something that already exists.

(iii) It should be noted that any exercise in critical thinking will result in more questions than originally anticipated. This is illustrated in the flowchart above, where one will look for interesting products/reactions, and begin to think about the mechanism of the process. Such a "first-principle" understanding of the process is essential in order to increase yields of the material, and scale-up the technology for industrial applications.

(iv) After the new technology is protected by filing patents, publication in scientific literature is also important to foster continual investigations and new/improved materials. Top journals such as *Nature*, *Science, The Journal of the American Chemical Society*, *The Chemistry of Materials*, *Advanced Materials, Nano Letters,* and *Small* publish articles every week related to new developments in the most active areas of science. In recent years, the number of materials-related papers has increased exponentially. The continual compounding of knowledge fosters further development related to the synthesis, characterization, and modeling of materials. However, this may only take place as active researchers share their results with their worldwide colleagues.

The objective of *Materials Chemistry* is to provide an overview of the various types of materials, with a focus on synthetic methodologies and relationships between the structure of a material and its overall properties. Each chapter will feature a section entitled "Important Materials Applications" that will describe an interesting current/future application related to a particular class of material. Topics for these sections include fuel cells, depleted uranium, solar cells, "self-healing" plastics, and molecular machines (*e.g.*, artificial muscles).

The following major classes of materials will be investigated:

- Metals
- Semiconductors
- Superconductors
- Glasses and ceramics
- Magnetic materials
- *Soft* materials, such as polymers and composites
- Nanostructural materials
- Thin films

The atomic and molecular architectures of each of these materials will be intimately described, to understand their respective properties. Without such an appreciation of these relationships, we will not be able to progress our civilization with new and improved materials, further improving our way of life.

Any field of chemistry must make use of extensive characterization techniques. For instance, following an organic synthesis, one must use nuclear magnetic resonance (NMR) or spectroscopic techniques to determine if the correct compound has been produced. The world of materials chemistry is no different; characterization techniques must also be used to verify the identity of a material, or to determine why a certain material has failed in order to guide the developments of improving technologies. Hence, characterization techniques will also be provided in this text, which will illustrate the sophisticated techniques that are used to assess the structures/properties of modern materials. Since common techniques such as UV–visible

absorption spectroscopy, atomic absorption/emission spectroscopy, infrared spectroscopy, mass spectrometry, and NMR are covered in a variety of other textbooks,[5] *Materials Chemistry* will focus on the techniques that are frequently used by modern materials chemists, such as:

- Surface/nanoscale analysis (partial list)
  - Photoelectron spectroscopy (PES)
  - Auger electron spectroscopy (AES)
  - Scanning electron microscopy (SEM)
  - Transmission electron microscopy (TEM)
  - Energy-dispersive spectroscopy (EDS/EDX)
  - Secondary ion mass spectrometry (SIMS)
  - Scanning probe microscopy (SPM)
  - X-ray absorption fine structure (XAFS)
  - Electron energy-loss spectroscopy (EELS)
- Bulk characterization
  - X-ray diffraction (powder and single crystal)
  - Gel-permeation chromatography (GPC)
  - Thermogravimetric analysis (TGA)
  - Differential scanning calorimetry (DSC)
  - Small-angle X-ray scattering (SAXS)

References and Notes

1 There is often a "grey area" concerning the best definition for small particulate matter. In particular, most structures are automatically referred to as "nanoscale materials," fueled by the popularity of the nanotechnology revolution. However, the most precise use of the "nano" prefix (*e.g.*, nanoparticles) is only for materials with architectural dimensions (*e.g.*, diameters, thicknesses, *etc.*) of less than 100 nm; intermediate dimensions between 100 and 1,000 nm should instead be referred to as "submicron."

2 *CRC Handbook of Chemistry and Physics*, 84th ed., CRC Press: New York, 2004.

3 For a thorough background on alchemy, refer to the following website: http://www.levity.com/alchemy/index.html

4 Balata golf ball covers may be fabricated from sap extracts of balata/bully trees in South America to produce a thin, resilient material. It is worth mentioning that balata-based materials are now produced artificially.

5 Most sophomore organic chemistry undergraduate textbooks provide a thorough coverage of these techniques, including the interpretation of example spectra. For more specialized books on solid-state NMR and MALDI-MS, see (respectively): Duer, M. J. *Introduction to Solid-State NMR Spectroscopy*, Blackwell: New York, 2005 and Pasch, H.; Schrepp, W. *MALDI-TOF Mass Spectrometry of Synthetic Polymers*, Springer: New York, 2003.

Topics for Further Discussion

1. What are the differences between "Materials Chemistry" and "Solid-State Chemistry"?
2. What is meant by "top-down" or "bottom-up" synthetic approaches? Provide applications of each (either man-made or natural materials).
3. Are complex liquids such as crude oil or detergents considered materials? Explain your reasoning.
4. After reviewing Appendix A, is there a relationship between the major materials-related discoveries and societal foci/needs? Elaborate.

5. Current problems/needs in our world are related to drinking water availability, "green" renewable energy sources, increasing computational speeds, and many others. What are some current areas of research focus that are attempting to address these issues? One should search Chemical Abstracts, as well as books and Web sites for this information.
6. Science is rapidly becoming an interdisciplinary. Describe the origin of chemistry disciplines (inorganic, organic, physical), to their current status including multidisciplinary programs.
7. When a new technology is introduced, the consumer price is astronomical. What are the factors that govern when and how much this price will be lowered? Cite specific examples.

Further Reading

1. Brock, W. H. *The Norton History of Chemistry*, W. W. Norton: New York, 1992.
2. Brush, S. G. *The History of Modern Science: A Guide to the Second Scientific Revolution, 1800–1950*, Iowa State University Press: Ames, IA, 1988.
3. *Oxford Companion to the History of Modern Science*, Heilbron, J. L. ed., Oxford University Press: New York, 2003.
4. Mount, E. *Milestones in Science and Technology: The Ready Reference Guide to Discoveries, Inventions, and Facts*, 2nd ed., Oryx Press: Phoenix, 1994.
5. Rojas, P. *Encyclopedia of Computers and Computer History*, Fitzroy Dearborn: Chicago, 2001 (2 volumes).
6. Bierman, A. K.; Assali, R. N. *The Critical Thinking Handbook*, Prentice Hall: New Jersey, 1996.
7. Browne, M. N.; Keeley, S. M. *Asking the Right Questions: A Guide to Critical Thinking*, 5th ed., Prentice Hall: New Jersey, 1998.
8. Paul, R. *Critical Thinking: What Every Person Needs to Survive in a Rapidly Changing World*, Foundation for Critical Thinking, 1993.

CHAPTER 2

SOLID-STATE CHEMISTRY

Of the three states of matter, solids possess the most structural diversity. Whereas gases and liquids consist of discrete molecules that are randomly distributed due to thermal motion, solids consist of molecules, atoms, or ions that are statically positioned. To fully understand the properties of solid materials, one must have a thorough knowledge of the structural interactions between the subunit atoms, ions, and molecules. This chapter will outline the various types of solids, including structural classifications and nomenclature for both crystalline and amorphous solids. The material in this key chapter will set the groundwork for the rest of this textbook, which describes a variety of materials classes.

2.1. AMORPHOUS *VS*. CRYSTALLINE SOLIDS

A solid is a material that retains both its shape and volume over time. If a solid possesses long range, regularly repeating units, it is classified as a *crystalline* material. Crystalline solids are only produced when the atoms, ions, or molecules have an opportunity to organize themselves into regular arrangements, or *lattices*. For example, crystalline minerals found in nature have been formed through many years of extreme temperature and pressure, or slow evaporation processes. Most naturally occurring crystalline solids comprise an agglomeration of individual microcrystalline units; single crystals without significant defects are extremely rare in nature, and require special growth techniques (see p. 24).

If there is no long-range structural order throughout the solid, the material is best described as *amorphous*. Quite often, these materials possess considerable short-range order. However, the lack of long-range translational order (periodicity) separates this class of materials from their crystalline counterparts. Since the majority of studies have been addressed to study crystalline solids relative to their amorphous counterparts, there is a common misconception that most solids are crystalline in nature. In fact, a solid product generated from many chemical reactions will be amorphous by default, unless special procedures are used to facilitate molecular ordering (*i.e.*, crystal formation). Although the crystalline state is more thermodynamically favorable than the disordered state, the formation of amorphous materials is favored in kinetically bound processes (*e.g.*, chemical vapor deposition, sol–gel, solid precipitation, *etc.*).[1]

Table 2.1. Glass Transition Temperatures

Material	Intermolecular bonding	T_g(°C)
SiO_2	Covalent	1,430
Borosilicate glass	Covalent	550
$Pd_{0.4}Ni_{0.4}P_{0.2}$	Metallic	580
BeF_2	Ionic	570
As_2S_3	Covalent	470
Polystyrene	Van der Waal	370
Se_∞	Covalent	310
Poly(vinyl chloride)	Van der Waal	81
Polyethylene	Van der Waal	−30

Some materials featuring extended networks of molecules such as glasses may never exist in the crystalline state. In these solids, the molecules are so entangled or structurally complex that crystallization may not occur as the temperature is slowly decreased. Due to the rigidity of the solid, but proclivity to remain in the amorphous state, these compounds are best referred to as *supercooled liquids*. It was once thought that a slow flow of glass over 100s of years has caused 19th century stained glass windows to have a proportionately thicker base. However, it is now well understood that the glass structure remains in tact unless its threshold transition temperature is exceeded. This parameter is known as the glass transition temperature, T_g, and corresponds to the temperature below which molecules have very little mobility.

Other amorphous solids such as polymers, being rigid and brittle below. T_g, and elastic above it, also exhibit this behavior. Table 2.1 lists the glass transition temperatures of common solid materials. Although most solid-state textbooks deal almost exclusively with crystalline materials, this text will attempt to address both the crystalline and amorphous states, describing the structure/property relationships of major amorphous classes such as polymers and glasses.

2.2. TYPES OF BONDING IN SOLIDS

Every amorphous and crystalline solid possesses certain types of inter- and intramolecular interactions between its subunits that govern its overall properties. Depending on the nature and strength of these interactions, a variety of physical, optical, and electronic properties are observed. As expected, these associations not only govern the behavior of a material in the solid state, but also for the less ordered liquid phase. For example, the hydrogen bonding interactions between neighboring water molecules within an ice lattice are also important in the liquid phase, resulting in high surface tension and finite viscosity. For the gaseous state, intermolecular forces have been overcome, and no longer have an impact on its properties.

2.2.1. Ionic Solids

These solids are characterized by cationic and anionic species that are associated through electrostatic interactions. All purely ionic salts possess crystalline structures,

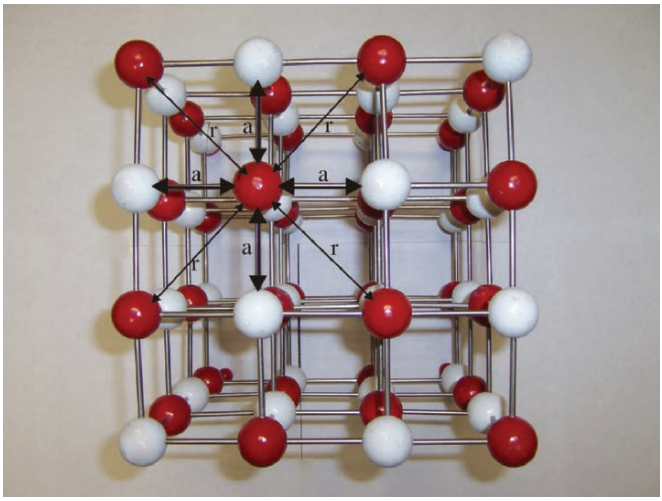


Figure 2.1. Ionic model for sodium chloride. This is a face-centered arrangement of chloride ions (white), with sodium ions occupying the octahedral interstitial sites (red). The attractive electrostatic forces, a, between adjacent Na^+ and Cl^- ions, and repulsive forces, r, between Na^+ ions are indicated.

as exhibited by common Group 1–17 or 2–17 binary salts such as NaCl and $CaCl_2$ (Figure 2.1). The melting points of these solids are extremely high, as very strong electrostatic attractions between counterions must be overcome. Although oppositely charged ions have attractive interactions, like charges repel one another. In the determination of the lattice energy, U, the sizes and charges of the ions are most important (Eq. 1). That is, the lattice energy for MgO would be much greater than BaO, since the ionic bonding is much stronger for the magnesium salt due to its high charge/small size (large charge density). By contrast, the salt MgN does not exist, even though Mg^{3+} and N^{3-} would be very strongly attracted through electrostatic interactions. The ionization energy required to produce the trivalent magnesium ion is too prohibitive.

$$(1) \qquad U = \frac{N M Z_{\text{cation}} Z_{\text{anion}}}{r_0} \left(\frac{e^2}{4\pi\varepsilon_0}\right)\left(1 - \frac{\rho}{r_0}\right),$$

where: N = Avagadro's number (6.02×10^{23} molecules mol^{-1})
$Z_{\text{cation, anion}}$= magnitude of ionic charges;
r_0= average ionic bond length
e= electronic charge (1.602×10^{-19} C)
$4\pi\varepsilon_0$, permittivity of a vacuum (1.11×10^{-10} C^2 J^{-1} m^{-1})
M= the Madelung constant
ρ= the Born exponent

The Madelung constant and Born exponent appearing in Eq. 1 are related to the specific arrangement of ions in the crystal lattice. The Madelung constant may be considered as a decreasing series, which takes into account the repulsions among

ions of similar charge, as well as attractions among oppositely charged ions. For example, in the NaCl lattice illustrated in Figure 2.1, each sodium or chloride ion is surrounded by six ions of opposite charge, which corresponds to a large attractive force. However, farther away there are 12 ions of the same charge that results in a weaker repulsive interaction. As one considers all ions throughout the infinite crystal lattice, the number of possible interactions will increase exponentially, but the magnitudes of these forces diminishes to zero.

Ionic solids are only soluble in extremely polar solvents, due to dipole–dipole interactions between component ions and the solvent. Since the lattice energy of the crystal must be overcome in this process, the solvation of the ions (*i.e.*, formation of $[(H_2O)_nNa]^+$) represents a significant exothermic process that is the driving force for this to occur.

2.2.2. Metallic Solids

Metallic solids are characterized by physical properties such as high thermal and electrical conductivities, malleability, and ductility (*i.e.*, able to be drawn into a thin wire). Chemically, metals tend to have low ionization energies that often result in metals being easily oxidized by the surrounding environmental conditions. This explains why metals are found in nature as complex geological formations of oxides, sulfates, silicates, aluminates, *etc*. It should be noted that metals or alloys may also exist as liquids. Mercury represents the only example of a pure metal that exists as a liquid at STP. The liquid state of Hg is a consequence of the electronic configurations of its individual atoms. The 6s valence electrons are shielded from the nuclear charge by a filled shell of 4f electrons. This shielding causes the effective nuclear charge (Z_{eff}) to be higher for these electrons, resulting in less sharing/delocalization of valence electrons relative to other metals. Further, relativistic contraction of the 6s orbital causes these electrons to be situated closer to the nucleus, making them less available to share with neighboring Hg atoms.[2] In fact, mercury is the only metal that does not form diatomic molecules in the gas phase. Energetically, the individual atoms do not pack into a solid lattice since the lattice energy does not compensate for the energy required to remove electrons from the valence shell.

Metallic bonding

The bonding in metals is best described as a close-packed array of atoms, with valence electrons delocalized throughout the extended structure. This close chemical association among neighboring atoms in the solid gives rise to physical properties such as high melting points and malleability. The nondirectional bonding in metals allows for two modes of deformation to occur when a metal is bent. Either the atomic spacing between neighboring metal atoms in the crystal lattice may change (*elastic deformation*), or planes of metal atoms may slide past one another (*plastic deformation*). Whereas elastic deformation results in a material with "positional memory" (*e.g.*, springs), plastic deformation results in a material that stays malformed.

Chemists are familiar with the traditional linear combination of atomic orbitals–molecular orbital (LCAO–MO) diagrams for diatomic metals such as Li or Na. The

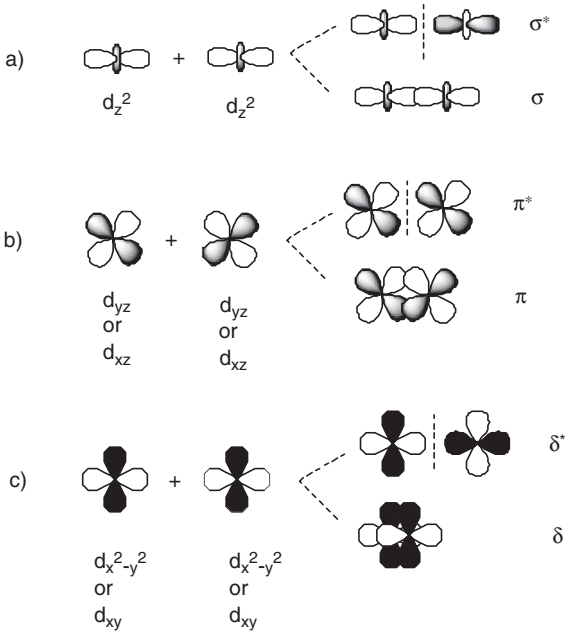


Figure 2.2. Illustration of d-orbital overlap between adjacent metal atoms in an extended metallic network.

overlap for metals containing d orbitals (Figure 2.2) is more complex than those involving simple s and p orbitals. In the LCAO–MO bonding theory, there must be the same number of molecular orbitals formed as the number of atoms that are combined (*e.g.*, there are two bonding and two antibonding MOs formed when four atoms interact covalently). As a real-world example, in 63.5 g (1 mol) of copper, there will be an Avagadro's number of copper atoms, corresponding to 6.022×10^{23} closely spaced molecular orbitals!

As the number of atoms increases to infinity, as within a crystal lattice, the $\Delta E \approx 0$ between energy levels within bonding and antibonding regions. This is due to more interactions among neighboring atoms in a crystal lattice, relative to less static atoms in gas or liquid phase molecules. Figure 2.3 shows the energy levels for a homonuclear diatomic Ti_2 molecule; as the number of titanium atoms increases to infinity, the description of the energy levels in solids is best described as bands, rather than discrete MO energy levels. For metals, there is no gap between the highest occupied molecular orbitals (HOMOs) and lowest unoccupied molecular orbitals (LUMOs). The electron-occupied band is known as the valence band, whereas the unfilled band is referred to as the conduction band. The energy gap between these levels is known as the bandgap, E_g. Although $E_g < 0.01$ eV (*ca.* < 1kJ mol^{-1}) for metals and alloys, the values for semiconductors and insulators are on the order of 190 kJ mol^{-1} and >290 kJ mol^{-1}, respectively. Specific details of the decreased electrical conductivity of these other materials will be presented in subsequent chapters (*e.g.*, semiconductors in Chapter 4).

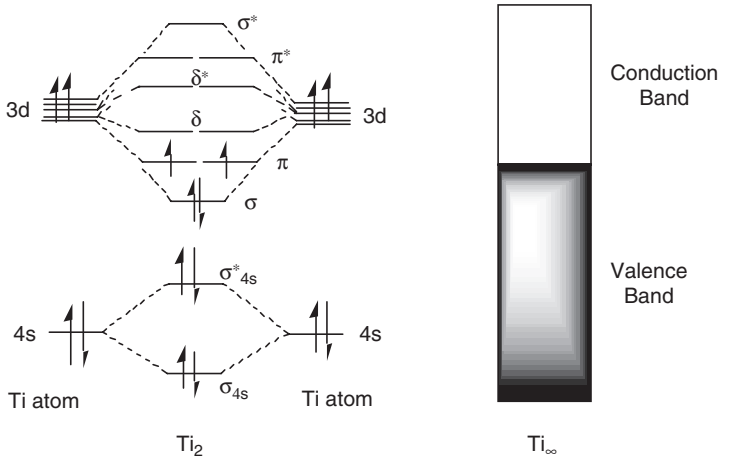


Figure 2.3. Band diagram for titanium metal, illustrating the continuum between valence and conduction bands (*i.e.*, no bandgap). The Fermi level corresponds to the highest occupied energy state at absolute zero (think of pouring coffee (electrons) into a mug (valence band) – the top of the fluid level represents the Fermi level/energy).

Since the delocalization of electrons occurs more readily for valence electrons farther from the nucleus (experiencing a lesser Z_{eff}), "metallic character" increases going down a Group of the Periodic Table. Perhaps the best example of this phenomenon is observed for the Group 14 congeners. As you move from carbon to lead, the elemental properties vary from insulating to metallic, through a transitional range of semiconducting behavior for Si and Ge.

Although metals are mostly characterized by crystalline structures, amorphous alloys may also be produced, known as metallic glasses. A recent example of such a material is the multicomponent alloy $Zr_{41}Ti_{14}Ni_{12}Cu_{10}Be_{23}$.[3] These materials combine properties of both plastics and metals, and are currently used within electric transformers, armor-piercing projectiles, and even sports equipment. The media have recently been focused on this latter application, for LiquidMetal^{TM} golf clubs, tennis racquets, and baseball bats. Unlike window glass, metallic glass is not brittle. Many traditional metals are relatively easy to deform, or bend permanently out of shape, because their crystal lattices are riddled with defects. A metallic glass, in contrast, will spring back to its original shape much more readily.

Amorphous metallic materials may be produced through a variety of procedures. Pure individual metal powders may be melted at high temperatures in an arc furnace. Depending on the composition of the melt, supercooling may be possible, resulting in a vitreous solid rather than a crystalline form. Although facile glass-forming solids such as B_2O_3 will form amorphous solids even upon relatively slow cooling (*e.g.*, $1\,\text{K s}^{-1}$), metals generally require very high rates ($>10^6\,\text{K s}^{-1}$) to prevent crystal growth. These high rates are accomplished by placing a material with high thermal conductivity (*e.g.*, copper) in contact with a molten metal or alloy. This method is

referred to as "melt spinning" or "melt extraction," and results in metallic ribbons up to 15 cm wide and 30 μm thick.

Another common procedure uses a vapor deposition technique to form amorphous metallic thin films (see Chapter 4). Upon thermal annealing, the irregularly deposited atoms in the film have an opportunity to assemble into a crystalline array. A procedure referred to as *ball-milling* may also be used to create amorphous metal alloy powders. This method uses a mixture of crystalline powders that is vigorously agitated with a stainless steel ball within a round vessel. This results in regions of high pressure that cause local melting of the crystalline powders, breaking apart metallic bonds, and facilitating atomic diffusion along preferential crystallite interfaces.

2.2.3. Molecular Solids

This class of solids features discrete molecules that are held together by rather weak intermolecular forces such as dipole–dipole, London Dispersion, and hydrogen bonding. Since these forces are much weaker than ionic or metallic bonding interactions, molecular solids are usually characterized by low melting points. Examples include dry ice (CO_2), ice (H_2O), solid methane (CH_4), sugar (comprising various arrangements/conformations of $C_6H_{12}O_6$ molecules), and polymers. For polymeric materials, the melting points vary significantly depending on the nature of interactions among the polymer subunits. As Chapter 5 will delineate, these interactions also greatly affect many other physical properties of these materials.

Molecular solids may exhibit either crystalline or amorphous structures, depending on the complexity of the individual molecules comprising the bulk material. As with all solids, the more complex the subunits are, the harder it is for them to organize themselves in a repeatable fashion, resulting in an amorphous structure. Unlike purely ionic solids, molecular compounds may be soluble in either nonpolar or polar solvents, as long as the solvent polarity between solute and solvent is matched ("like dissolves like").

Both dipole–dipole and London Dispersion forces are subclasses of van der Waal interactions. When two polar molecules approach one another, a natural attraction known as dipole–dipole forces is created between oppositely charged ends. The relative intensity of dipole–dipole forces may be represented by Eq. 2:

$$\text{dipole–dipole} = \frac{\mu^4}{kTR^6}, \tag{2}$$

where μ is the molecular dipole moment (Debyes); R, the average distance of separation (Å); T, the temperature (K); and k is the Boltzmann constant (1.38065×10^{-23} J K^{-1}).

In addition to the mutual attraction between polar molecules, there may also be an interaction between the solute molecules and liquid or gaseous solvent. In highly polar solvents such as water or alcohols, a dense shell of solvent molecules will surround the polar molecules. Although this solute/solvent interaction assists in the solubility of the molecules in the solvent, the dipole–dipole interactions between individual molecules are suppressed.

In contrast to dipole–dipole forces, London Dispersion interactions are much weaker in nature since they involve nonpolar molecules that do not possess permanent dipole moments. The only modes for molecular attraction are through *polarization* of electrons, which leads to the creation of small dipole–dipole interactions and mutual attractive forces. Since electron polarization occurs much more readily for electrons farther from the nucleus, this effect is more pronounced for molecules that are larger with a greater number of electrons, especially positioned on atoms with a high atomic number, consisting of more diffuse orbitals. These "induced dipole" forces are responsible for the liquefaction of gases such as He and Ar at low temperatures and pressures. The relative strength of London Dispersion forces is described by Eq. 3:

$$\text{(3)} \qquad \text{London Dispersion} = \frac{I\alpha^2}{R^6},$$

where I is the ionization potential of the molecule; and α is the polarizability of the molecules.

If both polar and nonpolar molecules are present, a dipole-induced dipole interaction may occur. For this situation, the strength of association may be represented by Eq. 4, which is dependent on both the dipole moment of the polar molecule, and the polarizability of the nonpolar component. Once again, this relation does not include the interactions between the polar molecule and solvent molecules:

$$\text{(4)} \qquad \text{Dipole–induced dipole} = \frac{\mu\alpha^2}{R^6}.$$

Hydrogen bonding may be considered a special case of dipole–dipole forces, where there exist relatively strong interactions between extremely polar molecules. This interaction is often designated by $A - H \cdots B$, where the hydrogen bond is formed between a Lewis basic group (B) and the hydrogen covalently bonded to an electronegative group (A). In general, the magnitudes of these interactions (*ca.* 12–30 kJ mol^{-1}) are much less than a covalent bond. However, the linear F–H–F anion present in concentrated hydrofluoric acid has a bond energy of *ca.* 50 kJ mol^{-1}, representing the strongest hydrogen bond ever discovered (Figure 2.4). The degree

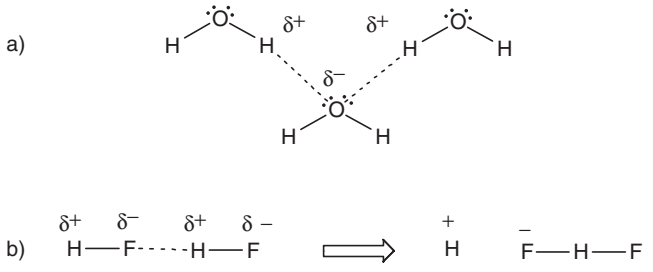


Figure 2.4. Examples of hydrogen bonding in (a) water/ice and (b) the HF_2^- ion in liquid hydrogen fluoride. As with all dipole–dipole interactions, the intermolecular attraction may occur through simplistic linear or more complex modes.

of hydrogen bonding has an influence on many physical properties of a compound such as melting and boiling points, dielectric constants, vapor pressure, thermal conductivity, index of refraction, viscosity, and solubility behavior.

2.2.4. Covalent Network Solids

These solids are characterized by very strong, directional covalent bonds between their constituent atoms. This bonding array generally leads to high melting points and bulk hardness. Due to the arrangement of the atoms comprising these solids, a variety of physical properties may be observed, as evidenced by the very different properties exhibited by the three *allotropes* (*i.e.*, discrete structural forms) of carbon. For instance, diamond is an extremely hard, insulating material that is transparent to light, whereas graphite is a soft, black solid that is capable of conducting electricity along the graphitic layers of the extended solid. Buckminsterfullerene (C_{60}) is very different from either of these carbon forms, being soluble in aromatic solvents, and thereby capable of undergoing chemical reactions. Other examples of covalent network solids are quartz $(SiO_2)_x$, $(BN)_x$, $(ZnS)_x$, $(HgS)_x$, and the two allotropes of selenium – grey (Se_∞) and red $(Se_8)_x$.

2.3. THE CRYSTALLINE STATE

Single crystals comprise an infinite array of ions, atoms, or molecules, known as a *crystal lattice*. The strength of the interactions between the species comprising the crystal is known as the *lattice energy*, and is dependent on the nature and degree of interactions between adjacent species. For example, the extremely high melting points of salts are directly associated with the strength of the ionic bonds between adjacent ions. For molecular species, it is the degree of intermolecular interactions such as van der Waal and hydrogen bonding forces that controls the lattice energy. Ionic and covalent crystals have similar lattice energies (*ca.* 700–900 kJ mol^{-1}), followed by metallic crystals (*ca.* 400–500 kJ mol^{-1}). By contrast, molecular crystals such as solid carbon dioxide are much more readily broken apart (*ca.* 5–20 kJ mol^{-1}); a consequence of the weak Van der Waals interactions consisting between the discrete molecules that comprise the lattice.

The ions, molecules, or atoms pack in an arrangement that minimizes the total free energy of the crystal lattice. For ionic crystals, there is an overall balance of charge among all ions throughout the lattice. Nonionic crystals exhibit a greater variety of packing interactions between constituent molecules. One of the most influential forces found in these lattices is hydrogen bonding. The molecules will pack in such a manner to balance the number of hydrogen bond donor and acceptor groups. Often, a residual polar solvent, capable of participating in hydrogen bonding, will play an important role in the observed packing arrangement. Depending on the polarity of the encapsulated solvent, a variety of arrangements of molecules will be observed in the crystal lattice, with hydrophobic and hydrophilic groups being preferentially aligned with respect to each other and the solvent. The arrangements of molecules throughout the crystal may be observed empirically. While X-ray diffraction is used

to determine the locations of most atoms, neutron diffraction is more diagnostic for locating the positions of hydrogen atoms. Diffraction techniques will be discussed in more detail in Chapter 7, dealing with materials characterization.

Depending on how strongly a solvent is contained within the crystal lattice, sometimes the encapsulated solvent is lost, an occurrence referred to as *efflorescence*. By contrast, if the solid contains ions with a high charge density (high charge/size ratio) and is soluble in water, the crystals will readily adsorb water from the atmosphere and may even be transformed to a solution. An example of such a *deliquescent* crystal is calcium chloride, which is employed as a dehydrating agent for removal of moisture from a flow of inert gases.

The overall shape or form of a crystal is known as the *morphology*. Often, there is more than one crystalline form of the same substance. Each form is known as a *polymorph*, differing in both the arrangement of constituents as well as unit cell dimensions. Although polymorphs differ in both the shape and size of the unit cell, most compounds may exhibit this behavior, under appropriate experimental conditions. Common reasons for a varying crystal structure are similar ionic ratios for anions and cations in ionic crystals, or variations in temperature or pressure during crystal growth. These latter effects alter the amount of disorder within the crystal lattice, allowing for the migration of atoms/ions/molecules into lattice positions that are thermodynamically disfavored at lower temperatures and/or pressures.

If the solid is an element, polymorphs are known as *allotropes*. One of the best-known examples for elemental polymorphism (allotropy) is observed for Group 16 elements (Chalcogens). For instance, sulfur may exist in the common yellow powder form, consisting of arrangements of 8-membered rings. However, at higher temperatures a polymeric substance is formed that may be quenched to yield infinite chains of disordered sulfur atoms (*catenasulfur*, or "plastic sulfur"). The other Group 16 congeners also possess this structural diversity, with the relative thermodynamic stability of a particular allotrope being governed by the structure with the lowest overall free energy. Most often, the energy for the interconversion between polymorphs is small, resulting in phase changes that occur after only moderate changes in temperature or pressure. In general, exposing a crystal to an applied pressure forces neighboring atoms closer together, causing a decrease in the volume of the unit cell, and an increase in the coordination number of individual atoms. For instance, silicon is transformed from a 4-coordinate polymorph at ambient pressure to the 12-coordinate "SI VII" phase at pressures in excess of 50 GPa.

If the chemical contents of a polymorph are different than other forms, it is designated as a *pseudopolymorph*.[4] Most often this occurs due to the presence of differing amounts of solvent, and may alter physical properties of the crystals such as melting points and solubilities. Polymorphism and pseudopolymorphism may be observed when different experimental conditions are used for synthesis. For example, if crystals are grown by sublimation, changing the temperature will often yield different crystal structures, possibly even metastable phases that are kinetically favored.

A subclass of polymorphism known as *polytypism* is found for one-dimensional close-packed and layered structures such as SiC, CdI_2, GaSe, micas and clay

minerals such as kaolins. These crystal structures exhibit differing three-dimensional lattices that vary depending on the stacking order of the two-dimensional sheets that comprise the crystal, known as modular layers.

When different compounds yield almost identical crystals, the forms are referred to as *isomorphs*. The word "almost" is indicated here, as isomorphs are not exactly the same; although the arrangement of atoms/ions in the lattices are identical, one or more of the atoms in the lattice have been replaced with another component. For example, alums of the general formula $(M)_2(SO_4) \cdot (M)_2(SO_4)_3 \cdot 24H_2O$ may crystallize as isomorphs where one of the monovalent or trivalent metals is substituted with another.

The conversion between polymorphs may be observed by differential scanning calorimetry (DSC), which shows peaks corresponding to endothermic or exothermic events associated with structural changes. Since there is an activation energy that must be overcome, an increase in temperature often accompanies such a conversion. The rate of a polymorphic transition depends on the mobility of atoms in the solid state. When steel is tempered, it is heated to high temperatures and then rapidly cooled to temperatures far below its melting point. This low temperature is necessary to slow down the rate of transition to the stable phase at room temperature. Sometimes pressure is used to convert one form into another. When this medium is used, a polymorph with higher density will result, due to the relative confinement of lattice species through the externally applied pressure.

2.3.1. Crystal Growth Techniques

Crystal growth involves a phase change from liquid or gas to a solid, such as the precipitation of a solute from solution or the formation of a solid from condensation of a gas. This occurs through two processes, nucleation and growth, being favored by using supersaturated solutions and/or temperature gradients. When several molecules in the gas phase or in solution approach each other in appropriate orientations, they form a submicroscopic nucleus upon which additional molecules may adsorb *en route* toward an ordered extended crystal structure. The probability that a crystal will form depends on the nature and concentration of the solute (*i.e.*, the distance between solutes), as well as solvent conditions such as temperature, pH, ionic strength, viscosity, polarity, *etc*. To grow single crystals suitable for X-ray diffraction analysis, relatively few nuclei should be formed rather than multiple sites of nucleation that will yield microcrystalline solids.

High quality crystals may only be obtained when the rate of deposition onto a nucleation site is kept at a rate sufficiently low to allow oriented growth. A high growth rate may lead to defects in the crystal, forming multibranched or dendritic crystallites through rapid growth in too many directions. As molecules in the gas-phase or solvent interact with the surface of the growing crystal, they may or may not be preferentially adsorbed. That is, a nucleation site that contains steps, ledges, or surface depressions is able to provide more efficient crystal growth due to the prolonged interaction of suspended molecules with a greater surface area.

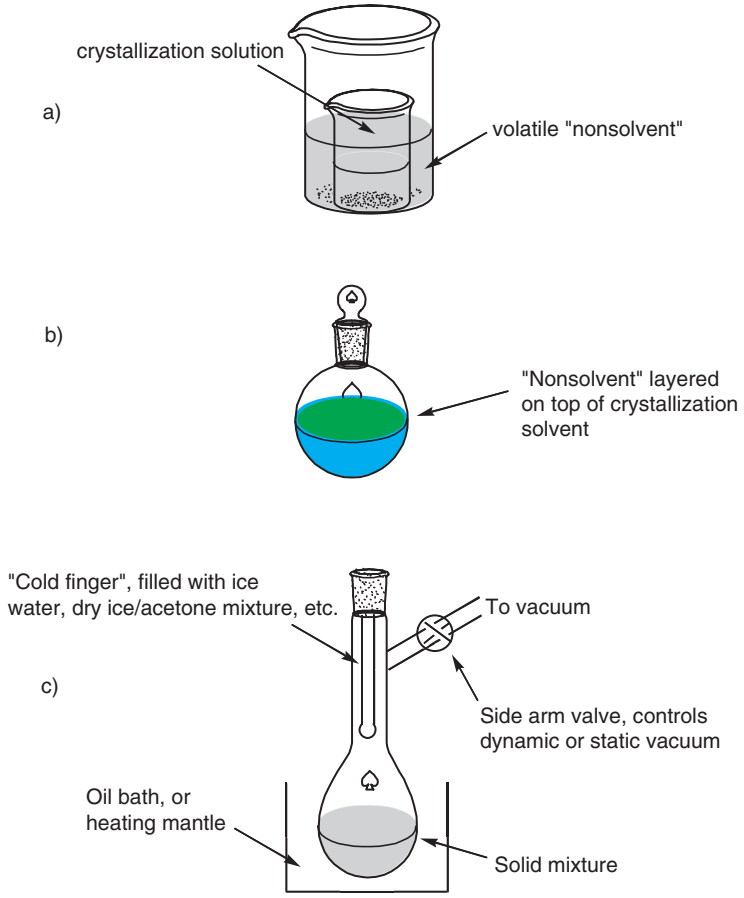


Figure 2.5. Schematic of three common methods used to grow single crystals. Shown is (a) *diffusion* where vapors from a volatile "nonsolvent" meet the crystallization solvent; (b) *interfacial* where the nonsolvent is layered on top of the crystallization solvent; (c) *sublimation* where the solid mixture is heated and the vapors form crystallites on the surface of a cold finger. For this latter technique, the crystallization flask may be opened to vacuum throughout the sublimation process (dynamic vacuum), or closed after maintaining initial vacuum to allow slower crystal growth (static vacuum).

Experimentally, the successful growth of single crystals on the order of 0.01–0.1 mm^2 is not trivial, and has long been considered as a "black art"! Figure 2.5 illustrates common techniques that may be applied for crystal growth via sublimation or from solution. Perhaps the most important starting point is with a solution that is filtered to remove most suspended nuclei. For air-sensitive solutions, this requires careful manipulation using filtering cannulas and Schlenk techniques. Most of the solvent is then removed to create a nearly supersaturated solution, and then left undisturbed. Another method that is used to grow single crystals from saturated solutions consists of layering a "nonsolvent" onto the top of the saturated solution.

Since the compound of interest is not soluble in the layered nonsolvent, crystal formation may begin at the interfacial region. If the nonsolvent is volatile, vapor diffusion may provide another route for the growth of crystals.

Depending on the nature of the suspended molecules, crystal formation may begin immediately, or may even take months to occur. Many organometallic chemists have been surprised to find large crystals at the bottom of flasks placed in the back of the freezer, after months of observation and concluding that no crystals would ever be realized. Sometimes, fortuitous crystal growth may also be realized from unexpected sources. Crystals may be formed from the incorporation of impurities such as dust or vacuum grease, or from surface scratches on the inside walls of the flask. Surprisingly, NMR tubes are notorious for the formation of large crystals, discovered only as the tubes are about to be cleaned! Quite often, chemists set these tubes aside for weeks after the analysis, creating an undisturbed environment for crystal growth. NMR tubes are long and narrow, suppressing convection currents, and solvents very slowly evaporate through the low-permeable cap. Hence, the overall take-home message for crystal growth is to exercise patience; in the process of impatiently checking for crystal growth, additional nucleation sites are often introduced, resulting in the formation of small crystals. Fortunately, many institutions now possess CCD X-ray diffractometers that allow for enough data to be obtained from even microcrystalline solids, in a fraction of the time required for older 4-circle instruments.

Although much crystallization from a solution is performed at low temperatures, crystals may also be formed from molten solids. For example, the Czochralski method for purification of silicon uses a seed crystal on the surface of the melt maintained slightly above its melting point. As the crystal is slowly pulled from the melt, there is extension of the crystal due to preferred alignment of other silicon atoms present in the melt (see Chapter 4). The empirical use of seed crystals is frequently used for crystal growth at all temperature regimes. Scratching the inside of the flask, or even using boiling stones has been successful at inducing crystal formation, through the introduction of sites for nucleation. It must be noted that pH is often not an important factor for inorganic or organic crystals; however, for protein crystallization, which is beyond the scope of this book, many such atypical conditions must also be considered.

2.3.2. The Unit Cell

The smallest repeating unit of an extended crystal lattice is known as the *unit cell*, which governs the symmetry and structure of the overall bulk crystal. Oftentimes, it is possible to define a number of possible repeat units for a crystal lattice (Figure 2.6). The proper selection of a unit cell represents the smallest repeatable unit that possesses the same symmetry elements of the bulk crystal, and if translated along the x, y, and z axes, will generate the entire extended crystal lattice.

Figure 2.7 provides a schematic of a general structure for a unit cell. It is convenient to describe these units as having three vectors ($\vec{a}$, $\vec{b}$, and $\vec{c}$) that may or may not be aligned along the Cartesian axes, depending on the values of unit cell angles.

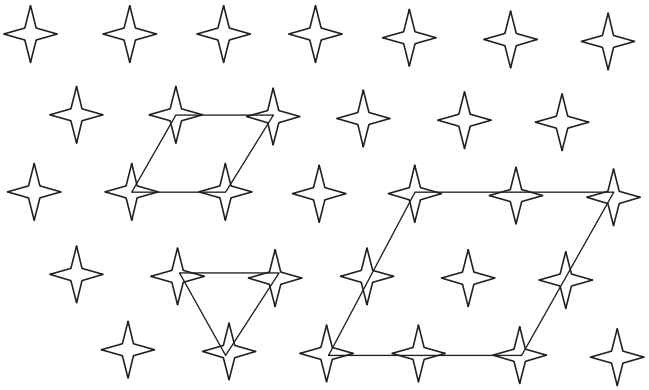


Figure 2.6. Representation of some unit cell selections from a two-dimension lattice arrangement. The triangle and smaller rhombus selections are known as "primitive" unit cells, as each contains one object per unit cell ($3 \times 1/3$ for the triangle, $4 \times 1/4$ for the rhombus). By contrast, there are four objects per unit cell for the larger rhombus ($4 \times 1/4$ for corner objects; $4 \times 1/2$ for those on the edges; one in the center). When each of these selections is translated along the two-dimensional axes, they reproduce the positions of all objects in the lattice.

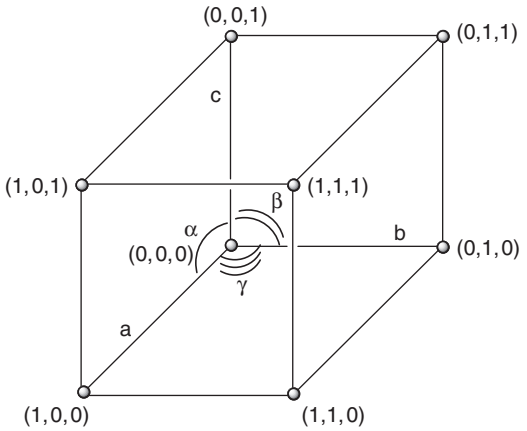


Figure 2.7. Schematic of a unit cell. The angles and side lengths shown above are not representative of all unit cells; seven types of cell dimensions are possible (see text).

Depending on the geometry and volume of the unit cell, there are seven crystal systems that may be generated (Table 2.2).

For simplicity, fractional coordinates are used to describe the lattice positions in terms of crystallographic axes, a, b, and c. For instance, the fractional coordinates are (1/2,1/2,1/2) for an object perfectly in the middle of a unit cell, midway between all three crystallographic axes. To characterize crystallographic planes, integers known as *Miller indices* are used. These numbers are in the format (*hkl*), and correspond to the interception of unit cell vectors at $(a/h, b/k, c/l)$. Figure 2.8 illustrates

Table 2.2. Unit Cell Definitions for the Seven Crystal Systems

Crystal system	Unit cell vector lengths	Unit cell vector angles
Cubic	$\lvert\vec{a}\rvert = \lvert\vec{b}\rvert = \lvert\vec{c}\rvert$	$\alpha = \beta = \gamma = 90°$
Tetragonal	$\lvert\vec{a}\rvert = \lvert\vec{b}\rvert \neq \lvert\vec{c}\rvert$	$\alpha = \beta = \gamma = 90°$
Orthorhombic	$\lvert\vec{a}\rvert \neq \lvert\vec{b}\rvert \neq \lvert\vec{c}\rvert$	$\alpha = \beta = \gamma = 90°$
Trigonal	$\lvert\vec{a}\rvert = \lvert\vec{b}\rvert = \lvert\vec{c}\rvert$	$\alpha = \beta = \gamma \neq 90°$, $\gamma < 120$
Hexagonal	$\lvert\vec{a}\rvert = \lvert\vec{b}\rvert \neq \lvert\vec{c}\rvert$	$\alpha = \beta = 90°, \gamma = 120°$
Monoclinic	$\lvert\vec{a}\rvert \neq \lvert\vec{b}\rvert \neq \lvert\vec{c}\rvert$	$\alpha = \gamma = 90°, \beta \neq 90°$
Triclinic	$\lvert\vec{a}\rvert \neq \lvert\vec{b}\rvert \neq \lvert\vec{c}\rvert$	$\alpha \neq \beta \neq \gamma$

examples of the (011) and (001) plane/face of a crystal; since (*hkl*) intercepts the unit cell at $\{(a, b, c) : (1/h, 1/k, 1/l\}$, a zero indicates that the plane is parallel to the particular axis, with no interception along $\pm\infty$.[5] A Miller index with capped integers such as $(01\bar{1})$ indicates that the crystallographic axis is intercepted in the negative region (interception of the c axis at -1, in this case).

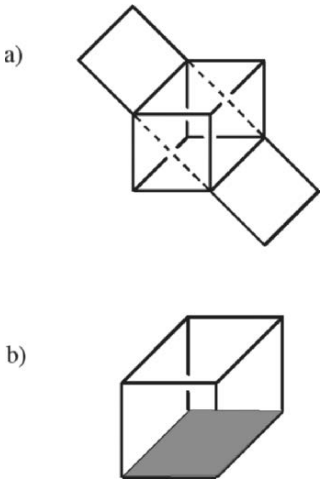


Figure 2.8. (a) Representation of the (011) plane in a unit cell. Based on the Miller indices, the plane does not intersect the a axis, and passes through (0,1,0), and (0,0,1). (b) Illustration of a (001) plane in a unit cell, which does not intersect either the *a* or *b* axes.

2.3.3. Crystal Lattices

Definitions and nomenclature

For crystals comprising ions and metal atoms, the lattice may be thought of as regular packing of hard spheres. In three dimensions, there are only two ways that spheres may pack: either simple cubic or close-packed (Figure 2.9). Whereas the coordination number (*i.e.*, number of nearest neighbors) of individual species for cubic packing is 6, the coordination number for close-packing is 12.

The Periodic Table shown in Figure 2.10 demonstrates that the close-packing motifs are the most prevalent for natural forms of the elements. That is, think about dropping marbles into a large beaker partially filled with water. As the marbles settle at the bottom of the container, they will seek their most stable resting position. These positions are within the voids formed between three spheres (*cf.* B or C sites in Figure 2.9c), rather than on top of the crests of individual spheres. Since a simple cubic arrangement has a significant amount of excess void-space, this would present a much less efficient manner to pack individual spheres. This larger volume of empty space in simple packing would result in a lower density for the solid relative to its close-packing analog. It should also be noted that the sliding of one layer over another would be more preferred for simple cubic relative to close-packing. This corresponds to a greater degree of crystal defects that are possible for these solids.

Although marble migration is only mentioned as an example, this process is not far removed from the crystallization process. In the formation of single crystals, individual ions/atoms/molecules slowly come into contact with one another, and nucleate from thermodynamically favored positions. Sometimes metastable phases may be obtained if one does not allow such preferential migration to occur, through rapid cooling events, for instance.

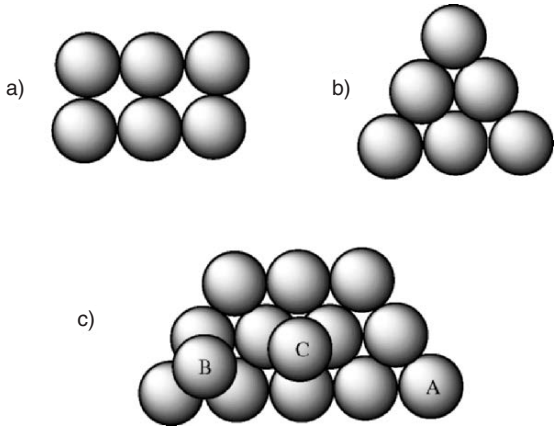


Figure 2.9. Representation of (a) simple cubic packing and (b, c) close-packing. The two sites that may be occupied by subsequent layers in close-packing, B and C sites, are shown. This results in two types of packing: ABABAB… (hexagonal close-packing, hcp), and ABCABC… (cubic close-packing, ccp).

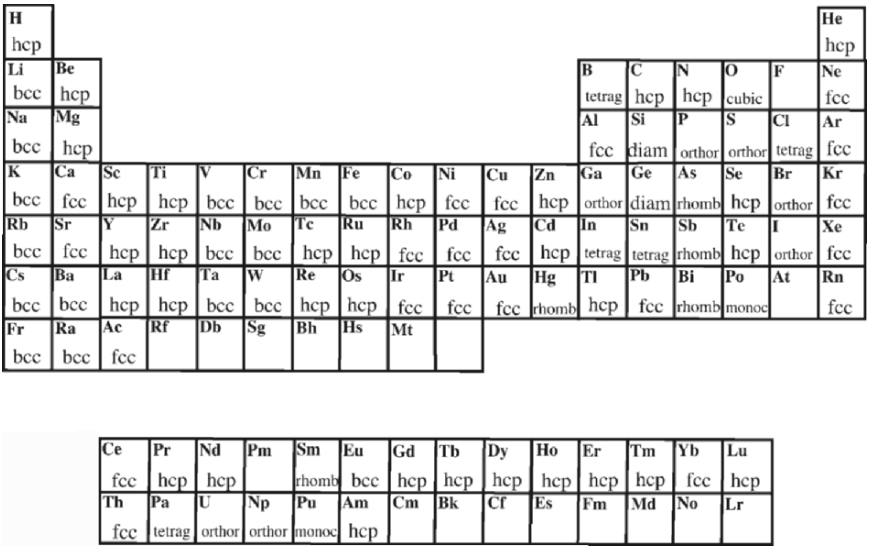


Figure 2.10. Preferred crystal structures of the elements. Shown are the ordinary forms of each element in its solid state (*e.g.*, graphite, α-iron, *etc.*). The following abbreviations are used: fcc, bcc (body-centered cubic), hcp, cubic (simple cubic), diam (diamond or zinc blende), monoc (monoclinic), rhomb (rhombohedral), tetrag (tetragonal), and orthor (orthorhombic).

To aid in the visualization of sphere packing sequences in three-dimensions, consider the following arrangement of something more tangible such as baseballs or oranges. After a few spheres (in A sites) are arranged in a close-packed arrangement (Figure 2.9b), the next layer packs into B sites, formed as voids above three A spheres (Figure 2.9c). For the third layer of spheres, there are two possibilities that may occur. The atoms/ions in the third layer may sit directly above the spheres in the A layer, or may be placed in C sites (Figure 2.9c). If the packing is described by the former case (ABABAB…), the crystal lattice is described as a hexagonal close-packed (hcp) arrangement. By contrast, if the packing scheme occupies all possible sites (ABCABCABC…), then the arrangement is termed cubic close-packed, or face-centered cubic (fcc).

Since hcp crystals have vacant C sites, there are uniaxial channels through the channels that may influence physical properties of these crystals. For example, Mg and Ti that crystallize in hcp arrangements possess extremely low densities, while other hcp metals (*e.g.*, Co) have the possibility for anisotropic magnetic properties.

Although it is easy to visualize a fcc unit cell, with atoms/ions on each corner and on each face of a cube, the ABCABC… packing scheme is more difficult to unravel. Once the unit cell is viewed along the cell corners, perpendicular to the (110) plane, the packing layers are readily observed (Figure 2.11a). When viewed along a vector perpendicular to the packing planes, the atoms/ions occupying the B and C sites are observed as inverted triangles (Figure 2.11b). Common examples of this packing scheme are diamond, ZnS (zinc blende form), HgS, CuI, AlSb, BeSe, *etc*.

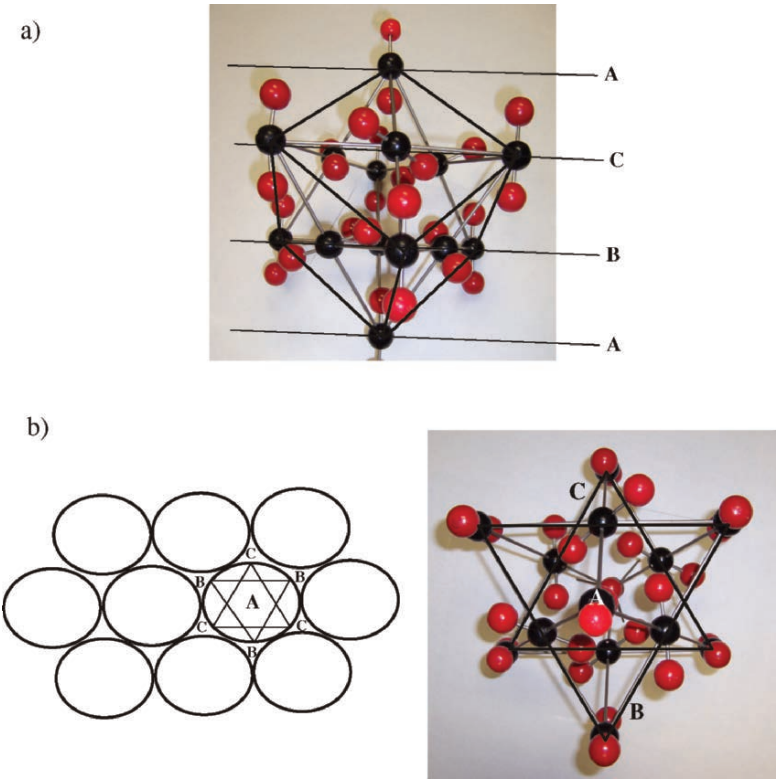


Figure 2.11. Schematic of sphere packing for cubic close-packed (face-centered cubic, fcc). Shown is solid carbon dioxide (dry ice), that consists of individual CO_2 molecules that pack in an fcc lattice. The black carbon atoms shown in (a) illustrate a traditional representation of the lattice, consisting of a cube with atoms on each of the faces. The packing layers are shown from two different views; (a) parallel to the ABC… layers (*i.e.*, (111) planes) and (b) perpendicular to these layers.

Hexagonal close-packing is illustrated in Figure 2.12, showing a unit cell, with translation to form a hexagonal arrangement with a sixfold rotation axis. Using simple geometry, it may be proven that the coordinates of the B sites are $\{(a, b, c) = (2/3, 1/3, 1/2)$ or $(1/3, 2/3, 1/2)\}$. There are two possibilities due to lattice equivalency; however, once one of these is designated for B sites, the other positions are designated as C sites and remain vacant in hcp crystal lattices. Although the packing sequence is different between hcp and fcc lattices, there is a similarity between these close-packed structures. The atoms in the (111) planes of B and C layers shown in Figure 2.11 are arranged in a hexagonal pattern, analogous to the stacking planes for hcp shown in Figure 2.9c.

For hexagonal crystal planes, a slightly different indexing nomenclature is used relative to cubic crystals. To index a plane in the hexagonal system, four axes are used, called Miller–Bravais indices. In addition to both a and b axes, another axis

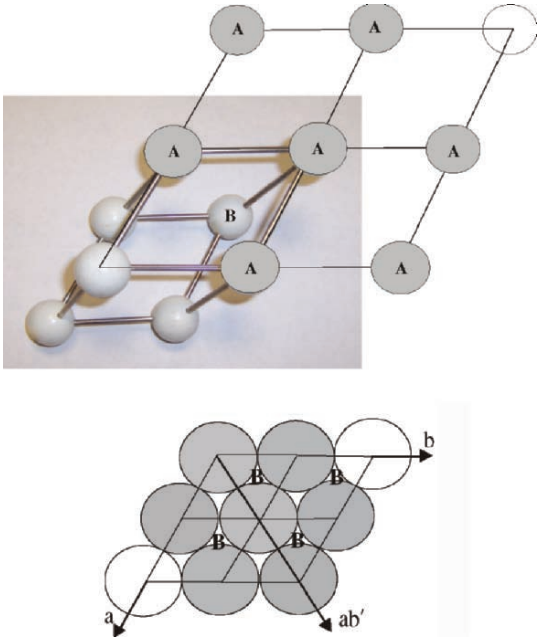


Figure 2.12. Schematic of the packing in an hcp unit cell, showing equivalent close-packed B sites. The hexagon of A atoms is referred to as the basal plane, as this plane marks the top and bottom of a unit cell. The fractional coordinates of each set of B sites are {(a, b, c): (1/3,2/3,1/2) – shown, or (2/3,1/3,1/2)}.

is used, that bisects the γ-angle on the hexagonal *basal plane* (Figure 2.12). This additional axis is designated ab' in Figure 2.12.[6] The use of three axes on the basal plane is required due to the threefold rotation axis present in the lattice. Examples of hexagonal plane indices are shown in Figure 2.13; lattice directions are indexed analogously to cubic crystals, only using a, b, and c crystallographic axes.

Although ccp and hcp arrangements have been shown in detail, there are many other possibilities for the packing of species in a crystal lattice. In 1849, Bravais discovered that there are only 14 possible arrangements of points in three-dimensional space, assuming that each point has an identical environment. Hence, regardless of the identity of the species comprising an infinite crystal lattice, the unit cell must be described by one of these Bravais lattices (Figure 2.14). Other designations that are not listed are not oversights, but may be further simplified. For instance, a "base-centered tetragonal" unit cell does not appear in Figure 2.14, as it may be reduced to a simple (primitive) tetragonal unit cell by repositioning the base-centered atoms to the unit cell corners.

Interstitial crystal lattices

Thus far, we have considered crystal structures consisting of one type of identical atom, ion, or molecule. For crystal lattices consisting of more than one type

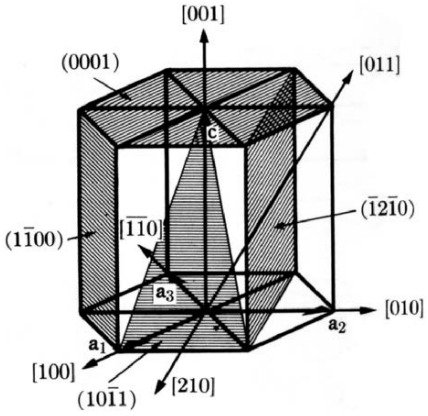


Figure 2.13. Examples of planes (in parentheses) and directions [in brackets] of a hexagonal unit cell. Reproduced with permission from Cullity, B. D. *Elements of X-Ray Diffraction*, 2nd ed., Addison-Wesley: Reading, MA. Copyright 1978 Prentice-Hall.

Table 2.3. Ionic Radii Ratios Corresponding to Interstitial Sites

Radii ratio range (r_+/r_-)	Geometry of interstitial site (coordination number)
<0.225	Trigonal (3)
<0.414	Tetrahedral (4)
<0.732	Octahedral (6)
>0.732	Cubic (8)

of species, the unit cell is best described as one of the 14 Bravais lattices for the larger component, with the smaller species occupying vacant sites within the lattice, known as *interstitial sites*. Based on the number of nearest neighbors immediately surrounding these positions, interstitial sites may exist with coordination numbers of 3 (trigonal), 4 (tetrahedral), 6 (octahedral), and 8 (cubic). An example of a cubic site is a species that is housed within the (1/2,1/2,1/2) position of a unit cell, surrounded by eight nearest neighbors.

For ionic crystals, the preference of a cation to occupy a certain interstitial site is primarily governed by the ionic radius ratio of the cation/anion (r_+/r_-). Since anions are most often larger than cations, this ratio is usually less than 1 (Table 2.3). An exception may be found for unusually large cations with small anions such as CsCl ($r_+/r_- = 1.08$). As the value of this ratio decreases, the size of the anions become significantly larger than the cations, and the cation will prefer to occupy a smaller interstitial site. To rationalize this preference, consider a simple cubic arrangement of bowling balls (*cf.* large anions) with a small golf ball (*cf.* small cation) in the middle. Due to the large size difference, this arrangement would not be stable, as the golf ball would rattle around the cubic "cage" formed by the bowling balls. Rather, a smaller close-packed interstitial site such as trigonal or tetrahedral would best contain the smaller golf ball.

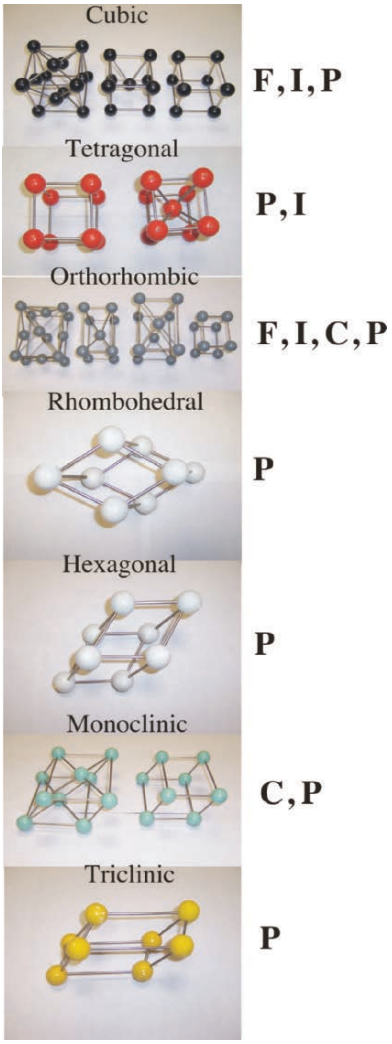


Figure 2.14. Models of the 14 Bravais lattices. The various types of Bravais centering are given the symbols *P* (primitive/simple), *F* (face-centered), *I* (body-centered), and *C* (base-centered). The primitive rhombohedral Bravais lattice is often given its own symbol, *R*, and corresponds to a primitive unit cell possessing trigonal symmetry.

For compound unit cells, it is important to point out the occupation of the atoms/ions occupying the Bravais framework and the other species in interstitial sites. For example, the CsCl structure is best described as consisting of a simple cubic arrangement of Cl^- ions, and a Cs^+ ion in a cubic interstitial site. Alternatively, one could also designate this structure as having Cs^+ ions at the corners of a cube, and Cl^- in the interstitial cubic site. Even though the overall arrangement

of ions in this structure is bcc, such a compound unit cell cannot be assigned to this arrangement since the ions are not equivalent.

It is not immediately apparent why many different values for individual ionic radii appear in reference books. The value for an ionic radius is dependent on its lattice arrangement, and is determined from electron density maps and empirical X-ray diffraction data. Some general trends for cationic radii are:

1. For a given species and charge, the radius increases as the coordination number increases.
2. For a given charge, the radius decreases with increasing effective nuclear charge, Z_{eff}.[7]
3. For a given species, the radius decreases with increasing ionic charge.
4. For a given species and charge, the radius is larger for high-spin (weak field) ions than for low-spin (strong field) ions.

Most inorganic chemistry texts list cut-off values for the r_+/r_- ratios corresponding to the various geometries of interstitial sites (Table 2.3). However, it should also be pointed out that deviations in these predictions are found for many crystals due to covalent bonding character. An example for such a deviation is observed for zinc sulfide (ZnS). The ionic radius ratio for this structure is 0.52, which indicates that the cations should occupy octahedral interstitial sites. However, due to partial covalent bonding character, the anions are closer together than would occur from purely electrostatic attraction. This results in an "effective radius ratio" that is decreased, and a cation preference for tetrahedral sites rather than octahedral.

The crystal structure for this complex lattice (wurtzite structure) is shown in Figure 2.15. This is best described as an hcp lattice of sulfide ions, with zinc ions occupying one-half of the available tetrahedral interstitial sites. For this lattice, there are two units of ZnS per unit cell:

$2S^{2-}$ ions/unit cell $(8(1/8) + 1)$ {at corners (*e.g.*, (0,0,0) and (2/3,1/3,1/2)}
$2Zn^{2+}$ ions/unit cell $(4(1/4)+1)$ { at (0,0,5/8), (1,0,5/8), (0,1,5/8), (1,1,5/8), and (2/3,1/3,7/8)}

It is noteworthy to point out that bonding character for compounds is rarely purely covalent or ionic in nature, especially for inorganic species. Rather, a combination of both bonding modes provides the best description of the electron density for these solids. As you might expect, this will also affect the physical properties of the solid; for instance, the hardness of ZnS is significantly greater than what would be expected for a purely ionic solid.

If the cation in the crystal lattice exhibits a cubic environment (coordination number of 8), the *fluorite* structure is commonly observed (Figure 2.16). Lattices of this variety consist of an fcc arrangement of cations, with all 8 tetrahedral interstitial sites (*e.g.*, (1/4,1/4,1/4), etc.) occupied by the anionic species. Of course, this will only be prevalent when the size of the anion is much smaller than the cation, such as CaF_2. For structures with relatively smaller cations, the anions will form the fcc lattice, with cations situated within the interstitials. Since the relative positions of cations and anions are reversed in the latter case, the *anti* prefix is used, designating the structure as *antifluorite*.

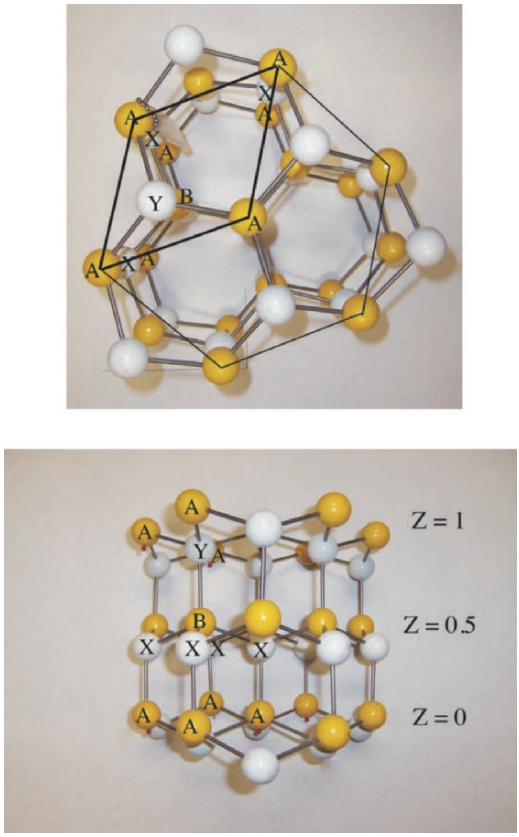


Figure 2.15. Model of the wurtzite (ZnS) crystal structure. The framework is based on an hcp lattice of S^{2-} anions (yellow; the unit cell consists of A and B ions) with zinc ions occupying tetrahedral interstitial sites (white, labeled as X and Y ions).

Metal oxide lattices

The vast majority of catalysts used in heterogeneous catalytic processes are based on mixed metal oxides. There is ongoing interest in the preparation of these catalysts with specific reproducible properties; a challenge that has been possible through increasing knowledge regarding the structure/property relationships of these materials.

There are a number of important crystals that comprise a close-packed array of oxide anions, with cations situated in vacant interstitial sites. Often, there are two or more different types of cations that occupy the vacancies. Some useful iron oxides with this architecture are used in magnetic tapes, containing octahedrally coordinated Fe^{3+} ions (*e.g.*, MFe_2O_4, $MFe_{12}O_{19}$, or $M_2M'_2Fe_{12}O_{22}$, where M, M′ = divalent

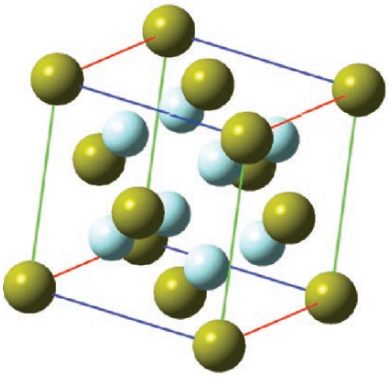


Figure 2.16. Unit cell representation for the fluorite structure of CaF_2.

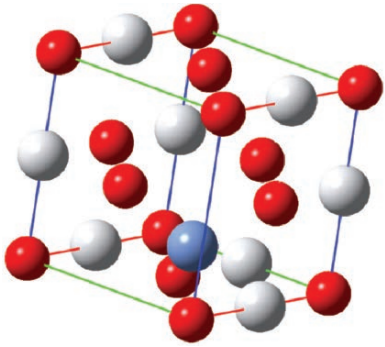


Figure 2.17. Unit cell of a normal spinel lattice for $CrFe_2O_4$.

metals such as Ba^{2+}, Sr^{2+}, Ca^{2+}, Mn^{2+}). These magnetic iron oxides have the *spinel* structure, consisting of a ccp array of oxide ions, with 1/8 of the tetrahedral holes occupied by M^{2+} ions, and 1/2 of the octahedral holes occupied with Fe^{3+} ions (Figure 2.17).

It should be noted that there are three types of spinel lattices:

$$M^{2+}M_2{}^{3+}O_4 - e.g., CrFe_2O_4, CoCo_2O_4$$
$$M^{4+}M_2{}^{2+}O_4 - e.g., TiZn_2O_4, SnCo_2O_4$$
$$M_2^{+}M^{6+}O_4 - e.g., Na_2MoO_4, Ag_2MoO_4$$

In addition to the simple binary spinels, more complex ternary ferrite spinels have also been synthesized, of the general structure:

$$[M_{1-i}{}^{2+}Fe_i{}^{3+}](M_i{}^{2+}Fe_{2-i}{}^{3+})O_4 \quad (5)$$

where the cations in brackets are in tetrahedral sites, and the cations in parentheses are in octahedral sites. In the case of $i = 0$, the normal spinel structure will be obtained. In the other extreme case, $i = 1$, the structure is known as the *inverse spinel* lattice.

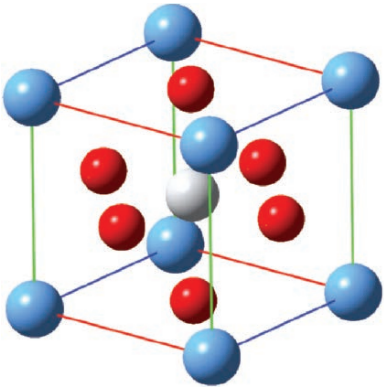


Figure 2.18. Unit cell of the perovskite $BaTiO_3$, illustrating the ABO_3 stoichiometry.

The equilibrium distribution of cations in the spinel lattice depends on the size of the ionic radii, electrostatic energies, and polarization effects. As one would expect, the degree of cation disorder will significantly influence the magnetic properties of magnetic ferrite spinels.[8] One may intentionally prepare spinels containing nonequilibrium cation distributions through rapid quenching techniques of sintered powders. For instance, when ground mixtures of MnO and Al_2O_3 powders are sintered and quenched the resulting $(Mn_xAl_{1-x})_3O_4$ spinel has an oxygen-deficient metal:oxygen ratio of 3:3.7.[9]

The Boltzmann distribution (Eq. 6) relates the dependence of temperature on the amount of cation disorder, where E_i is the activation energy required to exchange interstitial sites of M^{2+} and M^{3+} cations (*i.e.*, between octahedral and tetrahedral sites). Energy values for these interchanges generally fall in the range 0.11–0.14 eV:

$$\frac{i(1+i)}{(1-i)^2} = e^{-E_i/kT}. \tag{6}$$

Another prominent category of mixed metal oxide crystals is the *perovskites*. These oxides have the general formula ABO_3, where A is a cation of larger size than B. These lattices consist of a ccp arrangement of both oxide anions and the larger cation (Figure 2.18). The smaller cation occupies the octahedral hole at the position (1/2,1/2,1/2). The broad range of perovskites is attributed to the fact that around 90% of the metallic elements of the Periodic Table are known to form stable oxides with this structure. It is also possible to partially substitute A and B cations to yield a perovskite of the formula $A_{1-x}A'_xB_{1-x}B'_xO_3$. It should be noted that the perovskite structure is not only obtained for oxides, but also for some nitrides (*e.g.*, Ca_3GeN), halides (*e.g.*, $KMgF_3$), hydrides (*e.g.*, $BaLiH_3$), and carbides (*e.g.*, $MgCNi_3$). Recently, oxynitride perovskites (*e.g.*, $BaTaO_2N$) have received considerable attention due to their potential applications for nontoxic inorganic pigments and photocatalysts.[10] The reduced electronegativity of the nitride ion, relative to

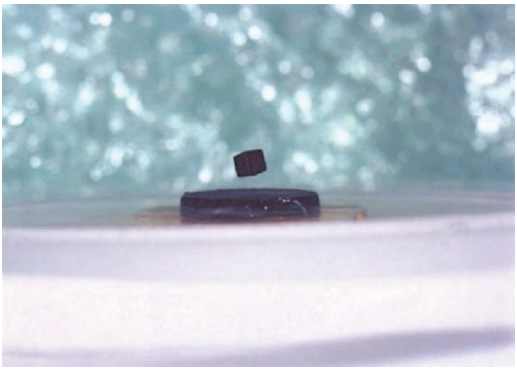


Figure 2.19. Photograph of the Meissner effect for a rare-earth magnet above a sample of YBCO immersed in liquid nitrogen (from http://www.physics.brown.edu/physics/demopages/Demo/em/demo/5G5050.htm). The onset of strong diamagnetism ("superdiamagnetism," as observed by the repulsion of an external magnetic field) is the most reliable method to determine superconductive behavior.

the oxide anion, increases the covalency of the cation–anion bonds thus affecting its overall structure and physical/optical properties.

In addition to catalytic applications, the perovskite backbone is a key component in modern high-temperature superconductive materials. By definition, a superconductor exhibits no resistance to electrical conductivity, and will oppose an external magnetic field, a phenomenon referred to as the Meissner effect (Figure 2.19). Many pure transition metals (*e.g.*, Ti, Zr, Hf, Mo, W, Ru, Os, Ir, Zn, Cd, Hg) and main group metals (*e.g.*, Al, Ga, In, Sn, Pb) exhibit superconductivity, many only when exposed to high-pressure conditions. These materials are referred to as Type I or *soft superconductors*.

Binary and ternary alloys and oxides of these elements, as well as pure V, Nb, Gd, and Tc are referred to as Type II or *high-field superconductors*. In contrast to Type I, these materials exhibit conductive characteristics varying from normal metallic to superconductive, depending on the magnitude of the external magnetic field. It is noteworthy to point out that metals with the highest electrical conductivity (*e.g.*, Cu, Au) do not naturally possess superconductivity. Although this behavior was first discovered in 1911 for supercooled liquid mercury, it was not until 1957 that a theory was developed for this phenomenon.[11]

In order to exhibit superconductive behavior, early Type I and II materials needed to be cooled below a critical temperature (T_c) ranging from 0.015 K (for W) to 23 K (for Nb_3Ge). An intriguing goal of current research is to increase the T_c to room temperature ("high-temperature superconductors," HTS), which would trivialize resistance-free applications such as power grid lines and widespread levitated trains. In 1986, Muller and Bednorz at IBM made an important discovery toward this goal – the first high-temperature superconductor, $La_{2-x}Sr_xCuO_4$ (LSCO), with a critical temperature of 35 K. A year later, the first material with

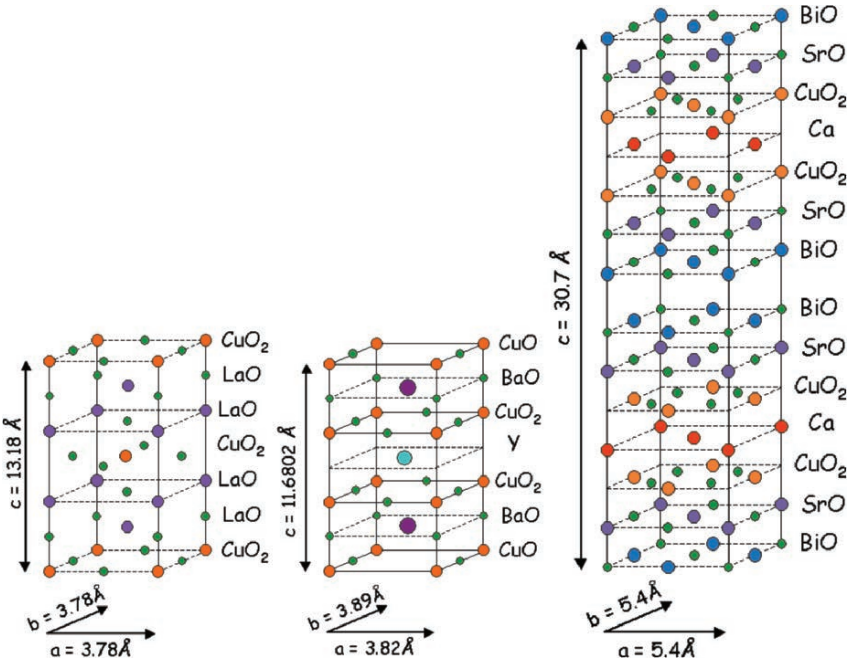


Figure 2.20. Comparison of the crystal structures of (left–right) LSCO, YBCO, and BSCCO superconductors, respectively. Reproduced with permission from Prof. Hoffman's webpage at Harvard University: http://hoffman.physics.harvard.edu/research/SCmaterials.php

a critical point above the boiling point of nitrogen (77 K) was discovered, known as $YBa_2Cu_3O_{7-\delta}$ (YBCO), with a critical point of 92 K. In more recent years, the highest-temperature cuprate based superconductors have been synthesized with a general formula $M_vN_wCa_xCu_yO_z$ (where M = Y, Bi, Tl, or Hg; N = Ba or Sr; $v = 1$ or 2; $w = 2$ or 4; $x = 0, 1,$ or 2; $y = 1, 2,$ or 3; $z = 3, 4, 6, 7, 9, 10,$ or 15). To date, the highest-temperature superconductive materials are thallium (*e.g.*, $TlBa_2Ca_2Cu_3O_9$, $T_c = 133$ K), mercury (*e.g.*, $Hg_{0.8}Tl_{0.2}Ba_2Ca_2Cu_3O_{8.33}$, $T_c = 138$ K), or lead-doped (*e.g.*, $(Hg_{0.75}Pb_{0.15}Tl_{0.1})Ba_2Ca_2Cu_3O_{8+}$, $T_c = 142$ K).[12]

Figure 2.20 shows a comparison between the major stacked cuprates that have played an important role in the ongoing development of a room-temperature superconductor. The presence of three or four metals in the crystal structure is referred to as a ternary or quaternary metal oxide, respectively. The lattices for these compounds are distorted perovskites, where "CuO_2" layers are responsible for the superconductive behavior, and the other metal and metal oxide layers act as charge reservoirs, inducing redox behavior in the copper oxide layers. If we consider the formal charges on the ions, YBCO may be written as: $Y^{3+}(Ba^{2+})_2(Cu^{2+})_2Cu^{+}(O^{2-})_6$. The Cu^{2+} ion is square planar with an electronic configuration of $[Ar]3d^9$ (one unpaired electron, $S = 1/2$). For $YBa_2Cu_3O_{7-\delta}$ at an oxygen concentration corresponding to

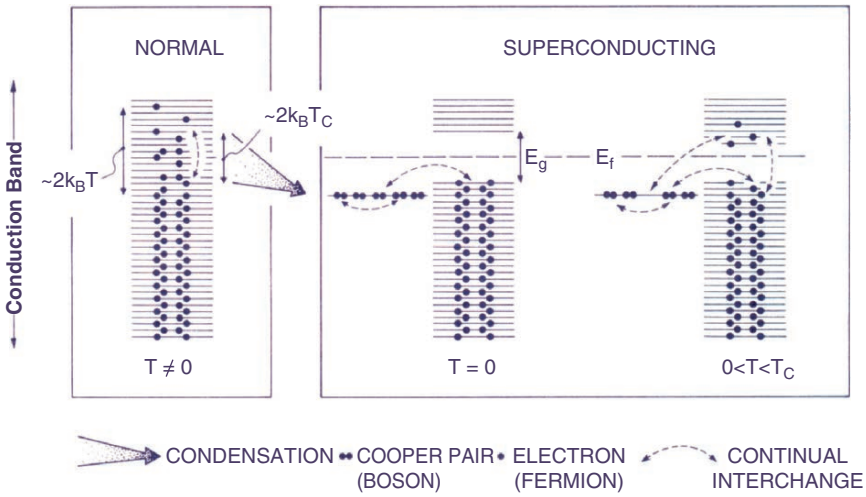


Figure 2.21. Schematic of Cooper pair formation for a metallic superconductor. At a certain critical temperature, T_c, the electrons within thermal energy (kT_c) of the Fermi level are able to correlate their motion in pairs (Cooper pairs). This process is referred to as condensation. At absolute zero, the condensed Cooper pairs form a single energy state that lies kT_c below the Fermi level. As a consequence of condensation, a number of forbidden energy levels (bandgap, E_g) appear within the conduction band, centered about the Fermi level. The energy of this bandgap corresponds to the minimum energy required to break up a Cooper pair and release the electrons into the vacant quantum levels. At $0 < T < T_c$, the superconductor metal is in an excited state, wherein a number of electrons, primarily from broken Cooper pairs, have been promoted across the bandgap into vacant energy states. Reproduced with permission from Hurd, C. M. *Electrons in Metals*, Wiley: New York, 1975. Copyright John Wiley & Sons Limited.

$\delta \geq 0.7$, the solid is insulating. This is due to effective antiferromagnetic ordering of the unpaired electron spins among neighboring Cu^{2+} ions in the CuO_2 layer.

However, as additional O is added to the lattice, YBCO undergoes the *Mott transition* of insulating to metallic conductivity – resulting from a change in the crystal symmetry from tetragonal to orthorhombic. The additional electronegative O atoms in the lattice serve to inject excess "holes," residing in states associated with the CuO_2 layers. At $\delta \leq 0.4$, the excess holes form *Cooper pairs*, which is responsible for superconductive behavior (v. strong diamagnetism). This is in contrast to a metal (Type I) superconductor, in which correlated electrons form Cooper pairs (Figure 2.21).

There are already commercial applications that employ superconductive materials; for example, MAGLEV trains have been operable for many years in Japan and England. However, reports of deleterious effects of radio waves may slow the widespread use of this technology. In 2001, three 400-ft HTS cables were installed at the Frisbie Substation of Detroit Edison, capable of delivering 100 million watts of power. This marked the first time commercial power has been delivered to customers of a US power utility through superconducting wire. Similar plans are underway to install an underground HTS power cable in Albany, New York, in Niagara Mohawk

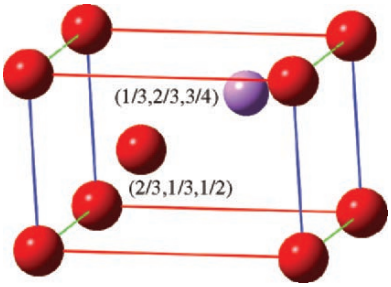


Figure 2.22. Unit cell of the hcp-based rutile allotrope of titanium oxide.

Power Corporation's power grid. The 350-m cable, believed to be four times the length of any previously installed HTS cable, will be designed to provide more power and operate at significantly lower loss levels than other HTS installations.

Another important oxide lattice is the *rutile* structure, common for both oxides and fluorides with the general formula MO_2 and MF_2 (*e.g.*, TiO_2, SnO_2, FeF_2, NiF_2, ZrO_2). This structure (Figure 2.22) consists of an hcp arrangement of oxide ions, with metal ions occupying 1/2 of the available octahedral holes. In a space-filling view of the lattice, the metal occupation causes the structure to expand so that the oxide ions are no longer in contact with one another. Although we have focused on the interstitial sites within ccp lattices, it should be noted that an hcp lattice, defined by atoms at (0,0,0) and (2/3,1/3,1/2), will have two octahedral sites at (1/3,2/3,3/4) and (1/3,2/3,1/4). Four tetrahedral sites will be located at (2/3,1/3,1/8), (2/3,1/3,7/8), and 8 additional sites (each 1/4 inside the unit cell) at positions (0,0,3/8), (1,0,3/8), (0,1,3/8), (1,1,3/8), (0,0,5/8), (1,0,5/8), (0,1,5/8), (1,1,5/8).

The nickel arsenide (NiAs) structure is also based on an hcp array; in this case, the cations form the backbone lattice, and the larger anions occupy both octahedral sites. This structure, also associated with metal sulfides, selenides, tellurides, and antimonides, is only adopted for weakly ionic compounds. Since the octahedral sites are extremely close to one another, purely ionic compounds would be much too unstable due to strong anion–anion repulsions.

2.3.4. Crystal Imperfections

The previous section focused on archetypical unit cells for ideal crystals. In reality, all crystals above absolute zero will possess a variety of defects in isolated or more extensive areas of the lattice. Surprisingly, even in solids with a purity of 99.9999%, there are *ca.* 6×10^{16} impurities cm^{-3}! Such impurities are not always a disadvantage. Often, these impurities are added deliberately to solids in order to improve its physical, electrical, or optical properties.

There are four main classifications of crystalline imperfections that exist in crystalline solids:

(a) *Point defects* – voids, interstitial/substitutional dopants, Schottky/Frenkel defects
(b) *Linear defects* – edge and screw dislocations

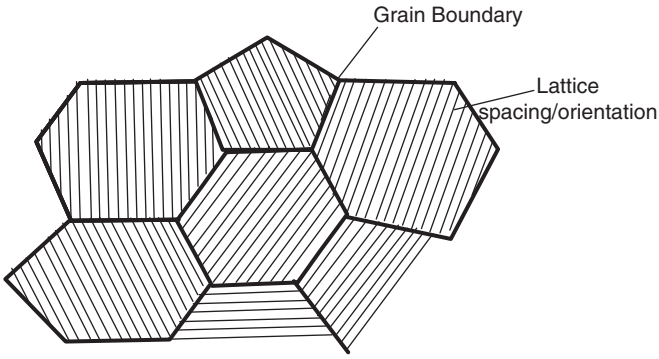


Figure 2.23. Illustration of grain boundaries between individual crystalline domains.

(c) *Planar defects* – grain boundaries, surfaces
(d) *Bulk defects* – pores, cracks

Of these four types, (a)–(c) are observed at the atomic level, whereas bulk defects are easily observed by the naked eye, or using a light microscope. These bulk defects are produced through the propagation of the microscopic flaws in the lattice. For crystals with a planar defect such as polycrystalline solids, the *grain boundary* marks the interface between two misaligned portions of the bulk crystal (Figure 2.23).

The size of the individual microcrystals (or grains) that comprise a larger aggregate greatly affects many properties of the bulk crystal. Both optical microscopy and X-ray diffraction are used to determine the grain sizes; most commercial metals and alloys consist of individual crystallites with diameters ranging from 10 to 100 μm, each corresponding to millions of individual metal atoms. A decrease in the size of these microscopic particles results in an increase in both strength and hardness of the bulk material, due to closer packing among neighboring grains. The density of atoms at a solid surface, or in the region surrounding a grain boundary is always smaller than the bulk value. This is due to atoms at these regions containing dangling bonds, known as coordinatively unsaturated (Figure 2.24). Hence, surfaces and interfaces are very reactive, often resulting in the concentration of impurities in these regions. Since energy is required to form a surface, grains tend to grow in size at the expense of smaller grains to minimize energy. This growth process occurs by diffusion, which is accelerated at high temperatures.

Although a solution is typically thought of as a solid solute dissolved in a liquid solvent, *solid solutions* are formed upon the placement of foreign atoms/molecules within a host crystal lattice. Solutions that contain two or more species in their crystal lattice may be either substitutional or interstitial in nature (Figure 2.25). If the regular crystal lattice comprises metal atoms, then this mixture is referred to as an *alloy*. Substitutional solids feature the actual replacement of atoms/ions in the regular lattice with solute species, known as *dopants*. In most substitutional solid solutions, the dopant species are arranged in a random fashion among the various unit cells of the extended lattice. Examples of these types of lattices are illustrated by metal-doped

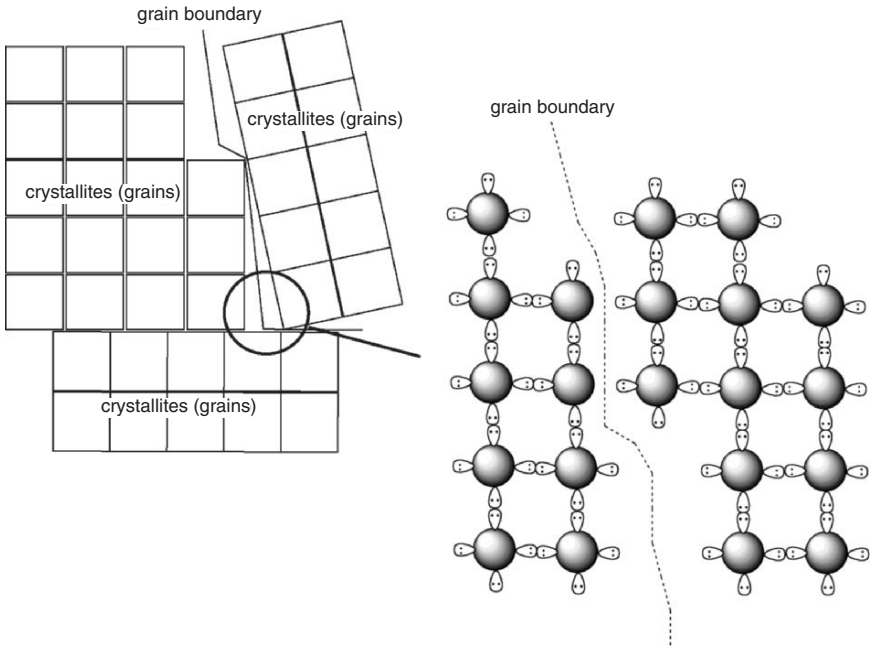


Figure 2.24. Schematic of a polycrystalline solid, with grain boundaries formed from dangling bonds between neighboring metal atoms.

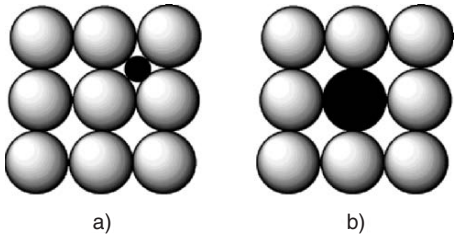


Figure 2.25. Illustration of the difference between (a) interstitial and (b) substitutional defects in crystalline solids.

aluminum oxide (alumina – Figure 2.26), constituting gemstones such as emeralds and rubies (both due to Cr-doping). For these solids, small numbers of formal Al^{3+} lattice sites are replaced with solute metal ions.

The movement of atoms within a crystal lattice occurs through a diffusion process, which is easiest when atomic vibrations are present at high temperatures. That is, atoms must move from their lattice positions in order to let the diffusing atoms pass. As one would also expect, smaller atoms diffuse more readily than larger ones. If the interstitial atom is identical to the lattice atoms, the occupancy is referred to as *self-interstitial*. Empirically, the likelihood of an atom occupying interstitial sites of the

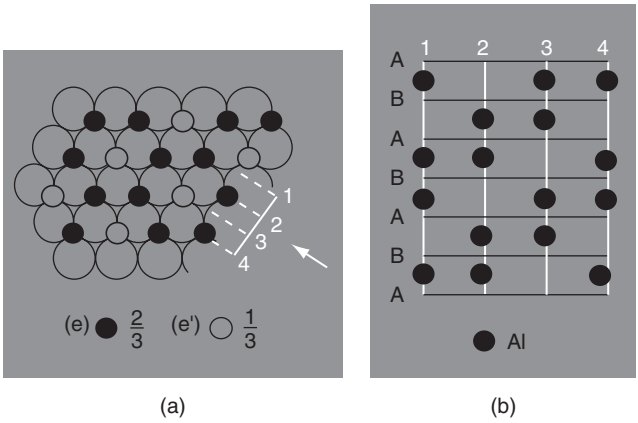


Figure 2.26. Representation of the structure of α-alumina (corundum). (a) Al^{3+} ions (filled circle) are shown to occupy the octahedral sites between the hcp layers of O^{2-} ions (open circle). (b) The stacking sequence of Al^{3+} ions as viewed in the direction of the arrow in (a). Reprinted from Greenwood, N. N.; Earnshaw, A. *Chemistry of the Elements*, 2nd ed., Copyright 1998, with permission from Elsevier.

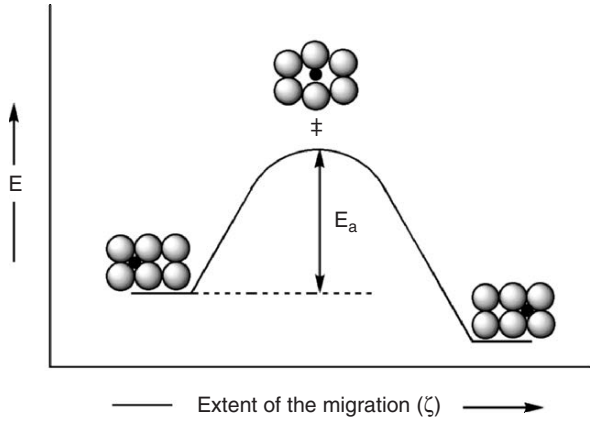


Figure 2.27. Illustration of the energetics involved for the atomic diffusion of interstitial impurities.

solvent lattice may be predicted by comparing the radii of the interstitial dopant and solvent atom (Eq. 7):

$$\text{(7)} \qquad \text{Interstitial Solubility} = \frac{r_{\text{solute}}}{r_{\text{solvent}}} \leq 0.59.$$

For an atom to migrate from one position to another within a lattice, an energy barrier must be overcome. The magnitude of this barrier, known as the activation energy (E_a), is directly proportional to the rate of atomic diffusion (Figure 2.27). As the temperature of the crystal is increased, the atoms in the crystal vibrate about their equilibrium positions. The Arrhenius equation (Eq. 8) is used to calculate the

probability of an atom having a suitable energy to overcome E_a. It should be noted that this equation may also be applied to predict the number of vacancies or voids in the crystal lattice created by thermal equilibrium. Since the activation energy is significantly greater than the thermally induced kinetic energy of lattice atoms, the most pronounced atomic migration occurs along dislocations and voids in the crystal, since fewer atoms are involved in the atomic displacement (*i.e.*, E_a much lower):

$$\frac{N_i}{N_T} \propto e^{-(E_a/kT)}, \tag{8}$$

where N_i is the number of impurities with $E > E_a$; N_T, the total number of impurity atoms in the crystal lattice; E_a, the activation energy for the diffusion process; k, the Boltzmann constant (1.38×10^{-23} J atom^{-1} K^{-1}); and T is the temperature (K).

In contrast to interstitial defects, the replacement of lattice atoms with dopant species is referred to as *substitutional disorder*. Since this type of displacement involves extensive diffusion of solvent and solute atoms, a number of requirements must be satisfied. These are known as the William Hume-Rothery Rules[1]:

1. The percentage difference between solute and solvent atomic radii should be less than 15% (Eq. 9). If there were a larger mismatch between the dopant and solvent atomic radii, the rate of diffusion would either be too slow (for large r_{solute}), or the dopant would position itself in an interstitial site rather than replacing a solvent atom (for small r_{solute}). For instance, although most gemstones feature the replacement of formally Al^{3+} ions in an aluminum oxide (alumina) based crystal with M^{3+} ions (M = Cr, Ti, Fe), the substitution of aluminum ions with Li^+ or W^{3+} ions would cause too drastic a perturbation of the crystal structure. Whereas small lithium atoms/ions would have an opportunity to diffuse into the lattice, forming interstitial solutions, the large tungsten ions would only adsorb to the surface of individual crystals:

$$\frac{|r_{solute} - r_{solvent}|}{r_{solvent}} \times 100\% \leq 15\%. \tag{9}$$

2. The crystal structures of the dopant and solvent atoms must be matched. That is, the density of the host solvent unit cell must be voluminous enough to accommodate the solute atom. In order of highest solubility,

$$FCC_{solute} - FCC_{solvent} > FCC_{solute} - BCC_{solvent} > FCC_{solute} - HCP_{solvent} > BCC_{solute} - HCP_{solvent}.$$

3. In order for dopant atoms to be stabilized within a host lattice, both types of atoms must have similar electronegativities. If this prerequisite were not met, electron density would transfer to the more electronegative atoms, forming a compound between lattice and dopant atoms. For instance, the exposure of an iron lattice to chlorine gas would result in a yellow compound of $FeCl_3$ instead of reddish crystals of Cl-doped iron.
4. The solute and solvent atoms should have similar valences in order for maximum solubility. For instance, if a divalent atom were to replace a hexavalent solvent atom, the solvent lattice would be incapable of stabilizing the dopant atom, since too few bonding interactions would be possible between solute and solvent atoms.

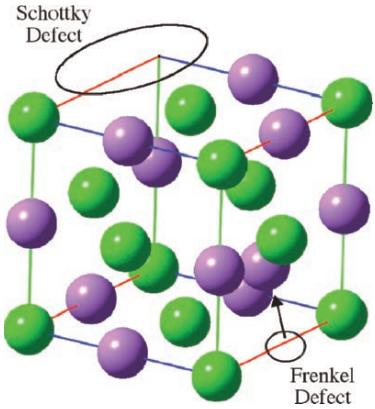


Figure 2.28. Illustration of a unit cell of an ionic crystal with Frenkel and Schottky defects.

Two point defects are observed for ionic crystals (Figure 2.28). *Schottky defects* are formed through the migration of a cation–anion pair to the surface, leaving behind a vacancy. The other type of imperfection is known as a *Frenkel defect*, occurring when a lattice ion migrates into an interstitial position, leaving behind a vacant site. For ionic crystals, the overall charge of the crystal must be charge-balanced. That is, if trivalent ions such as La^{3+} are substituted with divalent cations such as Ca^{2+}, there must be concomitant placements of divalent anions (*e.g.*, O^{2-}) to balance the crystal charge. This is an important consideration of solid electrolytes, the topic of this end-of-chapter "Important Materials Application."

When the composition of a crystal is defined by a distinct chemical formula (*e.g.*, SiO_2), it is known as a *stoichiometric compound*. If the composition of the crystal is altered upon doping or thermal treatment, the resulting solid may deviate from the original chemical formula, forming a *nonstoichiometric* solid. Nonstoichiometry and the existence of point defects in a solid are often closely related. For instance, the formation of x anion vacancies per each quartz unit cell will result in the nonstoichiometric compound SiO_{2-x}.

An external pressure (stress) that is exerted on a material causes its thickness to decrease. This change is referred to as *strain*, and is related to the bonding scheme of the atoms comprising the solid. For example, a rubbery material will exhibit a greater strain than a covalently bound solid such as diamond. Since steels contain similar atoms, most will behave similarly as a result of an applied stress. If a stress causes a material to bend, the resultant flex is referred to as shear strain. For small shear stresses, a material will generally deform elastically, involving no permanent displacement of atoms and reversal of the deformation upon removal of the shear stress. However, for a large shear stress, a material will deform plastically, involving the permanent displacement of atoms, known as *slip*.

Edge and screw *dislocations* are abrupt changes in the regular ordering of atoms along an axis in the crystal (Figure 2.29). These line defects are introduced during the crystallization process, and may not always be detrimental. As we will see in Chapter 3, it is the dislocation interactions within a metal that are responsible for

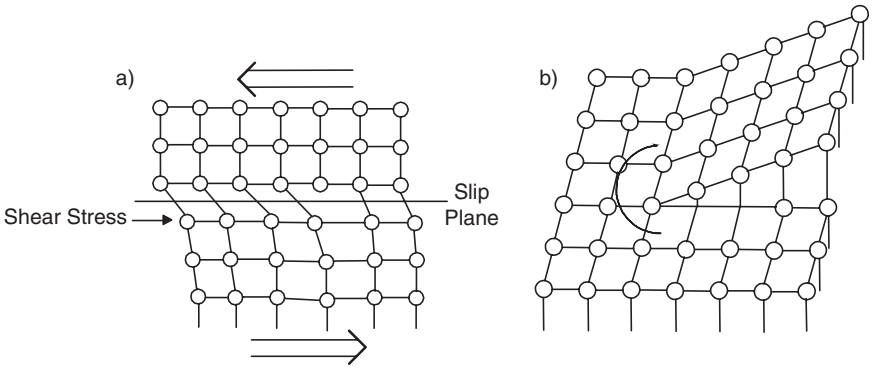


Figure 2.29. Illustration of (a) edge and (b) screw dislocations.

the deformation and strengthening of metals. In single crystals, there are preferred planes where dislocations may occur.

2.3.5. Phase-Transformation Diagrams

All matter consists of different arrangements, or phases, of its subunits upon changes in its ambient temperature and/or pressure. These phases each have distinct physical and/or chemical characteristics. *Phase diagrams* for a material illustrate such structural transformations between liquid, solid, and gaseous states, as well as more subtle changes within its solid state. Even for a pure substance such as water, the solid state becomes increasingly more complex under particular conditions, such as low temperature/high pressure. Figure 2.30 shows the phase diagram for water and ice, showing the regions of stability for the 11 distinct forms of ice, each comprising strong hydrogen bonding interactions among neighboring water molecules.

As you may recall from physical chemistry, the thermodynamic spontaneity of a reaction is governed by the Gibbs free energy, G. For a phase transformation from an initial to final state, the ΔG (Eq. 10; $G_{\text{final}} - G_{\text{initial}}$) must be < 0 for the process to be spontaneous:

$$\Delta G = \Delta H - T\,\Delta S \text{ (at cst. } P \text{ and } T). \quad (10)$$

This thermodynamic approach may be used to predict whether a system is in equilibrium ($\Delta G = 0$), and to assess whether a particular phase transformation will occur. For instance, the ΔG for the transformation between graphitic and diamond allotropes of carbon shown below is $-2.9\,\text{kJ mol}^{-1}$.

However, even though this transformation is thermodynamically favored, the diamond allotrope still exists at high pressures and over long time periods. That is, if a particular phase transformation is predicted as spontaneous, the actual rate of that process will depend on the kinetics of the transformation. Since the sp^3 carbon bonds in diamond are extremely strong, the kinetics governing the migration of carbon atoms between diamond–graphite is extremely slow at normal temperatures and

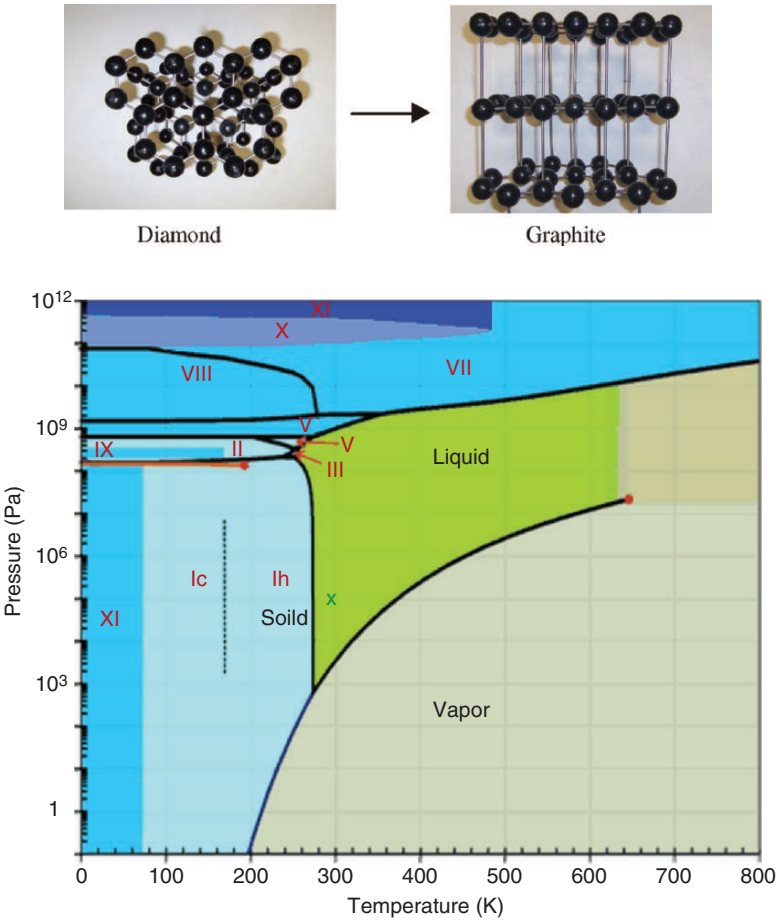


Figure 2.30. Phase diagram for water, showing the complexity of the solid region. Reproduced with permission from Prof. M. F. Chaplin (http://www.sbu.ac.uk/water/phase.html).

pressures. As one would expect, such a change is facilitated at higher temperatures and pressures when atoms exhibit their greatest vibrations within the lattice.

Phase transitions in solids progress through two processes: nucleation and growth. In response to changes in external conditions, atoms or molecules comprising the solid first migrate to thermodynamically favored positions in the lattice. Quite often, this occurs at crystal defect sites such as grain boundaries and dislocations. Once these nuclei are formed in the solid, they begin to grow through further atomic diffusion. Whereas nucleation may occur most easily at low temperatures, atomic migration and new phase growth may only exist at relatively high-temperature regimes. When a material is introduced to fast heating or cooling rates, intermediate phases known as *metastable* states are formed, since the phase change is slow relative to the

temperature change. Although these phases are often quite stable, they do not appear in phase diagrams since they are not formed at equilibrium.

2.3.6. Crystal Symmetry and Space Groups

Crystallography, or the nomenclature of crystalline solids, uses two terms to describe the symmetry of the crystal lattice: *point groups* and *space groups*. Chemists should be quite familiar with the notion of point groups, as this designation is used to predict molecular reactivities and IR absorption bands, among other diagnostics. A flowchart for the assignment of the molecular point group is shown in Figure 2.31. Examples of this notation, given by Schoenflies symbols, are C_{3v} for ammonia and $Os_3(CO)_9(C_6H_6)$, O_h for $[CoF_6]^{3-}$, and D_{3h} for $(CH_3)_8Si_5O_6$ molecules (Figure 2.32).

By definition, a symmetry operation is an event that results in the transposing of one item into another that is indistinguishable from the original. This operation may take place about a point, line, or plane of symmetry. When a set of symmetry operations is applied to components of the unit cell of a crystal lattice, the resulting symmetry is designated as a *crystallographic point group*. Whereas a molecular point group operation will reproduce an individual molecule, a crystallographic point group operation must leave the entire crystal unchanged.

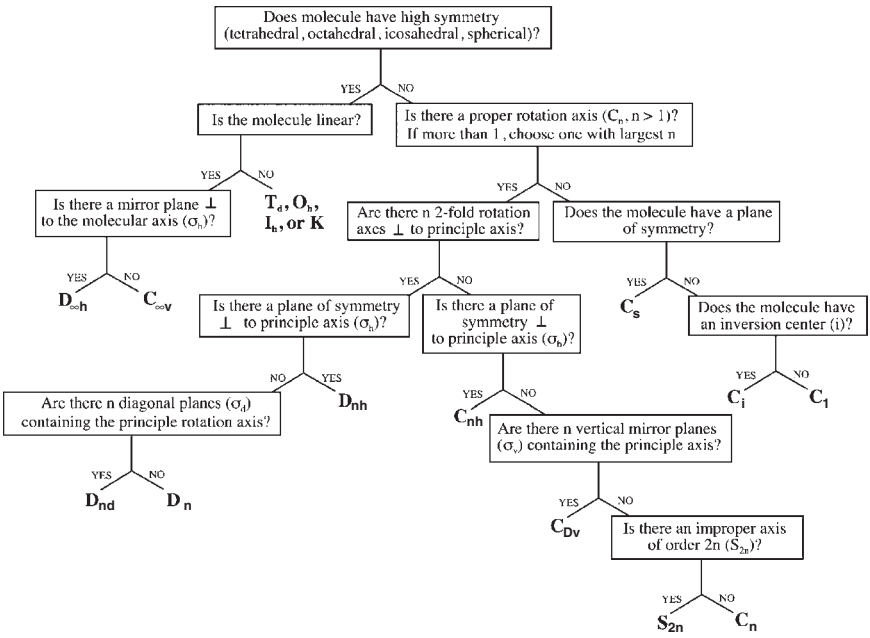


Figure 2.31. Scheme for determining the point group symmetry of a molecule.

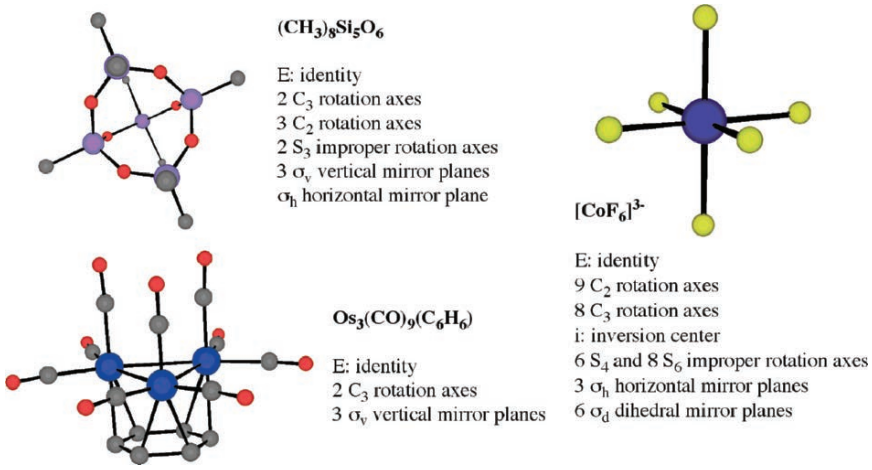


Figure 2.32. Examples of symmetry group elements and molecular point group assignments.

Illustrative examples of crystallographic symmetry operations are shown in Figure 2.33. An integer label, n, indicates the regeneration of an equivalent lattice point when an object in the crystal lattice is rotated $360°/n$ about an axis. A rotation-inversion axis is designated by $\overline{n}$, featuring rotation about an axis $(360°/n)$, followed by inversion about a center of symmetry. In contrast to axes of rotation, if individual lattice points are reflected through a plane of symmetry, the operation is symbolized by m, denoting the presence of a mirror plane. Although individual atomic and ionic lattice points are not affected by $\overline{n}$ and m operations, molecular lattice points exhibit a change in handedness following these operations (see Figure 2.33b,c). Point groups that include these operations, known as *improper symmetry* operations, must exclude all chiral molecules, as they would then be superimposible on their mirror images. Two important restrictions apply to crystallographic point group symmetry operations:

1. The symmetry operations must be compatible with infinite translational repeats in a crystal lattice.
2. A symmetry operation cannot induce a higher symmetry than the unit cell has.

The point group symmetry describes the nontranslational symmetry of the crystal. The rest of the crystal is then generated by translational symmetry. Only two-, three-, four-, and sixfold rotation axes are compatible with translational symmetry, so point groups containing other types of rotation axes (*i.e.*, 5 or $\overline{5}$) are not relevant. Furthermore, since there is no such thing as a linear three-dimensional crystal, the linear point groups are also irrelevant. The remaining point group symmetry operations yield a total of 32 crystallographic point groups, designated by Hermann–Mauguin (H–M) symbols (Table 2.4). As they can be deduced from the macroscopic crystal symmetry, they are also referred to as the *32 crystal classes*. For the same reason, the symmetry elements that give rise to the crystal classes are sometimes referred to as external symmetry elements.

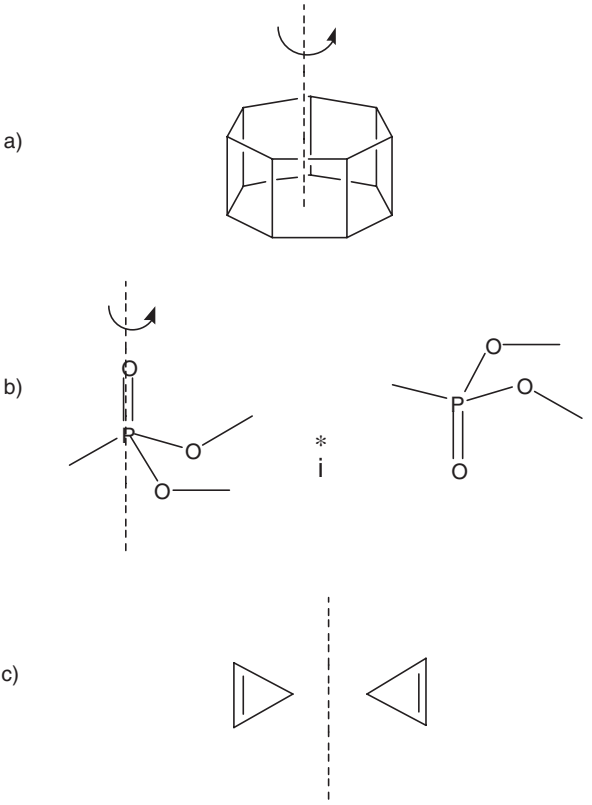


Figure 2.33. Illustration of crystallographic point group operations. Shown are (a) rotation axis, (b) rotation-inversion axis, and (c) mirror plane.

Both glide planes and screw axes feature translation in addition to mirror and rotation operations, respectively (Figure 2.34). For glide planes, the translation is designated as a, b, or c if movement is halfway along the a, b, or c unit cell axes, respectively. If the translation is along the diagonals $\frac{1}{2}(a+b)$, $\frac{1}{2}(a+c)$, or $\frac{1}{2}(b+c)$, the glide plane is designated as n. Lastly, if the translation is along $\frac{1}{4}(a+b)$, $\frac{1}{4}(a+c)$, or $\frac{1}{4}(b+c)$, the glide plane is given the symbol d.[13] For screw axes, the nomenclature is of the form n_x, indicating a $360/n$ rotation, followed by a $x/4$ translation along one of the unit cell vectors, a, b, or c. For example, if there is a sixfold axis of rotation followed by 3/4 translation along a, the symmetry designation is 6_3.

It is noteworthy to point out that two sequential screw-axis or glide-plane operations will yield the original object that has been translated along one of the unit cell vectors. For example, a 6_3 axis yields an identical orientation of the molecule only after 6 repeated applications – 4.5 unit cells away (*i.e.*, $6 \times 3/4 = 4.5$). However, since glide planes feature a mirror plane prior to translation, the first operation will

Table 2.4. The 32 Crystallographic Point Groups

Crystal system (bravais lattices)	Crystallographic point groups[a] (molecular point groups[b])
Cubic (P, I, F)	23, $m3$, 432, $\bar{4}3m$, $m3m$ (T, T_h, O, T_d, O_h)
Tetragonal (P, I)	4, $\bar{4}$, $4/m$, 422, $4mm$, $\bar{4}2m$, $4/mmm$ (C_4, S_4, C_{4h}, D_4, C_{4v}, D_{2d}, D_{4h})
Orthorhombic (P, C, I, F)	222, $mm2$, mmm (D_2, C_{2v}, D_{2h})
Trigonal/Rhombohedral (P)	3, $\bar{3}$, 32, $3m$, $\bar{3}m$ (C_3, C_{3i}, D_3, C_{3v}, D_{3d})
Hexagonal (P)	6, $\bar{6}$, $6/m$, 622, $6mm$, $\bar{6}m2$, $6/mmm$ (C_6, C_{3h}, C_{6h}, D_6, C_{6v}, D_{3h}, D_{6h})
Monoclinic (P, C)	2, m, $2/m$ (C_2, C_s, C_{2h})
Triclinic (P)	1, $\bar{1}$ (C_1, C_i)

[a]The H–M symbolism derived from crystal symmetry operations. For image and movie representations of each point group, see the website: http://neon.mems.cmu.edu/degraef/pg/index.html

[b]The analogous Schoenflies symbolism derived from molecular symmetry operations.

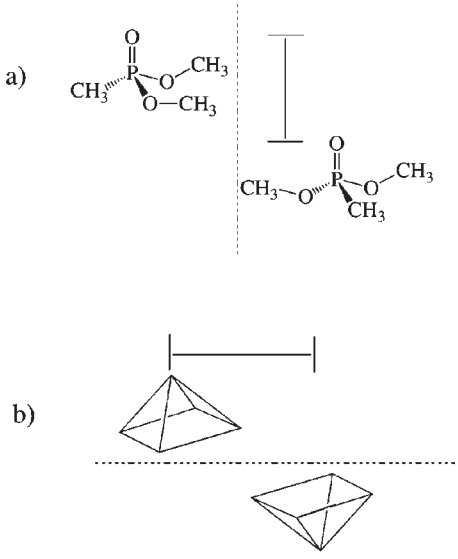


Figure 2.34. Illustration of translational symmetry elements. Shown is (a) screw axis and (b) glide plane.

cause a change in handedness of the molecule. By contrast, screw axis operations do not alter the stereoisomerism of the molecule.

Both glide and screw axes are not point group operations because they involve translations. That is, one cannot distinguish between analogous rotation and screw axes, or between glide and mirror planes, by just looking at the crystal faces. You may notice that of the symmetry elements discussed, both glide planes and screw axes are absent from the list of point group symbols, listed in Table 2.4. For the

purposes of determining the crystallographic point group, glide planes are treated as mirror planes, and screw axes as rotation axes.

When the symmetry elements are applied to species arranged periodically on a crystal lattice, the result is a *space group*. The combination of the 32 crystallographic point groups with the 14 Bravais lattices yields a total of 230 possible space groups for crystals, designated by the H–M symbol. The 73 different space groups that can be generated from point groups only, without using glide planes and/or screw axes, are called *symmorphic* space groups. The first letter of the H–M symbol is a single letter that refers to the Bravais lattice type. The letters used are P (primitive), A ((100) face centered), B ((010) face centered), C ((001) face centered), F (face centered), and I (body centered). The remaining three letters refer to the point group symmetry of the unit cell. The most common space groups are $P2_1/c$, $P2_1/n$, $P1$, $C2/c$, $C2/m$, C_c, *Pbcn*, and *Pnma*. For optical activity considerations, a space group with enantiomorphous symmetry must be present (*i.e.*, 222, 422, 32, 622, 432, 1, 2, 3, 6, 23).

Example 1: *Pnma* (also indicated by $P2_1/n2_1/m2_1/a$) indicates a primitive orthorhombic unit cell, with a glide plane $\perp$ to the a axis (translation along diagonal $\frac{1}{2}(b+c)$), a mirror plane $\perp$ to b, and another glide plane $\perp$ to c (translation of $a/2$).

Example 2: $P2_1/a$ indicates a primitive monoclinic unit cell with a twofold screw axis along b, and a glide plane $\perp$ to this axis with a translation of $a/2$.

Example 3: $C2/m$ indicates a c-centered unit cell with a twofold rotation axis along a, and a mirror plane $\perp$ to this axis.

2.3.7. Physical Properties of Crystals

Hardness

Thus far, we have examined the arrangements and associative interactions of atoms, ions, or molecules comprising a crystal lattice. The macroscopic physical properties of crystalline materials are related to these arrangements. The overall hardness of a crystal depends on the nature of the interactions between components of the crystal lattice. Those crystals possessing covalent interactions (*e.g.*, diamond) will have a high hardness, whereas those containing only van der Waals forces will be soft (*e.g.*, talc). A variety of scales (Table 2.5) may be used to assign the relative hardness of a material. The Mohs scale is generated by a qualitative assessment of how easily a surface is scratched by harder materials, with the hardest material (diamond) given a value of 10. The hardness of a material is directly proportional to its tensile strength; depending on which method is used, proportionality factors may be calculated. For instance, if the Brinell hardness value is known, the tensile strength is simply 500 times that value.

Tests such as Vickers, Knoop, and Brinell use an indentation technique that impinges a hard tip (*e.g.*, diamond) into the sample with a known load (Figure 2.35a). After a designated period of time, the load is removed, and the indentation area is measured. The hardness, H, is defined as the maximum load, L, divided by the residual indentation area, A_r (Eq. 11). The coefficient, F, varies depending on which indentation method is used. This value (14.229 for Knoop and 1.854 for Vickers)

Table 2.5. Hardness Scales

Solid	Mohs	Vickers	Knoop
Talc	1	27	N/A
Graphite	1.5	37	N/A
Gypsum	2	61	N/A
Fingernail	2.5	102	117
Calcite	3	157	169
Fluorite	4	315	327
Apatite	5	535	564
Knife blade	5.5	669	705
Feldspar	6	817	839
Pyrex glass	6.5	982	929
Quartz	7	1,161	N/A
Topaz/Porcelain	8	1,567	N/A
Sapphire/Corundum	9	2,035	N/A
Diamond	10	N/A	N/A

N/A indicates the hardness value is above/below the acceptable range of the particular hardness scale. Values were obtained from the conversion site: http://www.efunda.com/units/hardness/convert_hardness.cfm?HD=HM&Cat=Steel#ConvInto

is related to the geometry of the pyramidal probe, which will affect the penetration depth under the same load. Since a spherical probe is used for the Brinell test, a more complex formula is used to calculate the hardness (Eq. 12), where D is the diameter of the spherical indentor, and D_i is the diameter of the indentor impression (both in mm):

$$H = F\frac{L}{A_r}, \tag{11}$$

$$H = \frac{L}{\frac{\pi}{2}D\left(D - \sqrt{D^2 - D_i^2}\right)}. \tag{12}$$

Quite often, an indentation is so small that it is difficult to resolve with a normal light microscope. To circumvent these problems, software is now capable of monitoring the load and displacement of the probe during the measurement, and relating this to the contact area. Such an analysis without the need for visual confirmation is necessary for nanoindentation techniques for thin films and other surface hardness applications. As its name implies, the hardness of a material is evaluated by the depth and symmetry of the cavity created from controlled perforation of a surface with a *nanosized* tip (Figure 2.35b). It should be pointed out that although we have discussed crystalline solids in great detail thus far, hardness measurements are also easily performed on amorphous solids such as glasses and other ceramics.

Cleavage and fracturing

The intermolecular forces in a crystal lattice are often not homogeneous in all directions. If the solid consists of strong interactions among neighbors in specific layers, and weak interactions among molecules in neighboring layers (*e.g.*, graphite),

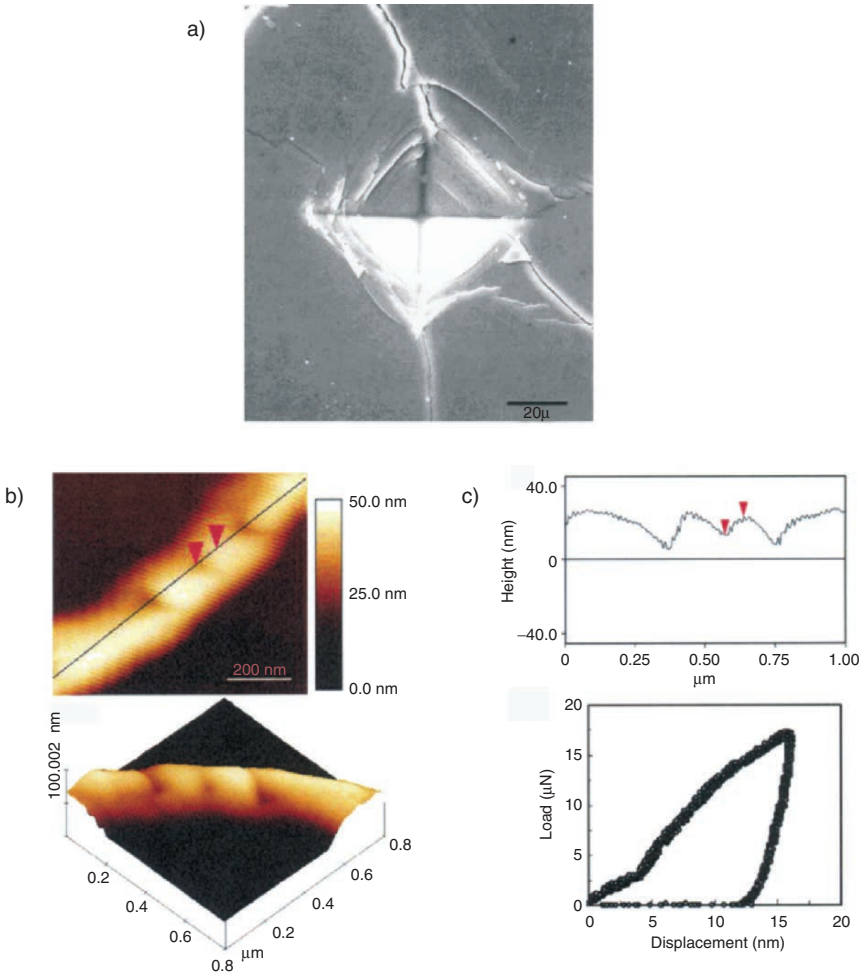


Figure 2.35. Examples of indentation processes to determine surface hardness. Shown are (a) Vickers indentation on a SiC–BN composite, (b) atomic force microscope images of the nanoindentation of a silver nanowire, and (c) height profile and load–displacement curve for an indent on the nanowire. Reproduced with permission from *Nano Lett.* **2003**, 3(11), 1495. Copyright 2003 American Chemical Society.

a *cleavage plane* is created where little force is needed to separate the crystal into two units (Figure 2.36). Since the cleavage planes are parallel to crystal faces, the fragments formed upon cleavage will retain the symmetry exhibited by the bulk crystal. Whereas *cleavage* describes the formation of a smooth piece of the original crystal when subjected to an external stress, a *fracture* refers to chipping a crystal into rough, jagged pieces. Figure 2.36c,d shows photos of cleavage and fracture of a NaCl crystal subjected to stress at oblique angles to the cleavage plane. Although preferential cracking will occur along the cleavage plane, smaller fractured pieces will also

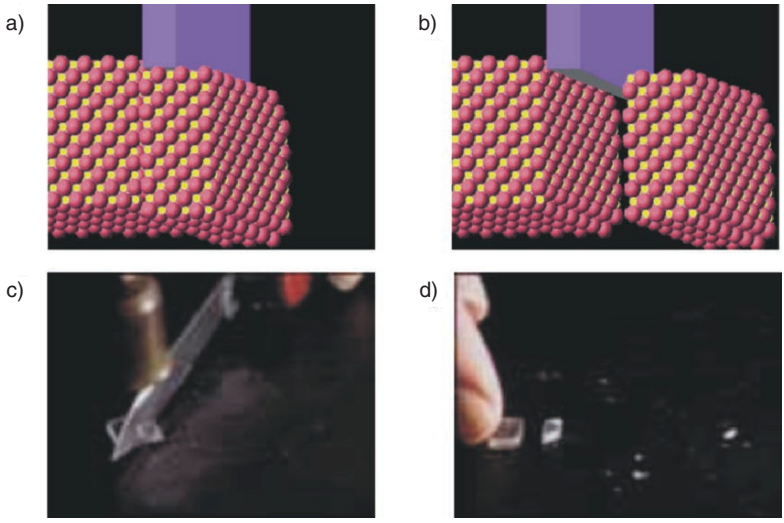


Figure 2.36. Example of cleavage and fracture at the atomic (a, b) and macroscopic (c, d) levels. In images (c, d), a crystal of NaCl is exposed to a stress along an oblique angle to the cleavage plane, resulting in both cleavage and fracturing. Images taken with permission from the Journal of Chemical Education online: http://jchemed.chem.wisc.edu/JCESoft/CCA/CCA2/SMHTM/CLEAVE.HTM

be formed. In general, as one increases both the magnitude and obliqueness of the applied stress, the amount of fracturing will increase, relative to cleavage.

On occasion, the cleavage plane may be easily observed due to a fibrous network lattice. One example of such a crystal is ulexite, sodium calcium borate of the chemical formula $NaCa(B_5O_6)(OH)_6{\cdot}5H_2O$. It is quite obvious that upon external stress, the crystal will preferentially cleave in directions parallel to the crystallite fibers (Figure 2.37a). However, more intriguing is the interesting optical properties exhibited by this crystal, commonly designated as the "T.V. rock" (Figure 2.37b, c). If the crystal is surrounded by a medium of a lower refractive index (*e.g.*, air), light is propagated through the individual fibers by internal reflection. This is analogous to fiber optic cables that will be described a bit later in this chapter. Figure 2.37b shows that light passing through the crystal will exhibit concentric circles. This is due to the difference in the effective path lengths of light passing either directly through individual fibers, or crossing grain boundaries *en route* through the crystal via adjacent fibers.[14]

Color

In a previous section, we discussed substitutional impurities within crystal lattices without mentioning the changes in physical properties that this creates. Perhaps the most obvious outcome of a crystal impurity is the resultant color. In this section, we will answer the common question: "why are some crystals colorless, and others colored?"

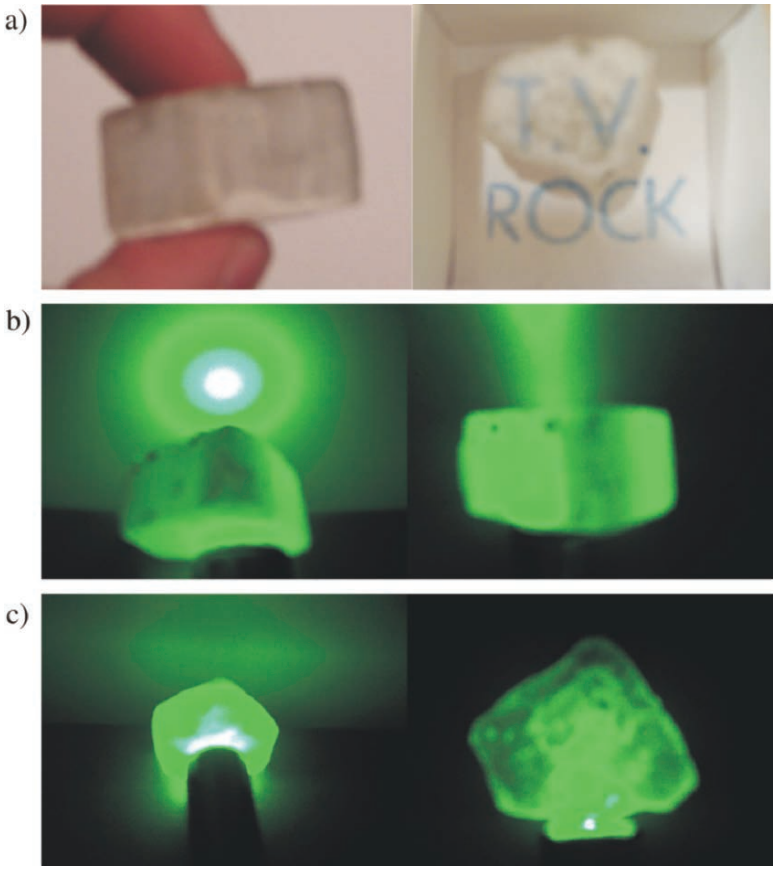
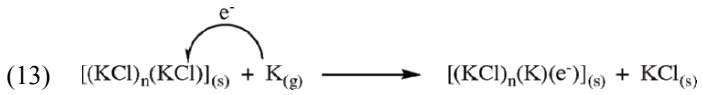


Figure 2.37. Optical properties of ulexite, the "T.V. rock." Shown are (a) the fibrous morphology of the crystal, and projection of the words "T.V." on the surface of the rock from transmission through parallel fibers,; (b) passage, and (c) blockage of a green laser beam when impinged on the crystal at angles parallel and perpendicular to the fibers, respectively.

Many crystals such as diamonds, quartz and corundum are normally colorless upon inspection. In these crystals, the constituent atoms form a rigid, regular framework of covalent or ionic interactions. Since visible light (350–700 nm) is not energetically sufficient to cause bond rupturing and/or electronic transitions of the constituent metal atoms/ions, this energy is not absorbed by pure crystals, giving rise to a colorless state. However, when an impurity is added to the lattice, visible radiation may be suitably energetic to cause lattice alterations and/or electronic transitions, yielding an observable color change.

Colored crystals need not be gemstones; in fact, a colorless crystal of potassium chloride may be suitably altered to exhibit color. When solid KCl is heated to 500°C in the presence of potassium vapor, the crystal becomes a violet color. This occurs due to the ionization of gaseous potassium atoms that abstract a Cl^- anion from the

crystal lattice. The electron formed in the oxidation process becomes trapped in the anion vacancy, as this will rebalance the overall charge of the crystal (Eq. 13).



Another process that may be used to generate an anion vacancy is through irradiation of the crystal with ionizing radiation such as X-rays. This high-energy radiation will cause the removal of a halide ion from the lattice and will excite some of the lattice electrons from valence to conduction bands. At this point, the electrons are free to diffuse through the crystal, where they remain mobile until they find an anion vacancy site. At low temperatures (*e.g.*, in liquid nitrogen), electrons may even become localized by polarizing their surroundings; that is, displacing the surrounding ions, to give self-trapped electrons.

For each type of electron trap, there is a characteristic activation energy that must be overcome for the release of the electron. As an irradiated crystal is heated, electrons are released from their traps by thermal activation, leading to visible light emission. The free electrons are able to migrate once again through the crystal until they recombine with an anion hole. Exposure to light of appropriate wavelengths may also cause release of electrons from traps and thus give rise to light emission. This is the active process that occurs in self-darkening sunglasses, in which the ultraviolet wavelengths present in sunlight produce a darkening effect, and room temperature leads to fading as soon as the UV source is removed.

Another example of the coloration phenomenon is demonstrated by transition metal doping of crystals exhibited by gemstones such as ruby. Table 2.6 lists some common gemstones, and the respective host crystal and dopants that give rise to their characteristic colors. Whereas crystals of pure corundum (α-alumina) are colorless,

Table 2.6. Active Dopants in Gemstone Crystals

Gemstone	Color	Host crystal	Impurity ion(s)
Ruby	Red	Aluminum oxide	Cr^{3+}
Sapphire	Blue	Aluminum oxide	Fe^{2+}, Ti^{4+}
Emerald	Green	Beryllium aluminosilicate	Cr^{3+}
Aquamarine	Blue-green	Beryllium aluminosilicate	Fe^{2+}
Garnet	Red	Calcium aluminosilicate	Fe^{3+}
Topaz	Yellow	Aluminum fluorosilicate	Fe^{3+}
Tourmaline	Pink/red	Calcium lithium boroaluminosilicate	Mn^{2+}
Turquoise	Blue-green	Copper phospho aluminate	Cu^{2+}

a small amount (<1%) of chromium doping yields the familiar reddish/pink color. This color change is only possible if the periodic framework of the crystal is altered, through the incorporation of additional dopant atoms/ions or vacancies in the lattice. For ruby, a transition metal ion, Cr^{3+}, replaces Al^{3+} yielding electronic d–d transitions that were unattainable for the original main-group ion.

In a pure crystal of Al_2O_3, each aluminum ion is surrounded by six oxygens in a slightly distorted octahedron. The formal electronic configuration of Al^{3+} ions is [Ne], indicating that all electrons are paired. Since the irradiation of the crystal with visible light is not energetic enough to cause promotion of electrons into empty excited-state orbitals, the crystal appears colorless. However, if Al^{3+} ions are replaced with Cr^{3+}, each of the transition metal ions will have a configuration of $[Ar]3d^3$. Although general chemistry tends to simplify the d-orbitals as being a set of five degenerate orbitals, transition metal complexes exhibit splitting of the d-orbital energy levels. This results in facile electronic transitions upon exposure to visible light, explaining the bright colors exhibited by many transition metal compounds.

A simple theory, referred to as *crystal field theory*,[15] is often used to account for the colors and magnetic properties of transition metal complexes. This theory is based on the electrostatic repulsions that occur between electrons in d-orbitals of a transition metal, and electrons contained in ligand orbitals. Figure 2.38 shows the splitting of the d-orbitals resulting from the electrostatic repulsions between the metal and ligand electrons for an octahedral Cr^{3+} complex. Since the d_{z^2} and $d_{z^2-y^2}$ orbitals are located directly on the internuclear bonding axis, a greater electrostatic repulsion will occur resulting in an increase in energy. Visible light is capable of being absorbed by the complex, causing the excitation of electrons into empty d_{z^2} or $d_{z^2-y^2}$ orbitals. Since the ΔE (10 Dq) between d-orbitals is in the violet region of the visible spectrum, the reflected complementary red color is observed for ruby crystals.

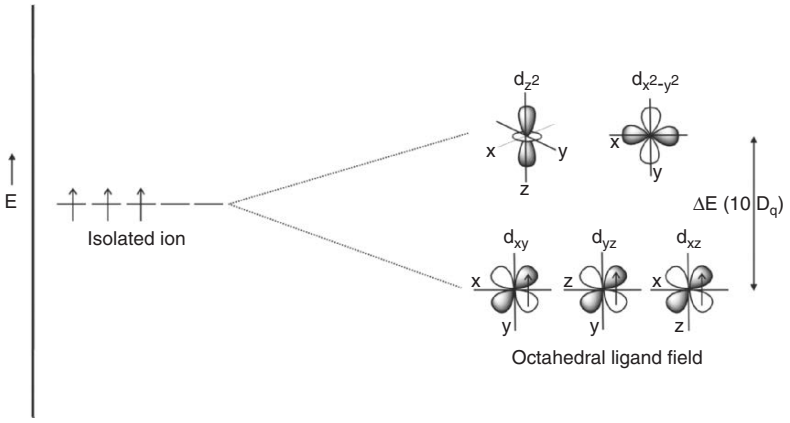


Figure 2.38. Energy splitting diagram for an octahedral Cr^{3+} ion in the ruby lattice.

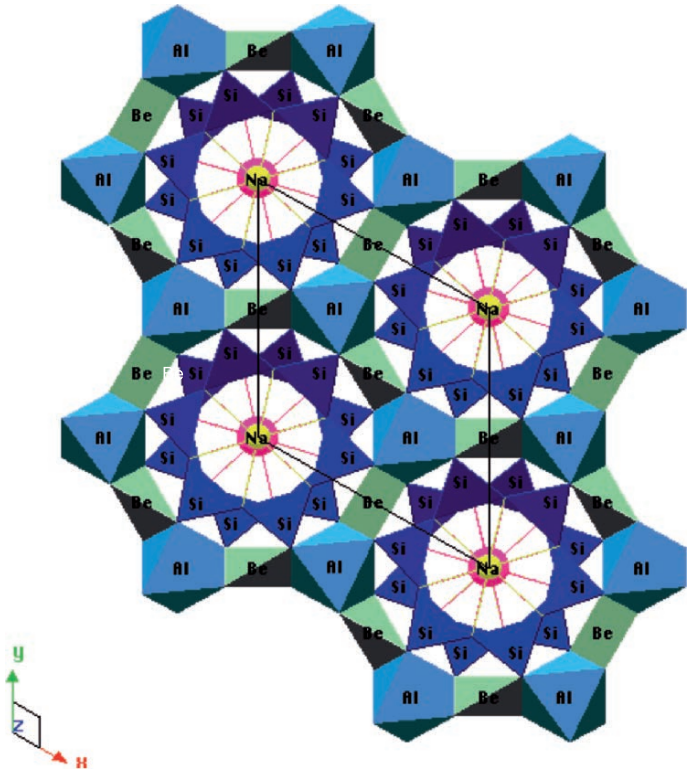


Figure 2.39. Crystal structure of the beryl lattice, showing the location of Al octahedra. Reproduced with permission from http://www.seismo.berkeley.edu/~jill/wisc/Lect9.html

Interestingly, if Cr^{3+} is substituted for Al^{3+} in the $Be_2Al_2Si_6O_{18}$ base lattice, the crystal will appear green rather than red for an Al_2O_3 lattice. Since both coordination spheres about the Cr^{3+} center are distorted octahedra, the shift in the absorption wavelength must result from the lattice structure. In the beryl lattice (Figure 2.39), the Be^{2+} ions will pull electron density away from the oxygen ions, which will cause less electron–electron repulsions between the Cr^{3+} d-orbitals and lone pairs of the oxygen ligands. This will correspond to a decrease in the 10 Dq value, the absorption of lower-energy wavelengths, and a shift of the reflected color from red to green.

Not only does the observed color depend on the nature of the transition metal impurity, but on the oxidation state of the dopant. For instance, the color of a beryl-based crystal changes from blue to yellow, upon doping with Fe^{2+} and Fe^{3+} (Aquamarine and Heliodor), respectively. For Mn^{2+} and Mn^{3+} impurities (Morganite and Red Beryl), the color changes from pink to red, respectively. As a general rule, as the oxidation state of the transition metal ion increases, the ligand ions are drawn in closer to the metal center. This will result in more electron–electron repulsions between the metal and ligand, and a larger 10 Dq. The increase in the energy gap

between d-orbitals causes the absorption of higher energy wavelengths, and a corresponding red shift for the observed/transmitted color. As you might expect, it should be possible to change the color of such a crystal through heating in an oxidizing or reducing environment. This is precisely the operating principle of "mood rings" that respond to differences in body temperature. The color change resulting from a temperature fluctuation is referred to as *thermochromism*. When the application of an external pressure causes a color change, the term *piezochromism* is used.

Another factor that must be mentioned relative to our discussion of color is the wavelength of light used to irradiate the crystal. For alexandrite, another Cr^{3+} doped beryl-based lattice, the crystal will appear green upon exposure to blue light or UV radiation. However, upon exposure to red light, it will appear a deep red color. Whereas the high-energy UV–Visible radiation is enough to promote a d-electron, red light is not sufficiently energetic for excitation, resulting in a simple reflection of the incoming red light through the internal reflection from the multifaceted crystal faces.

Thus far, we have only considered the colors responsible for doping a crystal with one type of metal ion. However, crystals such as blue sapphire contain two metal dopants that yield the desirable deep blue color due to charge-transfer effects. If two adjacent Al^{3+} sites in alumina are replaced with Fe^{2+} and Ti^{4+}, an internal redox reaction may occur, where the iron is oxidized and the titanium is reduced (Eq. 14).

$$(14) \quad Fe^{2+} + Ti^{4+} \rightarrow Fe^{3+} + Ti^{3+}$$

As this process requires energy corresponding to yellow wavelengths of visible light, the complementary color of blue is reflected. Although blue sapphire is an example of heteronuclear charge transfer involving two different transition metal ions, there are also precedents for homonuclear charge transfers involving two oxidation states of the same metal (*e.g.*, Fe^{2+} and Fe^{3+} in magnetite). Lastly, it should be noted that charge transfer may also occur between the metal and ligands. For instance, the blue color of the gemstone lapis lazuli (*i.e.*, $(Na, Ca)_8(Al, Si)_{12}O_{24}(S, SO_4)$), is due to ligand-metal charge transfer initiating from the sulfur atoms in the ligand.

Properties resulting from crystal anisotropy

Crystals are classified as being either *isotropic* or *anisotropic* depending on the equivalency of their crystallographic axes. All crystals that do not belong to the cubic crystal system possess anisotropic symmetry. Since electromagnetic radiation partially comprises an electrical component, the velocity of light passing through a material is partially dependent upon the electrical conductivity of the material. The relative speed at which electrical signals travel through a material varies with the type of signal and its interaction with the electronic structure of the solid, referred to as its *dielectric constant* (ε or κ).

Anisotropic crystals are composed of a complex crystal lattice orientation that has varying electrical properties depending upon the direction of the impinging light ray. As a result, the refractive index will vary with direction when light passes through an anisotropic crystal, giving rise to direction-specific trajectories and velocities. This effect is most easily observed in crystals when there are large differences in the

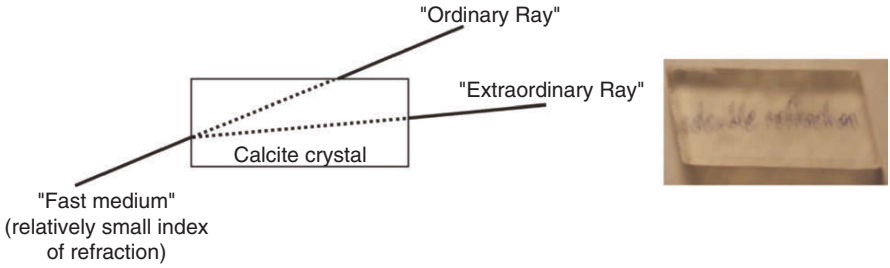


Figure 2.40. Illustration of the double refraction phenomenon exhibited by calcite crystals.

refractive indices of the crystallographic axes. This phenomenon is referred to as *birefringence* and is illustrated by the double refraction exhibited by optical calcite shown in Figure 2.40. Examples of the two distinct refractive indices for representative crystals are calcite (1.6584 and 1.4864), tourmaline (1.669 and 1.638), and rutile (2.616 and 2.903). The birefringence exhibited by a crystal is dependant on the difference in the refractive indices experienced by the extraordinary and ordinary rays as they propagate through a crystal.

Crystals that do not possess a center of symmetry (*i.e.*, noncentrosymmetric) exhibit interesting properties when exposed to pressure – a phenomenon known as *piezoelectricity*. As pressure is applied, the crystal changes shape slightly by the movement of ions. The ionic migration causes some of the positive and negative ions to move in opposite directions, causing a polarization of charge. Conversely, if a piezoelectric crystal is placed in an electric field, the ions move toward opposite electrodes, thereby changing the shape of the crystal. All noncentrosymmetric crystal classes will exhibit piezoelectricity, with the only exception of *432*.

Piezoelectricity is the operating principle of quartz watches. In these devices, a tiny crystal of quartz oscillates at a frequency of 32 kHz in response to an electrical charge generated from the battery. In general, the overall size and composition of a piezoelectric crystal will affect its oscillation frequency. Since quartz loses very little energy upon vibration, the integrated circuit (IC, see Chapter 4) within a watch is used to reduce the repeatable oscillations into electric pulses, which are displayed as hours, minutes, and seconds on the watch face.

Other noncentrosymmetric crystals that alter their shape in response to changes in temperature are referred to as *pyroelectric*. These crystals are used in infrared detectors; as an intruder passes the detector, the body warmth raises the temperature of the crystal, resulting in a voltage that actuates the alarm. Even for such miniscule temperature changes of a thousandth of a degree, a voltage on the order of 15 mV may result, which is readily measured by electronic components. Crystals that exhibit this effect are $BaTiO_3$ (barium titanate), $PbTi_{1-x}Zr_xO_3$ (lead zirconate titanate, PZT), and PVDF (polyvinylidene fluoride). Of the 32 crystal classes discussed earlier, 20 are piezoelectric and ten are pyroelectric.

2.4. THE AMORPHOUS STATE

Thus far, we have focused on solids that have a well-ordered crystalline structure. It is now time to turn our attention to some examples of amorphous solids. We already discussed the synthesis of amorphous metals; those obtained through fast nonequilibrium conditions. However, there are two other classes of materials that exhibit an amorphous structure, that our society is indebted to for countless applications – glasses, ceramics, and cementitious materials. Although the majority of ceramic materials exhibit an amorphous structure when synthesized at low temperatures, these materials are converted to crystalline phases as their temperature is increased, a process referred to as *annealing*. This results in a ceramic material that is extremely hard with a high melting point, desirable for structural applications or those occurring within extreme environments such as high temperatures and/or pressures. In this section, we will describe some important classes of amorphous glasses, ceramics, and cement with a particular focus on the conditions required for, and structural implications of, their conversion to a well-ordered array.

2.4.1. Sol–Gel Processing

The sol–gel (solution–gelation) process is a versatile solution-based process for making ceramic and glassy materials. In general, the sol–gel process involves the formation of a *sol* (colloidal suspension of *ca.* $\geq$ 200 nm solid particles) and subsequent crosslinking to form a viscous *gel*. Though this technique has been in practice since the 1930s, until only recently have the complex mechanisms involved in sol–gel been investigated.[16] The most common starting materials, or *precursors*, used in the preparation of the sol are water-sensitive metal alkoxide complexes ($M(OR)_x$, where R = alkyl group (*e.g.*, CH_3, C_2H_5, CF_3, etc.). Although original formulations used sodium silicates, the use of alkoxide precursors avoids undesirable salt byproducts that may only be removed through long, repetitive washing procedures. In addition, the nature of the metal and associated R groups may be altered to affect the rate and properties of the ultimate oxide material.

Sol–gel syntheses are typically carried out in the presence of polar solvents such as alcohol or water media, which facilitate the two primary reactions of *hydrolysis* and *condensation* (Eqs. 15 and 16, respectively). During the sol–gel process, the molecular weight of the oxide product continuously increases, eventually forming a highly viscous three-dimensional network (step-growth polymerization – Chapter 5).

(15) $\text{M-OR} + \text{H}_2\text{O} \rightarrow \text{M-OH} + \text{ROH}$

(16) $\text{M-OR} + \text{M-OH} \rightarrow [\text{M-O-M}]_n + \text{ROH}$

The most widely used metal alkoxides are $Si(OR)_4$ compounds, such as tetramethoxysilane (TMOS) and tetraethoxysilane (TEOS). However, other alkoxides of Al, Ti, and B are also commonly used in the sol–gel process, often mixed with TEOS. For instance, aluminum silicates may be generated through hydrolysis/condensation of siloxides (Eq. 17), which proceed through an intermediate Al–O–Al network known as *alumoxanes*. Alumoxanes are important for applications in antiperspirants,

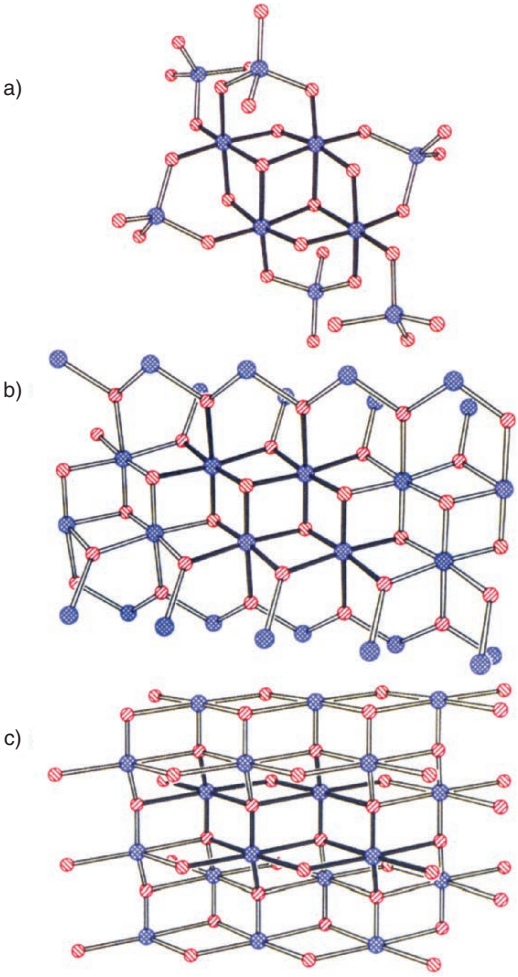


Figure 2.41. Comparison of the core structure of (a) siloxy- substituted alumoxane gels with (b) diaspore, and (c) boehmite minerals. The aluminum and oxygen atoms are shown in blue and red, respectively. Reproduced with permission from *Chem. Mater.* **1992**, 4, 167. Copyright 1992 American Chemical Society.

catalysts, and paint additives. Their structure is best described as Al(O)(OH) particles with a core structure analogous to the minerals boehmite and diaspore (Figure 2.41), and organic substituents on the surface.

$$\mathrm{Al(OSiR_3)_3 + H_2O \rightarrow [Al(O)(OH)_x - (OSiR_3)_{1-x}]_n(gel)}$$

$$\xrightarrow{\Delta} \mathrm{(Al_2O_3)_m(SiO_2)_n} \qquad (17)$$

As one would expect from the similar electronegativities of Si and O, the hydrolysis of silicon alkoxides are significantly slower than other metal analogues. For identical metal coordination spheres and reaction conditions, the general order or reactivity

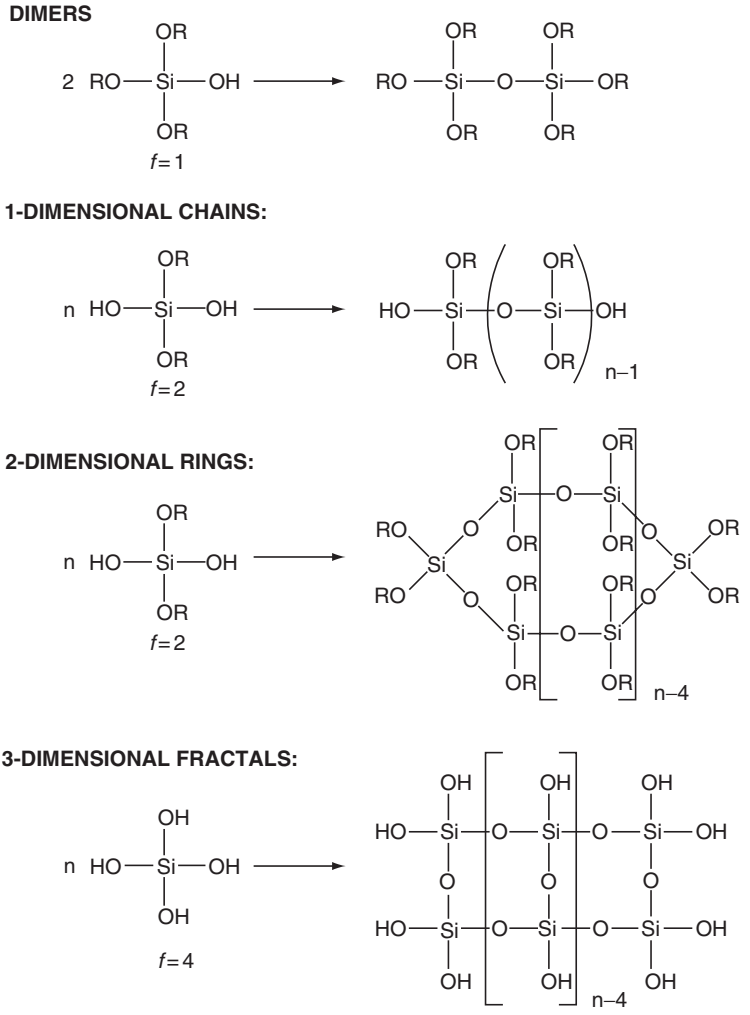


Figure 2.42. Relationship between a siloxy precursor and its control over the three-dimensional shape of a gel. Reproduced with permission from *Chem. Rev.* **2004**, 104, 3893. Copyright 2004 American Chemical Society.

for some common alkoxides is: $Si(OR)_4 \ll Sn(OR)_4 \sim Al(OR)_3 < Zr(OR)_4 < Ti(OR)_4$. That is, larger and more electropositive metals are more susceptible to nucleophilic attack by water. As a result, the hydrolysis of most metal alkoxides is too rapid, leading to uncontrolled precipitation. Although the ratio of $H_2O/M(OR)_n$ may be tuned to control the hydrolysis rate, the nature of the metal alkoxide (*e.g.*, altering the OR groups or metal coordination number) is the most powerful way to control the rate of hydrolysis. It is also possible to control the stepwise hydrolytic pathway, which governs the ultimate three-dimensional structure of the gel (Figure 2.42).

It should be noted that a sol–gel process may also take place through nonhydrolytic pathways. In these systems, a metal halide reacts with an oxygen donor such as ethers, alkoxides, etc. to yield a crosslinked metal oxide product (Eq. 18).

(18) $\text{M-OR} + \text{MX} \rightarrow [\text{M-O-M}]_n + \text{RX}$

In stark contrast to other metal alkoxides, the kinetics for the hydrolysis of $Si(OR)_4$ compounds often require several days for completion. As a result, acid (*e.g.*, HCl, HF) or base (*e.g.*, KOH, amines, NH_3) catalysts are generally added to the mixture, which also greatly affects the physical properties of the final product. Under most conditions, condensation reactions begin while the hydrolytic processes are underway. However, altering the pH, $[H_2O/M(OR)_n]$ molar ratio, and catalyst may force the completion of hydrolysis prior to condensation.

A likely mechanism for an acid-catalyzed system is shown in Figure 2.43. The protonation of the alkoxide group causes electron density to be withdrawn from Si, allowing the nucleophilic attack from water. In contrast, the base-catalyzed hydrolysis of silicon alkoxides proceeds through the attack of a nucleophilic deprotonated silanol on a neutral silicic acid (Figure 2.44). In general, silicon oxide networks obtained via acid-catalyzed conditions consist of linear or randomly branched polymers; by contrast, base-catalyzed systems result in highly branched clusters (Figure 2.45).

As condensation reactions progress, the sol will set into a rigid gel. Since the reactions occur within a liquid alcoholic solvent, condensation reactions result in a three-dimensional oxide network $[\text{M-O-M}]_n$ that contains solvent molecules within

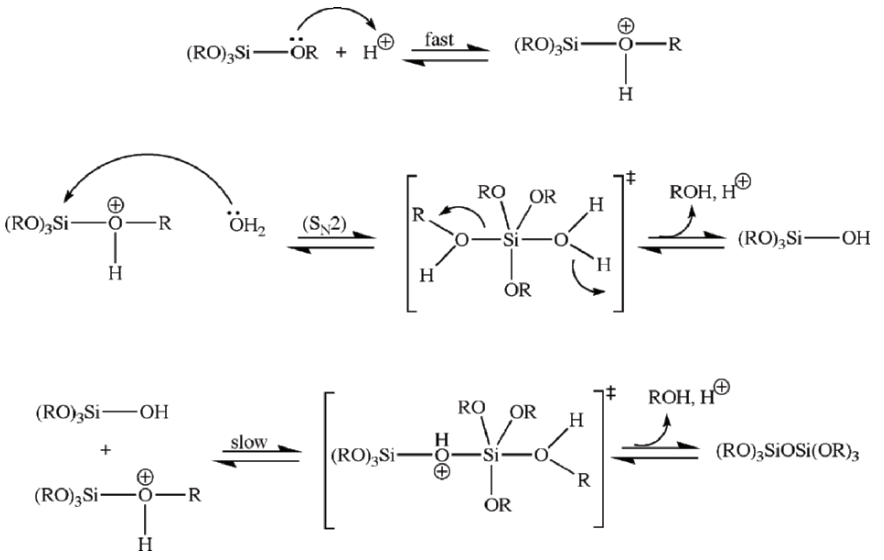


Figure 2.43. Reaction schemes for the acid-catalyzed hydrolysis and condensation of a silicon alkoxide precursor.

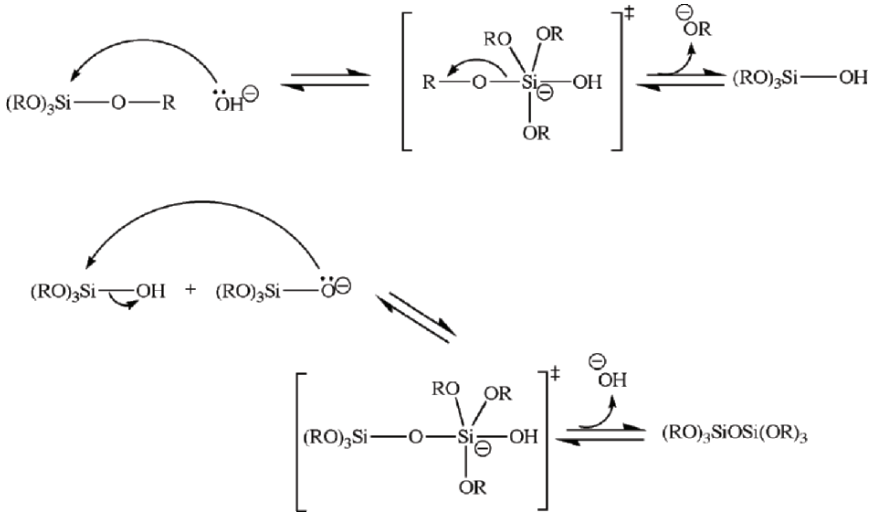


Figure 2.44. Reaction schemes for the base-catalyzed hydrolysis and condensation of a silicon alkoxide precursor.

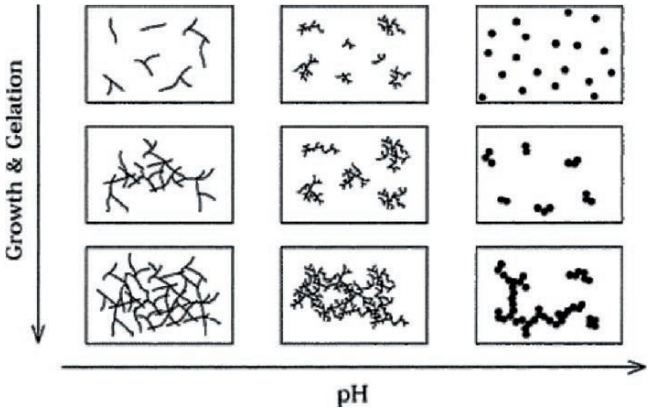


Figure 2.45. Comparison of the morphology with the pH of the sol–gel process. Reproduced with permission from *Chem. Rev.* **2004**, 104, 3893. Copyright 2004 American Chemical Society.

its pores. The product at its gel point is often termed an *alcogel*, recognizing the trapped solvent. At this stage, the gel is typically removed from its original container, but remains saturated with the solvent to prevent damage to the gel through evaporation. It is worthwhile to note that at the gel point, the –O–Si–O– framework still contains a number of unreacted alkoxide moieties. Hence, sufficient time (typically 48 h+) must be allotted to allow for complete hydrolysis and polycondensation, so the network is suitably strengthened to prevent cracking – a process referred to as *aging*, or *syneresis*.

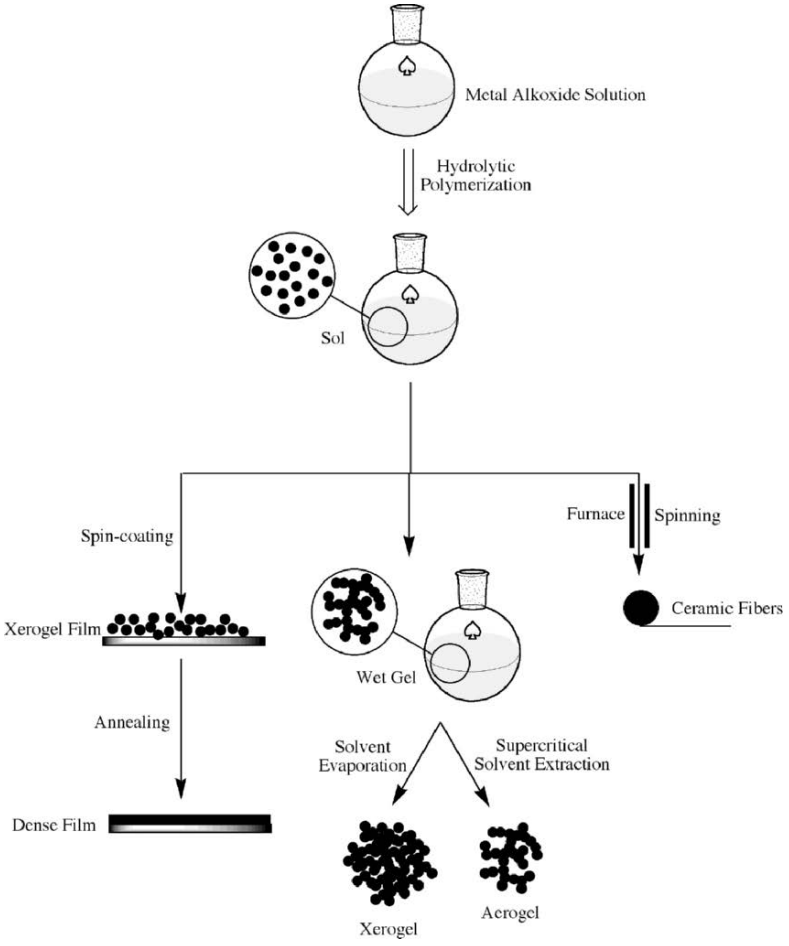


Figure 2.46. Illustration of the products obtained through sol–gel processing.

Depending on the post-treatment used for the sol, a wide variety of materials may be synthesized: ultra-fine powders, thin film coatings, ceramic fibers, microporous inorganic membranes, ceramics and glasses, or extremely porous materials (Figure 2.46). Thin films are easily generated on a substrate through simple spin-coating or dip-coating of the gel, followed by slow evaporation to prevent extensive cracking. Alternatively, the gel may be retained in a mold and heat-treated to convert the material into a dense ceramic or glass. If the solvent of an alcogel is removed through slow evaporation, a porous material known as a *xerogel* is formed. By contrast, if supercritical CO_2[17] is used to remove the solvent, a highly foam-like, porous, transparent material called an *aerogel* is formed. Silica aerogels consist of 99.8% air, and have a density of 1.9×10^{-3} mg L^{-1} and a thermal conductivity of

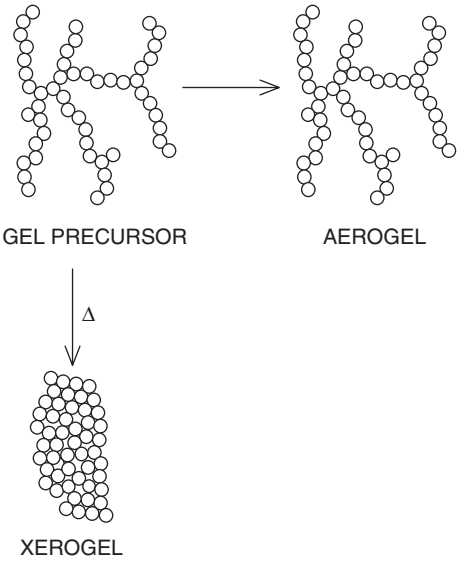


Figure 2.47. Comparison of the three-dimensional shape of an aerogel and xerogel formed from a gel. Reproduced with permission from *Chem. Rev.* **2004**, 104, 3893. Copyright 2004 American Chemical Society.

1.7×10^{-2} W m^{-1} K^{-1}. The properties of aerogels afford a range of applications, among which include sound dampening, catalysis, desiccation, and thermal insulating (*e.g.*, windows, refrigerators, walls).[18] Due to the complex, crosslinked structure of aerogels, the insulating ability is an order of magnitude greater than commonly used fiberglass. As a testament to the unique properties of silica aerogels, the Guinness Book of World Records recognizes this material as the best insulator and least-dense solid.

Aerogels retain the original shape and volume of the alcogel – typically > 85% of the original volume. By contrast, xerogels exhibit significant shrinking and cracking during drying, even under room-temperature conditions (Figure 2.47). It is important that the water be removed prior to the drying event. This is easily accomplished through soaking the alcogel in pure alcohol. The soaking time is dependent on the thickness of the gel. Any water left in the gel will not be removed by supercritical drying, and will lead to a dense, opaque aerogel. Similarly, water will not be removed as readily as alcohol by simple evaporation; hence, water-containing gels will result in heavily cracked and heterogeneous xerogels.

Inorganic gels are rarely used in their as-dried state. The gel is first dehydrated through thermal removal of surface –OH groups, thus preventing rehydration reactions. Most often, this step is followed by high-temperature annealing ($T > 800$°C), in order to convert the amorphous material into a desired crystalline phase. As an example for alumoxanes, it has been shown that the gel is initially transformed to γ-Al_2O_3, *en route* to its highest crystalline form, α-Al_2O_3.[19] During the sintering

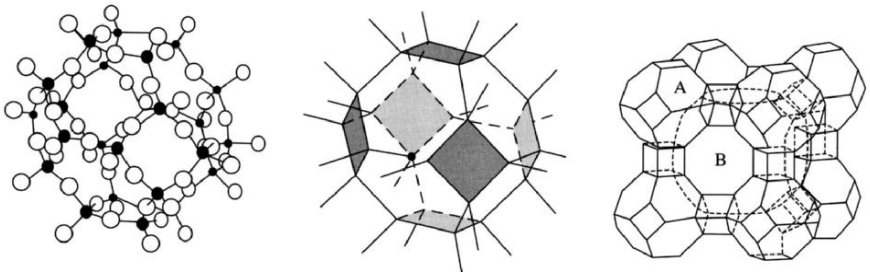


Figure 2.48. Representations of zeolite structures. Shown are molecular and crystal representations of a polyhedron (A) formed from 24 SiO_4 tetrahedra. Also shown is the three- dimensional array of LTA, Linde A: $[Na_{12}(Al_{12}Si_{12}O_{48})]\cdot 27H_2O$ formed from interlocking SiO_4 and AlO_4 polyhedra of (pore size, B: 4.1 Å). Reprinted from Greenwood, N. N.; Earnshaw, A. *Chemistry of the Elements*, 2nd ed., Copyright 1998, with permission from Elsevier.

(or firing) process, the pore size is also diminished and organic moieties are removed. Hence, this results in significantly better mechanical properties, as desired for dense ceramics and glasses.

An extremely useful class of sol–gel synthesized porous materials is called *zeolites*. These materials are best described as hydrated aluminosilicate minerals that consist of interlocking tetrahedrons of SiO_4^{4-} and AlO_4^{5-}. The three-dimensional structure of zeolites is characterized by a complex system of interconnected channels, resulting in low density. By definition, the (Si + Al)/O ratio in zeolites must equal 0.5, resulting in a negatively charged polyhedron structure. Therefore, other cations (*e.g.*, Na^+, K^+, Ca^{2+}, Mg^{2+}, etc.) must occupy the large spaces or cages of the zeolite structure in order to maintain overall charge neutrality. Depending on the type of zeolite, these voids may also be large enough to accommodate larger molecular species such as water, ammonia, and anions such as CO_3^{2-} and NO_3^-. The structural rigidity and channel system of zeolites allow facile movement of ions and molecules into and out of the structure. Hence, zeolites are able to adsorb and desorb water without damage to the crystal structure, and ion-exchange takes place readily among the chelated cations.

There are 48 naturally occurring zeolites (even found on Mars!), and more than 150 synthetic varieties (Figure 2.48).[20] The natural varieties are mostly used in applications such as concrete, soil treatment ("zeoponics" – *e.g.*, controlled release of fertilizer or nutrients such as K^+ or N_2), and even kitty litter that are not affected by their high levels of compositional and structural heterogeneity. However, synthetic zeolites possess a uniform and precise composition and structure, which is best suited for applications such as catalysis, molecular sieves, laundry detergents, and water purification. Both natural and synthetic zeolites are being explored for an intriguing adsorptive application as a blood clotting facilitator (*e.g.*, QuikClot^{TM} and Hemosorb^{TM}).

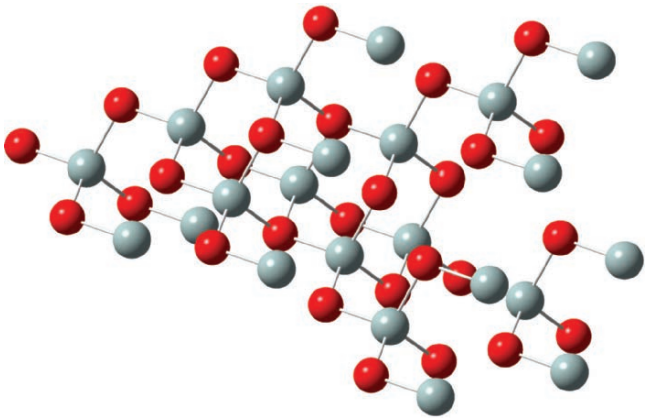


Figure 2.49. Molecular structure of amorphous SiO_2, comprising randomly corner-linked SiO_4 tetrahedra.

2.4.2. Glasses

From drinking vessels and windows to eyeglass lenses, materials comprising glass have long played an important role in our society. In fact, it is estimated that applications for glass date back to Egypt in *ca*. 3,500 B.C. Though we are most familiar with transparent silica-based (SiO_2) glass (Figure 2.49), there are many other types of glass that may be fabricated for various applications. For instance, infrared- transmitting chalcogenide glasses such as As_2E_3 (E = S, Se, Te) are suitable for specialized applications such as optical storage, sensors, and infrared lasers. As we will see in Chapter 3, even metals may be suitably synthesized to possess a bulk-disordered structure. By definition, the term *glass* is actually not a specific material, but a general architectural type – an amorphous solid that has cooled to rigidity without crystallizing. Glasses are most commonly made by rapidly quenching a melt; accordingly, the constituent atoms are not allowed to migrate into regular crystalline lattice positions.

It is noteworthy to point out why a material as disordered as glass is transparent. That is, the amorphous structure of glass should actually facilitate *opacity*, which is the extent to which visible radiation is blocked by the material it is passing through. If we think about single crystals, these will appear translucent if the lattice spacings are smaller than the visible range of wavelengths (*ca*. 300–700 nm). By the same measure, glass also appears transparent since the degree of disorder actually covers a distance less than a wavelength of visible light (*Rayleigh scattering*, Eq. 19). However, if the glass contains imperfections, and/or inclusions of metals or larger particles, it will become increasingly opaque.

$$\text{scattering} \propto \frac{(\Delta\eta)(d^3)}{\lambda^4}, \tag{19}$$

where $\Delta\eta$ is the change in the refractive index and d is the spatial distance covered by the disorder ($d < \lambda$).

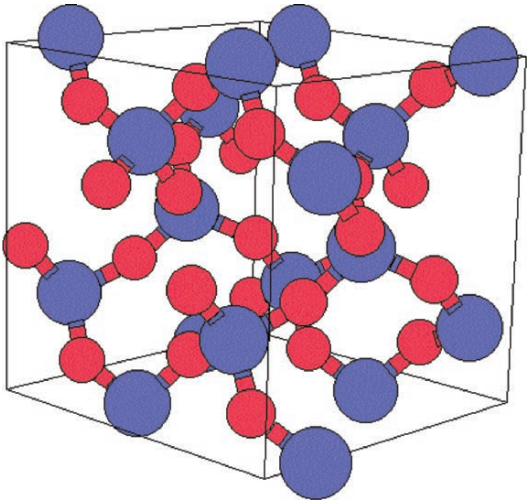


Figure 2.50. Unit cell of the α-quartz crystal lattice. Reproduced with permission from the Naval Research Laboratory – Center for Computational Materials Science website: http://cst-www.nrl.navy.mil/lattice/struk.picts/sio2a.s.png

The most straightforward method to make common SiO_2 glass, known as *fused silica* or *quartz glass*, is through melting sand at a temperature of 1,800–2,000°C. Unlike other glasses, that require a rapid quenching event, quartz will automatically form a glassy solid at all but the slowest cooling rates – a consequence of its complex crystal structure (Figure 2.50). For example, it is estimated to take 100,000 years to form natural quartz! Fused silica is thermally stable at temperatures up to *ca.* 1,665°C. Further, the coefficient of linear expansion is 5.5×10^{-7} cm cm^{-1} K^{-1}; by comparison, the softening point and coefficient of linear expansion for normal window pane glass are *ca.* 500°C and 9.0×10^{-6} cm cm^{-1} K^{-1}, respectively:

Although the mechanical properties of quartz are desirable for high-temperature applications, this glass is relatively difficult to mold into desired shapes through conventional glass-blowing techniques. Quartz glass is transparent toward ultraviolet radiation ($\lambda = ca.$ 190–300 nm), indicating that the spatial range of structural disorder is less relative to other glasses that contain additional additives (*vide infra*). As a result, quartz windows are used for ultraviolet lamps that are employed in a number of important applications in chemistry, biology, engineering, and materials science.

The chemistry of glass making is now a mature field, with many types available through combination of oxides. In order to decrease the prohibitively high melting point of SiO_2, *ca.* 18% of sodium carbonate ("soda," Na_2CO_3) is often added to sand. This is another useful exploitation of the freezing-point depression *colligative property*, as taught in introductory physical chemistry (*e.g.*, adding salt to icy roads in the winter). The Na^+ ions are detrimental to the finished glass, since they are easily solvated by water leading to corrosion. To prevent such weathering, *ca.* 10% of limestone ($CaCO_3$) is added to effectively replace the Na^+ ions with Ca^{2+}. When

this mixture is heated to its melting point (*ca*. 1,000°C), a mixture of calcium silicate ($CaSiO_3$) and sodium silicate (Na_2SiO_3) results. Upon cooling, the most prevalent type of glass, called "crown glass" or *soda–lime* glass, is generated. This type of glass accounts for over 90% of the glass used worldwide. Interestingly, our current synthetic procedure has not deviated from the earliest glassmaker's recipe, which also used white stone pebbles (quartz, SiO_2) and plant ash containing sodium- and calcium-based additives (Na_2O and lime (CaO)).

There are a number of other glass recipe variations that may be used to yield desired properties. Most likely, these formulations were discovered by accident or in a trial-and-error manner, using materials from their locale and measuring the resultant properties. For instance, the Europeans were the first to discover that K_2O, obtained locally from plant ash, could also be combined with lime and quartz to yield a potash-lime glass, later exploited for stained-glass windows. Another popular variation substitutes boric oxide (B_2O_3) for lime and soda to yield *borosilicate glass*. The physical properties of this glass resemble fused silica (*e.g.*, coefficient of thermal expansion: 3.3×10^{-7} cm cm^{-1} K^{-1}), except that its softening temperature is only *ca*. 700°C. Borosilicate glass is the variety that is sold in stores as Pyrex^{TM} cookware and laboratory equipment. These applications demand a glass that resists thermal expansion (*i.e.*, cracking) as a result of significant changes in temperature.

It was not until the late 17th century that PbO was substituted for lime in glass formulations. This "soda–lead" glass is what we know as *crystal* (referred to as *flint glass* in pre-Civil War America), and has always been a symbol of wealth and extravagance such as expensive glassware and chandeliers. In order for crystal to be legally given the "full lead" designation, at least 24% of lead oxide must be present in its structure. The addition of the heavy element lead adds significant weight to the glass, while increasing its refractive index. This latter property results in the familiar clear, sparkling appearance of crystal glassware. The presence of lead also makes the glass softer than regular types that must be cut with a diamond saw. Black crystal is truly one of the most fabulous materials for modern artistic design. The lack of transparency is caused by a combination of additives – typically Fe_2O_3, CuO, MnO_2, and Co_2O_3.

Colored glass has been used since the construction of the first churches, prior to the 10th century. Although decorative applications represent the majority of uses for colored glass, there are some other recent functional applications such as traffic light signals. The colors imparted by glass are a result of dopant species that are added during its fabrication (Table 2.7). Both transition metal ions and colloidal suspensions yield an observable color, with the hue dependent on the concentration used. Variation of the color and intensity is also extremely sensitive toward the heating regime (both temperatures and exposure times) used during the glassmaking process. In general, the observed color is the complement of the color that is absorbed by the ion. That is, the absorption of short wavelengths will result in an observable red color.

For colloidal dopants, the particle size must be smaller than a wavelength of visible light, or an opaque glass will result. If one would prefer a cloudy glass, a number of additives (*e.g.*, SnO_2, TiO_2, CaF_2) may be used that result in a suspension that

Table 2.7. Colors of Glass Resulting From Doping

Additive	Color
Co_2O_3	Blue
Fe_2O_3	Yellow-green
FeO	Bluish-green
Colloidal Se[a]	Red
Colloidal Au[a]	Red
Colloidal Cu[a]	Red
CuO	Turquoise
NiO	Blue/violet/black
SnO_2	White
Sb_2O_3, As_2O_3	White
TiO_2	Yellow-brown
UO_2	Fluorescent yellow/green
$AgNO_3$	Orange-red
PbO/Sb_2O_3	Yellow (opaque)
$K_2Cr_2O_7$	Dark green/black
Mn_2O_3	Purple

[a]With average particle diameters of *ca.* 50–100 nm.

changes the overall index of refraction. Colloidal metals yield a deep red color, with colloidal gold first used in the late 17th century. Alternatively, a metal salt such as $AuCl_3$ may be added to glass followed by thermal or chemical (*e.g.*, using $NaBH_4$) reduction to metallic Au. It is important to note that a red color will only result if an agent is also added to prevent particle agglomeration. In general, the observed color will shift toward the blue portion of the spectrum as the average particle size decreases (*e.g.*, blue color results from diameters of <50 nm). Chapter 6 will provide more details related to the scattering properties and other applications of nanoparticles.

In comparison, transition metals are added to a molten glass matrix as soluble oxides. As you may see from Table 2.7, the observed color is a consequence of the metal ion type/concentration, as well as its oxidation state. To obtain a desired color, oxidizing agents such as $NaNO_3$, or reducing agents such as carbon powder may be added to afford the desired oxidation state. An intriguing form of glass, referred to as *vaseline glass* contains UO_2 and is slightly radioactive. Since UV radiation is sufficient to excite the weakly bound outer electrons of U, this additive results in a fluorescent green color. Although this is observable under normal light, it is most pronounced upon irradiation with a UV lamp. Interestingly, UO_2 was also added to ceramic glaze to yield bright orange dinner plates and tableware in the 1930s. However, it was later discovered that heat and acidic foods caused uranium to leach from the glaze, resulting in an immediate disband of this application. As one might expect, UO_2-doped materials are not currently manufactured for decorative applications, making such acquisitions a collector's item.[21]

Thus far, we have considered the varying chemical compositions and properties of glasses. In this last section, we will examine an important architecture – glass fibers, of paramount importance in our society. The synthesis of glass fibers dates back to

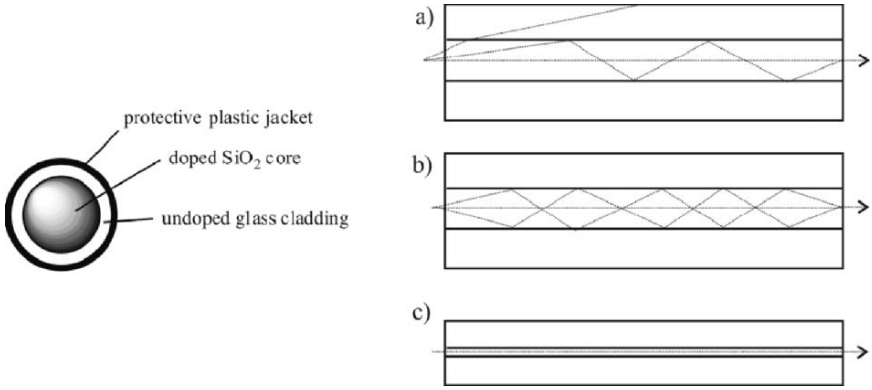


Figure 2.51. Cross-section image of an optical fiber. Shown is the propagation of light waves through an (a) step-index multimode fiber, (b) graded-index multimode fiber, and (c) single-mode fiber.

the early 18th century, and applications for surgical lamps were prevalent as early as the 19th century. We are all familiar with the bright pink bags of fiberglass insulation that may be purchased from home improvement stores. In fact, insulation represents the leading application for fibrous glass materials.

In contrast to fiberglass, which consists of a disordered array of needle-like fibers, extremely long, one-dimensional cylindrical glass structures may be carefully fabricated to allow the transmission of light from one end to the other. Although the diameters of optical fibers are less than a human hair (*i.e.*, 8–10 μm), these fibers are stronger than steel and are able to withstand being buried underground. The fiber consists of a *core* surrounded by a *cladding* layer (Figure 2.51). The silica core is doped with other oxides such as B_2O_3, GeO_2, and P_2O_5, resulting in a slightly higher refractive index than the cladding. The boundary between the core and cladding may either be abrupt (*step-index fiber*), or gradual (*graded-index fiber*).

There are two types of optical fibers: single-mode or multiple-mode, referring to the simultaneous transmission of single or multiple light rays, respectively. Single-mode fibers have much smaller diameters than multimode analogues (Figure 2.51a–c), resulting in a simpler pathway for light through the fiber. Whereas multimode fibers are used to transmit information short distances (*e.g.*, LAN applications), single-mode fibers are used for the long-distance transmission of cable television and telephone signals. It is not hard to see why fiber optics are desirable for transmission applications. The amount of information transmitted through 0.25 lb of optical fiber would take 33 tons of copper wire! Further, optical fibers are able to transmit signals at lightning speeds – unmatched by any other material. For example, three half-hour T.V. episodes may be transmitted in 1 s.

In a step-index fiber, wavelengths of light are transmitted through the fiber core by total internal reflection (Figure 2.51a). Depending on the angle of incidence of the light, the rays may either be transmitted through the fiber to the detector, or refracted from the core into the cladding resulting in signal loss. The desired angle (*acceptance*

angle) required for total internal reflection is determined by the difference in index of refraction between the core and cladding materials. In a graded-index fiber, the refractive index in the core decreases continuously between the axis and the cladding. This causes the light rays to bend smoothly as they approach the cladding, rather than reflect abruptly from the core-cladding boundary.

A fiber optic system typically consists of a transmitting device that generates the light signal, a fiber cable that transmits the light, and a receiver. The information (voice, data, or video) is encoded into electrical signals. At the light source, these electrical signals are converted into either digital or analog light signals. Once the signals are converted to light, they travel down the fiber until they reach a detector, which changes the light signals back into electrical signals. Finally, the electrical signals are decoded into the original voice, data, and/or video information.

The most common method to make optical fibers is heating a rod (preform), of the desired refractive index, to temperatures of *ca.* 2,000°C. The preform is made from the high-temperature reaction of $SiCl_4$ in the presence of dopant gases such as BCl_3, $GeCl_4$. Once the tip of the preform is melted, it falls by gravity to form a thin strand. This wire is threaded through a coating reel, and then pulled into an optical fiber of the desired diameter. The draw towers used for this process are impressive buildings, often 8–10 stories in height. The speed of the pulling process (typically 10–20 m s^{-1}) governs the ultimate diameter of the fibers. For subsequent applications, the fiber is spooled onto shipping reels and cut to the desired length.

Another interesting application for glasses is for light control, referred to as "smart glass." We are all familiar with movie scenes where a top-secret meeting takes place, and a flip of the switch instantly darkens the windows. More routinely, it is now commonplace to have self-dimming mirrors that react to trailing vehicle headlights. Two main technologies are responsible for these intriguing materials applications: *electrochromic devices* (ECDs) and *suspended-particle devices* (SPDs).

As its name implies, electrochromic materials change color as a result of an injection of electrons. The typical ECD has a number of layers, sandwiched between glass (Figure 2.52a). When no voltage is applied to the device, the incoming light

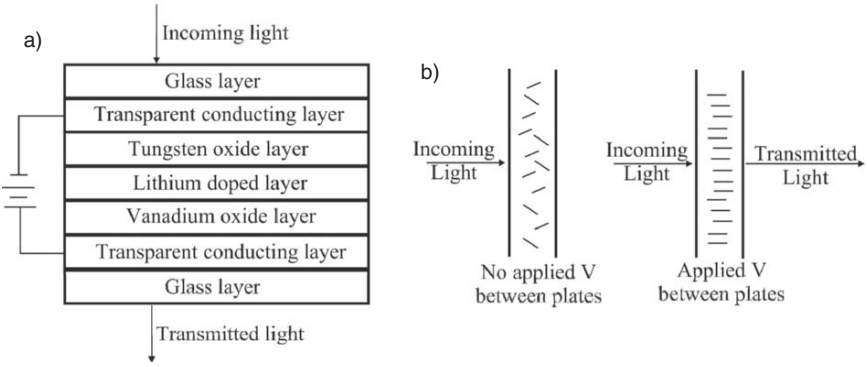


Figure 2.52. Cross-section schematic of an (a) electrochromic device and (b) suspended-particle device.

will pass through undisturbed (*ca*. 70–80% transmittance). However, when a negative voltage is applied, the positive Li^+ ions are injected into the WO_3 layer of the distorted perovskite structure. A redox reaction takes place, where some of the tungsten sites are reduced from W^{6+} to W^{5+} and an electron is placed into the conduction band. Charge balance is maintained through the interstitial placement of Li^+ ions in the lattice. Since the electron becomes delocalized, metallic behavior is induced in the tungsten oxide layer changing the transparent layer to a dark reflective color (*ca*. 10% transmittance of incoming light through the device). The dark color will remain even if the applied voltage is removed, since the reverse reactions are not spontaneous. If the reverse bias is applied to the device, lithium ions flow from the WO_3 layer reoxidizing the W^{5+} ions and restoring transparency.

More recently, thin films of Ni/Mg hydride alloys have also been developed for light attenuation using electrochromic or gas-chromic (injection of H_2 and O_2 gases) technology.[22] Although they can technically be classified as electrochromic materials, the new reflective hydrides that are being developed behave in a noticeably different way. Instead of absorbing light, they reflect it. Thin films made of nickel–magnesium alloy are able to switch back and forth from a transparent to a reflective state. The switch can be powered by low voltage electricity (electrochromic technology) or by the injection of hydrogen and oxygen gases (gas-chromic technology). Furthermore, this material has the potential to be even more energy efficient than other electrochromic materials.

By comparison, SPDs operate through the behavior of rod-like particles (*e.g.*, liquid crystals, see Appendix C.3) toward an applied voltage (Figure 2.52b). When no voltage is applied, the particles are randomly aligned, and do not allow light to pass through the device. However, an electric charge will polarize the particles to align with the field. We will describe the molecular behavior of polarizable particles in more detail later (Chapter 4), related to dielectric materials placed in a parallel plate capacitor.

2.4.3. Cementitious Materials

The use of cementitious materials for structural applications dates back to ancient Egypt. A type of cement was used to hold together the limestone blocks of the great pyramids that still stand today. During the time of the Roman Empire, an improvement of the cement formulation was developed, which used a fine, volcanic ash known as Pozzolana found in various parts of Italy. Although they did not realize it at the time, the hardening process occurred due to the reaction of the aluminosilicate-based ash with $Ca(OH)_2$ in the presence of water to yield a calcium–silicate–hydrate (CSH) rigid gel. Amazingly, thousands of years later, the CSH structure is not yet completely understood – it is likely a disordered form of the hydrated calcium silicate mineral tobermorite (Figure 2.53).

The last major development in cement technology occurred in the early 19th century in England. Bricklayer Joseph Aspdin first made a variety of cement known as *Portland cement* – not in a laboratory, but on his kitchen stove! His patent in

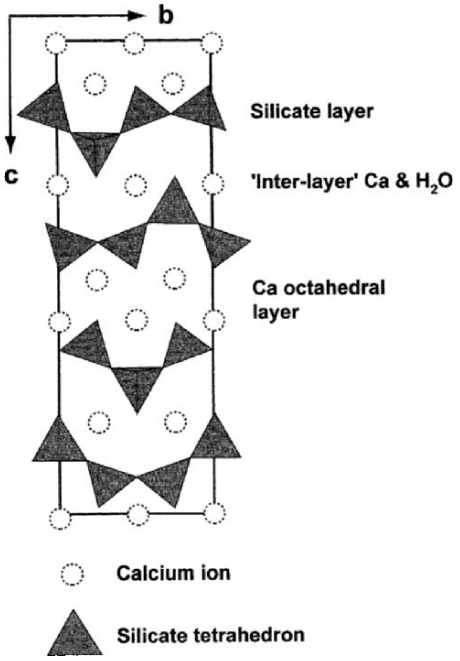


Figure 2.53. The crystal structure of tobermorite, viewed along the bc plane. The amorphous gel produced during cement formation likely contains defects such as missing/disordered silicate tetrahedra and/or water sites. Reprinted from *Chem. Geol.* **2000**, 167, 129, Copyright 2000, with permission from Elsevier.

1824 changed the world forever, as this form of cement is the basic ingredient in concrete – essential for the erection of virtually all buildings and many roads throughout the world. In fact, concrete is the most heavily used man-made material in our world – in 2005, it was estimated that the worldwide annual production of concrete amounts to 1 ton for every man, woman, and child on earth!

It is interesting to note the development timeline for cement/concrete, which has addressed many important societal needs. For example, if we consider road construction, stones were used as early as 4,000 B.C., and were still prevalent in early America – still evident in some historical cities such as Boston, MA. As motorcars became more abundant and faster in the early 20th century, replacements were sought for dirt, gravel, and stone roads. Asphalt[23] and simple road tarring were the first alternatives that addressed the environmental problem of dust as well as road smoothness, and the majority of highways across North America still comprise asphalt pavement. Comparatively, a large proportion of city streets and highways, especially in the warmer locales, are made from concrete. Since concrete is more apt to crack relative to asphalt under significant temperature changes, this material is not greatly used to construct roadways in northern climates (*e.g.*, North Dakota, Montana, Minnesota, Michigan, Canadian Provinces).

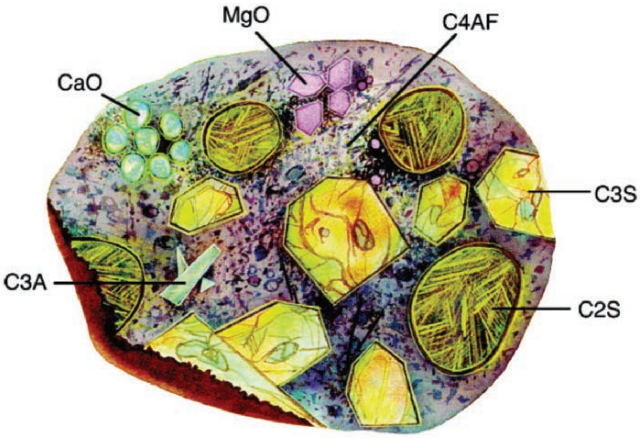


Figure 2.54. Cross-section representation of a powdered cement particle. Dicalcium silicate (Ca_2SiO_4), tricalcium silicate (Ca_3SiO_5), tricalcium aluminate ($Ca_3Al_2O_6$), and tetracalcium aluminoferrite ($Ca_4Al_nFe_{(2-n)}O_7$) crystallites are abbreviated as C2S, C3S, C3A, and C4AF, respectively. Reproduced with permission from *Chem. Mater.* **2003**, 15, 3074. Copyright 2003 American Chemical Society.

Portland cement is produced from the sintering of minerals containing $CaCO_3$, SiO_2, Al_2O_3, Fe_2O_3, and MgO in a ceramic kiln, held at a temperature of *ca.* 1,500°C. Equations 20–24 show the reactions that occur during the processing of cement. The resulting complex material is referred to as *clinker* (Figure 2.54), and may be stored for many years under anhydrous conditions before its use in concrete. As one can see from Eq. 20, this process releases the greenhouse gas CO_2, causing environmental concerns over the world's increasing production of cement. It is estimated that Portland cement manufacturing accounts for over 10% of the world's total emission of CO_2. As a result, there is an increasing focus on cements fabricated with fly ash (*e.g.*, by-product from coal-fired power plants), that possess the same qualities as Portland cement clinker without the need for the calcination step (Eq. 20). When water is mixed with Portland cement, the product sets in a few hours and hardens over a period of 3–4 weeks. The initial setting reaction is caused through reaction between water, gypsum ($CaSO_4 \cdot 2H_2O$ – added to clinker to control the hardening rate), and C3A forming calcium and aluminum hydroxides. The crystallization of these hydroxides results in an observable hardening within the first 24 h. The subsequent reactions between the formed hydroxides and C3S result in formation of a crosslinked CSH gel (*vide supra*); this provides further strengthening over the first week. The reaction between hydroxides and C2S proceed the slowest, and result in hardening/strengthening of the material in latter stages of setting.[24] The hardening of concrete is actually a consequence of individual grains of *aggregates* being cemented together by the clinker–water byproducts. Aggregates that are present in concrete are of two or more size distributions – typically coarse gravel/stones and fine sand. As a general rule, 1 m^3 of concrete contains over 4,400 lb of gravel and sand.

(20) $CaCO_3 \rightarrow CaO + CO_2$

(21) $3CaO + SiO_2 \rightarrow Ca_3(SiO_5)$"C3S"

(22) $2CaO + SiO_2 \rightarrow Ca_2(SiO_4)$"C2S"

(23) $3CaO + Al_2O_3 \rightarrow Ca_3Al_2O_6$"C3A"

(24) $4CaO + Al_2O_3 + Fe_2O_3 \rightarrow Ca_4Al_2Fe_2O_{10}$"C4AF"

IMPORTANT MATERIALS APPLICATIONS I: FUEL CELLS

By definition, a *fuel cell* is any device that generates electricity by chemical reactions that occur at the cathode (positively charged) and anode (negatively charged). Figure 2.55 illustrates the general operating principle of fuel cells. As opposed to batteries that store a limited amount of energy, fuel cells operate with a continuous fuel flow that allows prolonged periods of electricity generation. In addition, these systems may be easily scaled-up to power large electrical grids.

An electrolyte is an essential component within fuel cells, used to facilitate the selective migration of ions between the electrodes. Fuel cells are typically classified according to the electrolytes used: alkaline fuel cell (AFC), polymer electrolyte (or proton exchange membrane) fuel cell (PEMFC), phosphoric acid fuel cell (PAFC),

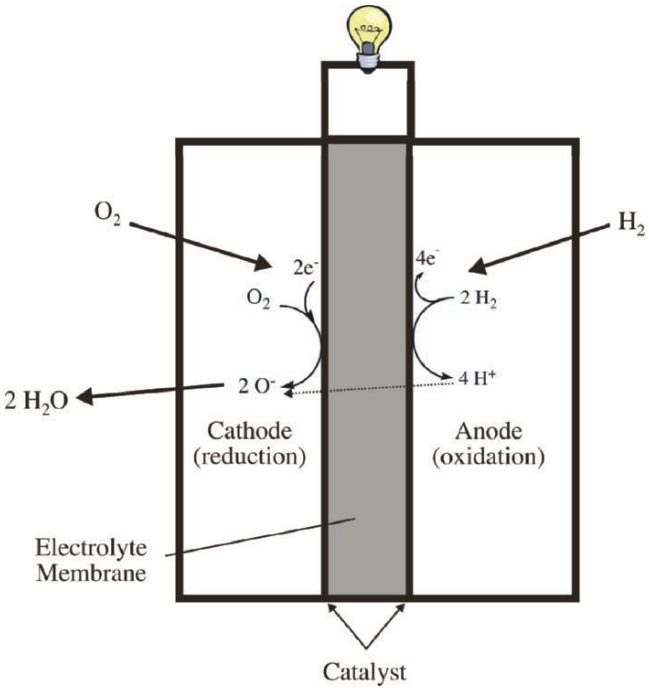


Figure 2.55. Illustration of the operation of *proton exchange membrane* (PEM) fuel cell. In this design, the electrolyte facilitates the transfer of protons across its membrane.

Table 2.8. Comparative Data for Fuel Cells

Fuel cell	Operating temperature (°C)	Efficiency (%)	Output current (kW)
Alkaline	100–200	70	0.3–5
Polymer	80	40–50	50–250
Phosphoric acid	150–200	40–80	200–11,000
Molten carbonate	650	60–80	2,000+
Solid oxide	800–1,000+	60	100

molten carbonate fuel cell (MCFC), and solid oxide fuel cell (SOFC). Typical efficiencies, operating temperatures and output voltage for the various types of fuel cells are shown in Table 2.8. It should be noted that none of these fuel cells are yet cheap/efficient enough to widely replace traditional ways of generating power, such as coal- and natural gas-fired, hydroelectric, or nuclear power plants. Fuel cells are currently being tested and marketed for the replacement of traditional internal combustion engines in automobiles. The high efficiencies and low emissions of fuel cells are extremely intriguing, but problems with emission-free production, and safe storage of hydrogen gas remain the primary stumbling blocks for widespread incorporation of this technology.

Alkaline fuel cells have been used the longest, since the 1960s by NASA for space shuttles. In fact, this application illustrates the utility of fuel cells. Hydrogen and oxygen gases are used to power the fuel cell, which powered the electrical components of the space shuttle. Water, the only byproduct of the reaction, was used as onboard drinking water for the crew. Although AFCs are the most inexpensive to produce and have some of the highest efficiencies, they require high-purity oxygen to prevent catalyst poisoning by carbon dioxide. PEMFC designs also suffer from catalyst poisoning. The presence of small concentrations of CO (>1 ppm) in the reformate of fuels drastically alters the performance of the anodic catalyst (*e.g.*, Pt, Pt/Ru).

In this section, we will consider SOFCs, the most relevant to our discussion of crystal structures. These ceramic-based fuel cells are perhaps best suited for large-scale stationary power generators that could provide electricity for entire cities. This is a consequence of its high operating temperature, often greater than 1,000°C. In addition to the primary electricity generated from the electrodes within the fuel cell, the steam byproduct may be used to turn turbines, generating secondary electricity. The SOFC can operate on most any hydrocarbon fuel, as well as hydrogen gas:

(25) $H_2 + O^{2-} \rightarrow H_2O + 2e^-$ (at anode),

(26) $1/2O_2 + 2e^- \rightarrow O^{2-}$ (at cathode).

The electrochemical reactions occurring within a SOFC are shown in Equations 25 and 26. The anode consists of a porous mixture of a Ni or Co catalyst on yttria-stabilized zirconia. Such a mixture of metal and ceramic is referred to as a *cermet*. The zirconia acts to inhibit grain growth of the catalyst particles of nickel or cobalt and protects against thermal expansion. The cathode is generally a Sr-doped

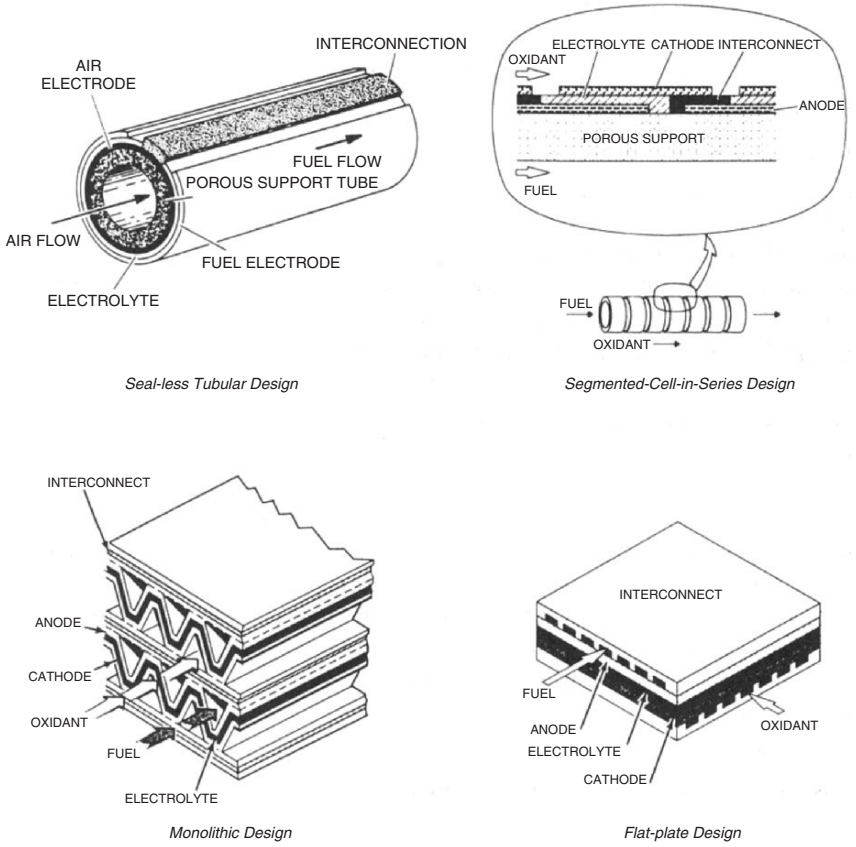


Figure 2.56. Types of solid oxide fuel cell (SOFC) designs. Reproduced with permission from *Electronic Materials Chemistry*, Bernhard Pogge, H. ed., Marcel Dekker: New York. Copyright 1996 Taylor & Francis.

$LaMnO_3$ perovskite. The Sr dopant provides for oxygen transfer to the cathode–electrolyte interface.

Figure 2.56 shows a variety of stacked cell designs employed by SOFCs. Since an individual fuel cell produces a low voltage (typically <1 V), a number of cells are connected in series forming a fuel cell *stack*. An interconnect comprising a high-density material is used between the repeating {anode–electrolyte–cathode} units of the stack.

The electrolyte in a SOFC must have the following characteristics:

1. High oxide conductivity
2. Stability in both oxidizing and reducing environments
3. Chemical compatibility with other cell components
4. High density to prevent mixing of fuel and oxidant gases
5. Desirable thermal expansion properties, to prevent cracking of the fuel cell at high temperatures

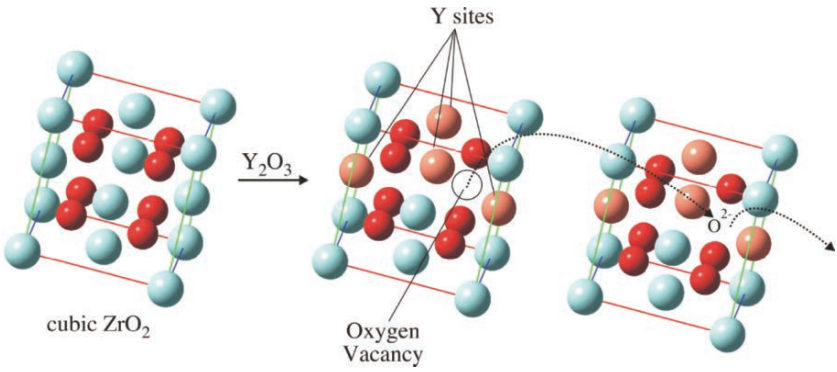


Figure 2.57. Comparison of unit cells of cubic zirconia and yttrium stabilized zirconia (YSZ), showing transport of oxide ions through a lattice via vacant sites in neighboring YSZ unit cells.

One such material that satisfies all of the above requirements is yttrium-stabilized ZrO_2. In a SOFC, the electrolyte must allow oxide-ion transport between the anode and cathode. In its high-temperature cubic form, zirconia is not able to conduct oxide ions. That is, there are no suitable interstitial sites in the lattice to trap oxide ions. However, when a lower valent metal oxide such as Y_2O_3 (*i.e.*, Y^{3+}) is substituted for ZrO_2 (*i.e.*, Zr^{4+}), a vacancy is created in the unit cell of the extended lattice (Figure 2.57). This site is able to accept an O^{2-} ion generated at the cathode, and deliver it to the anode where it is transformed to H_2O. As illustrated in Figure 2.57, the oxide ion is transferred among adjacent zirconia unit cells en route from the cathode to anode compartments of the fuel cell. By contrast, the "holes" created by Y^{3+} substitution may be envisioned to move in the opposite direction, from the anode to cathode.

References and Notes

1 Zallen, R. *The Physics of Amorphous Solids*, John Wiley and Sons: New York, 1983.

2 Electrons in an s-orbital have a finite probability of being found at the nucleus. As the principal quantum number increases, the s-orbitals become more diffuse, leading to electrons being found at distances further from the nucleus. With less attraction toward the nucleus, these electrons are able to orbit the nucleus at speeds approaching the speed of light. When objects move at such high speeds, an increase in relativistic mass occurs, whereby the s-electrons behave as though they were more massive than electrons moving at slower speeds. This mass increase causes the orbiting electrons to be slightly contracted toward the nucleus, decreasing their availability to participate in chemical reactions.

3 Schroers, J.; Johnson, W. L. "History Dependent Crystallization of $Zr_{41}Ti_{14}Cu_{12}Ni_{10}Be_{23}$ Melts" *J. Appl. Phys.* **2000**, *88*(1), 44–48, and references therein. More information may be obtained from http://www.liquidmetal.com.

4 There is an ongoing debate whether the term "pseudopolymorph" should be abandoned, instead designating these compounds as "solvates". Two viewpoints may be found at: (a) Bernstein, J. *Cryst. Growth Design* **2005**, *5*, 1661. (b) Nangia, A. *Cryst. Growth Design* **2006**, *6*, 2.

5 Note: the Miller indices for the (211) plane may also be visualized by extending the unit cell beyond a cell volume of 1 cubic unit. For instance, this plane would also pass through (2,0,0), (0,2,0), and

(0,0,2), as well as other extended coordinates. For the (001) plane, the zeroes indicate that the plane does not intercept either the *a* or *b* axes).

6 Cullity, B. D. *Elements of X-ray Diffraction*, 2nd ed., Addison-Wesley: Reading, Massachusetts, 1978.

7 The effective nuclear charge is defined as the actual nuclear charge felt by a particular valence electron. It is expressed as $Z_{\text{eff}} = Z - \sigma$, where Z is the nuclear charge for the atom and σ is the "screening constant." This latter term corresponds to the number of core electrons, and the effectiveness of the orbitals to shield core electron density.

8 Willard, M. A.; Nakamura, Y.; Laughlin, D. E.; McHenry, M. E. *J. Am. Ceram. Soc.* **1999**, *82*(12), 3342.

9 Nakamura, Y.; Smith, P. A.; Laughlin, D. E.; De Graef, M.; McHenry, M. E. *IEEE Trans. Magn.* **1995**, *31*(6), 4154.

10 (a) Jansen, M.; Letschert, H. P. *Nature* **2002**, *404*, 980.
(b) Kasahara, A.; Nukumizu, K.; Hitoki, G.; Takata, T.; Kondo, J. N.; Hara, M.; Kobayashi, H.; Domen, K. *J. Phys. Chem. A* **2002**, *106*, 6750.
(c) Hitoki, G.; Takata, T.; Kondo, J. N.; Hara, M.; Kobayashi, H.; Domen, K. *Chem. Commun.* **2002**, 1698.

11 (a) Honle, W. *J. Solid State Chem.* **1983**, *49*, 157.
(b) Perrin, et al. *Acta Crystallogr.* **1983**, *C39*, 415.

12 A nice updated website for past/recent superconductor discoveries is found at http://superconductors.org/type2.htm

13 The *d* notation indicates a diamond glide plane, found in diamond or zinc blende extended crystal structures. Whereas glide planes are found in many inorganic-based crystals, screw axes are found predominantly in protein structures.

14 Garlick, G. D.; Kamb, W. B. *J. Geol. Educ.* **1991**, 39, 398.

15 For a thorough treatment of crystal field theory, see Cotton, F. A.. Wilkinson, G.; Murillo, C. A.; Bochmann, M. *Advanced Inorganic Chemistry*, 6th ed., Wiley: New York, 1999.

16 For a recent review, see Cheetham, A. K.; Mellot, C. F. *Chem. Mater.* **1997**, *9*, 2269. It should be noted that the hydrolytic condensation of trifunctional organosilicon monomers (*e.g.*, $RSiCl_3$ or $RSi(OMe)_3$) results in polyhedral oligomeric silsesquioxanes (POSS) – see: http://www.azonano.com/details.asp?ArticleID=1342#_POSSTM_Polymers_Polymerization/Gr. These structures represent the smallest forms of silica, often being denoted as "molecular silica". Since particle diameters range from 0.07 to 3 nm, these are important architectures for nanoapplications (*e.g.*, see http://www.reade.com/Products/Polymeric/poss.html

17 A supercritical fluid has intermediate properties of liquid and gas. Typically, the alcogel is placed in an autoclave filled with ethanol. The system is pressurized to 750–850 psi with CO_2 and cooled to 5–10°C. Liquid CO_2 is then flushed through the vessel until all the ethanol has been removed from the vessel and from within the gels. When the gels are ethanol-free, the vessel is heated to a temperature above the critical temperature of CO_2(31°C). As the vessel is heated, the pressure of the system rises. The pressure of CO_2 is carefully monitored to maintain a pressure slightly above the critical pressure of CO_2 (1050 psi). The system is held at these conditions for a short time, followed by the slow, controlled release of CO_2 to ambient pressure. The length of time required for this process is dependent on the thickness of the gels; this process may last anywhere from 12 h to 6 days.

18 For a recent review on the synthesis, properties, and applications of aerogels see: Pierre, A. C.; Pajonk, G. M. "Chemistry of Aerogels and Their Applications", *Chem. Rev.* **2002**, *102*, 4243.

19 The γ-Al_2O_3 crystal structure is best described as a defect spinel structure; an FCC array of O^{2-} ions, with Al^{3+} ions in $21\frac{1}{3}$ of the 16 octahedral and eight tetrahedral sites. By contrast, α-Al_2O_3 is an HCP array of O^{2-}, with Al^{3+} in 2/3 of the octahedral sites.

20 For a comprehensive database of zeolite structures refer to: http://www.iza-structure.org/databases/

21 Some remaining stock of safe, weakly radioactive glass items such as ceramic plates, ore, marbles, etc. may still be acquired online from http://www.unitednuclear.com/

22 Richardson, T. J.; Slack, J. L.; Armitage, R. D.; Kostecki, R.; Farangis, B.; Rubin, M. D.; *Appl. Phys. Lett.* **2001,** *78*, 3047.

23 Asphalt is a black, sticky, viscous liquid that is obtained from crude petroleum. It comprises almost entirely a form of tar called bitumen. The structure of asphalt is actually a colloidal suspension, with

small particulates called asphaltenes dispersed through the petroleum matrix. More environmentally friendly aqueous-based asphalt emulsions are currently being used for road repair applications.

[24] For more details regarding the role of C4AF in the hardening mechanisms of Portland cement, see: Meller, N.; Hall, C.; Jupe, A. C.; Colston, S. L.; Jacques, S. D. M.; Barnes, P.; Phipps, J. *J. Mater. Chem.* **2004**, 14, 428, and references therein.

Topics for Further Discussion

1. Considering that Si has the zinc blende crystal structure, draw the (111), (110), and (100) planes of Si. Place these planes in order of highest atomic density, from least to greatest. What impact would the structures of these planes have on their relative surface reactivities?
2. Show the relationship between 4_1 and 4_3 crystallographic point group operations.
3. What are the differences between amorphous and crystalline materials? Cite examples of each.
4. Describe some techniques used to fabricate amorphous or crystalline materials.
5. What is the difference between point groups and space groups?
6. Is it possible to have a material that is both an ionic solid and a molecular solid? Explain your reasoning, and cite examples of such hybrids (if possible).
7. Consulting a Periodic Table and Table of Atomic Radii, what atoms would be suitable (a) interstitial dopants and (b) substitutional dopants within a Mn lattice? Show your calculations and rationale.
8. Using diagrams, determine whether (111) or (110) planes are more densely packed with atoms for FCC crystals.
9. What are the benefits of fuel cells, as compared to batteries and other "standard" power sources such as fossil fuel power plants?
10. Why do acid-catalyzed and base-catalyzed sol–gel processes result in linear and branched metal oxides, respectively?
11. We have all seen cement trucks that have a distinctive rotating tank. Why is this constant rotation necessary, and what would happen if the tank stopped rotating?
12. Why are coarse and fine aggregates needed for the overall strength of concrete?
13. Explain why the reaction of C2S with hydroxides occurs much slower than the C3S/hydroxide reaction during cement hardening.

Further Reading

1. Pogge, H. B. *Electronic Materials Chemistry*, Marcel-Dekker: New York, 1996.
2. Glusker, J. P.; Lewis, M.; Rossi, M. *Crystal Structure Analysis for Chemists and Biologists*, VCH: New York, 1994.
3. Larminie, J.; Dicks, A. *Fuel Cell Systems Explained*, 2nd ed., Wiley: New York, 2003.
4. West, A. R. *Solid State Chemistry and Its Applications*, Wiley: New York, 1987.
5. West, A. R. *Basic Solid State Chemistry*, 2nd ed., Wiley: New York, 1999.
6. Moore, E.; Smart, L. *Solid State Chemistry: An Introduction*, 2nd ed., CRC Press: New York, 1996.
7. Hurd, C. M. *Electrons in Metals*, Wiley: New York, 1975.
8. Elliott, S. *The Physics and Chemistry of Solids*, Wiley: New York, 1998.

CHAPTER 3

METALS

Of all the 115 elements listed in the Periodic Table, 70% exhibit metallic character. Since the discovery of copper and bronze by early civilizations, the study of metals (*i.e.*, metallurgy) contributed to most of the early investigations related to materials science. Whereas iron-based alloys have long been exploited for a variety of applications, there is a constant search for new metallic compositions that have increasing structural durability, but also possess sufficiently less density. The recent exploitation of titanium-based alloys results from this effort, and has resulted in very useful materials that are used for applications ranging from aircraft bodies to golf clubs. Indeed, there are many yet undiscovered metallic compositions that will undoubtedly prove invaluable for future applications.

In Chapter 2, you learned how individual atoms pack in crystal lattices. Moreover, the nature of metallic bonding was described, which is responsible for characteristic physical properties of these materials. This chapter will continue this discussion, focusing on the relationship between specific metallic lattice structures and their impact on overall physical properties.

3.1. MINING AND PROCESSING OF METALS

Before we examine the structures and properties of metallic classes in further detail, it is useful to consider the natural sources of the metals, generally as oxide and/or silicate-based mineral formations. If the mineral deposit contains an economically recoverable amount of a metal, it is referred to as an *ore*. The waste material of the rock formation is known as *gangue*, which must be separated from the desired portion of the ore through a variety of processing steps.

There are three main types of rocks, grouped according to their form of origin. *Igneous* rocks are those formed from the solidification of molten mass following volcanic activity. Common examples include granite, feldsar, and quartz; metals such as the alkali and alkaline earths, gold, platinum, and chromium are isolated from these formations. *Sedimentary* rocks are those formed through compaction of small grains deposited as sediment in a riverbed or sea. Common examples include limestone, sandstone, and dolomite. Metals such as copper, iron, zinc, lead, nickel, molybdenum, and gold may all be found together within sulfur-based sedimentary deposits. *Metamorphic* rocks are those formations that have changed composition

and morphology over time through the influence of temperature and pressure. Examples of these rocks include marble, slate, and gneiss, and yield similar metals as igneous formations.

Most metals are obtained through physical and chemical modification of ores that contain a sufficiently high concentration of the desired metal. Table 3.1 lists some metals, regions of their primary occurrence, and some applications for the metallic materials. Quite often, the processing steps used to isolate and purify metals from ores have a significant impact on the overall microstructure, which will affect the physical properties of the materials fabricated from them.

To obtain the ores in sufficient quantities needed to recover a desirable amount of metal, either open pit or shaft mining is used depending on how accessible the ore is located. We are all familiar with the picture of treacherous mine shafts that have even been featured in movies such as *Coal Miner's Daughter*. However, such "brute force" is sometimes not required. For instance, to recover coal that is used in power plants to generate electricity, there are often sufficient resources near the surface that may be obtained using large cranes. Pit mining for metals such as copper, iron, and precious metals is commonly used throughout the US and Canada.

Of course, once the ore is obtained from its deposit, the actual work of extracting the desired metal has yet to be accomplished. In addition to metals, a variety of other substances comprise natural minerals. Since aluminum and silicon are the most prevalent elements in the Earth's crust, most of the metals exist naturally as aluminates, silicates, or aluminosilicates. The most common minerals are feldspars and clays. These materials have been used since ancient times for the production of materials such as pottery, brick, and china. An example of a feldspar is $K_2Al_2Si_6O_{16}$, which corresponds to a mixture of potassium superoxide, alumina, and silica ($K_2O{\cdot}Al_2O_3{\cdot}6SiO_2$). Upon contact with water and carbon dioxide, a weathering reaction results in kaolinite, an aluminosilicate clay (Eq. 1). However, in addition to these oxidized sources of metals, there are substances such as alkaline carbonates, sulfates, phosphates, as well as organic matter that need to be removed to yield the desired metal. As you would expect, the yield for this process is quite low; ores typically possess less than 1% of the desired metal!

$$(1) \qquad K_2Al_2Si_6O_{16} + 3H_2O + 2CO_2 \rightarrow \underset{\text{(kaolinite)}}{Al_2Si_2O_7} + 2H_2O + 2KHCO_3 + 4SiO_2$$

The variety of procedures that are used to obtain metals from their ores is known as extractive metallurgy. Usually, a *flotation* process is first used where the ore is ground into a powder, mixed with water, oil, and surfactants (*i.e.*, collectors) and vigorously stirred into a frothy mixture. The most common collector molecules are alkali salts of the *O*-alkyl dithiocarbonates (**I**) commonly known as xanthates, which are widely used for the concentration and separation of metals from sulfide-based ores. The coated grains then attach themselves to air bubbles that float to the solution surface, where they are removed through a skimming/filtration process. This technique is very useful in the separation of lead and zinc components, often found together in natural sulfide deposits.

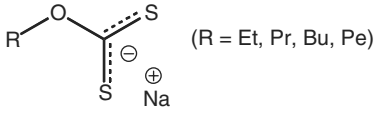
Table 3.1. Natural Occurrence of Selected Metals and Materials-Based Applications

Metal	Selected applications[a]	Primary locations[b]
Li	Batteries, glasses, aerospace components	US, Chile, China, Argent
Na	Dyes, pigments, lamps, photoelectric cells	Chile, Peru, US
K	Nuclear reactors, respiratory equipment	Ger., Can., France, Spain
Be	X-ray tubes, ship/aircraft navig. systems	US, Africa, Brazil, Rs.
Mg	Auto parts, coatings, photoengraving	US, Can., S. Amer., Asia
Ca	Lighter flints, lights, plastics stabilizer	US, Can., S. Amer., Asia
Sr	Glasses, ceramics, paints, TV tubes	Mexico
Sc	High-intensity lamps, nuclear reactors	China, US, Aus., Ind.
Y	Coatings, capacitors, superconductors, lasers	China, US, Aus., Ind.
Ti	Golf clubs, surgical implants, jet engines	US, Aus., S.A., Norway
Zr	Prosthetic devices, nuclear reactors	Ind., US, Aus., S.A.
V	Tools, ceramics, batteries, magnets	US, Mex., Peru, Rs., S.A.
Nb	Heat shields, electromagnets, spaceships	Japan, Rs., Can., Brazil
Ta	Aircraft turbines, rocket nozzles, HT vessels	Aus., Brazil, Thailand, Can.
Cr	Corrosion-resistant coatings, tools	S.A., Turk., Rs., Cuba, Phil.
Mo	Lubricants, lamp filaments, integ. circuits	US, Aus., Swe.
W	Filaments, coatings, missiles, tools, paints	US, Asia, S. Amer., Rs.
Mn	Steels, paints, batteries	US, Can., France, Ger., Ind.
Fe	Autos, tools, structural materials	US, Can., China, Brazil, Rs.
Ru	Coatings, superconductors, electrodes	S.A., Rs., US
Os	Fountain pen tips, pacemakers, forensics	S.A., Rs., US
Co	Magnets, batteries, recording media	Can., N. Africa, Swe.
Rh	Catalytic converters, coatings, crucibles	S.A., Rs., US
Ir	HT crucibles, spark plugs	S.A., Rs., US
Ni	Batteries, coatings, currency	Can., Rs., Aus.
Pd	Jewelry, dental crowns, surgical instruments	S.A., Rs., US
Pt	Jewelry, coatings, surgical implants, lighters	S.A., Rs., US
Cu	Conductive wire, pipes, currency	Chile, US, Can.
Ag	Jewelry, dental fillings, photography	Mex., Peru, US, Can.
Au	Jewelry, integ. circuits, heat shields, coatings	S.A., US, Aus., Can.

Table 3.1. Continued

Metal	Selected applications[a]	Primary locations[b]
Zn	Coatings, pipes, highway guard rails	US, Can., Aus., China
Cd	Alloys, batteries, nuclear reactors	Canada, Aus., Mex.
Al	Auto parts, ceramics, coatings, paints	Brazil, Jamaica, Asia
Sn	Alloys, plating, flat glass	Bolivia, Brazil, Malaysia
Pb	HT vessels, pipes, sound abs. materials	Aus., US, China, Peru
Ce	Lighter flints, ceramic capacitors, magnets	China, US, Aus., Ind.
Eu	Television screens, nuclear reactors	China, US, Aus., Ind.
Er	Lasers, catalysts, phosphors	China, US, Aus., Ind.

[a]These selected applications may be for the pure metal or other compounds/alloys.
[b]Can. = Canada, Rs. = Russia, Aus. = Australia, Ind. = India, S.A. = South Africa.



I

Once the bulk ore is concentrated from its many impurities, the metal (M) is reduced from its natural oxides. The most common method that is used is referred to as *smelting*, where the ore is heated in the presence of a reducing agent such as a carbonaceous material, or carbon monoxide (Eqs. 2 and 3, respectively). Although carbonaceous material was used exclusively by early civilizations, modern extractive methods commonly employ other reducing agents such as Na, Mg, Ca, Al, or alloys such as Zn/Mg. However, with the advent of biomass as an alternative fuel, the organic-based waste materials are now being used as an effective reducing agent.

$$\text{(2)} \quad \mathrm{M}_x\mathrm{O}_y + y\mathrm{C} \xrightarrow{1500\text{-}2000^\circ\mathrm{C}} x\mathrm{M} + y\mathrm{CO}$$

$$\text{(3)} \quad \mathrm{M}_x\mathrm{O}_y + y\mathrm{CO} \xrightarrow{1500\text{-}2000^\circ\mathrm{C}} x\mathrm{M} + y\mathrm{CO}_2$$

If the ore consists of metal sulfides, it must first be converted to the oxide through reaction with oxygen at high temperature (Eq. 4):

$$\text{(4)} \quad 2\mathrm{MS} + 3\mathrm{O}_2 \xrightarrow{1500\text{-}2000^\circ\mathrm{C}} 2\mathrm{MO} + 2\mathrm{SO}_2$$

The above techniques that involve high-temperature processes are known as *pyrometallurgy*. Another common technique involves the electrolytic reduction of metal compounds, often referred to as *hydrometallurgy* or *electrorefining*, depending on whether the procedure is carried out before or after the metal has already been separated from its ore, respectively.

Figure 3.1 illustrates examples of electrolytic purification reactors for the isolation of sodium and aluminum metals. For the purification of sodium, a fused salt is used at high temperatures. As is often necessary for ionic salts, a solid solution is necessary to reduce the melting point of the salt. The addition of calcium chloride effectively

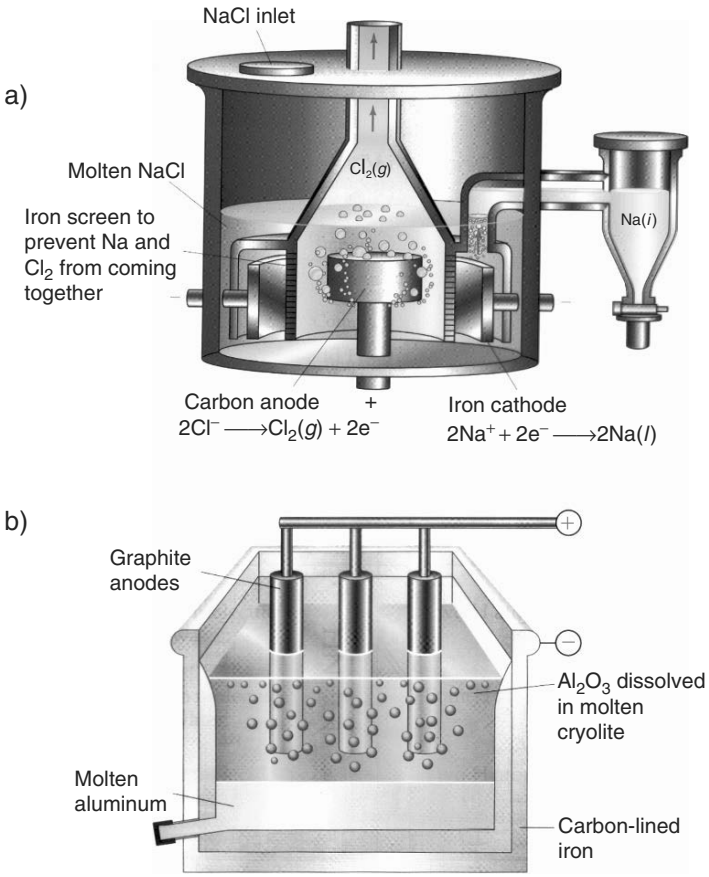


Figure 3.1. Schematic of electrometallurgical processes used to purify metals. Shown is the (a) Downs cell used to purify sodium metal and (b) Hall cell used to purify aluminum metal. Reproduced with permission from *Chemistry: The Central Science*, 8th ed., Brown, LeMay, Bursten. Copyright 2002 Prentice-Hall.

reduces the melting point of the solution from 800 (for NaCl alone) to 600°C. To circumvent the high melting point for aluminum (2,030°C), anhydrous aluminum oxide (*i.e.*, alumina, m.p. 2,050°C) is dissolved in molten cryolite (Na_3AlF_6, m.p. 1,009°C) and excess AlF_3, resulting in a more reasonable operating temperature (920–980°C) for the electrolytic process.

It is estimated that iron constitutes 90% of all applications that involve metals. Hence, it is not surprising that the purification and postprocessing for iron is the most widely practiced. The most primitive method that was used in the 19th century to purify iron from its ore is called bloomery. This method used pure carbon in the form of charcoal to reduce the metal (Eq. 2). In this process, the temperature is not sufficient to completely melt the iron, so a spongy mass consisting of iron and silicates are formed. Through heat/hammering treatments, the silicates mix into

the iron lattice, creating *wrought iron*. This form of iron was used exclusively by early blacksmiths, since the heating of wrought iron yields a malleable, bendable, and extremely easy compound to work with.

Figure 3.2 shows a flowchart for the various types of procedures that are used for iron metallurgy. Modern techniques employ a blast oven that uses large jets of air that mix with iron ore, charcoal, and calcium sources (from limestone and dolomite), in combination with the silicates in the ore, to form *slag*. The composition of slag is 30–40 wt% SiO_2, 5–10 wt% Al_2O_3, 35–45 wt% CaO, and 5–15 wt% MgO. The molten iron collects at the bottom of the blast furnace and is cooled into a form known as *pig iron*. This form of iron contains between 4 and 5 wt% C, making it too brittle and hard for any application. If pig iron is reheated, remixed with slag, and hammered to remove most of the carbon, *wrought iron* is created. However, it is often most useful to convert this form into high-strength steel. While impurities such as silicon, calcium, aluminum, and magnesium were removed in the slag, phosphorus and sulfur impurities remain behind, and are dissolved into the molten iron. A basic oxygen furnace is used to oxidize these impurities, which would greatly weaken steel if not removed during processing. Another process known as the Bessemer process is used in modern steel-making plants that use a specially lined reactor that provides an efficient method at removing impurities, all in approximately 15–20 min!

3.1.1. Powder Metallurgy

Although the origin of making metallic materials through flame sintering dates back to *ca*. 3,000 B.C., this method was not widely applied until the late 18th century. Until only recently, metallurgy focused on doping and strengthening bulk metallic materials. Nowadays, powders are frequently used as precursors for metallic materials. For instance, tantalum powder is used in the fabrication of capacitors for electronics and telecommunications applications, including cell phones. Iron powder is used as a carrier for toner in electrostatic copying machines; over 2 million pounds of iron powder is incorporated each year in iron-enriched cereals! Copper powder is used in antifouling paints for boat hulls and in metallic pigmented inks for printing and packaging. Indeed, the list of applications for metal powders goes on and on, and must constantly be updated as new discoveries occur.

Modern *powder metallurgy* consists of placing metal powder(s) into a closed metal cavity, or die, repeatedly compacting under high pressure (typically 200–300+ MPa), and sintering in a furnace to yield a metal with desired porosity and hardness characteristics. The *sintering* process effectively results in the welding together of powder particles to form a mechanically strong finished material.

Metal and alloy powders may be produced through the following routes, with the last three accounting for the most common methods currently employed:

1. Grinding and pulverization of a metallic solid or oxide-based ore
2. Reductive precipitation from a salt solution
3. Thermal decomposition of a chemical compound
4. Electrodeposition
5. Atomization of molten metal

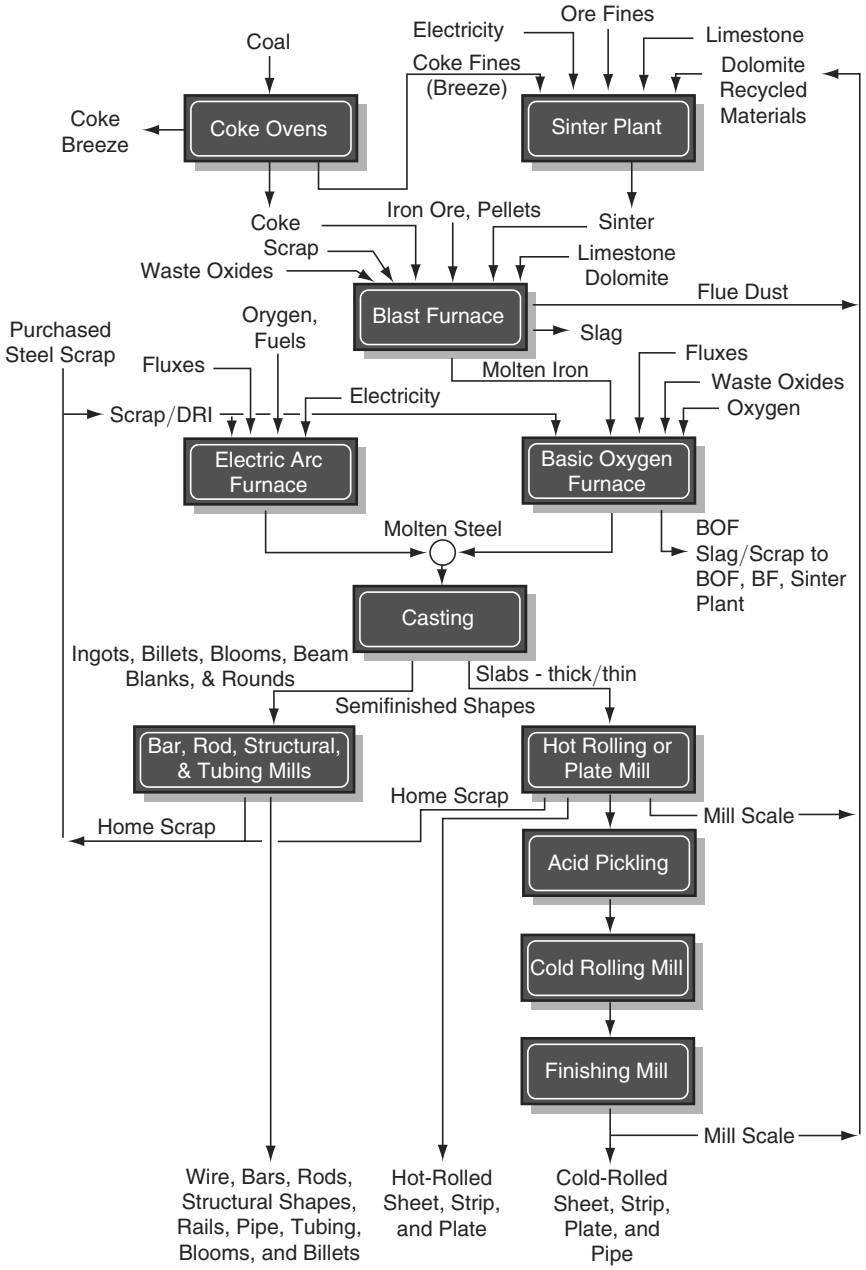


Figure 3.2. Flowchart illustrating the various steps involved in steelmaking and processing. Reproduced with permission from the American Iron and Steel Institute. May be found online at http://steel.org/AM/Template.cfm?Section = How_Steel_is_Made&CONTENTID = 4800&TEMPLATE=/CM/ContentDisplay.cfm

For relatively brittle materials such as intermetallic compounds, and ferro-alloys, mechanical pulverization is sufficient to produce metallic powders. This process uses a ball or rod mill, a cylindrical-shaped steel container filled with ceramic balls or rods, respectively. As the grinding mill is rotated, the grinding media collides with the ore/metallic compound effectively grinding the material into a fine powder. Either alumina or zirconia represents the most common ceramic materials used within grinding mills. This procedure is commonplace for the fabrication of iron powder, from the co-grinding and postannealing of the ore with carbon (Eq. 2). Refractory metals are normally produced through the reduction of oxides with hydrogen gas.

Chemical precipitation of metal from a solution of a soluble salt may also be used to form metallic powders. In this procedure, a reducing agent such as sodium borohydride is added to the metal ions in solution (Eq. 5). As we will see in Chapter 6, this is the most widely used procedure for the synthesis of nanoparticulate metallic powders.

$$(5) \qquad M^{n+} + 2X^{-} + 2NaBH_4 \rightarrow M^0 + 2NaX + B_2H_6 + H_2$$

Another useful means of producing metal powders is through thermolysis of a chemical precursor, such as metal carbonyl complexes. This process was originally developed to refine nickel from its crude metal extracted from its ore. Carbon monoxide gas readily reacts with late transition metals, due to the synergistic effects of σ electron donation from the ligand to metal, and π back-donation from the metal to the ligand (Figure 3.3). Hence, by passing CO gas over impure nickel at 50°C, $Ni(CO)_4$ gas is formed, leaving the impurities behind. The carbonyl decomposes upon heating to *ca.* 250°C, forming the pure nickel powder. Industrially, the Mond process uses the same chemistry, using nickel oxides from the natural ore. Upon reaction with a mixture of H_2 and CO gases, the nickel is first reduced to impure metal, and then converted to high-purity metal through the $Ni(CO)_4$ byproduct.

Electrolysis may also be used to produce metallic powders, through redox reactions at electrode surfaces. By choosing suitable reaction conditions – composition and strength of the electrolyte, temperature, current density, *etc.* – many metals can be deposited in a spongy or powdery state. However, most often a brittle deposit is formed, requiring extensive postprocessing such as washing/drying, reduc-

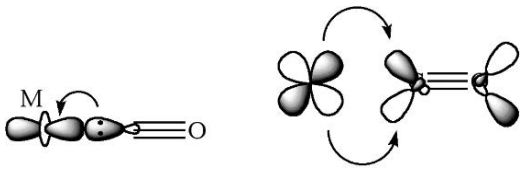


Figure 3.3. The synergistic stabilizing effect of metal carbonyl complexes. Shown is (a) ligand-to-metal σ donation from the carbon lone pair to the metal d_{z^2} orbital and (b) metal-to-ligand back-donation from the $d_{x^2-y^2}$ orbital to the empty π^* orbital on CO. This weakens the C–O bond, while concomitantly strengthening the M–C interaction.

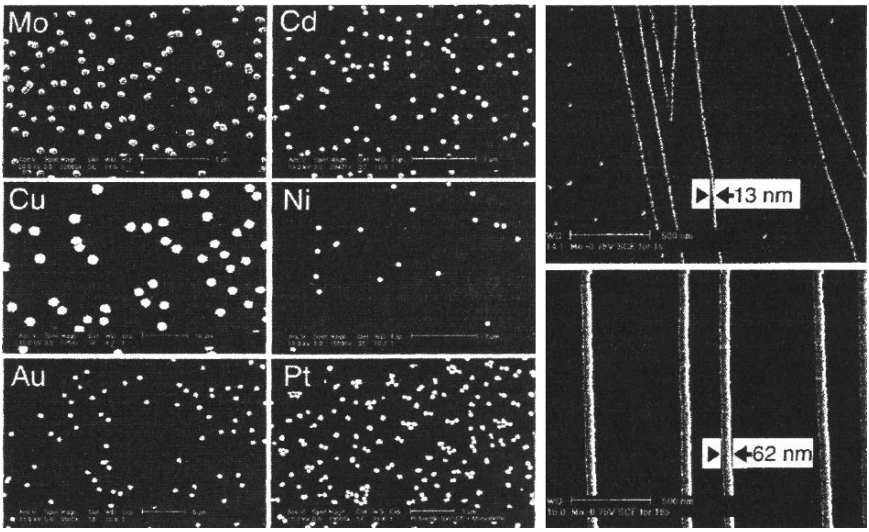


Figure 3.4. Electrodeposition of (a) silver nanoparticles and (b) silver nanowires. Reproduced with permission from *J. Phys. Chem. B.* **2002**, *106*, 3339. Copyright 2002 American Chemical Society.

tion, annealing, and crushing. Although this technique could be used for virtually all metals, it has been replaced with other less expensive methods such as solution reduction. Nevertheless, metallic powders of copper, chromium, and manganese are still mostly produced through electrolytic means. Toward the ongoing search for structures at the nanoregime (Chapter 6), electrodeposition has recently been applied for the intriguing synthesis of metal nanoparticles and nanowires (Figure 3.4).

The last method for generation of metallic powders that we will consider is atomization. In this high-temperature process, molten metal is broken up into small droplets and rapidly quenched to prevent wide-scale agglomeration (Figure 3.5). The atomization process occurs through the bombardment of a stream of molten metal with a high-energy jet of gas (*e.g.*, air, N_2, Ar) or liquid (*e.g.*, H_2O, hydrocarbons). Argon gas is used extensively to prevent the oxidation of reactive metals and alloys such as chromium or tungsten. Atomization is very different than ionization. Whereas the former consists of gaseous ground-state and excited-state metallic atoms, the latter contains electrons and metallic ions that are much more reactive.

By varying parameters such as jet design, pressure and volume of the atomizing fluid, and density of the liquid metal stream, it is possible to control the overall particle size and shape. In principle, atomization is applicable to all metals that can be melted, and is commercially used for the production of iron, steels, alloy steels, copper, brass, bronze, and other low-melting-point metals such as aluminium, tin, lead, zinc, and cadmium.

Atomization is particularly useful for the production of homogeneous powdered alloys, since the constituent metals are intimately mixed in the molten state. Further, this process is also useful to produce powders of difficult compositions. For instance,

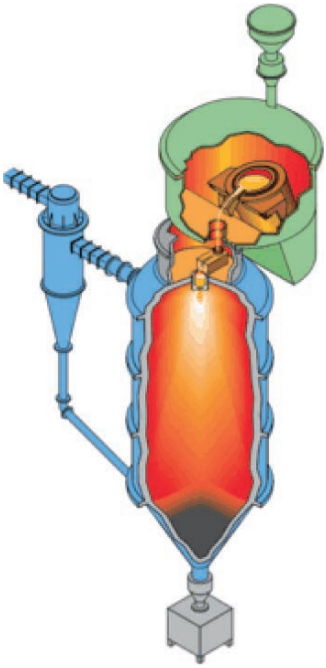


Figure 3.5. Illustration of an atomizer for the production of metallic powders. The molten metal/alloy is sprayed into a cooling tower under the flow of an atomizing gas. The particulates are allowed to cool as they descend downward, and are collected in a hopper at the bottom of the tower. Reproduced with permission from Crucible Materials Corporation (http://www.cruciblecompaction.com/process/atomization.cfm).

copper–lead powders may not be formed through simple precipitation from liquid solutions. Upon solidification, the lead will preferentially precipitate, resulting in a copper-rich metallic powder. By comparison, atomization of a Cu/Pb molten solution results in a copper powder containing a very fine and uniform distribution of lead inclusions within each powder particle.

For powder metallurgy, the density of the powder strongly influences the strength of the material obtained from compaction. As one would expect, the density of the powder depends on both the shape and porosity of individual micron-sized particulates. We saw in Chapter 2 that close-packed metals will have higher densities than simple cubic materials (Figure 3.6). Among the close-packed metals, the theoretical percentages of total space occupied by atoms, relative to voidspace for bcc (coordination number 8), fcc (coordination number 12), and hcp (coordination number 12) unit cells are 68%, 74%, and 74%, respectively. Even if the metal particulates have the same diameter and are completely spherical, the actual packing density is typically on the order of 55–60%. This value may be improved by introducing nanosized particles that will fill the voids among the larger particles. During subsequent high-temperature sintering, the larger particles will grow at the expense of

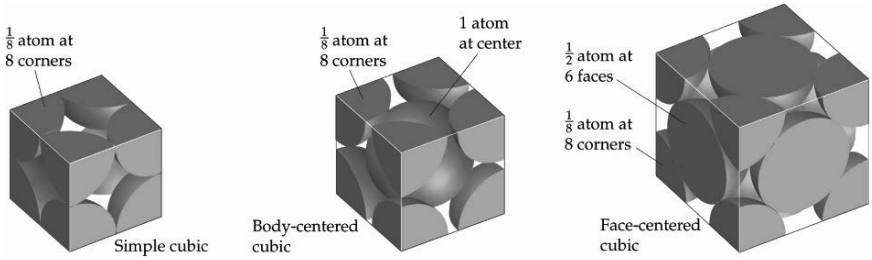


Figure 3.6. Space-filling models showing the occupancy and available interstitial sites within cubic unit cells. Reproduced with permission from *Chemistry: The Central Science*, 8th ed., Brown, LeMay, Bursten. Copyright 2002 Prentice-Hall.

the nanoparticles, leaving behind relatively small voids that are closed during the thermal treatment.

A lubricant may also be mixed together for co-pressing with the composite powders. The most common lubricants are stearic acid (octadecanoic acid), stearin, zinc stearate, and other waxy organic compounds (*e.g.*, palmates). The name stearate should be vaguely familiar, as the sodium salt is often employed as the active ingredient in soap. The primary use for the lubricant is to reduce friction between the powder mass and the surface of the die walls. For this purpose, it is often sufficient to apply lubrication to the walls of the die, rather than introducing the organic compound to the metallic powders. If a significant amount of organic residue is left following compaction, it will be removed upon sintering, leaving behind large voids that will greatly detract from the finished material's overall strength.

The density of the bulk material following the pressing event is referred to as the *green density*, coined more frequently for ceramic processing. It is most desirable to have a powder with lower density, as this will undergo a greater change in volume during compaction (Figure 3.7). The intimate pressing together, or alloying, of metals during the pressing process is known as *cold-welding*. Sometimes, the powder is too dense for efficient cold-welding; for these samples, such as heavy metal alloys, a greater pressure (*i.e.*, larger presses and stronger dies) is required. It should be noted that powders under pressure do not behave as liquids; the pressure is not uniformly transmitted and very little lateral flow takes place within the die.

The compacted powder will only be as pure as the initial components. The addition of small impurities will cause dramatic differences in the resultant metallic material, following the pressing and sintering steps. The presence of bound *vs.* free impurities may also result in observable differences in the compaction behavior for powders. For iron powders, the presence of iron carbide (Fe_3C) will increase the hardness of the matrix, requiring higher pressures for compaction. However, free graphite particles will act as a lubricant, increasing the pressing efficiency at lower pressures.

Unless handled under an inert atmosphere, metal powder grains will be coated with a thin oxide film. Unless excessively strong SiO_2 or Al_2O_3 films are produced, the coating will rupture during the pressing process, exposing the clean underlying metal surfaces. It should be noted that for alloying metals such as copper and zinc,

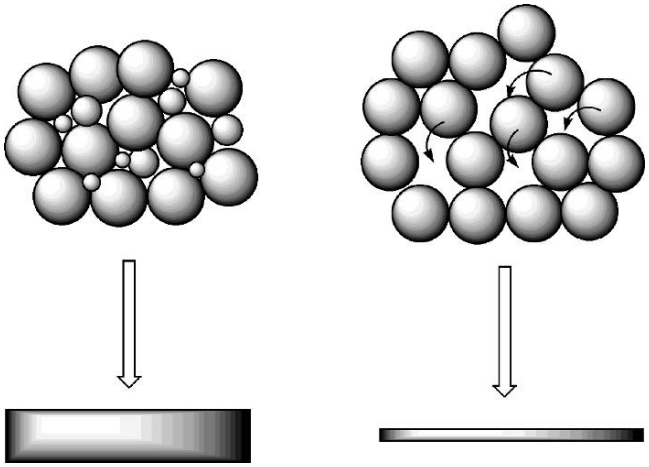


Figure 3.7. The effect of matrix density on the compaction volume yielded from pressing.

pressing is not necessary. The powders are simply placed in a mold and sintered, a process aptly referred to as loose-powder sintering.

The major applications for powder metallurgical products center around the automotive industry, specifically for engines, transmissions, and brake/steering systems. Following pressing, the compacted metals may be injected into a mold, or pressed under vacuum at high temperature within a hot isostatic press. A great deal of consumer products are fabricated using these techniques; high-tech plastics and other composite parts represent a significant market share for powder metallurgical materials. This is especially the case since the soaring gas prices and stringent environmental regulations dictate the design of lighter vehicles, to improve gas consumption. Outside of automotive applications, other uses are prevalent; for instance, Canadian nickels are made from strips rolled from pure nickel powder.

There are a number of attractive benefits for powder metallurgy:

1. Lack of machining eliminates scrap losses
2. Facile alloying of metals
3. *In situ* heat treatment is useful for increasing the wear resistance of the finished material
4. Facile control over porosity and density of the green and sintered material
5. Fabrication of complex/unique shapes which would be impractical/impossible with other metalworking processes
6. Rapid solidification process extends solubility limits, often resulting in novel phases

Although we have described powder metallurgy as being an ideal process, without limitations or hazards, it does pose some serious safety risks and limitations. The majority of metallic powders and other finely divided solids are *pyrophoric*, meaning

that they will spontaneously ignite in air, at temperatures below 55°C. Unlike black powder, which contains both the fuel (C and S) and oxidizer (potassium nitrate), it is not immediately apparent why metallic powders would ignite, since both key components are not present within the powder matrix.

There are two primary reasons for this pronounced reactivity. The extremely large exposed surface area of powders relative to the bulk results in rapid oxidation upon exposure to air, especially for metals that form stable oxides such as aluminum, potassium, zirconium, *etc.* Also, there is enhanced internal friction among the individual micron- or nanosized individual particulates comprising the powder. Simply pouring the powder onto a table will yield sparks that may or may not be visible to the naked eye. Indeed, if one does not physically see the spark, he or she will soon know if there was one! As you would imagine, both the pulverizing and pressing steps in powder metallurgy are especially dangerous, as the particles are forced into contact with one another and the equipment surfaces (another purpose for an added lubricant during compaction). NASA recently published a technical paper that describes the production of rocket propellants[1]; this is definitely worth a read, to find out how one prepares mixtures of such reactive components.

3.2. METALLIC STRUCTURES AND PROPERTIES

We are now in a position to investigate a question that will be assessed throughout this textbook: *What is the relationship between the microstructure of a material, and its overall properties*? If our world wishes to stay on our current path of unprecedented growth in areas of electronics, building materials, homeland security devices, and future "smart" materials, it is essential that we become familiar with the properties of individual classes of materials and current applications. Only then will we be able to extrapolate these properties into new and exciting applications for the future.

In Section 3.1, we saw that a wide variety of applications employ metallic substances (Table 3.1). In this section, we will examine the various classes of metals and alloys in more detail, focusing on phase transitions, changes in the microstructure, and atomic packing of the materials. With this insight, you will be in a good position to evaluate why a particular metal is more suited than others for an existing or future application. It should be noted that, analogous to metals, certain organic polymers may also exhibit high electrical conductivities. However, this chapter will only discuss inorganic-based metallic classes; organic-based electrical conductors will be detailed in Chapter 5.

3.2.1. Phase Behavior of Iron–Carbon Alloys

In general, for a mixture of two or more pure elements, there are two types of solid–solution alloys that may be obtained. Type I alloys are completely miscible with one another in both liquid and solid states. As long as the Hume-Rothery rules are satisfied, a random substitutional alloy will be produced. We will see many examples of these alloys in this section for a variety of metal dopants in stainless steels. By comparison, type II alloys are only miscible in the molten state, and will separate from

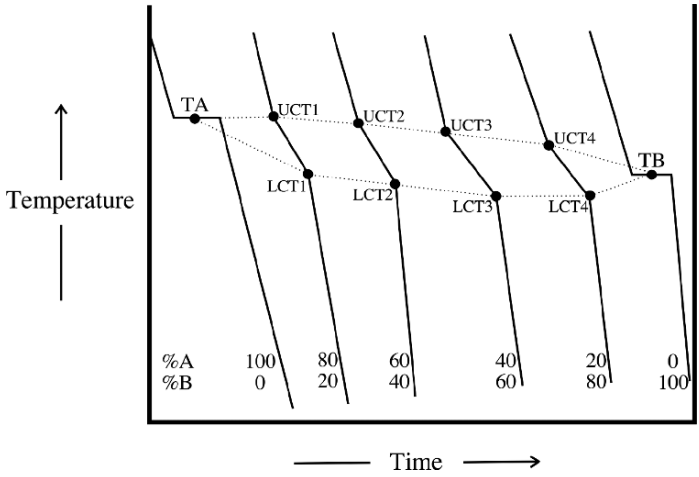


Figure 3.8. Representative cooling curves for an alloy, A–B. The terms LCT and UCT refer to lower and upper critical temperatures, respectively. TA and TB designate the melting points of pure A and pure B, respectively.

one another upon cooling. These alloys are usually associated with compound formation from the alloying of metals or metals/nonmetals that are too dissimilar in their reactivities (*e.g.*, Cu and Al to form $CuAl_2$ precipitates). The eutectic composition is the lowest melting point of type II alloys.

Type I alloys contain two types of atoms that are arranged within a single lattice. When solidification of the solution begins, the temperature may be higher or lower than the freezing point of the pure solvent. Unlike a pure molten metal, most solid solutions will solidify over a temperature range due to differing diffusion rates of the metals *en route* toward their preferred crystal arrangement (Figure 3.8).

Pure iron exists as a variety of allotropes depending on the external temperature or pressure. As the temperature is increased, iron undergoes allotropic transformation from α-Fe (ferrite, bcc) to γ-Fe (austenite, fcc), and finally to a narrow region of δ-Fe (bcc) before melting. As the temperature of the standard bcc crystal lattice is increased, thermally induced atomic motion increases, and it becomes more energetically favorable for atoms in the center of lattice unit cells to migrate into face-centered positions of neighboring unit cells (Figure 3.9).

As seen earlier, the steps used to purify iron involves carbonaceous material to remove the oxide-based impurities, through exothermic formation of CO and CO_2. Hence, carbon will be pervasive in a variety of concentrations throughout all phases of iron and steels, present as an interstitial dopant within these lattices. Experimental evidence shows that carbon-doped iron polymorphs are indeed interstitial solid solutions. For instance, the carbon atoms in bcc ferrite are located only on empty face-centered positions. However, very few of these positions are occupied throughout the lattice, as the maximum solubility of carbon in α-Fe is only between 0.01 and 0.02 wt%. From a metallic-bonding standpoint, the addition of carbon in the lattice acts as an "electron sink," that is able to accept some of the delocalized electron

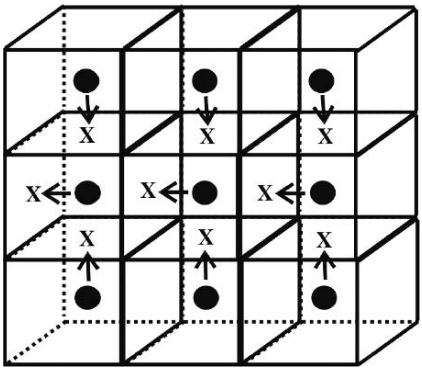


Figure 3.9. Simplified schematic of the transformation from BCC to FCC, exhibited between the three allotropes of iron. Corner atoms have been omitted for clarity.

Table 3.2. Unit Cell Dimensions of Iron Allotropes and Fe–C Alloys[a]

Fe–C composition (crystal structure)	Unit cell parameters (Å)
α-Fe (ferrite, BCC)	$a = 2.8665$
γ-Fe (austenite, FCC)	$a = 3.555 + 0.044x$[b]
δ-Fe (BCC)	$a = 2.9323$
Martensite (tetragonal)	$a = 2.867 - 0.013x$[b]
	$c = 2.867 + 0.116x$[b]
Cementite (orthorhombic)	$a = 4.525$
	$b = 5.088$
	$c = 6.740$

[a]Values taken from Cullity, B. D. *Elements of X-Ray Diffraction*, 2nd ed., Addison Wesley: Reading, MA, 1978.
[b]x = wt% C in interstitial sites of the iron lattice.

density from the metallic lattice. This results in a stronger interaction among all atoms in the lattice that adds to physical hardness, but detracts from the overall electrical conductivity, relative to pure (undoped) iron.

Table 3.2 compares unit cell dimensions for the various allotropes and Fe–C alloys. Since the dopant species are entrained within individual unit cells, the volume of each unit cell will increase concomitantly with the concentration of carbon. Although the δ-Fe lattice is isomorphous with ferrite, the difference in volume between these allotropes corresponds to different concentrations of carbon in each solid solution. That is, δ-Fe contains an order of magnitude greater concentration of carbon than ferrite. Since a greater number of interstitial sites may be occupied by fcc unit cells relative to bcc, austenite may contain an even greater concentration of C in the lattice, up to 2.1 wt%. It must be noted that the trend of increasing volume with dopant concentration is not only exhibited by the Fe–C system, but is also followed by all other interstitial alloys that we will examine later.

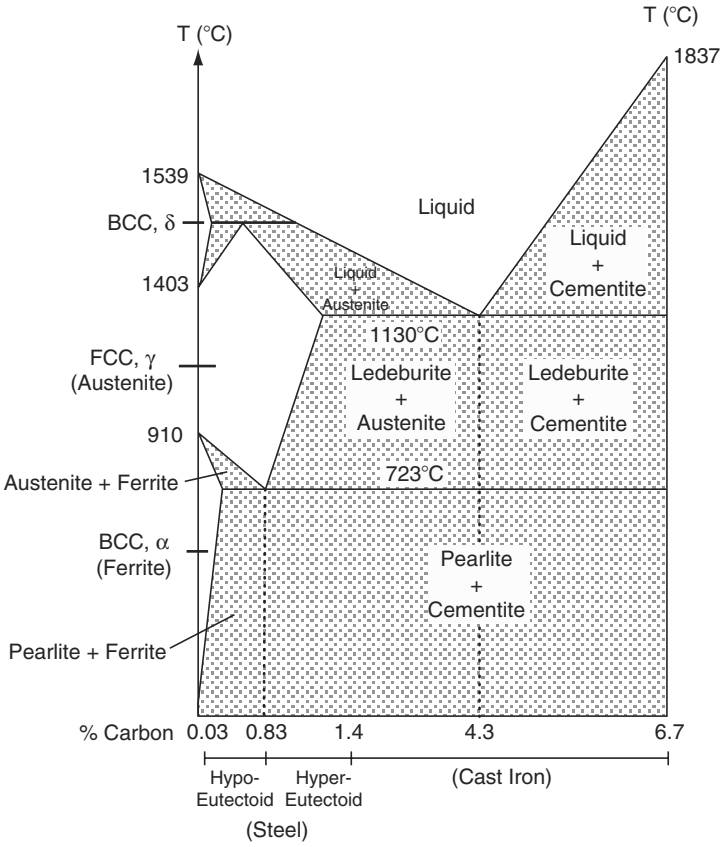


Figure 3.10. The equilibrium phase diagram for the iron–carbon system.

In general, the density of interstitial solid solutions is given by Eq. 6. Since the change in volume is usually more significant than the increase in number of unit cell atoms, interstitial solids usually exhibit a decrease in density, relative to the pure allotrope. For instance, the density of pure iron (7,874 kg m^{-3}) shows a significant decrease upon interstitial placement of carbon in cast irons (*ca.* 7,400 kg m^{-3}).

$$\rho = \frac{1.6604 \sum (n_l A_l + n_i A_i)}{V} \tag{6}$$

where n_l, n_i are the number of regular lattice and dopant atoms, respectively, per unit cell and A_l, A_i are the atomic weights of the regular lattice and dopant atoms, respectively.

The complex binary phase diagram for the Fe–C system is shown in Figure 3.10, and illustrates a number of important transitions. In particular, as the temperature is increased from ambient to its melting point, pure iron exhibits a variety of allotropic changes. At room temperature, the ferrite form is most stable; conversion to austenite

occurs at 910°C. Austenite is much softer, being more easily formed into desired shapes, relative to ferrite. At a temperature of 1,403°C, the fcc γ-Fe allotrope converts back to the bcc form of δ-Fe before melting at 1,539°C. This behavior is quite atypical; most metals do not alter their crystal structure *en route* toward their melting points.

By definition, iron containing between 0.15 and 1.4 wt% C is typically referred to as *steel*.[2] Hence, although we typically think of steel as containing chromium and other metal dopants, some γ- and δ-phases of pure iron could also be considered forms of steel. Steels with a carbon concentration of 0.83 wt% undergo a transformation from austenite into two intimately mixed solid phases at a temperature of 723°C. Although this phase transformation looks like a eutectic, the material above this point is a solid rather than a liquid. Hence, the transition is referred to as the *eutectoid*. As its name implies, alloys with carbon concentrations greater or less than the eutectoid are known as hypereutectoid or hypoeutectoid steels, respectively.

The eutectoid mixture of steel consists of *pearlite* in association with either ferrite or cementite regions, depending on the carbon concentration of the matrix. Since the pearlite phase is a lamellar mixture of ferrite and iron carbide, the interaction of light gives rise to a "pearl-like" multicolored pattern when viewed through a light microscope. Figure 3.11 illustrates the microstructural changes when austenitic steel is slowly cooled. For hypoeutectoid steel, ferrite begins to form along the austenite grain boundaries. Further cooling results in a ferrite-rich phase, with some remaining austenite crystals. At the eutectoid point of 723°C, the residual austenite is converted to pearlite, yielding a phase that contains both ferrite and pearlite crystals upon further cooling.

By comparison, hypereutectoid steel contains significantly greater carbon concentrations; cooling results in the precipitation of the excess carbon in the form of hard/brittle *cementite*, Fe_3C, nuclei that form along austenite grain boundaries. It should be noted that cementite is actually a metastable phase, with graphite representing the most stable form of carbon at equilibrium. However, it is difficult to obtain stable graphite nuclei in steels (with $<2\%$ C) since the Fe/Fe_3C metastable equilibrium is preferentially formed.

In Chapter 2, we showed how polycrystalline aggregates, always found in pure metals and alloys, form grain boundaries due to misaligned crystallites. Since the bonding character of neighboring atoms is broken across the grain boundary, the diffusion of impurities occurs more readily in these areas. Considering the atomic weights of Fe and C, pure cementite corresponds to 6.7% C by weight. It has been determined experimentally that the strength of steel increases with carbon content up to the eutectoid composition, and then begins to drop as cementite nuclei are formed in the material. It should be noted that other Fe–C phases exist with greater carbon concentrations than cementite. However, Figure 3.10 shows only the phases to the left of cementite that are technically useful for materials applications.

The phases of austenite, pearlite, and ferrite are relatively soft; hence, the observed high hardness of steels is obtained through processing of these materials. For instance, hypoeutectoid steel is first heated to form austenite and then slowly cooled so the pearlite/ferrite phases may be worked into desired shapes. If the material is

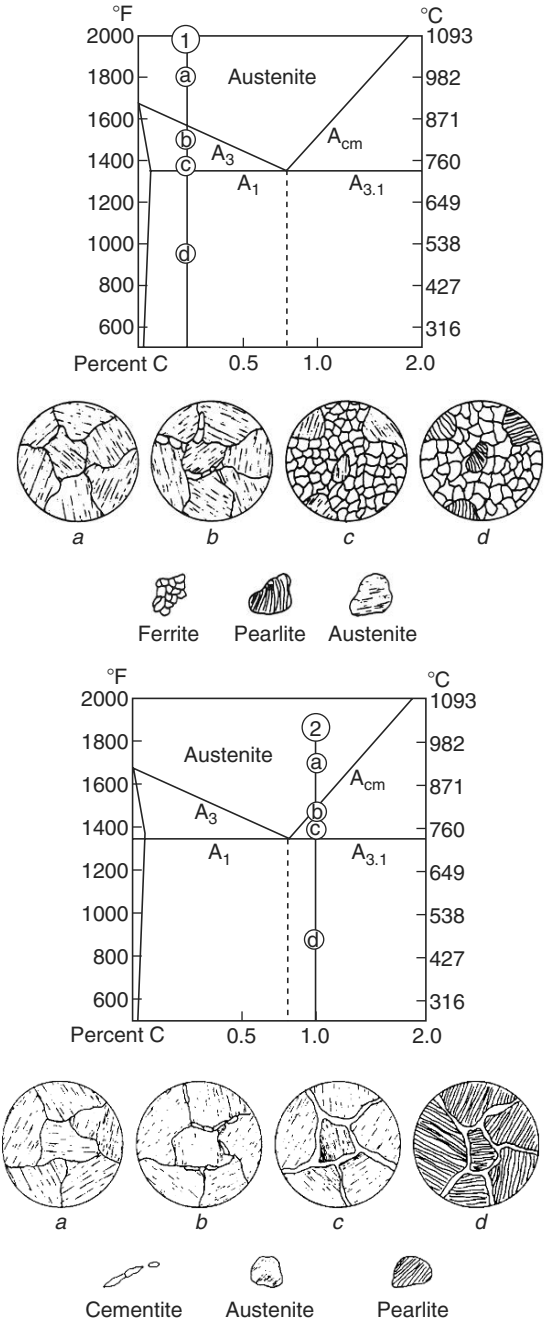


Figure 3.11. Comparison of the microstructural changes upon very slow cooling of hypoeutectoid (upper) and hypereutectoid (lower) steel. Reproduced with permission from *Machine Tools and Machining Practices*, White, W.; Wiley: New Jersey, 1977. Copyright John Wiley & Sons Limited.

reaustenized and quickly quenched to room temperature, a very hard phase known as *martensite* is formed. Some of the remaining pearlite and ferrite phases (if present) would still remain in the matrix. Hence, only when the steel has been heated to temperatures sufficiently high to convert all of the ferrite into austenite, that quenching will result in pure martensitic steel. It should be noted that the martensite phase does not appear in the above Fe–C phase diagram since it is a nonequilibrium phase.

As a final note regarding the Fe–C phase diagram, the eutectic temperature corresponding to the minimum melting point of the Fe–C system is 1,130°C. As the liquid is cooled at the eutectic temperature, solidification of ledeburite will occur. The microstructure of ledeburite is tiny austenite crystals embedded in a matrix of cementite. At carbon concentrations less than the eutectic (*i.e.*, 4.3 wt% C), ledeburite and austenite will form a solid solution. By contrast, increasing carbon concentrations will result in ledeburite/cementite solutions.

The incorporation of carbon into an iron lattice affects the interactions between neighboring iron atoms. As carbon is introduced at relatively low concentrations, the carbon atoms rearrange themselves within interstitial sites of the iron lattice, creating a strengthening effect. That is, metal atoms have a lesser range of movement due to the "glue" formed by interstitial carbon atoms. As a result, external forces such as temperature and pressure will not as readily cause atomic movement and surface/bulk deformation or fracturing. This is the reason why pure iron is not particularly hard or physically durable, but steels are significantly improved in these properties. Slow cooling of carbon-rich iron will yield a supersaturated solid solution. The carbon solubility in austenite decreases from about 1.7% at 1,150°C to about 0.7% at 715°C, causing the precipitation of the excess carbon in the form of microscopic carbides or graphitic nuclei.

Supersatured iron lattices yield a material known as *cast iron*, a ternary Fe–C–Si alloy containing much higher carbon than steel, typically around 3–5 wt% C. As the name implies, these materials are cast from their molten states into molds to yield the desired shapes. Due to the oversaturation of carbon present in these solids, cast iron is not suitable for structural applications. However, cast iron is extremely inexpensive to produce, making this material one of the most heavily used materials in industry for the manufacture of tools, valves, and automotive parts. Cast iron cookware has been employed for culinary applications since the late 19th century. However, with the advent of nonstick coatings such as Teflon in the 1940s, this application has largely been abandoned in favor of coated aluminum pans. A useful form of cast iron known as "Duriron," features a high silicon concentration (13–16 wt% Si, relative to standard cast irons with 1–3 wt% Si), and is resistant to strong acids and high temperatures.

There are a variety of cast irons, each differing in the nature of the carbon impurity associated with austenite within the iron lattice. For instance, white and gray cast irons contain cementite and graphite nuclei within the microstructure, respectively. The graphitic suspensions may be present as flakes (gray cast iron), or as spheres (ductile and malleable cast iron) depending on the cooling conditions employed. For gray cast irons, the formation of iron carbide must be minimized in order to prevent localized hard spots that would degrade ductility and machinability.

A number of dopants may be added to facilitate the preferential formation of graphite rather than cementite. As we have discussed earlier, the excess carbon precipitated from supersaturated iron will most often yield cementite. This is especially intriguing, since the formation of graphite actually represents the lowest-energy alternative for the Fe–C system. As a carbon-rich pure Fe/C alloy is cooled, the localized density of carbon atoms is never enough to serve as a nucleus for graphite formation. Rather, since the carbon is distributed throughout the lattice, the intimate combination of iron and carbon atoms makes it relatively easy to form Fe_3C nuclei, relieving the supersaturation and lowering the overall energy of the system. On the other hand, if a dopant is added to serve as a nucleation site, the formation of graphite will occur due to the more favorable thermodynamics.

It is proposed that the major nucleation mechanism in cast iron doping, known as *inoculation*, is the formation of sulfide species upon the addition of strong sulfide formers such as calcium, barium, cerium, or strontium. These sulfides possess lattice parameters very similar to the graphite crystal structure, serving as substrates for nucleation/epitaxial growth of graphite. Other common graphitizer dopants are Si (typically added as metal ferrosilicon compounds), Ni, and Cu; by contrast, Cr, Mo, V, and W are antigraphitizers, promoting the formation of carbides ($Cr_7C_3/Cr_{23}C_6$, Mo_2C/Fe_3Mo_3C, VC–V_4C_3, W_2C/Fe_3W_3C, respectively).

3.2.2. Hardening Mechanisms of Steels

As we have seen, it is not simply the carbon concentration, but rather the microstructure of Fe–C alloys that affect its physical properties. The size of the individual microcrystals (or grains) that comprise these aggregates greatly affects many properties of the bulk crystal. Both optical microscopy and X-ray diffraction are used to determine the grain sizes; most commercial metals and alloys consist of individual crystallites with diameters ranging from 10 to 100 μm, each corresponding to millions of individual metal atoms. A decrease in the size of these microscopic particles results in an increase in both strength and hardness of the bulk material. This can be understood by the tighter packing of smaller spheres relative to larger ones, effectively resisting atomic repositioning as a result of an external stress such as bending.

An external pressure (stress) that is exerted on a material will cause its thickness to decrease. A *shear stress* is applied parallel to the surface of a material, and may cause the sliding of atomic layers over one another. The resultant deformation in the size/shape of the material is referred to as *strain*, related to the bonding scheme of the atoms comprising the solid. For example, a rubbery material will exhibit a greater strain than a covalently bound solid such as diamond. Since steels contain similar atoms, most will behave similarly as a result of an applied stress. If a stress causes a material to bend, the resultant flex is referred to as *shear strain*. For small shear stresses, steel deforms elastically, involving no permanent displacement of atoms. The deformation vanishes when shear stress is removed. However, for a large shear stress, steel will deform plastically, involving the permanent displacement of atoms, known as *slip*.

Dislocation defects and thermal energy assist slip, allowing a sheet of atoms to slip gradually past one another. To stop slip, one can either lower the temperature, or spoil the crystal structure. A process referred to as *work hardening* (bending/hammering the cold material) is used to break up crystallites, introducing dislocations in the material. The more dislocations exist, the more they will interact with one another, becoming pinned or entangled with one another. This will impede the movement of dislocations and strengthening of the material. This is done at low temperature so the metal atoms cannot rearrange themselves, which would negate the effect.

The introduction of other metals into the lattice through alloy formation introduces alien crystallites that will also impede slip in steel crystals. In particular, if the alloying agent is carbon, hard crystallites of iron carbide may form, changing the microstructure to impede slip. By comparison, austenite usually does not contain iron carbide, and is quite susceptible to slip.

Heat treatment of iron alloys will also affect the slip characteristics of the material through changes in crystallite sizes. As the grain diameters become larger through annealing, more grain boundaries will tend to form, resulting in a greater proclivity for slip deformations. In general, as the grain size is decreased, the strength of the alloy will increase. As an illustration, consider a bag filled with sand *vs.* a bag filled with marbles. The extremely small granules of sand will be packed more efficiently with respect to one another, providing a solid with significantly greater density and less opportunity for individual granules to slide past one another, or change their positions in response to a bending force (Figure 3.12).

We are all familiar with the picture of a blacksmith withdrawing red-hot iron from a furnace and hammering it into the desired shape. Although these early laborers were not familiar with the microcrystalline changes they were imposing, they knew through experience that heating/cooling regimes were effective means to im-

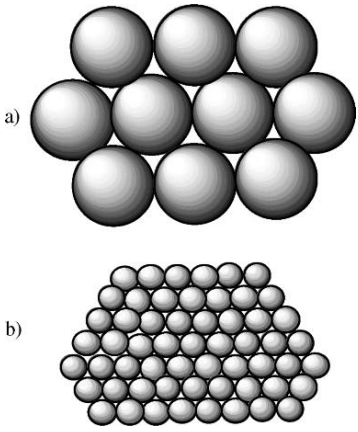


Figure 3.12. Schematic of the effect of grain size on ease of atomic movement. If a metal consists of large grains shown in (a), much less external force would be necessary to cause atomic movement (*i.e.*, bending), relative to a metal comprising small grains shown in (b).

prove many properties of steel such as hardness, toughness, ductility, machinability, and wear/stress resistance. The softening of steel at elevated temperatures is due to the formation of large iron carbide crystallites that may undergo facile slip deformations. The prolonged high temperature environment affects the microstructure through interrupting Fe–Fe and Fe–C bonds. This allows for cementite (Fe_3C) regions to agglomerate into spheres, yielding a new morphological structure referred to as *spheroidite*. This process, referred to as *precipitation hardening* induces an increase in hardness through the formation of homogeneous dispersions that reduce slippage between grains. It should be noted that the heating of iron surfaces is now often achieved by more energetic sources such as electron beams or lasers. Such a focused thermal treatment allows for localized surface hardening to improve its wear resistance.

The temperature range is not the only crucial variable affecting the properties of the material, but rather the *rate* of heating and cooling. For example, when austenitic steel is very slowly cooled to room temperature, the resulting solid will be soft and malleable. However, when the same steel is rapidly quenched in cold water to temperatures less than 250°F, the normal phase transformations to ferrite or pearlite (depending on %C present) are suppressed. Rather, the γ-Fe phase is converted to *martensite*, the hardest and strongest of all possible Fe–C microstructures. Interestingly, the formation of martensite may also occur through fast quenching of other austenitic mixtures, such as ferrite- or cementite-rich austenite (Eq. 7).

$$
\begin{aligned}
&\{\gamma\text{-Fe} + \alpha\text{-Fe}\} \xrightarrow{\text{rapid cooling}} \text{martensite} + (\alpha\text{-Fe}) \\
&\{\gamma\text{-Fe} + \text{Fe}_3\text{C}\} \xrightarrow{\text{rapid cooling}} \text{martensite} + (\text{Fe}_3\text{C})
\end{aligned}
\tag{7}
$$

The slowest cooling rate that prevents formation of the above nonmartensitic decomposition products is referred to as the critical cooling rate. Figure 3.13 shows the general procedure to plot a time–temperature transformation curve. Thin metal specimens are suitably heated to form austenitic steel. This temperature is held at varying temperatures to ensure full conversion of the microstructure to austenite. That is, to allow for metal carbides to dissolve in austenite; incomplete conversion of carbides will result in ferrite grains that will ultimately weaken the material. The austenized specimens are removed at specific times and quenched in cold water. Using optical microscopy, the nature of the decomposition products is determined (*i.e.*, pure martensite, ferrite, pearlite, *etc.*). A large number of samples are required to determine the time intervals required for the initial and full transformation of austenite to other phases, making this process an extremely labor-intensive exercise.

A slow rate will give the greatest opportunity for controlled atomic migration within the lattice, and growth of large ordered crystallites. However, in the rapid nonequilibrium conditions used to form martensite, there is no time for carbon diffusion to occur; this yields a supersaturated solution, with >2wt% C present as an interstitial impurity. During this phase transition, the fcc lattice of austenite is transformed to a distorted bcc lattice, commonly referred to as body-centered tetragonal bct (Figure 3.14). The degree of distortion from a perfect bcc lattice (cubic lattice axes ratio, $c/a = 1$) is amplified with increasing carbon concentration (Eq. 8).

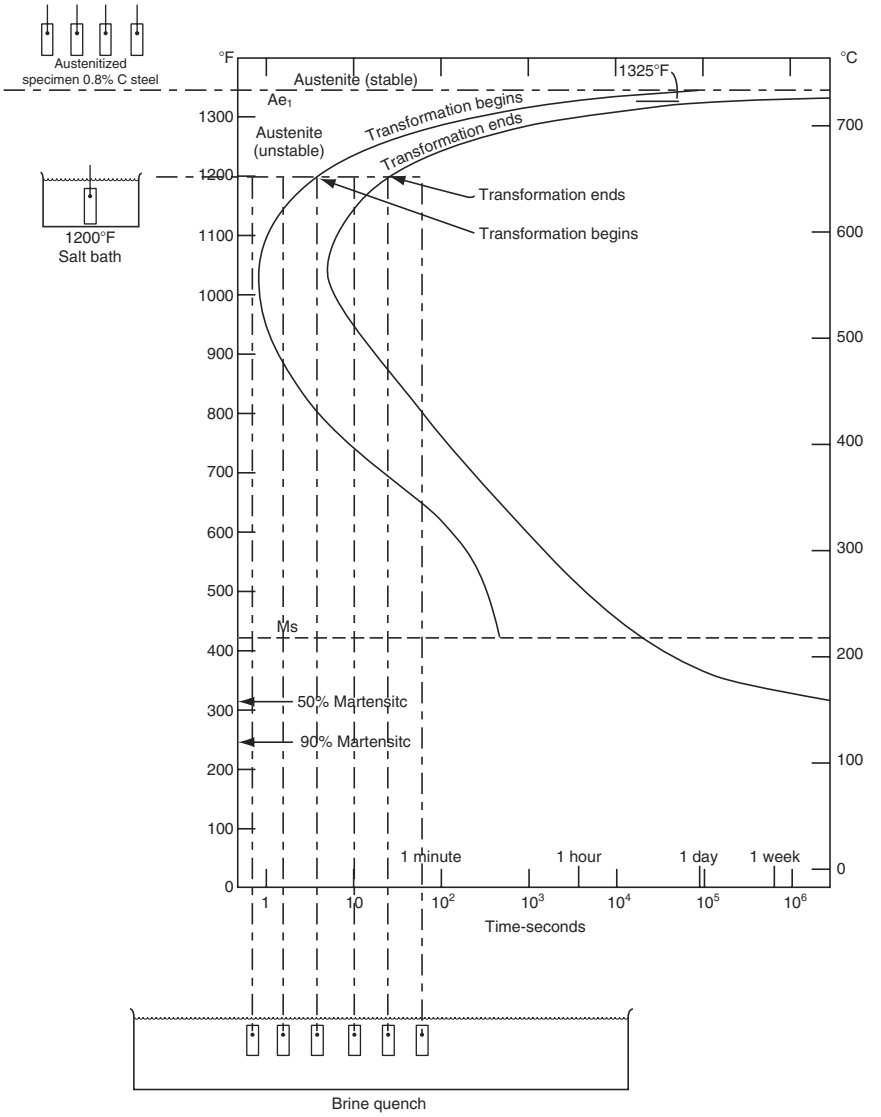


Figure 3.13. Method of generating a time–temperature transformation (TTT) diagram. Reproduced with permission from *Practical Metallurgy and Materials of Industry*, Neely, J. E.; Bertone, T. J.; 5th ed.; Prentice-Hall: New Jersey, 2000.

$$\frac{c}{a} = 1 + 0.045(\text{wt\% C}) \tag{8}$$

The fcc–bct conversion, known as the *Bain transformation*, is a diffusionless process. That is, unlike the previous high-temperature conversions we saw earlier (*e.g.*, austenite to ferrite), martensite can form at temperatures significantly below room

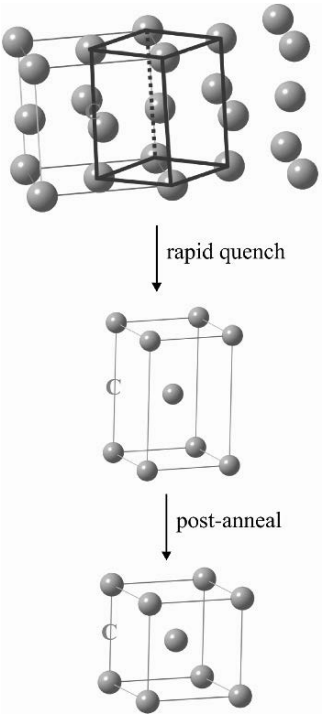


Figure 3.14. Crystallographic representation of the phase transformation from austenite to martensite. Two neighboring fcc unit cells of austenite associate, resulting in a body-centered tetragonal (bct) unit cell. A postanneal known as tempering converts the bct structure to α-Fe. Also shown is the placement of an interstitial carbon atom, remaining in an octahedral site among lattice iron atoms.

temperature, within 1×10^{-7} s. Such a fast growth rate precludes the deciphering of the exact mechanism for the nucleation and growth of martensite. However, leading theories suggest that the growth initiates from dislocations in the solid.

It is interesting to note that alloying metals that are present in austenite will decrease the temperature required to start martensite growth. The effect of alloying elements may be understood by examining the austenite/ferrite regions of the Fe–C phase diagram with respect to the alloy concentrations (Figure 3.15). Ferrite-forming elements such as Al, Si, W, Cr, and Mo result in a contraction of the austenite region, forming a gamma loop. By contrast, austenite-stabilizing elements such as C, N, Mn, Ni, and Cu cause an expansion of the austenitic phase boundary. Austenite stabilizers inhibit the nucleation and growth of ferrite and pearlite phases, assisting in the formation of pure martensite upon quenching. In general, since bcc ferrite contains more voidspace than fcc austenite, larger interstitial dopants may be incorporated into these lattices. Hence, ferrite stabilizers tend to be larger in contrast to the smaller size of austenite stabilizers. Most importantly, in accord with the "like dissolves like" principle, bcc and fcc dopants will tend to stabilize ferrite and austenite, respectively.

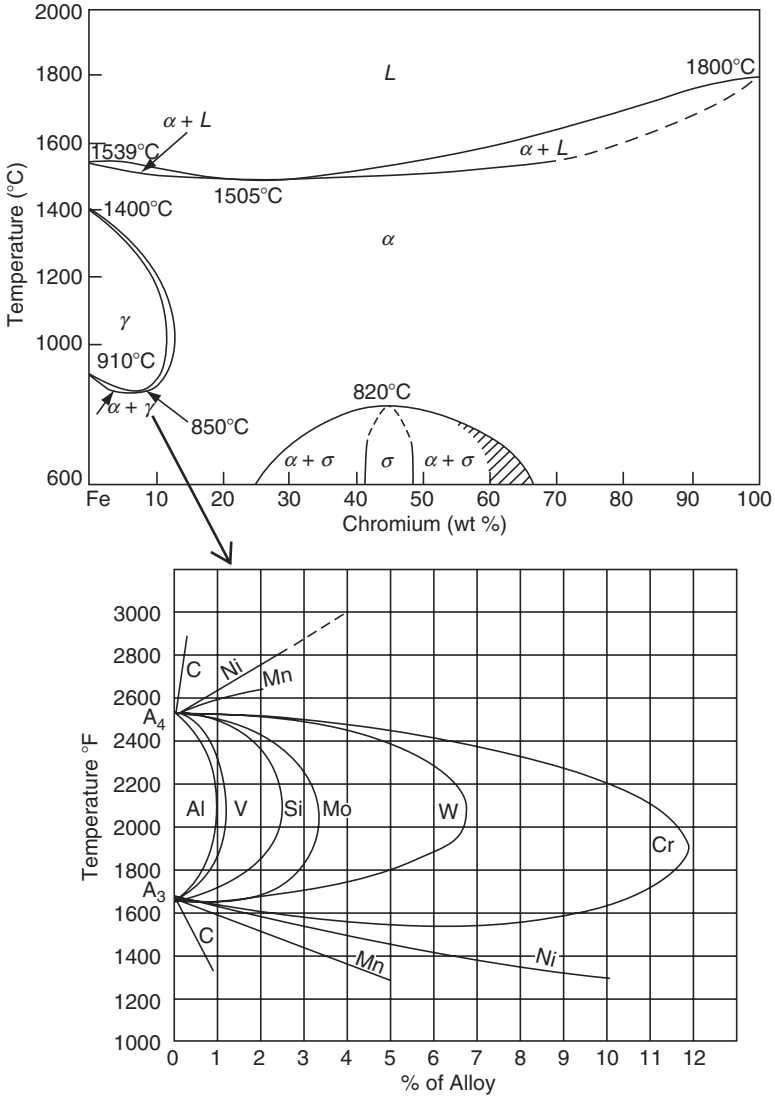


Figure 3.15. The phase diagram for the Fe–Cr system, illustrating the expanded ferrite region, relative to the normal Fe–C diagram. Below is shown the "gamma loop" illustrating the size of the austenite region from the presence of various dopants. Reproduced with permission from (top) *Steels: Microstructure and Properties*, Honeycombe, R. W. K.; Bhadeshia, H. K. D. H.; 2nd ed.; Copyright 1995 Elsevier, and (bottom) *Basic Metallurgy: Volume I, Principles*, Grosvenor, A. W.; 3rd ed.; American Society for Metals: Cleveland, OH, 1958. All rights reserved (http://www.asminternational.org).

Equation 9 quantifies the effects of dopant concentrations on the temperature required for the onset of martensite formation, M_s. The greatest effects are seen for the austenite-forming elements of C, Mn, and Ni where even small concentrations result in a sharp decrease in M_s. Whereas pure γ-iron may be converted to martensite at temperatures in excess of 500°C, hypereutectoid steel is not transformed to martensite until a temperature of *ca.* 160°C is reached during quenching. At carbon concentrations above 0.7%, martensite is still being formed at temperatures well below 0°C. Hence, high-C steels must be quenched in low-temperature media (*e.g.*, dry ice/acetone, liquid nitrogen) to ensure full conversion of austenite to martensite.

$$M_s(^\circ C) = 539 - 423(\%C) - 30.4(\%Mn) - 17.7(\%Ni) - 12.1(\%Cr) - 7.5(\%Mo) \quad (9)$$

Even though martensite has a pronounced hardness, the as-quenched material is much too brittle and highly stressed for structural applications. The ductility and toughness of martensite is greatly improved through postannealing, a process known as *tempering*. This process relieves stresses in the solid through conversion of bct martensite into bcc ferrite, with precipitation of iron carbide particulates. It is important to note that the annealed structure is not simply pearlite/ferrite, but is best referred to as tempered martensite. During the annealing process for martensite, a number of key transformations occur:

(i) 50–250°C: Interstitial carbon atoms in martensite begin to diffuse within the BCT lattice. This results in precipitation hardening from the formation of hexagonal ε-iron carbide crystallites, $Fe_xC (2<x<3)$. The martensite bct crystal lattice begins to lose its tetragonality.

(ii) 250–350°C: Decomposition of retained austenite to fine aggregates of ferrite plates (acicular ferrite) and cementite crystallites. This intermediate microstructure is referred to as *Bainite*, nucleating from the surface of ferrite crystallites. As the temperature approaches 300°C, cementite begins to dominate the microstructure at the expense of the ε-Fe particles. At temperatures approaching 350°C, the bct lattice is transformed to ferritic bcc.

(iii) 350–700°C: The cementite particles undergo a coarsening process (between 300 and 400°C), and spheroidization (near 700°C). These processes drastically reduce the hardness of the material, but improve overall ductility and brittleness characteristics (more desirable for particular applications).

Hence, to obtain high-strength tempered steels, it is essential to anneal at low temperatures (<350°C). As one might expect, the presence of alloying elements will have a dramatic effect on the microstructure during annealing. For instance, if 1–2 wt% Si is present, ε-iron carbide particles are stabilized up to temperatures of 400°C, yielding a much harder material at elevated temperatures. Further, transition metals such as Cr, Mo, V, W, and Ti will form stable carbides with higher enthalpies of formation than Fe_3C, typically at temperatures between 500 and 600°C. A high temperature is required due to the relatively low diffusivity of the alloying elements that must substitutionally diffuse through the iron lattice. By contrast, interstitial dopants such as C, N, and B move between the iron lattice sites with a much greater diffusivity. It is important to note that the alloy carbides remain as fine suspensions

even after prolonged tempering. This results in substantial strengthening referred to as *secondary hardening*.

From an analysis of various types of steels, only the following carbides will be present: Fe_3C, Mn_3C, $Cr_{23}C_6$, Cr_7C_3, Fe_3Mo_3C, Fe_3W_3C, Mo_2C, W_2C, WC, VC, TiC, NbC, TaC, Ta_2C, and ZrC. The occurrence of these species will depend on the type and concentration of the transition metal dopants within the iron lattice. For interstitial carbides, the size of the metal atoms will govern the type of carbide formed. In general, the metal radius must be >1.35 Å (*e.g.*, Ti, Zr, Hf, V, Nb, Ta, Mo, and W) to generate an interstitial vacancy large enough to accommodate C atoms. Metals with smaller radii (*e.g.*, Cr, Mn, Fe, Co, Ni) do not form MC species, and form carbides with relatively complex crystal structures (Figure 3.16). It should be noted that metal carbides do not generally exist as isolated pure species. That is, carbides of all alloying elements will exist as clusters that also contain iron. Further, when several carbide-forming dopants are present that share the same crystal structure, the resultant carbide will be present as a combination of those elements. As an example, steel containing Cr and Mn dopants will contain particulates of the complex carbide $(Cr, Mn, Fe)_{23}C_6$, rather than isolated $Cr_{23}C_6$ and Mn_3C species.

We have seen that only certain transition metals will form stable carbides; as a relevant digression, let us consider the chemical rationale behind such reactivity. The general trend for increasing carbide-forming ability of transition metals is

$$Fe < Mn < Cr < Mo < W < V < Nb < Ta < Ti < Zr < Hf$$

If one follows this sequence using the Periodic Table, this grouping consists of early transition metals that are relatively electron deficient. As you may recall, the valence shell of zero-valent transition metals in a crystal lattice is $[(ns^2)((n-1)d^x))]$. In the bulk solid state, the outer *s* electrons are completely delocalized, whereas the wave functions of the *d* electrons remain localized on the respective metal atoms. When a carbon atom enters the crystal field it behaves as a ligand toward the metal, with the ligand and metal electrons electrostatically interacting causing the d orbitals to lose their original degeneracy.[3] Since this is an electrostatic effect, stronger metal–carbon bonds will result from more diffuse metal d orbitals (5d *vs.* 4d *vs.* 3d), and metals with fewer *d* electrons (*i.e.*, both corresponding to less electron–electron repulsions between ligands and the metal).

The transference of electron density from the metal to carbon will result in the formation of a strongly polar covalent bond, between carbide ions ($C_x{}^{n-}$) and transition metal ions. It should be noted that only the portions of the alloying elements and carbon that cannot be dissolved in austenite at a given temperature may be used in carbide formation. Two types of carbides are possible. If the atomic radius ratio of carbon/metal is <0.59, an interstitial phase will result; otherwise, carbides of complex compositions will form, having a different crystal lattice from the host metal. The latter carbides are characteristic of Cr, Mn, and Fe, which have five or more *d* electrons and a correspondingly weaker interaction with ligands such as carbon. Hence, such complex carbides (*e.g.*, Figure 3.16a–d) will have lower melting points and hardness than analogous interstitial carbides of the 5d and/or early transition metals (*e.g.*, W_2C used as cutting blades).

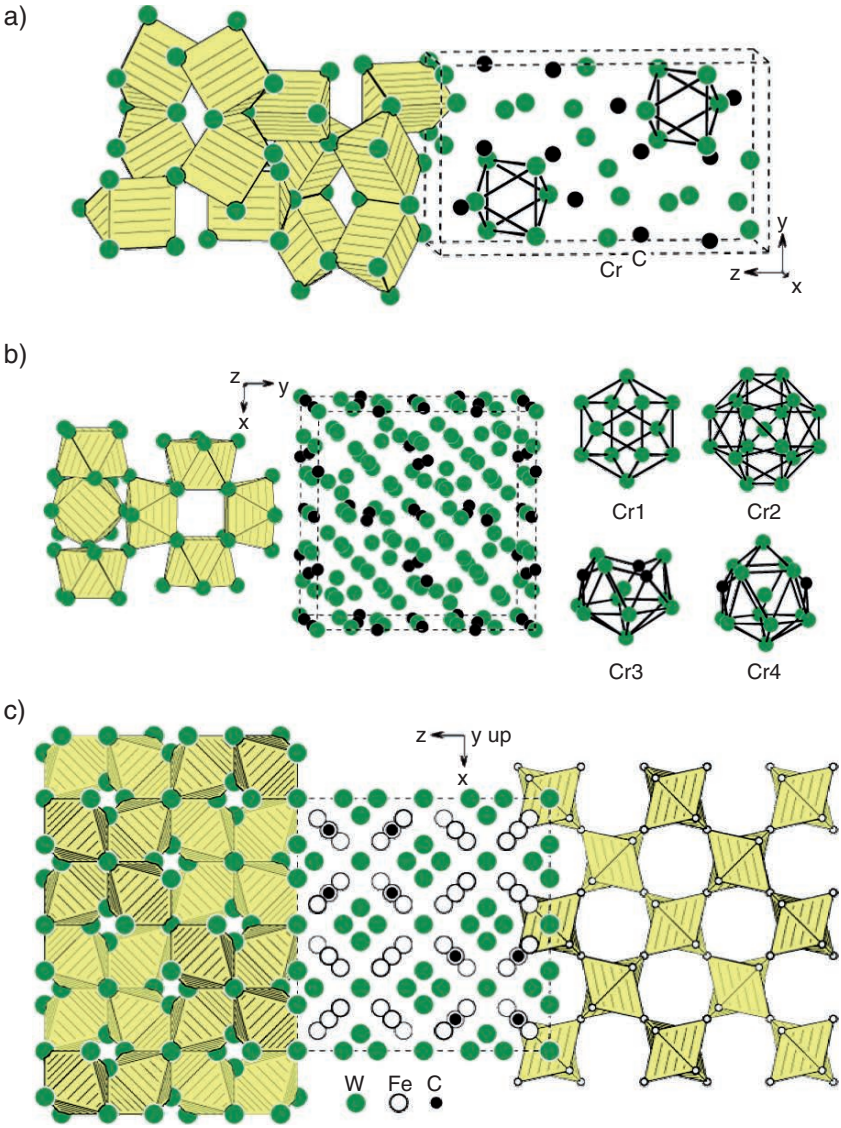


Figure 3.16. Crystal structures of typical carbides in steels. Shown are (a) orthorhombic Cr_7C_3 (space group: *Pnma*), (b) cubic $Cr_{23}C_6$ (space group: *Fm3m*), (c) cubic Fe_3W_3C (space group: *Fd3m*), (d) orthorhombic Fe_3C (space group: *Pnma*), and (e) cubic TiC (space group: NaCl) and hexagonal W_2C (space group: CdI_2). Reproduced with permission from *Handbook of Ceramic Hard Materials*, Riedel, R. ed., Vol. 1. Copyright 2000 Wiley-VCH. (http://www.hardmaterials.de/html/_crystal_structures.html).

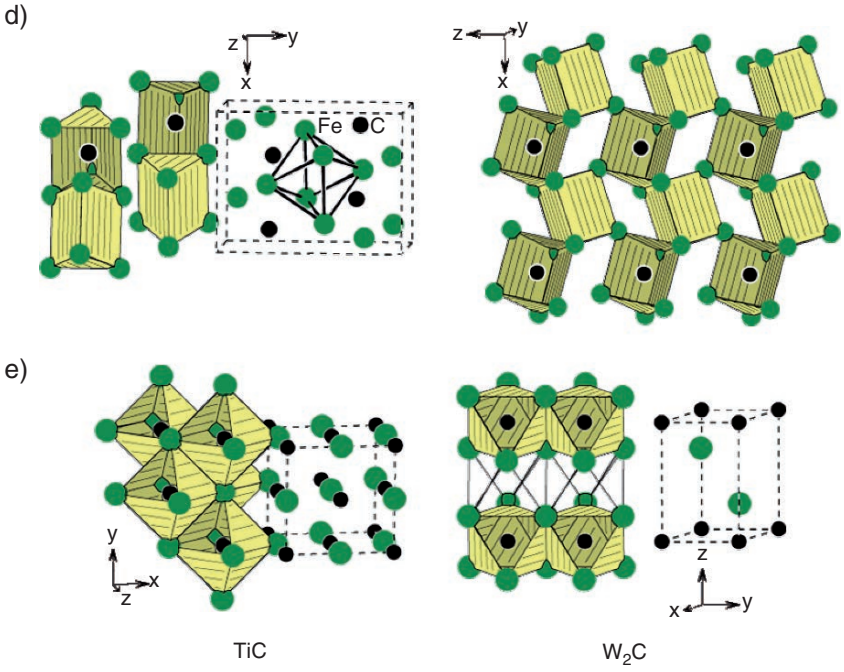


Figure 3.16. Continued

One serious drawback of the above austenize/rapid quench method for martensite formation is the possibility of distorting and cracking the metal due to the rapid cooling event. During the quenching process, thermal stresses arise from the varying cooling rates experienced by outer and interior areas of the steel. In addition, there is a volume change when austenite is transformed to martensite. Two methods that have been used to reduce quenching stresses are martempering and austempering (Figure 3.17). Martempering allows the transformation of austenite to martensite to take place at the same time throughout the structure of the metal part. By using interrupted quench, the cooling is stopped at a point above the martensite transformation region to allow sufficient time for the center to cool to the same temperature as the surface. Then cooling is continued through the martensite region, followed by the usual tempering process. By comparison, in austempering, the austenized steel is quenched at a rate faster than that required for pearlite formation, but above the temperature required for martensite growth. Hence, rather than transforming to martensite, the center and surface are converted to bainite – a strong material that shares the hardness of martensite with the toughness of pearlite. Bainitic steels are frequently used for the fabrication of gardening tools such as shovels.

The above changes in the microstructure upon annealing do not only apply to the bulk material, but also for the surface. If an iron material is placed at high temperature in the presence of carbon vapor, a procedure known as *carburization* occurs,

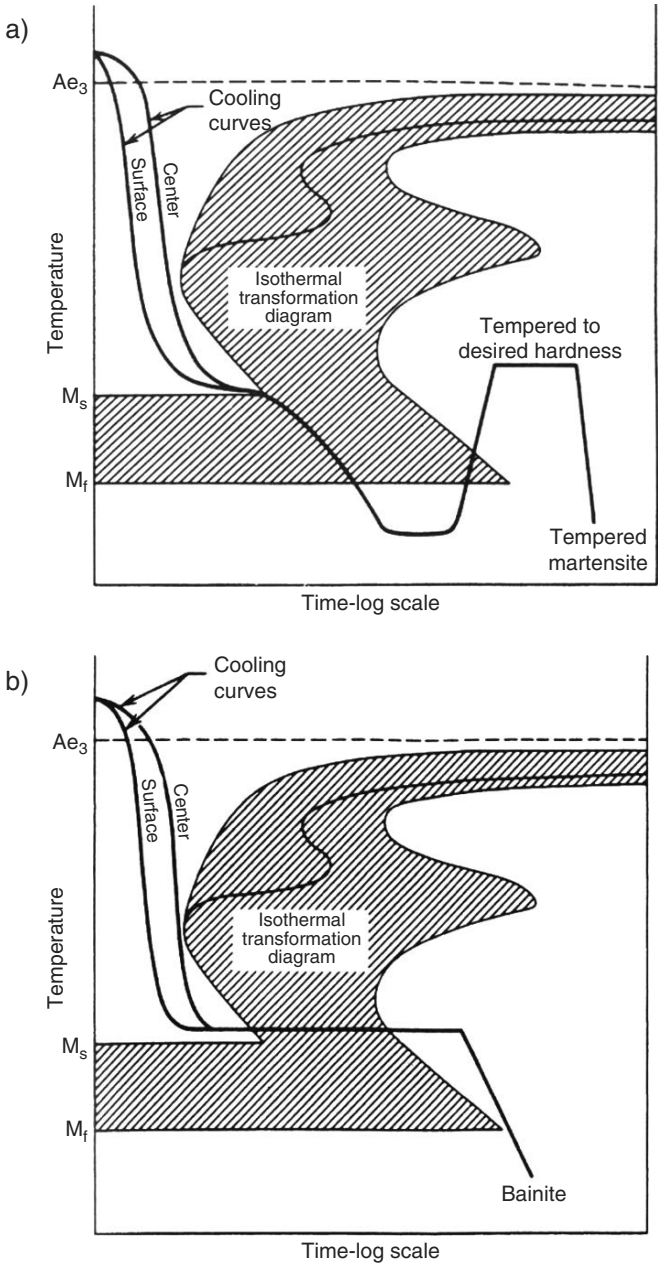


Figure 3.17. Schematic of (a) martempering and (b) austempering processes for steel. Courtesy of the International Steel Group and Mittal Steel Company (http://www.mittalsteel.com).

where carbon atoms diffuse into the surface of the steel, increasing the surface hardness. There must be careful control of the annealing atmosphere; if the steel is brought into contact with an oxidizing atmosphere, decarburization of the surface will occur through preferential formation of CO_2.

Other surface hardening techniques introduce nitrogen to steels containing metals such as Al, Cr, and V that form stable nitrides (Eq. 10). Surface hardening techniques add a variety of attractive properties to steel components such as increasing wear and stress resistances, decreasing the odds of fracturing, and increasing corrosion resistance.

$$4Cr + 3NH_{3(g)} \xrightarrow{600^{\circ}C} Cr_4N_3 + \frac{9}{2}H_{2(g)} \quad (10)$$

To summarize, there are five techniques that may be used to strengthen a metal. It should be noted that these methods are applicable to all metal classes, not just the iron alloys discussed in this section.

(i) *Strain hardening* – plastic deformation through cold or hot working of the material increases the density of dislocations. This obstructs the movement of neighboring dislocations in the solid.

(ii) *Precipitation hardening* – the formation of homogeneous suspensions of small, finely dispersed particles in a matrix prevents grain slippage (*e.g.*, formation of carbide crystallites upon annealing martensite).

(iii) *Grain size hardening* – the reduction in grain boundaries through prolonged thermal treatment also prevents movement among neighboring crystallites in the lattice.

(iv) *Solute hardening* – the addition of alloying dopants such as Mn or Cu, or interstitial atoms such as C, N result in a distortion of the metal lattice, increasing the lattice energy in the vicinity of the dopant. Strengthening occurs since more work is required to propagate a dislocation through these areas.

(v) *Surface hardening* – strengthening of the exterior of a material through either diffusional incorporation of dopants (*e.g.*, B, C, N), or selective portions of the surface using flame, induction, laser, electron beam, or ion bombardment. Strengthening occurs through either grain size reduction or solute hardening, but occurs only on the periphery of the material due to the controlled, limited exposure.

The typical method used to assess the strength of a metal is a tensile test, where the metal is clamped to upper and lower jaws, and pulled until it fractures. Figure 3.18 illustrates typical stress *vs.* strain curves for iron, steel, and hardened steel. The elastic limit (EL) is the greatest stress that a material may withstand and still revert back to its original shape when the stress is removed. In general, the closer a material is to its elastic limit, the longer it will take for the material to subsequently return to its original size/shape. For a metal within its EL the crystal lattice will be elastically lengthened, while becoming thinner at right angles to the applied stress. The ratio of lateral change to the change in length is referred to as *Poisson's ratio*.

When the elastic limit of a metal has been exceeded, it will undergo plastic flow beginning at the yield point. It is this property of metals that is exploited for cold and hot working into desired shapes. When a metal is deformed permanently from the tension force, it exhibits a property known as *ductility*. By comparison, the term

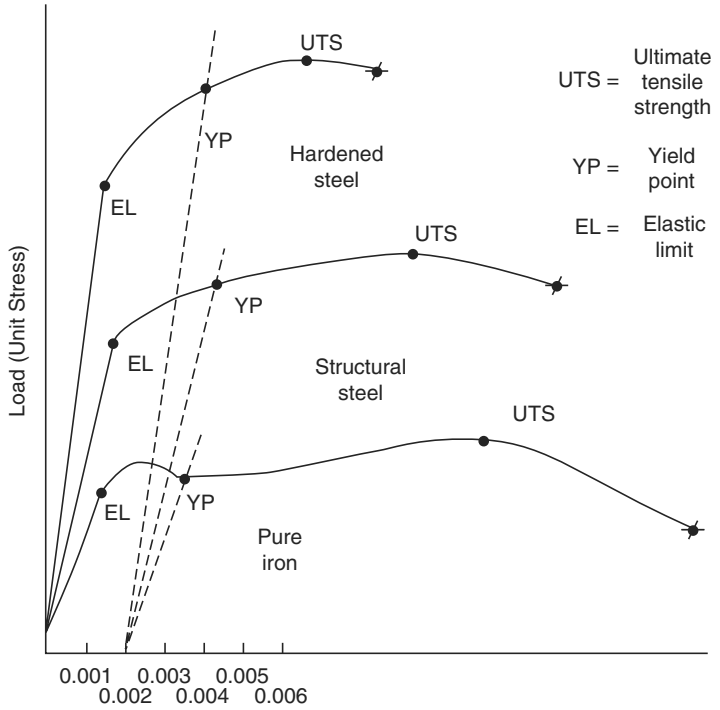


Figure 3.18. Comparative *stress vs. strain* curves for pure iron, steel, and hardened steel. Reproduced with permission from *Practical Metallurgy and Materials of Industry*, Neely, J. E.; Bertone, T. J., 5th ed., Prentice-Hall: New Jersey, 2000.

malleability refers to the permanent deformation of a metal under a compression force (*e.g.*, hammering, cold-rolling, *etc.*). Although most ductile metals are also malleable, the reverse is not always true. For example, lead is extremely malleable but is not easily drawn into a wire without the use of die-extrusion techniques. It should be noted that excessive cold working of metals may cause brittleness, where the metal will fracture rather than exhibiting plastic flow under stress. As previously mentioned, a high concentration of carbon in the metal lattice (*e.g.*, cast iron) will also cause brittleness, which explains the lack of structural applications for these metals.

3.2.3. Stainless Steels

Technological advancements in iron ore processing and metal doping have resulted in the fabrication of many types of high-strength steels for diverse applications. By contrast, earlier generations worked exclusively with wrought iron, an inferior material containing >20 wt% C, formed through simple annealing of the ore with coal. We have seen that the concentration and form (*e.g.*, Fe_3C, graphite, *etc.*)

of dopant species will alter the physical properties of the material. Hence, the bulk and/or surface thermal and physical processing of steels is the most important consideration for ultimate material performance, as these treatments greatly affect its microstructure.

Thus far, we have focused primarily on Fe–C alloys, with carbon atoms positioned within vacant interstitial sites within the iron lattice. As you may expect, a variety of other elements may also be present in steel that will alter its overall physical properties. For example, all steels contain manganese that assists in hardening mechanisms, as well as removing sulfur and oxygen atoms from the matrix. This prevents FeS formation and removes bubbles in the molten state of steels, both of which would greatly contribute to brittleness of the final product.

Typically, large transition metal dopants will exist as substitutional alloys, randomly replacing iron sites throughout the lattice. Steel containing <0.30 wt% C and chromium concentrations >10.5 wt% are referred to as *stainless steels*. The addition of Cr results in the formation of a native layer of Cr_2O_3, providing corrosion resistance. As the concentration of Cr is increased, the material is concomitantly less predisposed to rust. Such protection occurs from the comparative oxidation potentials between Cr and Fe (Eq. 11). For redox processes, the spontaneity (Gibbs free energy) is governed by Eq. 12, where a negative ΔG indicates a spontaneous reaction at equilibrium. If both chromium and iron are present together, chromium atoms are considered as a *sacrificial anode*, being preferentially oxidized leaving the iron untouched. Other metals with large positive oxidation potentials such as Zn (+0.763 V), Al (+1.68 V), Ni (+0.257 V), and Ti (+2.00 V) are also useful additives that serve as corrosion barriers.

$$\begin{aligned} &\mathrm{Fe} \rightarrow \mathrm{Fe}^{3+} + 3\mathrm{e}^- \quad E^\circ = -0.331\,\mathrm{V} \\ &\mathrm{Cr} \rightarrow \mathrm{Cr}^{3+} + 3\mathrm{e}^- \quad E^\circ = +1.32\,\mathrm{V} \end{aligned} \tag{11}$$

$$\Delta \mathrm{G}^\circ = -nFE^\circ \tag{12}$$

where n is the number of electrons involved in the redox process, F is Faraday's constant (9.64853×10^4 C mol^{-1}) and E° is the reaction potential, measured at STP.

A major difference between stainless and plated steels is the former will actually self-repair itself when scratched. Since the chromium is homogeneously dispersed throughout stainless steel, a scratch will serve to expose additional Cr sites forming additional layers of the protective oxide. By contrast, the application of a protective coating over steel will only be an effective barrier as long as it remains intact. When this coating is penetrated by a scratch/crack, the bare steel is exposed to the surrounding environment allowing the possibility for corrosion. Often, aluminum and silicon are also added to steel that also form native oxides that are effective in preventing surface corrosion of the underlying metal.

When some stainless steels are overheated (*ca.* 400–800°C) for a prolonged period, there exists the possibility for chromium carbide formation. Most often this results from an attempt to weld steels that are not suitable for such high-temperature treatment. If such a precipitous reaction causes the bulk Cr concentration to fall below 10.5 wt%, corrosion protection is drastically reduced. To make the situation

worse, the carbide usually forms at grain boundaries, leading to intergranular corrosion and stress cracking. Amazingly, this process is reversible, by reheating the steel to temperatures in excess of 1,000°C for a period long enough to redissolve the chromium carbide particles and form a homogeneous solid solution. Rapid cooling must then be introduced to suppress the reformation of carbide. Hence, if one wishes to use a stainless steel at high temperatures, either low C compositions must be used, or doping with carbide-forming metals such as V, Ti, or Ta that are more easily oxidized than Cr. As a general rule of thumb, more chromium must be added as the concentration of carbon is increased to ensure effective corrosion resistance.

There are currently over 200 commercially available types of stainless steels. Hence, there is an exact composition of stainless steel for virtually any application. As we have already seen, a tremendous number of substitutional and interstitial dopants may be alloyed with iron, resulting in significant changes in their physical properties. In addition, varying the heat treatment of the bulk or surface of steels will change these properties even further. It is truly mind-boggling to think of all the combinations of dopant composition/postprocessing that are possible! Fortunately, all of these combinations fall under the umbrella of four general types of stainless steels, classified according to their microstructural phases/compositions (Table 3.3).

The industrial applications for austenitic stainless steels far outweigh the other types due to their facile work hardenability and high corrosion resistance. As we have seen, the fcc austenite phase is not stable at temperatures below 723°C; however, austenite-stabilizers such as Ni, Mn, Cu, C, or N may be added to extend the stability of this phase down to room temperature. Figure 3.19 shows the stable phases that exist at room temperature, as a function of the Cr and Ni concentrations. An easy way to think about the effect upon Ni alloying is the replacement of an increasing number of iron atoms in the lattice with Ni (stable fcc lattice at room temperature), results in the solid solution being "fooled" into crystallizing in an fcc array rather than bcc ferrite. Due to high concentrations of easily oxidizable elements such as Cr, Ni, and Mn, the corrosion resistance is the greatest for austenitic stainless steels. However, their Achilles' heel is their reaction with chloride ions. Due to the large concentration

Table 3.3. General Types and Properties of Stainless Steels

Type	Concentration	Properties/applications
Martensitic	11–20 wt% Cr 0.15–0.75 wt% C	High hardness, magnetic/cutlery, blades, surgical instruments, valves, springs
Austenitic	16–26 wt% Cr 35 wt% Ni 20 wt% Mn	High and low temperature resistance, ductility, superior corrosion resistance/kitchen sinks, ovens, reaction vessels, food processors, gutters
Ferritic	10.5–30 wt% Cr <1 wt% C, N, Ni	Magnetic, inexpensive/automotive exhaust and fuel lines, cooking utensils, bank vaults, washing machines, dishwashers
Duplex (austenitic–ferritic)	18–26 wt% Cr 4–7 wt% Ni 2–3 wt% Mo	Weldable, high tensile strength, Cl^- ion resistance (acidic environments)/desalination plants, food pickling plants, petrochemical plants, pulp and paper industries

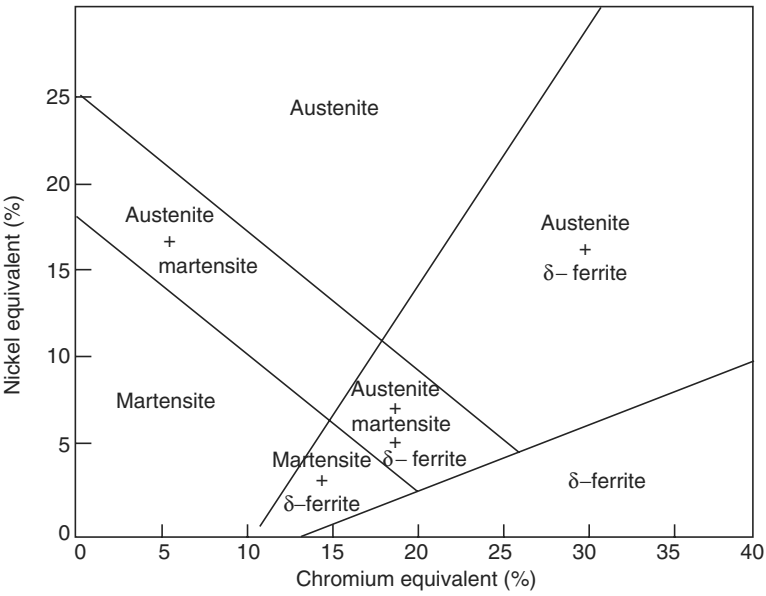


Figure 3.19. Relative phase stabilities of Ni–Cr steels. Reproduced with permission from *Steels: Microstructure and Properties*, Honeycombe, R. W. K.; Bhadeshia, H. K. D. H.; 2nd ed., Wiley: New York, 1995.

of transition metals, Cl^- will preferentially react with the metal centers, forming MCl_x rather than a protective coating of M_xO_y.

Since ferritic stainless steels contain more carbon than other classes, they are relatively harder to weld and shape than other varieties, which have historically limited their applications. However, since the 1960s, processes such as "argon oxygen decarburization" have resulted in steels with less carbon, allowing for smaller concentrations of chromium to be used. As a result, the price for ferritic stainless steel has dramatically dropped, and a number of applications now employ these materials – more than 2/3 of which include automotive exhaust systems.

Duplex stainless steels feature the "best of both worlds" since they contain both ferritic and austenitic phases. The ferritic phase helps to circumvent the problems associated with stress-corrosion cracking in chloride environments, while the austenitic component helps to improve the generally low strength and ductility of purely ferritic steels. This biphasic steel is generated through careful heating/cooling of the Fe–Cr–Ni alloy. When these materials initially solidify, they are 100% ferritic in nature; subsequent cooling causes the precipitation of austenite within individual ferrite crystals.

Strengthening of stainless steels is carried out through cold-working processes – rolling into sheets or drawing into wires/rods at temperatures around 25°C. This generates a strong material through formation of a distorted bcc lattice that is roughly analogous to martensite. Although we indicated that austenite may be stabilized at room temperature through alloying, this is only a metastable phase. In fact,

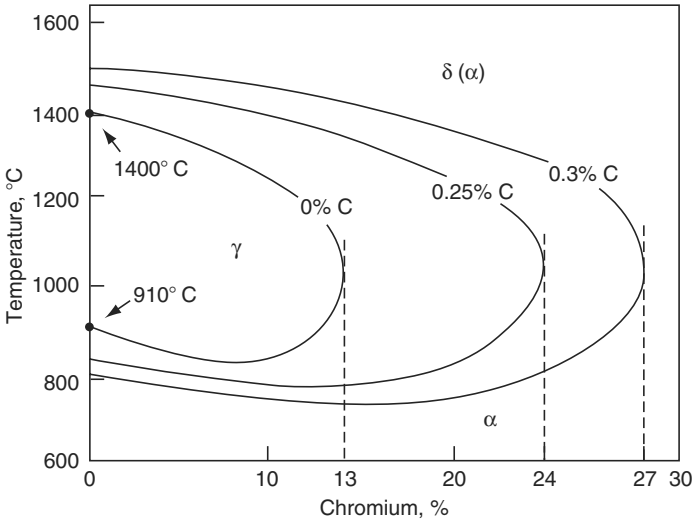


Figure 3.20. The effect of carbon on the gamma loop. Reproduced with permission from *Stainless Steels*, Lacombe, P.; Baroux, B.; Berauger, G. eds. Copyright 1993 EDP Sciences, Les Ulis, France.

if the steel is cooled to subzero temperatures or cold worked, the ferrite phase will be generated. Since the austenite would have contained at least a small amount of carbon, the resulting ferrite phase will also contain carbonaceous suspensions, *i.e.*, resulting in a martensite-like structure that will possess high hardness.

From a re-examination of the gamma loop in Figure 3.15, one can see that heating stainless steel with typical Cr concentrations >11 wt% will not yield austenite upon heating, as this region is outside the loop. Hence, it is not possible to quench harden these materials through transformation to the martensite phase. For example, austenitic, ferritic, and duplex stainless steels may not be hardened through heat treatment due to their high Cr/C ratios. Interestingly, increasing the carbon concentration will extend the gamma loop (Figure 3.20), allowing one to austenize the stainless steel and harden through fast quenching to martensite.

3.2.4. Nonferrous Metals and Alloys

Although we have focused on iron for the majority of this chapter, many of the materials we use on a daily basis comprise other metals. Applications that are particularly suitable for other metals include those that require more lightweight, highly conductive, and/or corrosion-resistant materials (often at a lower cost), relative to iron-based materials. Since metals comprise only one portion of this textbook, we do not intend on discussing all of the other metal classes in as much detail as the iron system. We may now expand our discussion a bit since the general structures and mechanisms involved for alloying, surface/bulk hardening, annealing, *etc.* also apply for other metals. For more information about the structure and processing of other metal

classes that are not discussed herein, refer to the Further Reading section at the end of this chapter.

The coinage metals

Let us begin our survey of other nonferrous metals with "*show me the money*!" The coinage metals consisting of copper, silver, and gold represent the first metals known to man. The first reports for copper purification date back to 3,500 B.C. in the Middle East; bronze alloys were first introduced in *ca.* 3,000 B.C. in India and Greece. However, it is likely that earliest use for copper may have been much earlier for weaponry applications. As a testament to the durability of the coinage metals, 5,000 years after an Egyptian Pharaoh had copper pipes installed in his bath, those same pipes were discovered, dug up, and were still in sufficient shape to carry water!

The first application for gold currency dates back to around 3,400 B.C. in Egypt; however, gold was probably employed for decorative applications much earlier – before 9,000 B.C. Although gold comprises an insignificant 0.004 ppm of the Earth's crust, its early widespread use was a consequence of its availability as the uncomplexed element. Hence, simple techniques such as panning along riverbeds were necessary to isolate the gold, requiring no previously developed knowledge of refining. Discovering any of the three coinage metals was as simple as noticing colors in rocks!

Needless to say, the surface deposits of the coinage metals have long been expired, and the metals must now be isolated in small quantities from sulfide-based ores. As early as 3,000 B.C., a process known as *cupellation* was used to isolate the precious metals from their ores – a method still in use today. In this technique, lead was added to the ore and heated to a temperature of *ca.* 800°C in air. Any gold or silver in the sample was dissolved in the liquid metal, separating it from the undissolved matter. The insoluble material primarily containing iron, copper, and zinc compounds was discarded, and the remaining precious metals were brought to a higher temperature by blowing the fire with bellows. This raised the temperature to the point where lead oxide formed rapidly. This was usually performed in a hearth comprising clay or crushed bones; the PbO would be absorbed into the hearth, while the precious metal deposited at the surface.

Another early process known as *amalgamation* was used by the Romans in the Middle Ages. This simple procedure consisted of combining a precious metal ore in mercury; the gold and silver content dissolved, forming liquid alloys. The metals were then obtained through simple distillation of mercury. However, the method of choice for gold production is cyanidation that usually follows a froth flotation process. The crushed ore is treated with an aqueous NaCN solution, along with enough CaO to neutralize any acid present in the rock that would generate highly toxic HCN. This results in the formation of a cyanoaurate complex (Eq. 13); any silver that is present also forms the analogous cyanoargentate complex. Finely divided zinc metal is then added to reduce the metal ions (Eq. 14). The addition of base regenerates cyanide through formation of zinc hydroxide, much more stable than $Zn(CN)_4$ (Eq. 15).

(13) $4Au + O_2 + 2H_2O + 8CN^- \rightarrow 4Au(CN)_2 + 4OH^-$

(14) $2Au(CN)_2 + Zn \rightarrow Zn(CN)_4{}^{2-} + 2Au$

(15) $Zn(CN)_4{}^{2-} + 4OH^- \rightarrow Zn(OH)_4{}^{2-} + 4CN^-$

The electronic configuration of the coinage metals is $n\mathrm{d}^{10}(n+1)\mathrm{s}^1$. Hence, one may suspect that Cu, Ag, and Au would share similar properties to the isoelectronic alkali metals. However, it should be noted that a filled d shell is far less effective at shielding an outer s electron from nuclear attraction than the less diffuse p shell. As a result, the first ionization energy of the coinage metals is much higher than the alkali metals, and their bonding is significantly more covalent in nature. This explains their relatively higher melting points, hardness, density, and inertness relative to the alkali metals.

The unreactivity of the coinage metals increases dramatically from Cu to Au; whereas copper and silver readily react with sulfur and halogens, gold is completely unreactive to all reagents except very strong oxidizing acids such as *aqua regia* (3:1 HCl/HNO_3). Acid rain produced from gaseous sulfur compounds (SO_x) react with copper surfaces, eventually resulting in the formation of a basic copper sulfate film, as evidenced by a green-blue-patina color. Another useful property of copper is its antifungal behavior; fine granules of ceramic-coated copper oxide are now placed within specialized asphalt shingles to prevent discoloration by algae. The activity can last as long as the shingles, 25–30 years, until all of the Cu^{2+} ions are leached from the porous ceramic granule.

Although all metals possess metallic luster, the only metals that exhibit colors in their bulk state are copper and gold. The familiar reddish and golden colors of these elements arise from the filled d shell near the top of the conduction band of the solid. By definition, the highest-occupied energy level at 0 K is referred to as the *Fermi level* (Figure 3.21). For copper, the gap from the top of the 3d-band to the Fermi surface is *ca.* 544 nm. Hence, energy in the green/blue region of the spectrum may be absorbed, resulting in an observed red/orange color. For gold, the gap is *ca.* 400 nm, corresponding to absorption in the blue region of the spectrum and an observed golden color. By contrast, the analogous energy gap for silver is 311 nm, resulting in UV absorption and an observed white/silver color (*i.e.*, equal reflection of all visible wavelengths).

In addition to high inertness, the coinage metals exhibit other desirable properties that are of tremendous use for materials applications. Silver and copper have the highest electrical and thermal conductivities of all metals in the Periodic Table. Gold is the most electronegative metal, behaving as a halogen through formation of Au^- ions in the presence of strongly electropositive metals such as cesium. These properties may all be rationalized by the band structures of the metals. As a consequence of the filled d shells of the coinage metals, there is facile thermal promotion of valence electrons into the delocalized s/p conduction band. In contrast, other metals such as iron do not have a filled d shell (Figure 3.21), and electrons remain within the partially filled d-band. This shell has more localized character than the s/p band, which results in a lower electrical conductivity of the bulk solid relative to Cu, Ag, or Au. In general, the population of the electronic energy levels (*i.e.*, *den-*

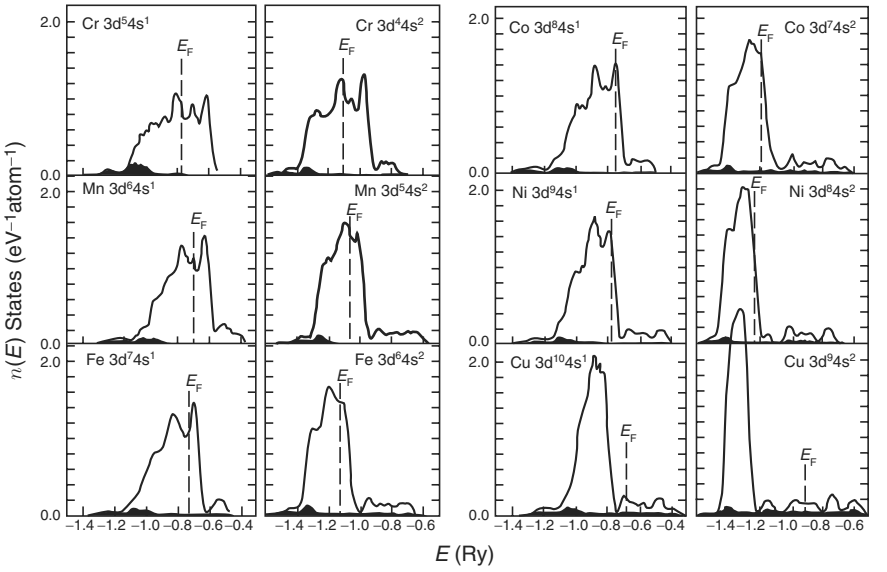


Figure 3.21. Calculated density of states (DOS) for the fcc phases of various transition metals. Reprinted from Snow, E. C.; Waber, J. T. *Acta Metall.* **1969**, *17*, 623. Copyright 1969, with permission of Elsevier.

sity of states) immediately surrounding the Fermi level, E_f, is most important, as typical thermal/electrical energy supplied to a solid is only sufficient to interact with a small fraction (*ca.* 0.4%) of electrons. That is, the overwhelming majority of the electrons are separated from the top of the Fermi surface by much more than thermal energy.

If you have purchased jewelry (or watched television commercials!), you will have heard the terms "carat," "karat," and "fineness." Whereas "carat" refers to the weight of precious stones (1 carat = 200 mg), the term "karat" is used to describe the purity of metals. For example, 24 karats is the pure, unalloyed metal that is almost never used for applications due to its softness. The term "fineness" refers to the weight portion of the precious metal in the alloy; 24 karats represents 100% purity, and a fineness of 1,000 whereas 18 karats represents a purity of 75% and a fineness of 750. A number of alloying agents may be added to gold such as Ag, Cu, Zn, Ni, Pt, and Pd. As required for solid solutions, each of these dopants is of a similar size, and has an fcc crystal lattice that matches that of gold. In addition to improving the strength, these dopants also impart colors to gold (Table 3.4).

Interestingly, "white gold" was developed in the 1920s as a substitute for platinum jewelry. In order to retain a grayish-white color, these alloys are only available up to a purity of 21 karats. Nickel and palladium are the most common dopants used for white golds, although copper may also be added to improve the strength and decrease the price. Due to cases of skin irritation through contact with nickel, European countries have already phased out the nickel whites from jewelry – not yet a policy in the US. It should also be noted that the bright white color is actually an artifact of the

Table 3.4. Commonly Used Alloys for Colored Golds

Composition (karat)	%Au	%Ag	%Cu	%Zn	%Ni	%Ti	%Pd	%Fe	%Si	%Co
Yellow gold										
23	99.0						0.9			
18	75.0	13.0	12.0							
18	75.0	15.0	10.0							
14	58.3	4.0	31.2	6.4	0.1			0.05	0.01	
14	58.3	24.8	26.8	0.14						
10	41.7	11.7	40.8	5.83					0.03	
10	41.7	5.5	43.8	9.0						
White gold										
18	75.0		2.23	5.47	17.8					
18	75.0	15.0					10.0			
14	58.3			28.3	4.8	8.6				
14	58.3	32.2					9.5			
10	41.7		29.2	12.1	15.1					2.0
10	41.7	47.4		0.9			10.0			
Rose gold										
18	75.0	5.0	20.0							
14	58.3	2.1	39.6							
10	41.7	2.8	55.5							

rhodium plating that is usually applied to white golds; however, this film will wear off over time diminishing the color and requiring reapplication of the coating.

As a more stringent application of the Hume-Rothery sizes of alloying metals, if the difference in radii is less than 8%, the metals will be soluble throughout the full range of compositions. This is the case for nickel and copper, whose radii are 1.49 and 1.45 Å, respectively. Hence, there are over 20 different alloys that are used in industry based on the mutual solubility of copper and nickel in all proportions. Monel (68% Ni, 32% Cu) is used to handle corrosive materials such as F_2 or HF. US currency coinage such as quarters and nickels are alloys of Cu/Ni, containing 91.7% Cu and 75% Cu, respectively. By comparison, Canadian quarters and nickels are predominantly steel, with only *ca.* 3.5% Cu and 2% Ni. Pennies are predominantly nickel, with a thin layer of copper deposited through electroplating. High-strength Cu/Ni alloys are produced from the addition of 1.5–2.5 wt% Al, which causes precipitate hardening through formation of Ni_3Al crystallites.

The atomic radii of Cu, Sn(1.45 Å), and Zn(1.42 Å) are also nearly identical, allowing for a full gamut of Cu/Sn and Cu/Zn alloy concentrations to be produced, known as *bronze* and *brass*, respectively. Although the use of bronze dates back to at least 3,000 B.C., there are also early examples of brass artifacts that date back to *ca.* 2,200 B.C. in India. Most likely, the discovery of bronze resulted from the inquisitive mixing of available metals at the time, only to discover that Au/Sn alloys possessed a greater strength than iron; steels were not developed until thousands of years later. Since zinc metal was not available until the mid-18th century, and tin was readily obtained, the widespread production of bronzes occurred at the expense of brasses. In the absence of pure zinc, early formulations of Cu/Zn alloys were most likely

made through heating a mixture containing zinc oxide, copper metal, and a reducing agent such as charcoal in a closed crucible. A temperature in excess of 950°C was necessary to reduce ZnO, and even a trace amount of oxygen would preclude the formation of the alloy.

The strength of a bronze increases with the tin content; however, its toughness and malleability decreases. The maximum strength of bronze occurs at *ca*. 30% Sn, but at this concentration the alloy is much too brittle for most applications due to the formation of Cu_3Sn particles. Recall that this phenomenon also occurred for the formation of Fe_3C in iron–carbon alloys – also involving a transition metal and Group 14 dopant. If more than 15% Sn is used, the alloy is called "bell metal," due to its resonating sound when tolled.

Although the radii of Cu and Zn satisfy the Hume-Rothery constraints for solid solutions, these metals do not share the same crystal lattice. Whereas the coinage metals are fcc, zinc crystallizes in an hcp array. Hence, as we introduce more Zn into the Cu lattice, there will be a shift in the overall structure. We may think of this change as occurring as a result of the change in electron concentrations of the solid. For instance, each Cu atom contributes one 4s electron to the valence shell of the extended lattice; by contrast, each Zn atom contributes two 4s electrons. For small concentrations of Zn, the fcc α-brass structure is formed. However, as the Zn concentration reaches 50%, the bcc β-brass (CuZn) phase predominates. The electron concentration, n, for the β-brass structure is 1.5 (*i.e.*, $1e^-$(for Cu) $+2e^-$(for Zn)/2 atoms). Due to the 1:1 combination of Cu and Zn, and the overpowering 2:1 electronic effect of Zn/Cu, the bcc structure becomes more stable. Further increase in the Zn concentration results in the complex γ-brass, Cu_5Zn_8, with $n = 1.615$ (*i.e.*, $[5 \times 1e^-(\text{for Cu}) + 8 \times 2e^-(\text{for Zn})]/13$ atoms). Additional zinc may continue to dissolve in this phase until a concentration of *ca*. 75% Zn is reached, which results in the final HCP phase referred to as ε-brass, $CuZn_3 (n = 1.75)$. Beyond this concentration, additional zinc results in the HCP η-brass phase which is no longer considered an alloy, but pure Zn with $n = 2$. Only the α and β phases are useful alloys; the others are too hard and/or brittle. The different alloy structures are often called *electron compounds* since they are governed by the ratio of # electrons/# atoms.

In general, the constituent atoms of solid solutions will be randomly positioned at any site in the lattice. However, as the temperature is lowered, each lattice position may no longer be equivalent, and ordered arrays known as *superlattices* may be formed. Examples of superlattice behavior are found for Au–Cu alloys used in jewelry, gold fillings, and other applications (Figure 3.22). For an 18-karat Au/Cu dental alloy, a superlattice will be present at temperatures below 350°C; at higher temperatures, a random substitutional alloy is formed. The AuCu I superlattice (Figure 3.22a) consists of alternate planes of copper and gold atoms, resulting in a tetragonal unit cell that has been elongated along both a and b axes. This is analogous to the hardening mechanism we saw for the austenite to martensite transformation, which also involved a tetragonal unit cell. In fact, the hardening of gold alloys is thought to arise from the superlattice ordering and precipitation hardening mechanisms. A more complex superlattice is also observed in Au–Cu alloys (Figure 3.22b) consisting of a periodic array of multiple unit cells, with Cu and Au atoms exchanging positions bet-

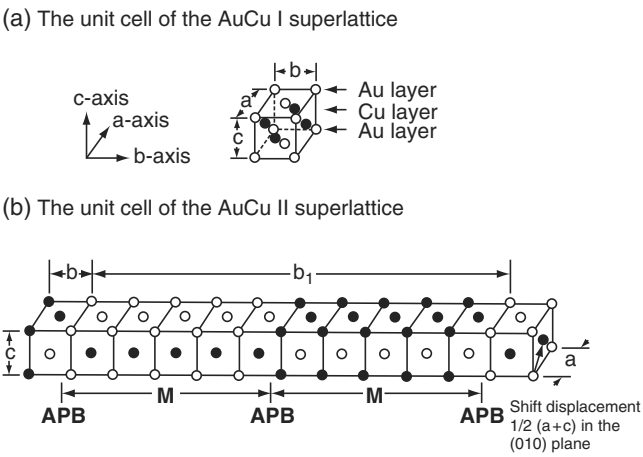


Figure 3.22. Unit cell representations of two varieties of AuCu superlattices. For the AuCu II superlattice, M refers to the length of repeat unit, and APB indicates the antiphase boundaries between adjacent periodic arrays. Republished with the permission of the International and American Associations for Dental Research, from "Determination of the AuCu Superlattice Formation Region in Gold–Copper–Silver Ternary System", Uzuka, T.; Kanzawa, Y.; Yasuda, K. *J. Dent. Res.* **1981**, *60*, 883; permission conveyed through Copyright Clearance Center, Inc.

ween corners and faces. In this case, hardening is thought to occur from the existence of relatively high-energy antiphase boundaries (APBs) between adjacent arrays.

As we will see more in Chapter 7, the resolution of electron microscopes is now suitable for the easy visualization of small atomic cluster arrays. Figure 3.22 illustrates a well-ordered array of Fe–Pd alloy nanoparticles. Interestingly, even though Fe (bcc) and Pd (fcc) do not share the same crystal structure, each nanoparticle crystallite comprises only one lattice, indicating that the Fe and Pd metals form a solid solution. It is actually common for the bcc lattice of iron to change to fcc when alloyed with metals such as Pt, Pd, Cu, or Ni (Figure 3.23).

Aluminum alloys

It is hard to imagine a world without aluminum-based materials. From the foil that we wrap leftovers with, to the cans that house beverages and deodorant aerosols, our world is inundated with applications for aluminum. The widespread use of this metal is a direct result of its availability – 8.3 wt% in the earth's crust, making it the most naturally abundant metal. The malleability of Al is second only to gold, and it possesses other desirable characteristics such as nonsparking, high thermal/electrical conductivity, corrosion resistance, and high ductility.

Unlike most metals, the strength of aluminum improves upon alloying, which extends its range of applications (Table 3.5). One of the most popular alloys is the binary Mg/Al type, which is sometimes referred to as magnelium. Only a maximum of 5 wt% of Mg may be dissolved in Al to provide solid-state strengthening. However, only 1.5 wt% of magnesium may be dissolved at room temperature, implying that

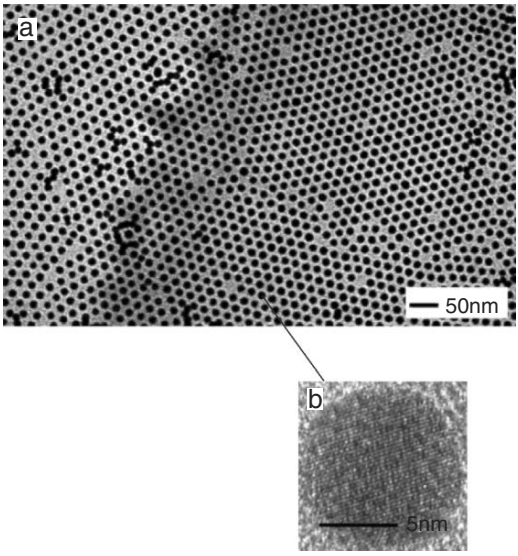


Figure 3.23. Transmission electron microscope image showing a 2D hexagonal superlattice of Fe–Pd alloy nanoparticles. Image (b) confirms that each nanoparticle consists of only one type of crystal lattice. Reproduced with permission from *Chem. Mater.* **2004**, *16*, 5149. Copyright 2004 American Chemical Society.

Table 3.5. Types and Properties of Aluminum Alloys

Metal additive	Resultant properties/applications
Cu (*ca.* 5%)	Heat treatable, high strength to weight ratio, limited corrosion resistance and weldability/autowheels and suspension components, aircraft fuselage, power lines
Mn (*ca.* 1.2%)	Moderate strength without heat treating, high workability/beverage cans, cooking utensils, heat exchangers, storage tanks, furniture, highway signs, roofing, side panels
Si (*ca.* $\leq$ 12%)	Low thermal expansion and m.p., high wear resistance/forged engine pistons, welding rod, brazing alloys, architectural products
Mg (*ca.* 0.3–5%)	Good weldability and strength, good corrosion resistance/ornamental trim, cans, household appliances, boats and ships, bridge railings, race cars
Zn (*ca.* 3–8%)	Heat treatable, moderate–very high strength/airframe structures, high-strength forgings

supersaturation will often occur, with precipitation of Mg species. This should be reminiscent of the Fe–C system, with precipitation hardening through dispersion of metal carbides. Frequently, iron and silicon will be present as processing impurities (or deliberately added), which will strengthen the alloy due to the formation of Mg_2Si and Fe_3Si precipitates upon cooling. In the presence of Mn, the hardening effect is even more pronounced, due to the formation of $FeMnAl_6$ crystallites.

Among the various Al-alloys at our disposal, many of them may not be heat-treated. In particular, alloys such as pure Al (*i.e.*, containing trace dopants), Al–Mn, Al–Si, and Al–Mg alloys deleteriously form precipitates along grain boundaries. However, Cu–Al and Al–Zn–Mg alloys are greatly strengthened by heat treatment, through formation of $CuAl_2$ and $MgZn_2$ precipitates, respectively. Even lithium may be added as a hardening agent – forming Al_3Li precipitates. As with all age-hardening techniques, the size and dispersion of the crystallites must be carefully controlled through the heating/cooling regime.

Whereas the solubility of Cu in aluminum metal is *ca.* 5 wt% at temperatures in excess of 500°C, the solubility drops to *ca.* 0.1 wt% at room temperature. Hence, a metastable alloy is present when the high temperature alloy is rapidly quenched. Subsequent annealing will result in further strengthening similar to what we discussed for martensite. The strengthening effect is thought to occur due to the formation of Cu-rich discs (approx. diameter of 100 atoms, and thickness of *ca.* 4 atoms) that align themselves preferentially with selected planes of the host Al lattice, causing coherency strains within the solid-state structure.

Refractory metals

By definition, refractory metals exhibit low thermal and electrical conductivities and have equally low thermal expansion properties (Table 3.6). As a relative benchmark, common metals such as iron and copper have coefficients of linear thermal expansion on the order of 12.1 and 17.7 $\mu m\ m^{-1}\ K^{-1}$, respectively. Also for comparative purposes, the electrical/thermal conductivities for Fe and Cu are $9.71\ \mu\Omega\ cm^{-1}$/$78.2\ W\ m^{-1}\ K^{-1}$ and $1.67\ \mu\Omega\ cm^{-1}$/$397\ W\ m^{-1}\ K^{-1}$, respectively.

However, it is the extremely high melting points (>1,650°C) of the refractory metals that separate this class from the others. As a result, these metals may not be processed through cold or hot working; powder metallurgy must be used to form the metals into desired shapes. Although we also generally associate the refractories with high hardness, it is worthwhile to point out that these metals are all soft and ductile in their pure states. However, the metals are rarely obtained in their pure forms, since they spontaneously react with C, B, O, N, and other nonmetals to form stable interstitial compounds. The incorporation of small main group elements, with large electron-rich refractory metals, results in solute hardening of the crystal lattice through localized covalent bonding within the material. A recent example is the formation of the refractory ceramic OsB_2, which is possible since boron is significantly smaller than Os (0.87 Å *vs.* 1.85 Å, respectively).[4]

The incompressibility (bulk modulus) of a material is directly related to its valence electron density, in units of electrons $Å^{-3}$. For example, diamond, the hardest known

Table 3.6. Properties of the Refractory Metals

Metal (Lattice)	Density (kg m^{-3})	Melting point (°C)	Resistivity ($\mu\Omega\,cm^{-1}$)	Thermal conduct. ($W\,m^{-1}\,K^{-1}$)	CLTE[a] ($\mu m\,m^{-1}K^{-1}$)
Ti (HCP)	4,540	1,668	42.0	21.9	8.35
Zr (HCP)	6,506	1,852	42.1	22.6	5.78
Hf (HCP)	13,310	2,233	35.5	22.3	5.90
V (BCC)	6,110	1,915	25.0	30.7	8.40
Nb (BCC)	8,570	2,230	15.2	53.7	7.10
Ta (BCC)	16,654	2,996	13.2	57.5	6.60
Cr (BCC)	7,140	1,900	13.0	93.9	4.90
Mo (BCC)	10,220	2,610	5.70	139	5.43
W (BCC)	19,300	3,407	5.65	174	4.59
Re (HCP)	21,010	3,270	13.5	48.0	6.70
Ir (FCC)	22,650	2,410	5.30	146	6.40
Os (HCP)	22,590	3,054	8.12	87.6	4.57

[a]Coefficient of linear thermal expansion.

substance, has a high valence electron density (0.705 electrons $Å^{-3}$) and an exceptionally high bulk modulus (442 GPa). By comparison, osmium has one of the highest valence electron densities for a pure metal (0.572 electrons $Å^{-3}$), resulting in an accompanying large bulk modulus (*ca.* 400 GPa). However, while the bulk moduli of diamond and osmium are equivalent, the hardness of diamond is more than an order of magnitude larger than Os – a consequence of covalent *vs.* metallic bonding. Upon doping Os with boron, the hardness improves dramatically, while retaining a high valence electron density (0.511 electrons $Å^{-3}$). There are current studies underway that are evaluating the hardness and bulk moduli of mixed solutions such as $Os_{1-x}M_xB_2$, which are likely harder than either OsB_2 or MB_2 compounds alone.

As we saw earlier, the facile reaction of Cr with oxygen is the acting principle behind the anticorrosive property of Cr-containing steels. This analogous reactivity also explains the high corrosion resistance exhibited by the refractory metals, especially those of Group 4 (Ti, Zr, Hf – often termed the "reactive metals"). Since refractories show a high reactivity toward constituent gases of the atmosphere, the metals in their finely divided forms are highly pyrophoric and must be handled within an inert-atmosphere glove box. In general, all metals that form stable oxides or nitrides (*e.g.*, iron, zinc, nickel, *etc.*) are dangerous as finely divided powders. The relatively large surface area of individual crystallites leads to simultaneous oxide formations (exothermic) that release enough heat to spontaneously catch the material on fire.

As one moves across the Groups of refractory metals, a number of trends are noteworthy and greatly affect their materials applications. For instance, moving left to right causes a decrease in atomic sizes due to ineffective shielding of the nuclear charge by d-electrons. As additional d-electrons are added to the valence shell, stronger metal–metal bonds are formed and the metals become increasingly dense/harder, with higher melting points as illustrated in Table 3.6. Also noteworthy

is the equal atomic sizes of 4d and 5d congeners due to the lanthanide contraction effect.

Among the refractory metals, perhaps none are as widely exploited for commercial applications as titanium. From golf clubs to shavers, titanium is now pervasive throughout our modern world. Although some of the proposed applications may be suitably classified as "hype," the broad appeal for titanium alloys is due to its favorable properties such as high strength/weight ratio and superior corrosion resistance. Titanium is also readily available from a number of mineral sources; it is the sixth most abundant metal, behind Al, Fe, Cu, Zn, and Mg.

As you will note from Table 3.6, the density of titanium is significantly less than the other refractories (midway between aluminum and iron); even so, the yield strength ranges up to 1,800 MPa. To put this in perspective, this strength is of the same magnitude as Ni-doped ultrahigh strength stainless steels – at a fraction of the weight. The low density of Ti (and other metals such as Be and Mg) is due to the hexagonal close-packed crystal structure which is much less dense than bcc or fcc arrays (as discussed in Chapter 2).

Shape-memory alloys

As their name implies, shape-memory alloys are able to revert back to their original shape, even if significantly deformed (Figure 3.24). This effect was discovered in 1932 for Au–Cd alloys. However, there were no applications for these materials until the discovery of Ni–Ti alloys (*e.g.*, NiTi, nitinol) in the late 1960s. As significant research has been devoted to the study of these materials, there are now over 15 different binary, ternary, and quaternary alloys that also exhibit this property. Other than the most common Ni–Ti system, other classes include Au–Cu–Zn, Cu–Al–Ni, Cu–Zn–Al, and Fe–Mn–Si alloys.

The shape-memory effect is observed when the temperature of a piece of alloy is cooled to below that required to form the martensite phase: M_s (initial martensite formation) until M_f (martensite formation complete), as seen in Figure 3.25. Upon heating the martensitic material, a reformation of austenite begins to occur at A_s

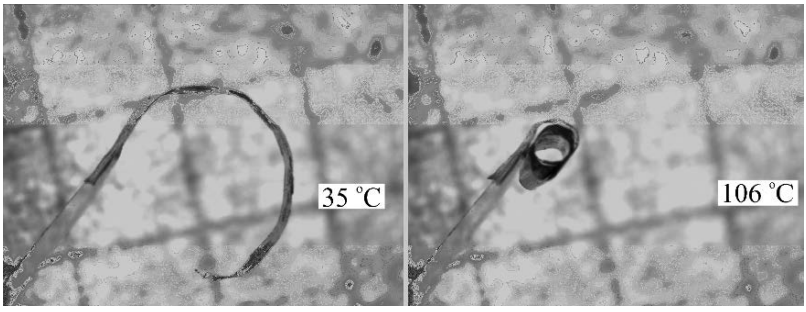


Figure 3.24. Photographs of the shape-memory effect at varying temperatures for a Ni–Ti wire. Reproduced with permission from the real-time video clip made by Rolf Gotthardt (rolf.gotthardt@epfl.ch) – found online at http://www.msm.cam.ac.uk/phasetrans/2002/memory.gif

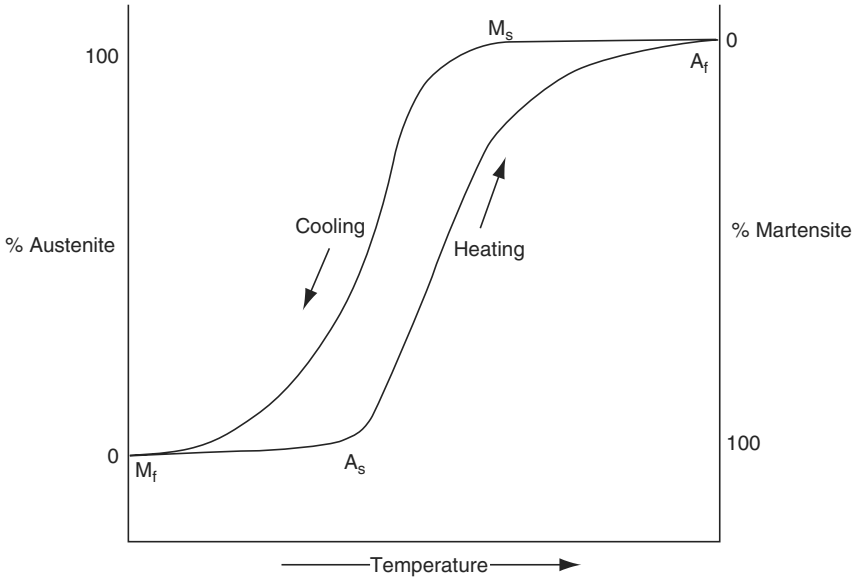


Figure 3.25. Hysteresis loop associated with the phase transitions of shape-memory alloys.

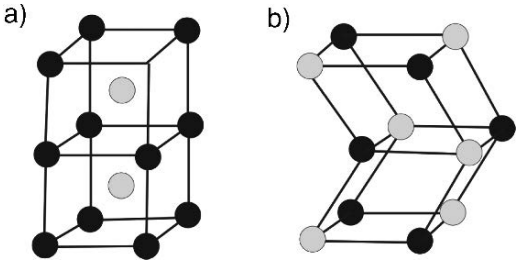


Figure 3.26. Unit cell representations of (a) the CsCl structure of austenitic TiNi and (b) the monoclinic structure of martensitic TiNi.

until a final temperature, A_f, is reached (Figure 3.25). Since the martensite phase is a highly distorted structure relative to fcc austenite (Figure 3.26), the alloy is highly soft and ductile and may be easily deformed while in its low-temperature phase. You may be thinking that this is the opposite of the Fe–C system that was previously discussed. That is, the martensite phase that was generated through fast quenching austenite resulted in an extremely hard material – much stronger than the native austenite phase. However, in that system, the martensite is associated with interstitially dissolved carbon that adds strength through solute hardening, but brittleness through the introduction of additional grain boundaries. It should be noted that there are 24 possible ways of accomplishing the austenite–martensite transformation.

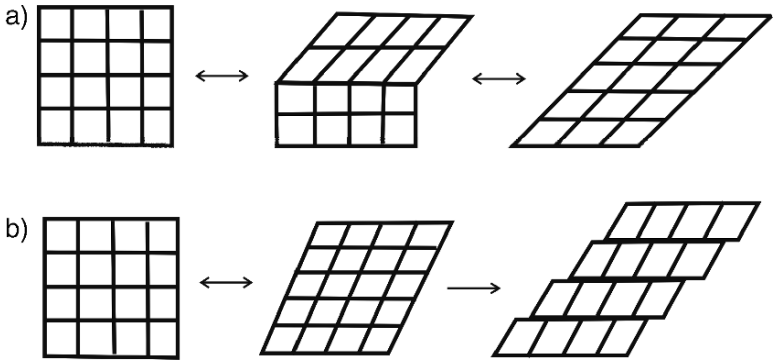


Figure 3.27. Deformation pathways for (a) shape-memory alloys, showing the reversible movement of twin boundaries. Shown in (b) is the irreversible slip deformation of other alloys, such as carbon steels.

That is, austenite has six equivalent facial planes, and each of these may shear along two perpendicular axes.

When austenite is cooled in the absence of applied stress, the material transforms into a twinned form of martensite (Figure 3.27a). Since both austenite and twinned martensite have the same macroscopic shape/size, reheating the material will not result in any observable shape change. However, if the material is plastically deformed through bending, *etc.* at low temperature, it will become detwinned and the new shape will prevail. The Ni–Ti alloys are preferred since they have a greater range of deformation (up to 8%), relative to other Cu-based alloys (4–5%). When the material is reheated, the deformed martensite structure will be converted to the original austenite phase with a different macroscopic structure. For comparative purposes, Figure 3.27b illustrates the irreversible slip deformation that other types of metals such as steel undergo as a result of the same stresses. Since these latter materials do not have suitable twin planes, shape-memory transitions are not possible resulting in a permanent shape alteration of the metal.

It is also possible to apply a stress to the material in its high-temperature austenitic phase. However, since the temperature is above A_f, the original shape will be reformed immediately after the load is removed. Such an immediate shape change is referred to as *pseudoelasticity* (or *superelasticity*), and is the active principle underlying cellular phone antennae that may be greatly distorted only to immediately return to their original shapes.

In addition to temperature- or stress-induced transitions, there are now a number of *ferromagnetic* shape-memory alloys that alter their shapes in response to a magnetic field. Examples of these systems include Fe–Pd, Fe–Pt, Co–Ni–Al, Co–Ni–Ga, and Ni–Mn–Ga. These materials are of great interest since the magnetic response time is faster and more reliable than temperature-based transitions. Whereas traditional alloys alter their structures as a result of the martensite–austenite transition, magnetic analogues exhibit a change in structure while remaining in the martensite phase.

The change in shape is a result of the detwinning of preferred planes based on their orientations with the applied magnetic field.

As one would expect, there are a number of applications that currently use shape-memory alloy materials; many more are projected for the future. The earliest application was for greenhouse window openers, with the metal serving as an actuator to provide temperature-sensitive ventilation. Some commercial faucets/showerheads are already equipped with this material that shuts off the water if a certain temperature is reached, which effectively prevents scalding. An intriguing future application will be for automobile frames; as we will see later, some plastics may also be designed with shape memory. Someday soon, your car may reshape itself in front of your eyes within minutes after an accident!

Since NiTi alloys have been shown to be biocompatible, perhaps the greatest use for these alloys has been for medical applications. In particular, for minimally invasive surgery where a metal wire is inserted through tiny incisions, and then reshaped into the original form (*e.g.*, tweezers, specialized probes, *etc.*) while *inside* the patient due to an increase in temperature! Likewise, metal probes may be bent into specific shapes for open surgeries and then returned to their original positions afterward through the heating/sterilization process. Other widespread applications include nitinol filters that are designed to trap blood clots in arteries, as well as suture anchors that are inserted directly into bone to facilitate the attachment of soft tissue such as ligaments and tendons.

Although we have not yet introduced nanotechnology, it is worthwhile to point out that shape-memory behavior has also been discovered in single-crystalline Cu nanowires.[5] This is interesting since this effect does not occur in bulk copper metal. Due to the extremely high surface/volume ratio of nanowires relative to bulk structures, these materials show reversible strains of *ca.* 50% – an order of magnitude larger than bulk shape-memory alloys. Further, the martensite–austenite transformation temperature changes dramatically with minute changes in nanowire diameter. For instance, increasing the diameter from 1.76 to 3.39 nm causes a *ca.* 800 K increase in the transition temperature! This will allow for the design of nanoscale components of varying sizes that will be functional over an extremely wide temperature range – not possible with bulk alloy systems. More importantly, the response time for these materials is also orders of magnitude faster than bulk alloys due to the extremely small dimensions of the nanoscale. We will see more of the "nanoworld" in Chapter 6 – this intriguing example was inserted to wet your appetite a bit...

3.3. METAL SURFACE TREATMENTS FOR CORROSION RESISTANCE

The corrosive deterioration of metal surfaces incurs a great cost to the worldwide economy. Accordingly, there have been many research efforts devoted to understanding the surface chemistry behind these reactions. As we have already seen, this has led to the development of a number of useful alloys that are sufficiently resistant to corrosion – through spontaneous formation of protective oxide layers. However, for other less resistant metals such as carbon steels, a protective layer must be postdeposited onto a metal surface in an effort to prevent corrosion. In this section,

we will discuss three strategies that may be used to protect a metal surface from its environment – inert-layer passivation (either native (*e.g.*, Cr_2O_3 on Cr) or purposefully deposited), sacrificial metal coatings (*e.g.*, galvanized steel), and organic-based coatings. It should be noted that a pretreatment process is often required for metal surfaces to allow the coatings to be strongly adsorbed to the surface. These steps effectively remove organic components such as oils, as well as inorganic species such as welding flux. The most common methods are either mechanical descaling (*e.g.*, abrasive blast techniques) or chemical pickling (*i.e.*, acid treatment).

The primary corrosive agents are oxidizing agents (*e.g.*, moist air, HNO_3, H_2SO_4), or halogenated species (*e.g.*, Cl_2, HCl, HF, CFCs). These agents degrade the metal by forming oxides, hydroxides, or halides that introduce embrittling grain boundaries on the surface. Other particularly detrimental gases for corrosion of metal surfaces are CO and H_2S. Both of these gases exhibit dissociative adsorption on metal surfaces, resulting in carbide or sulfide formation and concomitant embrittlement of the metal. In the presence of H_2S (or H_2O) at elevated temperatures, the hydrogen atoms may also interact with surface metal sites and cause surface cracking. Hydrogen-induced cracking is especially detrimental for iron surfaces.

For carbonaceous gases such as CO and CH_4 at relatively high temperatures (*ca.* $>800°C$), carburization of steel surfaces takes place in the form of brittle interstitial carbides that may cause surface cracking. Cementite may also form on the surface of steel; since its melting point is lower than the underlying metal, it may cause melting of the steel surface that is subsequently eroded by the gas stream.

One simple method used to deter the onset of corrosion is *phosphating*. This process is often used to chemically passivate a metal surface with a crystalline coating of zinc phosphate. The phosphating bath is an aqueous solution of dilute phosphoric acid, containing anionic and cationic elements that are capable of reacting with the metallic surface to yield a crystalline film on this surface. Other components of phosphating baths, known as accelerators, influence the kinetics of the reaction process and permit the control of redox reactions at the interface. The most common surface species present after phosphating are vivianite $[Fe_3(PO_4)_2 \cdot 4H_2O]$, hopeite $[Zn_3(PO_4)_2 \cdot 4H_2O]$, and phosphophyllite $[Zn_2Fe(PO_4)_2 \cdot 4H_2O]$. For more complex substrates such as steel coated with Zn/Fe, Zn/Ni, Zn/Al, and Zn/Cr alloys, tricationic phosphatings have been developed. For these systems, the surface is coated with crystalline pseudophosphophyllite $[(Zn, M, Ni)_3(PO_4)_2 \cdot 4H_2O]$, where M = Fe, Al, Cr, *etc.* For aluminum containing alloys, or hot-dipped galvanized steel (containing Al_2O_3 on the surface), it is necessary to use fluoride-based additives to cause surface crystallization.

Another useful passivation technique is *anodizing* or *anodic oxidation*. In this method the metallic surface acts as an anode, being oxidized during an electrochemical event. The most common metals/alloys are those containing aluminum, magnesium, and zinc. However, it is also possible to anodize other metals such as copper, steel, and cadmium for protective and decorative applications. The anodizing electrolytic solution consists of strong acids, generally combinations of chromic, sulfuric, oxalic, or boric acids. The anodic layer obtained in sulfuric acid baths consists of a relatively thin barrier layer, overlaid with a porous array. The density and

morphology of the pores may be varied through manipulation of the electrical current or nature of the electrolytic solution. Although this presents a sufficient limitation related to corrosion resistance, dyes may be added to the *ca.* 10–20 nm diameter pores to yield a colored film. However, for protective applications, hydration sealing is often required which consists of steam treatment in the presence of chromate or Ni/Co salts. In the case of aluminum, this post-treatment results in boehmite, AlO(OH), that sufficiently seals the pores.

A variety of other nonmetallic coatings may be used to impart corrosion resistance to the underlying metal surface. Common inorganic-based coatings include vitreous enamels, ceramics, glass, cements, carbides, and nitrides. By contrast, organic-based protectants are paint coatings, plastic coatings, adhesive tapes, and sheet linings. Whereas the inorganic layers are often used to coat internal surfaces of piping and reactors, organic films are most often used for external surface protection. Refractory coatings such as carbides (*e.g.*, TiC, B_4C, WC, and WCO), nitrides (*e.g.*, AlN and BN), oxides (*e.g.*, Al_2O_3, BeO, Cr_2O_3, ThO_2, and ZrO_2), silicides (*e.g.*, $NbSi_2$, WSi_2, and $MoSi_2$), and borides (*e.g.*, ZrB_2 and TiB_2) impart both corrosion/abrasive wear and temperature resistance to the underlying substrate.

More recently, an even greater corrosion resistance has been generated through use of composite coatings, comprising the above refractory compounds in association with a metal powder (*e.g.*, Cr + ZrB_2, Cr + HfO_2, Al + SiC, Ti + TiB_2, and PtRh+ZrB_2). For these coatings, the ceramic and metal powders are suspended in an aqueous solution with the assistance of surfactants, and sprayed onto the metal surface. This film is then allowed to dry at a temperature of *ca.* 70–90°C, and annealed with an energetic laser source (*e.g.*, Nd:YAG), resulting in formation of an interwoven matrix of metal and ceramic species.

Without question, the easiest and most inexpensive method to protect metal surfaces from corrosion is through simple painting. Paints comprise finely divided solid inorganic or organic pigments (Figure 3.28) that are suspended in association with binder molecules within a volatile solvent. In paints/varnishes, the nature of the binder defines the type of paint system such as oil or water based, epoxy, *etc.* By contrast, varnishes do not contain light-scattering pigments, resulting in a transparent coating. The volatile medium comprises solvents that are used to solubilize the binder and dilutants whose purpose is to place the paint at a suitable viscosity for its application. Solvents and dilutants normally comprise organic compounds such as hydrocarbons, alcohols, ketones, ethers, and esters. Additives such as antifungal agents, driers, *etc.* are often used to broaden the application of the coating.

Contrary to popular belief, a paint coating will not be completely impervious to environmental agents surrounding the material, though this may be limited through use of structurally complex pigments such as graphite, mica, aluminum oxide, *etc.* Due to the incomplete blockage of corrosive agents, sacrificially active paint pigments are often chosen specifically based on the corrosive agents that they will be in contact with. For instance, impeding the corrosive ability of oxidizing agents can be achieved through use of Zn-rich pigments; likewise, pigments consisting of strong oxidizing agents such as chromates or manganates will offset reductive corrosion pathways.

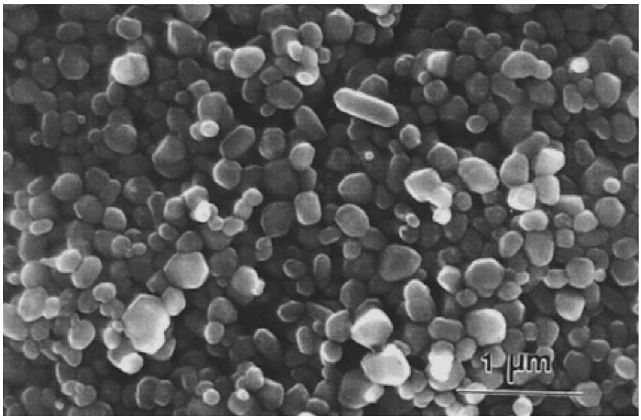


Figure 3.28. Scanning electron micrograph of TiO_2 particles, used as white pigments in paint formulations.

Passivation and painting are only effective when the coating completely encapsulates the metal surface. Sacrificial metal coatings are also problematic, as the protective metal will eventually degrade exposing the underlying substrate. Most recently, an alternative strategy has been developed, referred to as *photoelectrochemical* protection by semiconductor coatings. When photoactive semiconductor particles such as TiO_2 are deposited onto steel or copper substrates, the underlying metal exhibits significant corrosion resistance upon exposure to UV irradiation. Although the details of the photoactivity of semiconductor films will be discussed in Chapter 4, it is important to note here that the exposure to light of the appropriate energy causes electrons in the semiconductor to be excited. These electrons are transferred to the underlying metal, which creates a potential that is more negative than its corrosion potential.

Although photoelectrochemical effects will usually only occur in the presence of UV light, there are recent reports of using complex multilayered or composite films to yield anticorrosion properties even in the dark. Although the mechanisms are not presently well known, the general principle behind these films is the use of films that contain an intimate connection between semiconductor and electron-storage particles. After the UV light has been turned off, the stored electrons are injected to the metal so it is still protected from corrosion (Figure 3.29).

3.4. MAGNETISM IN METALS AND ALLOYS

Even as young children, we become familiar with the intriguing magnetic properties of iron, based on the strong attraction of the bulk material toward a permanent magnet. For instance, one of the most popular toys, MagnaDoodle^{TM}, operates solely through the attraction of magnetic iron oxide (magnetite, Fe_3O_4) particles toward a handheld stylus magnet. Regarding elemental Fe, not all allotropes are magnetic.

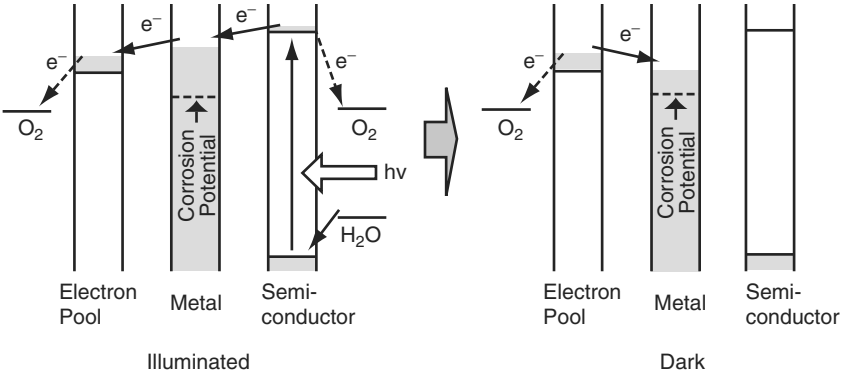


Figure 3.29. Schematic of an anticorrosion system containing both semiconductor and "electron pool" storage components. Reproduced with permission from *Chem. Mater.* **2001**, *13*(9), 2838. Copyright 2001 American Chemical Society.

That is, among the pure iron forms, only ferrite (α, bcc) is magnetic. This is intriguing, as the δ-Fe form also exhibits a body-centered cubic crystal structure. This must indicate that in addition to the simple 3D arrangement of lattice iron atoms, their individual magnetic dipoles must also be suitably aligned in order to yield a particular magnetic behavior.

In contrast to *diamagnetism*, where all valence electrons of each atom are spin paired, *paramagnetism* is found in solids where the constituent atoms contain an unpaired valence electron(s). In a simple paramagnetic substance, the unpaired electrons' spins are randomly oriented within the solid. Upon exposure to an external magnetic field, the spins become collectively oriented in a parallel direction to the applied field. However, the dipoles re-randomize when the field is removed.

The magnetic responses of diamagnetic and simple paramagnetic substances are small enough that a special instrument called a magnetic susceptibility balance is required to measure these effects. This technique measures the amount of repulsive (for diamagnetic) or attractive (for paramagnetic) force between the sample and a permanent magnet within the instrument. For paramagnetic substances, the magnitude of the attractive response is proportional to the number of unpaired electrons present in the sample. Hence, this technique provides an efficient means to determine the ground-state electron configuration of transition metal complexes.

As temperature increases, the magnetic susceptibility, χ, of a paramagnetic substance decreases. The increasing thermal motion of atoms comprising the solid disrupts the ordering among neighboring magnetic dipoles. Most often, the *effective magnetic moment*, μ_{eff}, is used to describe the paramagnetic behavior, since this quantity is independent of both the temperature and the magnitude of the external field. Qualitatively, the macroscopic magnetic moment of a solid may be thought of as a vector summation of all the microscopic magnetic dipole moments of each atom.

In a *ferromagnetic* material, the magnetic dipoles generated from unpaired electrons tend to align in the same direction, even in the absence of an external magnetic

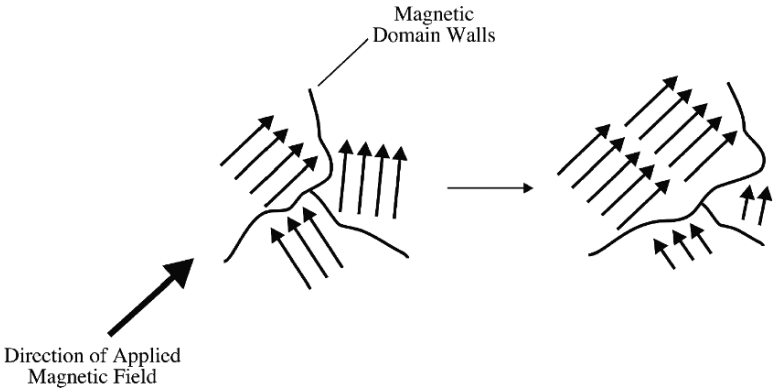


Figure 3.30. Representation of a Bloch wall expansion resulting from an applied magnetic field impinging on a ferromagnetic material.

field. This phenomenon is aptly referred to as ferromagnetic coupling. It should be noted that ferromagnetism is the direct opposite of superconductivity, where all electron spins pair to form a perfectly diamagnetic material. Examples of ferromagnetic behavior may be observed in bulk iron, cobalt, and nickel. The regions containing parallel-aligned magnetic spins are known as ferroelectric or Weiss domains, with a Bloch wall providing an interface between two adjacent domains. When an external magnetic field is applied, those domains that are aligned parallel to the direction of the field are energy favored over those magnetized in opposing directions. The favored domain walls then expand at the expense of the unfavored, resulting in a net magnetization (Figure 3.30).

The ground-state electronic configuration of iron in the solid-state metal lattice is [Ar] $4s^2 3d^6$. The magnetic moment per iron atom in a cm^3 solid is 2.2×10^{-20} emu, or approximately 2 Bohr magnetons ($\mu_B = 1 \times 10^{-20}$ emu). Therefore, there are two unpaired electrons in each iron atom throughout the lattice. By comparison, the μ_B values for cobalt ([Ar] $4s^2 3d^7$) and nickel ([Ar] $4s^2 3d^8$) are 1.72 and 0.61, respectively. Since there are five d orbitals and six d-electrons for iron, two separate d orbitals must house the unpaired electrons. In the bcc array of iron, two orbitals (d_{z^2} and $d_{x^2-y^2}$ – positioned *along* the cartesian axes), are not directed toward neighboring atoms in the lattice.[6] Hence, these orbitals will have nonbonding character, and may therefore accommodate two unpaired electrons. The remaining four electrons within d_{xy}, d_{xz}, and d_{yz} orbitals (having lobes directed *between* the Cartesian axes) participate in metallic bonding between neighboring iron atoms, forming a valence band of paired electrons.

The magnetic response of a ferromagnetic material exposed to an external field is typically represented by an S-shaped field-dependent *magnetization curve* (Figure 3.31). By definition, the applied and induced magnetic fields are given the symbols H and B_0, respectively. When the external magnetic field reaches a maximum value (H_s), the material will form a single domain with a net *saturation*

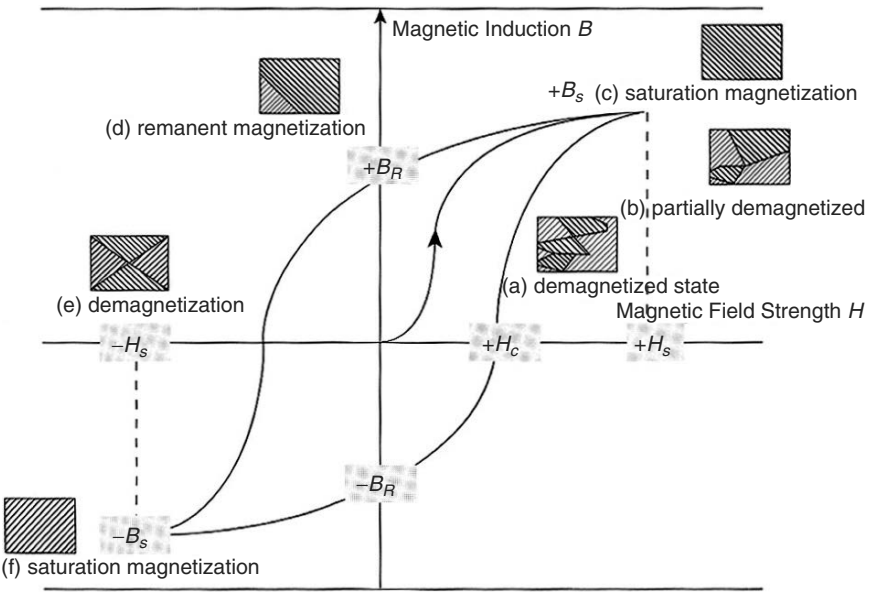


Figure 3.31. A *B–H* magnetization hysteresis curve for a ferromagnetic material. Reproduced with permission from Cardarelli, F. *Materials Handbook*, Springer: New York. Copyright 2000 Springer Science and Business Media.

magnetization (B_s) in a direction parallel to the applied field. Once the external field is removed from the material, the domain walls spring back toward their original positions, and the magnetization decreases to a value referred to as the *remanence* (B_R). This is the operating principle of magnetic storage devices such as audio/data cassette tapes. In order to remove the induced magnetism of the material, an opposite magnetic field is applied, known as the *coercive magnetic field* (H_c). If the magnitude of the opposed field is increased to a maximum value ($-H_s$), a saturation magnetization will again be found in the material, parallel to the new direction of the applied field.

It is noteworthy that the bulk size and shape of a ferromagnetic metal may also change as a result of reversible magnetization. This phenomenon, referred to as *magnetostriction*, is due to the coupling of electron spins between neighboring atoms, which affects the delocalized electrons involved in metallic bonding. This effect is responsible for the familiar hum of transformers and fluorescent lights, from the vibration of the iron components within these materials.

At low temperatures, spontaneous *antiferromagnetic coupling* between neighboring atoms results in an equal number of magnetic dipoles in opposite directions. However, as the temperature is increased, the dipoles are randomized resulting in paramagnetic behavior. Primary examples of antiferromagnetic behavior include transition metal compounds such as MnO, NiO, MnS, $FeCO_3$, MnF_2, as well as certain metal clusters (Figure 3.32b). In contrast to ferromagnetic behavior,

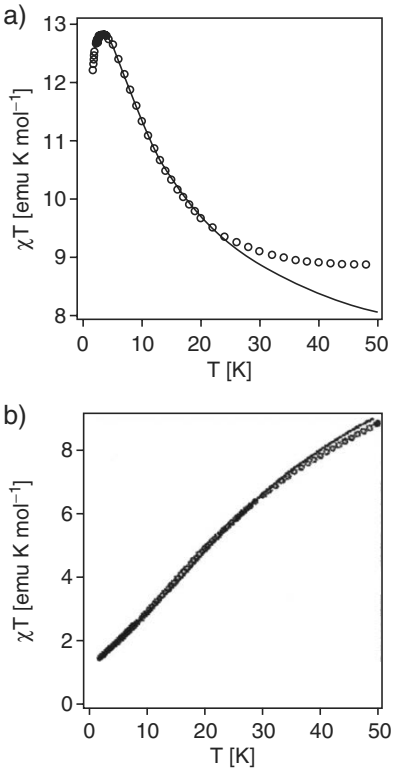


Figure 3.32. Comparison of magnetic susceptibility profiles for (a) $Na_{12}[Co_3W(H_2O)_2(ZnW_9O_{34})_2] \cdot 46H_2O$ and (b) $Na_{12}[Co_3W(D_2O)_2(CoW_9O_{34})_2] \cdot 46D_2O$. Whereas the profile of the Co_3 cluster is indicative of ferromagnetic coupling between Co(II) ions, the trend for the Co_5 cluster is representative of antiferromagnetic coupling between Co(II) centers. Reproduced with permission from *Inorg. Chem.* **2001**, *40*, 1943. Copyright 2001 American Chemical Society.

ferrimagnetic coupling results in magnetic spins in two opposite orientations, with more in one direction than in the other. Commonly, ferrimagnetic materials crystallize in a spinel lattice (*e.g.*, Fe^{2+} and Fe^{3+} ions within ferrites).

Vibrational motion of the molecules can disrupt the domain structure. Hence, the magnetic properties of ferro-, antiferro-, and ferrimagnetic materials are strongest at low temperatures. At sufficiently high temperatures, no domain structure is able to form, resulting in paramagnetic behavior. The cutoff temperature for the onset of paramagnetic behavior is referred to as the *Curie* (T_c) or *Neel* (T_N) temperature for ferro-/ferrimagnetic and antiferromagnetic materials, respectively. Curie temperatures range from 16°C for Gd, to 770°C and 1, 115°C for Fe and Co, respectively. By contrast, Neel temperatures range from −271°C for $MnCl_2 \cdot 4H_2O$ to 680°C for α-Fe_2O_3.

The larger the gap between B_0 and B_R, the more effective the material will be for magnetic storage applications. Magnetically "soft" materials such as pure iron, low

carbon steels, and alloys of ferromagnetic elements (*e.g.*, Fe, Co, Ni), consist of an ordered array of ferromagnetic atoms that easily revert back to their original domain structures following the removal of the external field. In contrast, a large remanence value indicates that the domain walls are not irreversibly transformed back to their original position – known as "hard" magnetic materials. As one would expect, the microstructure of the solid is paramount to the relaxation efficiency of the Bloch walls. That is, the domain walls will be less likely to relax to their original positions if the lattice contains trace amounts of interstitial dopants such as Si, C, O, or N, or dispersed particles from precipitation hardening processes (*i.e.*, "domain-wall pinning"). Such hard magnetic materials are known as *permanent magnets*, retaining their magnetism over prolonged periods of time after the external field is removed. In addition to high B_R, these materials also exhibit a relatively high coercive field, H_c.

Common examples of magnetically hard materials are high carbon steels, precipitation hardened alloys (*e.g.*, Alnico), and sintered or bonded fine-particle alloys (*e.g.*, ferrites, rare earth alloys). The earliest examples of rare earth magnets are $SmCo_5$ and Sm_2Co_{17}, with recent developments focused on the incorporation of Fe rather than other costly transition metals. Whereas iron–rare earth alloys such as R_2Fe_{17} have relatively low operating temperatures, the addition of boron results in the ternary compound $Nd_2Fe_{14}B$ with strong uniaxial magnetocrystalline anisotropy, and a higher operating temperature. The absorption of nitrogen to yield $Sm_2Fe_{17}N_3$ causes a further improvement in the magnetic properties; however, the applications are limited by its complex synthesis. It should be noted that the partial substitution of Co for Fe reduces the surface oxidation. This is important, since the processing of rare earth magnets involves the compaction of finely divided powders, which are more difficult to obtain through ball-milling if a hard oxide layer is present. Interestingly, since the Curie temperature of Co greater than Fe, the T_c of the ternary alloy $Nd_2(Co_xFe_{1-x})B$ increases at a rate of *ca.* 10°C per at.% Co that is substituted.

The origination of the desirable magnetic properties within the complex structures of rare earth alloys is not completely understood. Likely, the interaction of the 4f electrons with neighboring lattice atoms results in a preferential alignment of the rare earth magnetic moments along specific lattice directions. This allows for saturation magnetization to be achieved with only relatively small applied fields. A high intrinsic coercivity will also result, since significant energy is required to disrupt the preferential alignment, known as the *magnetocrystalline anisotropy energy*.

3.5. REVERSIBLE HYDROGEN STORAGE

With rising gasoline prices and increasing awareness/research related to renewables, mankind is facing an energy crisis, the likes of which could annihilate our entire population. It is predicted that in the next few years, fossil fuel use could become prohibitively expensive, leading to the necessity of using other fuel sources. Hydrogen is the most attractive alternative due to its nonpolluting nature, only yielding water as a byproduct of its combustion. However, before widespread utilization of this medium is possible, two key issues must be solved: hydrogen generation and

storage. At present, the fuel used in prototype vehicles designed by BMW, Toyota, and Honda is liquid hydrogen. Although H_2 contains more energy/mass than gasoline, a relatively large volume must be used due to its extremely low density. In addition, there are prohibitive costs involved in H_2 liquidification and cryogenic tank production. We are all familiar with the dangers associated with the storage of liquid and gaseous fuels. For instance, consider the major tragedies of the Challenger and Columbia explosions, as well as leveling of the twin towers in New York City on September 11, 2001 – all exacerbated by the presence of large volumes of liquid fuels onboard. Hence, there continues to be much interest in the search for solid-state materials that can reversibly store energetic fuels such as hydrogen.

Hydrogen combines with many elements to form binary hydrides, MH_n. There are three general classes of hydrides:

1. Saline or Binary (involving Group 1 and 2 metals; may be envisioned as an ionic lattice consisting of M^{n+} and H^- ions, *e.g.*, LiH, NaH, BaH_2)
2. Covalent (molecules containing covalently bound hydrogen to nonmetals, with individual molecules held together by intermolecular forces, *e.g.*, CH_4, SiH_4)
3. Metallic/interstitial (hydrogen molecules are contained in vacant interstitial sites of a transition-metal lattice, *e.g.*, $PdH_{0.6}$)

To be successful for hydrogen storage in energy devices, the following five parameters must be met:

1. The solid material must be able to adsorb/desorb at least 6.5 wt% and 65 g L^{-1} of hydrogen gas.
2. The decomposition temperature necessary for generation of hydrogen from the material should be in the range of 60–90°C.
3. The absorption/desorption of hydrogen from the material should be reversible.
4. The material should be low cost, precluding the use of noble-metal alloys.
5. The storage solid should be nontoxic and inert under environmental conditions. That is, the solid should not react with water, oxygen, nitrogen, *etc.*

Table 3.7 lists some important metals and alloys that have been studied for hydrogen storage applications. To date, no material (neither metals nor nonmetals) has been discovered that satisfies all of the above five constraints. Although hydrides exist for most elements of the Periodic Table, only the light elements (*e.g.*, Li, Mg, Al) are able to meet criterion 1 above.

The most widespread application for hydrogen storage materials continues to be for the negative electrode (cathode) in rechargeable alkaline nickel–metal hydride (Ni–MH) batteries – used extensively in portable electronic devices and electric vehicles. The most common metal hydrides used for battery applications are intermetallic species such as $LaNi_5$, $LaMg_{12}$, and complex AB_2 alloys (see Table 3.7). The development of complex alloys was necessary to circumvent the high equilibrium pressures that early batteries exhibited at room temperature. The composition of metal hydrides may now be fine-tuned to offer low operating pressures, corrosion resistance, and reversible H_2 storage.

The decomposition temperature, T_{dec}, for binary metal hydrides (MH_x) is found to correlate strongly with the standard reduction potential, E^o (Figure 3.33). In particular, the easier it is to reduce the metal (*i.e.*, a larger reduction potential), the lower

Table 3.7. Comparison of Metals and Alloys for Hydrogen Storage

Metal/alloy ($MH_{2(ads)}$ compound)	H_2 concentration stored (wt%)	Decomposition temp. (°C)	Reversible H_2 adsorption/desorption?
Pd ($PdH_{0.6}$)	0.6	25	Yes
"AB_2"[a]	1.7-3.3	<100	Yes
$LaNi_5$($LaNi_5H_6$)	2.5	25	Yes
FeTi ($FeTiH_{1.7}$)	2.5	25	Yes
BaRe ($BaReH_9$)	3.5	<100	Yes
Mg_2Ni(Mg_2NiH_4)	3.6	25	Yes
Na (NaH)	4.2	425	Yes
$LaMg_{12}$(LaH_3, MgH_2)	4.6	290	Yes
Ca (CaH_2)	4.8	600	Yes
NaAl:Ti ($NaAlH_4$: TiO_2)	5.5	125	Yes
Li_2N(Li_2NH)	6.7	285	Yes
Mg (MgH_2)	7.6	330	No
LiAl ($LiAlH_4$)	8.0	180	No
Li_3Be_2($Li_3Be_2H_7$)	8.7	300	Yes
LiB:Si ($LiBH_4$: SiO_2)	9.0	200–400	No
NaB ($NaBH_4$: H_2O)	9.2	25	No
Al (AlH_3)	10.0	150	No
Al:N ((NH_3)AlH_3)	12	150	No
Li (LiH)	12.6	720	No
NaB ($NaBH_4$)	13.0	400	No
LiB:N ($LiBH_4$: NH_4F)	13.6	25	No
Be (BeH_2)	18.2	250	No
LiB ($LiBH_4$)	19.6	380	No
BeB_2($Be(BH_4)_2$)	20.6	40	No

Note: This table does not include important nonmetals such as carbon allotropes or boron nitride compounds; These materials will be discussed in subsequent chapters.
[a]A = V, Ti; B = Zr, Ni. Also includes complex combinations (*e.g.*, $ZrNi_{1.2}Mn_{0.48}Cr_{0.28}V_{0.13}$).

the temperature that is required to decompose the solid into the metal and hydrogen gas (Eqs. (16)–(18)).

$$M^{n+} + ne^- \rightarrow M^0 \quad (16)$$

$$2nH^- \rightarrow 2ne^- + nH_2 \quad (17)$$

$$\text{Overall} : M^{n+} + 2nH^- \rightarrow M^0 + nH_2 \quad (18)$$

Ternary hydrides of the general formula $(MH_x)_a(EH_y)_b$, where E is either a metal or nonmetal, are also important candidates for hydrogen storage applications. For example, some of the highest wt% storage values are exhibited by reducing agents such as sodium metal or lithium borohydride – $NaBH_4$ and $LiBH_4$, respectively. Relevant for materials design, the decomposition temperature of ternary hydrides may be altered through choice of E^{y+}. For example, the T_{dec} of $LiGaH_4$ is *ca.* 50°C

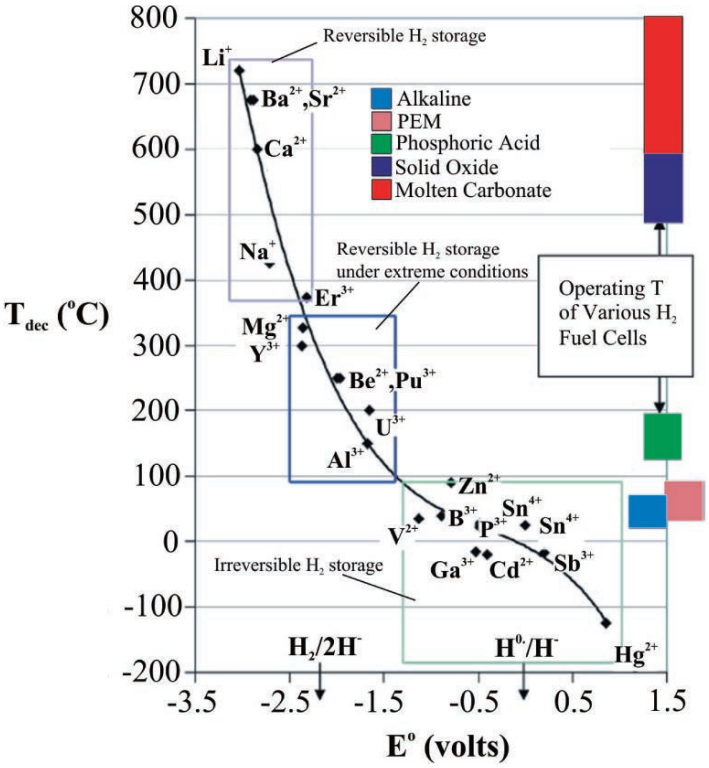


Figure 3.33. Relationship between the reduction potential and decomposition temperature for binary hydrides. Reproduced with permission from *Chem. Rev.* **2004**, *104*, 1283. Copyright 2004 American Chemical Society.

higher than that of the binary GaH_3; on the other hand, the T_{dec} of BeH_2 is *ca.* 225°C greater than that of $Be(BH_4)_2$.

The trend in T_{dec} may be rationalized by the relative difference in electronegativities of M^{x+} and E^{y+} species. For $LiGaH_4$, Ga^{3+} is a stronger Lewis base than Li^+, indicating that electron density will preferentially flow away from the gallium center, forming ionic Ga–H bonds. This causes a strengthening of the $Li^+ \cdots H^-$ interactions through donation of H^- to Li^+, resulting in a higher overall T_{dec}. In contrast for $Be(BH_4)_2$, the B–H bonds are covalent in nature, which causes H^- to be withdrawn from Be^{2+}. This results in a relatively low T_{dec} value that approaches room temperature.

A number of molecular transformations take place during the formation of metal hydrides. Once hydrogen gas is adsorbed on the metal surface, the diatomic hydrogen molecule is dissociated – a process that requires a great deal of energy. At that point, individual hydrogen atoms migrate from the surface to the bulk of the material where nucleation/growth of the hydride phase begins.

Among the possible ternary hydrides, $NaAlH_4$ is most attractive for hydrogen storage applications due to its relatively low H_2 desorption temperature (80°C *vs.* 300°C+ for magnesium compounds). The reactions involved in the thermal decomposition of complex hydrides of the general formula $MAlH_4$ (M = Li, Na) are shown by Eqs. 19–21. Whereas the first two reactions occur at temperatures around 200°C, Eq. 21 only occurs at very high temperatures and is thus not considered a useful route for H_2 generation.

$$3MAlH_4 \rightarrow M_3AlH_6 + 2Al + 3H_2 \quad (19)$$

$$M_3AlH_6 \rightarrow 3LiH + Al + \frac{3}{2}H_2 \quad (20)$$

$$3LiH \rightarrow 3Li + \frac{3}{2}H_2 \quad (21)$$

There continues to be significant research efforts devoted to the design of suitable catalysts that will improve the relatively slow kinetics associated with the reversible H_2 storage of $MAlH_4$ compounds. Intermetallic compounds doped with Ti and Zr have been successfully used to improve the rate of hydride formation/release in these compounds; typically, levels between 2–4 mol% Ti is sufficient to facilitate reversibility. The catalytic mechanism may be rationalized by the donation of H_2 s electron density to empty d orbitals on the Ti, along with synergistic donation of electrons from filled Ti d orbitals to the σ^* orbital of H_2. This two-way electron donation weakens the H–H bond, while strengthening the H $\cdots$ Ti interaction (Figure 3.34).

Sometimes catalytic dopants do not alloy with the host metal; for example, the addition of Nb and V to Mg form heterogeneous mixtures rather than intermetallic compounds. A new generation of catalysts are being developed that deliver hydrogen to the metal surface as H^0 radicals rather than H_2. It may then be possible to combine a high level of hydrogen storage with low T_{dec}, both associated with desirable adsorption/desorption kinetics.

It has recently been discovered that decreasing the particle size of metal alloy particles through ball-milling processes will increase the adsorption kinetics by an order of magnitude. This enhanced activity is due to the increased surface area of the ground particulates, and decreased surface reaction path length. For $LiAlH_4$, only prolonged milling is of sufficient energy to desorb H_2. When grinding is coupled with catalytic dopants, the H_2 storage kinetics increases even further. Upon milling bulk Mg_2NiH_4, the T_{dec} decreases by *ca.* 40°C. Such mechanical processing has also been used to synthesize H_2-storage compounds (*e.g.*, $La_{1.8}Ca_{0.2}Mg_{14}Ni_3$, $Li_xBe_yH_{x+2y}$, $MAlH_4$; M–Mg, Ca, Sr) that consist of a metastable amorphous or nanocrystalline structure.

IMPORTANT (AND CONTROVERSIAL!) MATERIALS APPLICATIONS II: DEPLETED URANIUM

Due to the high radioactivity of the actinides, we would expect that their use for materials applications would be limited. However, a relatively benign form of uranium, known as *depleted uranium* (DU), has been widely used in applications

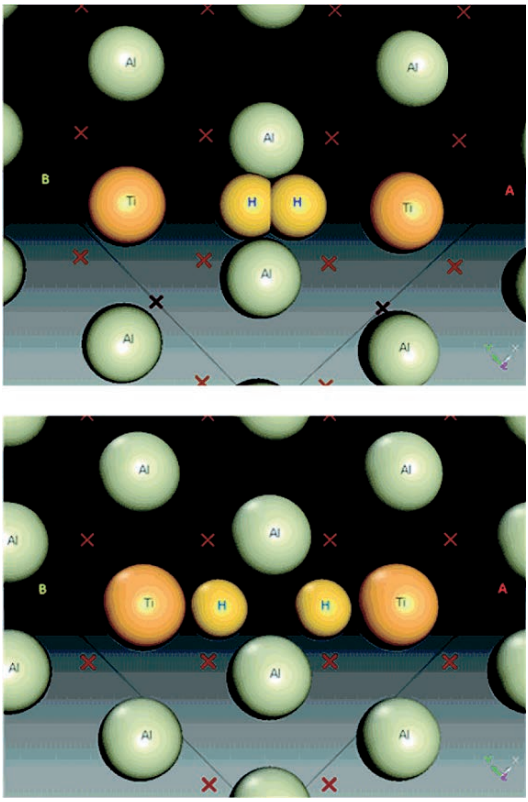


Figure 3.34. The influence of Ti doping on the dissociation H_2 on a (100) Al surface. For clarity, Al atoms from the layers below are shown as crosses. Reproduced with permission from *J. Phys. Chem. B* **2005**, *109*, 6952. Copyright 2005 American Chemical Society.

such as machinery ballast and counterweights, aircraft balancing/damping controls,[7] radiation and penetration shielding, oil-well drilling equipment, and high-impact weaponry.

Uranium is obtained from ores primarily located in New Mexico, Colorado, Wyoming, Utah, and Arizona as well as many other locations throughout the world. There are three isotopes of uranium – 99.28% of ^{238}U, 0.005% of ^{234}U, and 0.71% of ^{235}U. Only the ^{235}U and ^{234}U isotopes are used for nuclear power and weapon manufacturing. When ^{235}U and ^{234}U have been extracted from natural uranium, the remaining "depleted uranium" is primarily ^{238}U. Although the ^{238}U isotope of uranium is 40% less radioactive than ^{235}U, DU also contains *ca.* 0.3% ^{235}U and traces of other radioactive contaminants such as Np, Pu (yes, plutonium!), Am, and ^{99}Tc.

The major use of uranium is for nuclear power generation, which uses "low-enriched uranium" (LEU) of $\leq$20%^{235}U. To produce 1 kg of 5% LEU requires 11.8 kg of natural uranium; the remaining *ca.* 10.8 kg is DU. Hence, depleted

uranium is an extremely inexpensive waste product for which applications are actively sought to recycle the current stockpiles at enrichment plants. It is estimated that the US alone has more than 560,000 metric tons of depleted uranium currently stored as UF_6 in cylinders at various locations throughout the country.

When DU is alloyed with Mo or Ti and posttreated through heating/quenching regimes, the material is as strong as quench-hardened steel with a tensile strength >1,600 MPa. The combination of extreme hardness and density (19,050 kg m^{-3} – 1.6 times that of lead) makes DU very effective at piercing armor. This has resulted in military applications such as kinetic energy penetrators and protective armor. In fact, it has been reported that the US fired a total of 320 tons of DU projectiles during the Gulf War.

Even though the density of DU is essentially identical to Ti, its destructive force is much greater at a fraction of the cost. On impact with an armored target, the head of the projectile fractures in such a way that it self-sharpens. The heat released from its impact causes a disintegration of the surface resulting in the formation of pyrophoric dust, furthering its destructive ability. By comparison, tungsten penetrators are nonpyrophoric, and form a dull mushroom shape as they penetrate a hard target.

As you might imagine, there has been much public outcries about DU applications. Depleted uranium has been linked to everything from Gulf War syndrome to birth defects. Although these detrimental health effects have not been proven, there are efforts in the US to develop alternatives to DU for penetrator applications. Conventional tungsten-based projectiles comprise W particles embedded in a Ni alloy matrix. There is a focus to replace the Ni matrix with other metal alloys that will promote localized plastic deformation, through properties such as high hardness/density, low heat capacity, and low work hardening. However, none of the novel tungsten composites have yet to equal the performance of DU alloys currently used in ammunition.

Another[1] possibility for DU replacement materials is amorphous metal alloys. Amorphous tungsten alloys share many of the desirable properties of DU such as self-sharpening and pyrophoricity. Although preliminary studies show that fragments of tungsten metal also present health problems such as tumors through skin contact and inhalation hazards, these dangers are believed to be much less pronounced than DU, without any further problems associated with radioactivity.

References and Notes

1 Hohmann, C.; Tipton Jr., B.; Dutton, M. *Propellant for the NASA Standard Initiator* October 2000 (NASA/TP-2000-210186). May be downloaded for free at http://ston.jsc.nasa.gov/collections/TRS/_techrep/TP-2000-210186.pdf

2 These cutoff values for steels are arbitrary. Iron at the lower end of this range is referred to either mild steel, or low-carbon steel.

3 A good reference for "crystal field theory" is Cotton, F. A.; Wilkinson, G.; Gaus, P. L. *Basic Inorganic Chemistry*, 3rd ed., Wiley: New York, 1994.

4 Cumberland, R. W.; Weinberger, M. B.; Gilman, J. J.; Clark, S. M.; Tolbert, S. H.; Kaner, R. B. *J. Am. Chem. Soc.* **2005**, *127*, 7264.

5 Liang, W.; Zhou, M.; Ke, F. *Nano Lett.* **2005**, *5*, 2039.

[6] Note: think of the atom in the middle of the bcc unit cell – at lattice position (1/2, 1/2, 1/2). Since there are no atoms on the unit cell faces in a bcc array, there are no atoms that lie directly along the x, y, and z axes emanating from this central atom.

[7] Each Boeing 747 contains *ca.* 1,500 of depleted uranium for this application.

Topics for Further Discussion

1. For iron allotropes, why is the solubility of carbon greater in the austenite phase, relative to the ferrite phase?
2. Although iron is most stable in its bcc form, why are heavier Group 8 congeners Ru and Os most stable as hcp?
3. What is the difference between substitutional and interstitial dopants? Provide examples for each type.
4. Explain the atomic diffusion processes that occur when steel is heated and subsequently quenched by cold water. Use diagrams to illustrate your rationale.
5. Name the three types of hydrogen-storage metals/alloys and describe the placement of hydrogen within each lattice.
6. Consider the sintering process of compacted metal powders. Would the resulting sintered material be more or less desirable (from a mechanical standpoint), if an excessive amount of metal oxides were present in the presintered matrix? How would you design the sintering conditions (co-reactant gases, temperature, *etc.*) for these matrices?
7. How does precipitation hardening work to strengthen the material?
8. For surface phosphating of aluminum-containing metals, fluoride-based additives are needed to cause crystallization. Why?
9. Using redox potentials, explain the frequent occurrence of perforations in domestic hot tap water pipes manufactured from galvanized steel.
10. Explain how shape-memory metals are able to manipulate their shapes in response to temperature fluctuations. Are there other alloy candidates for this type of behavior?
11. Classify the various phases in the Fe–C system as type I or II alloys (or both).
12. For the density of states for transition metals (*e.g.*, Figure 3.21), the d-band is much narrower than the overlapping s/p band. Why is this so, and what physical properties does this govern?
13. Explain why many ferrimagnetic materials crystallize in a spinel lattice.
14. Why are finely divided metals pyrophoric?

Further Reading

1. Porter, D. A.; Easterling, K. E. *Phase Transformations in Metals and Alloys*, 2nd ed. CRC Press: New York, 1992.
2. Honeycombe, R. W. K.; Bhadeshia, H. K. D. H. *Steels: Microstructure and Properties*, 2nd ed., Wiley: New York, 1995.
3. Grosvenor, A. W. *Basic Metallurgy: Volume I, Principles*, 3rd ed., American Society for Metals: Cleveland, OH, 1958.
4. Beddoes, J.; Parr, J. G. *Introduction to Stainless Steels*, 3rd ed., ASM International: Materials Park, OH, 1999.
5. http://www.cobasys.com/pdf/tutorial/InsideNimhBattery/inside_nimh_battery_technology.html
6. *Magnetism: Fundamentals* Lacheisserie, E. T.; Gignoux, D.; Schlenker, M., eds., Springer: New York, 2004.

7. Mattis, D. C. *The Theory of Magnetism Made Simple: An Introduction to Physical Concepts and to Some Useful Mathematical Methods*, World Scientific Publishing Company: New York, 2006.
8. Lefteri, C.; Arad, R. *Metals: Materials for Inspirational Design*, Rotovision: London, 2004.
9. Sedriks, A. J. *Corrosion of Stainless Steel*, Wiley: New York, 1996.
10. *Damping Structural Vibrations with Shape-Memory Metals*, NASA Publication, University Press of the Pacific, 2004.

CHAPTER 4

SEMICONDUCTING MATERIALS

Our technologically advanced way of life would not be possible without the semiconductor industry. The first semiconductor device known as a *transistor* was discovered at Bell Labs in the late 1940s, and was widely used shortly thereafter for radio electronics. Today, transistors are still pervasive in every microelectronic component such as CD/DVD players, cellular phones, modes of transportation (*e.g.*, planes, automobiles, *etc*.), and computers. In fact, the dual-core chips released by Intel in early 2006 feature over 1.7 billion transistors – all on a surface that is smaller than a postage stamp!

This chapter will investigate the various types of semiconducting materials, focusing on the influence of their structure on overall properties. We will also detail the many applications for semiconductors, especially within the framework of microelectronic circuitry. It should be noted that nanostructural materials represent a new realm of semiconducting materials that are currently being investigated. However, these materials will not be considered in this chapter, but will instead be detailed in Chapter 6 that focuses solely on nanotechnology.

4.1. PROPERTIES AND TYPES OF SEMICONDUCTORS

As their name implies, semiconductors possess electrical conductivity intermediate between conductors such as metals and insulators such as ceramics. In Chapter 3, we discussed the band structure for the molecular orbitals of infinite lattices of metal atoms. The lack of an energy gap between the filled (valence) and empty (conduction) bands indicated that no thermal energy was required to facilitate electrical conductivity. That is, metals are able to transport electrons through their lattices even at absolute zero. By contrast, semiconductors possess a *bandgap* between valence and conduction bands (Figure 4.1). An appropriate amount of energy must be supplied to the material in order to promote an electron from the valence band to the conduction band where electron transport may occur. For semiconductors, any temperature greater than 0 K is sufficient for such electron promotion; if the bandgap is too large, the material is unable to conduct electricity, and is known as an *insulator*. The bandgap for semiconductors are typically in the range 150–290 kJ mol^{-1} (*i.e.*, 1–4 eV). Since the bandgap of Si is 1.12 eV, only radiation with wavelengths *ca.* $\leq$1,100 nm (near IR, visible, UV, X-ray, *etc*.) will be absorbed by silicon

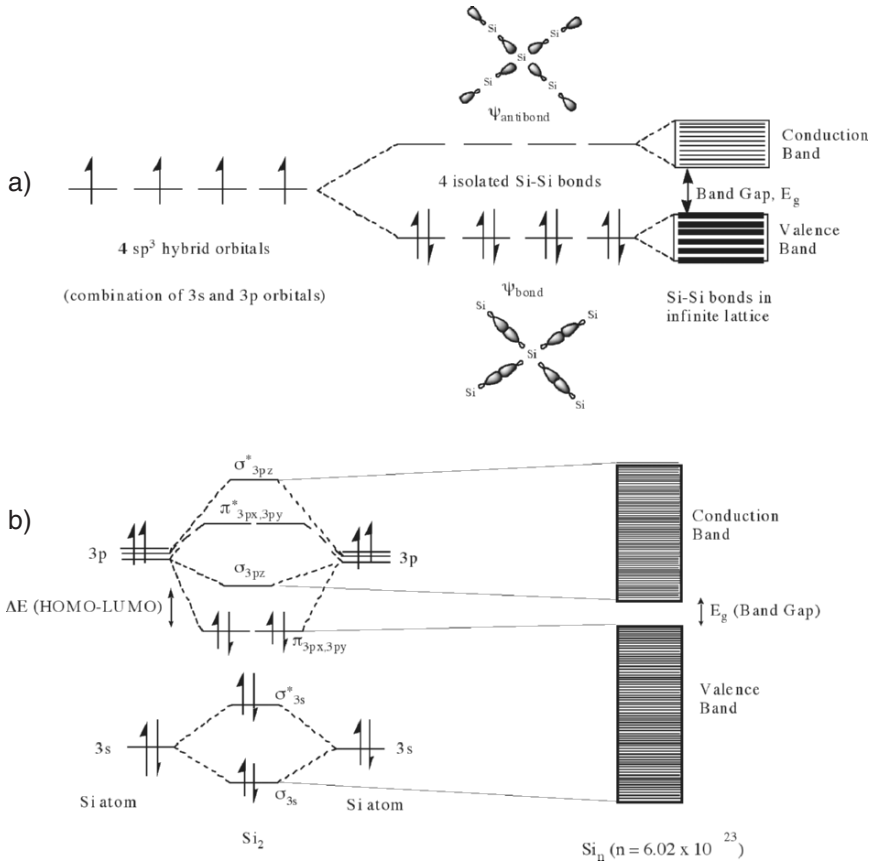


Figure 4.1. Electronic band diagrams for silicon. Shown is (a) bands resulting from overlap of sp^3 hybrid orbitals and (b) bands resulting from overlap of molecular orbitals.

semiconductors, as the absorption of less-energetic radiation will place the electron in the bandgap, where there are no available energy states.

Group 14 provides an interesting case-study for the most dramatic change in electrical conductivity among its congeners. In fact, all types of conductors are present in this group, from insulating carbon to metallic tin and lead. Since the electronic configurations of these atoms are all $ns^2\ np^2$, the change in electrical conductivities is related to the bond strength for the atoms comprising the crystal lattice. Recall that the individual atoms in diamond, Si, Ge, Sn, and Pb are tetrahedrally linked through the crystal lattice due to sp^3 hybridization. As the individual s and p orbitals that comprise the hybrid orbital become more diffuse (*e.g.*, Sn, Pb), the bonding electrons are less tightly bound to individual atoms, and become more polarizable. This results in delocalized metallic bonding in Sn and Pb relative to very strong localized bonding in diamond. For intermediate Si and Ge, the bonding between individual atoms is weaker than C, allowing the possibility for thermal motion to break bonds

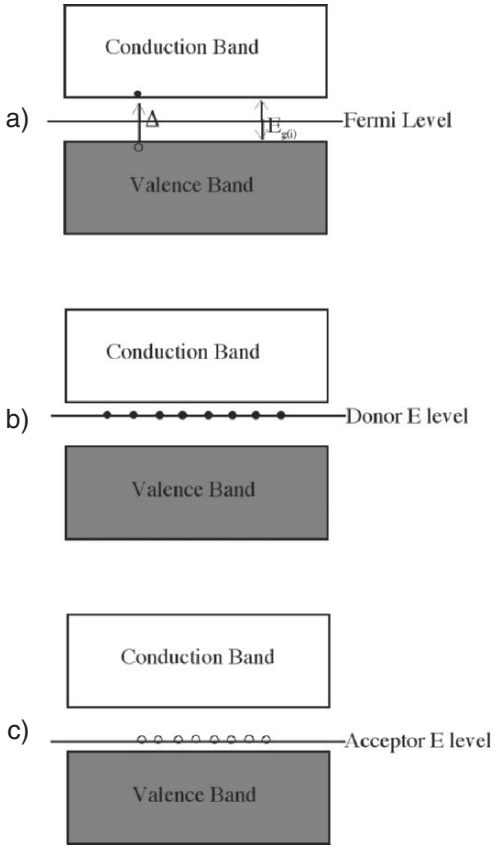


Figure 4.2. Band diagrams for intrinsic and extrinsic semiconductors. Shown are (a) an intrinsic semiconductor, with an equivalent number of free electrons and holes, (b) an n-type extrinsic semiconductor, with a greater number of electrons, and (c) a p-type extrinsic semiconductor, with an excess of holes.

in the solid-state lattice, by promoting bonding electrons into the conduction band and propagating electrical conductivity.

There are two types of semiconductors: *intrinsic* and *extrinsic* (Figure 4.2). Intrinsic semiconductors contain the same numbers of free bonding electrons, and "holes" created from the migration of electrons from the valence to conduction bands. The number of electrons that are able to traverse the bandgap is governed by the *Fermi function*, which gives the probability of an electron occupying an available energy state. The *density of states* (DOS), or number of available energy levels, is also paramount for the promotion of electrons from valence-conduction bands. The DOS begin at the bottom of the valence band and continue to the *Fermi level* – the highest occupied state at absolute zero (*i.e.*, −273.15°C). For metals that readily exhibit electrical conductivity, the Fermi level lies within the conduction band due to the lack of a bandgap. However, unlike metals, the DOS for conduction electrons in a

semiconductor do not coincide with the Fermi level, but rather begin at the top of the bandgap. Hence, the placement of the *Fermi level* represents the relative ease at which an electron is promoted from the valance band to the conduction band in a bulk semiconductor. Quantitatively, the density of occupied states per unit volume is the product of the DOS and the Fermi function, $f(E)$; the density of holes is the product of the DOS and $[1-f(E)]$.

For *extrinsic semiconductors*, the Fermi level corresponds to a level slightly above or below conduction or valence bands, depending on the dopants introduced into the lattice – either an excess of electrons or holes (Figure 4.2b,c). Since our frame of reference for semiconducting ability is the Group 14 element of Si, the terms *electron-deficient* and *electron-rich* dopants refer to atoms possessing <4 valence electrons (*e.g.*, B, Al) and >4 electrons (*e.g.*, N, P), respectively. By convention, if additional electrons are introduced into the lattice, the semiconductor is designated as *n-type*, whereas the doping of additional holes yields a *p-type* semiconductor designation. From the convenient "n" and "p" notation, one may immediately recognize whether there are excess *n*egative or *p*ositive carriers in the lattice.

As Figure 4.3 illustrates, when thermal energy promotes a bonding electron from the valence band to the conduction band, the released electrons are free to migrate throughout the lattice. However, the vacancies (*i.e.*, holes) left behind are also free to move – in the opposite direction as electrons. One may consider these holes as positively charged species formed from loss of an electron. Thus, electrons and holes represent the two types of carriers that correspond to electrical conductivity in semiconductors.

Since an electron that has been promoted to the conduction band will have a greater energy than those left in the valence band, there is a possibility for the electron to lose this excess energy. The spontaneous return of electrons in the conduction band to the valance band is known as *recombination*, and is usually accompanied by light emission and heat (Figure 4.4). This phenomenon happens all the time for excited-state molecules. For instance, consider what happens when one supplies sodium atoms with sufficient energy to promote an electron from the 3s energy level

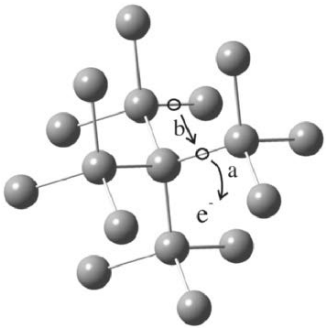


Figure 4.3. Illustration of creation/migration of electrons and holes created through Si–Si bond thermolysis. Shown are (a) the release of an electron and concomitant formation of a hole and (b) the migration of an electron from a nearby bond to fill the vacancy.

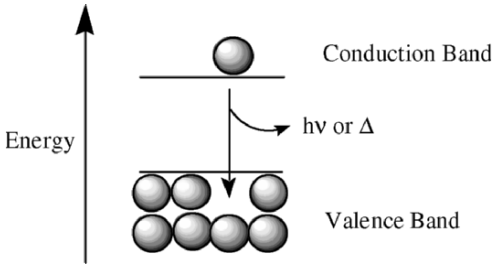


Figure 4.4. Schematic of recombination of electron-hole pairs generating either a photon of energy or heat.

into an empty 3p, or 4s orbital. The electron does not stay in the higher energy level very long (typically *ca.* 10^{-6} s) before it is spontaneously dropped to its original ground state, releasing heat and/or light.

To understand the recombination process in more detail, we must first consider some concepts of quantum physics. Recall that matter and light exhibit both wave- and particle-like behavior. This duality may be expressed by the de Broglie equation, which equates the wavelength and momentum, p, of a particle (Eq. 1). The potential energy of an electron in a crystal lattice depends on its location, and will be periodic due to the regular array of lattice atoms. The periodic wavefunctions that result from solving the Schrödinger equation are referred to as *Bloch wavefunctions*. Each wavefunction represents the energy of an electron at a specific location in the lattice, referred to as k-space. Hence, an *E-k diagram* may be constructed, with the potential energy of the electron on the y-axis, and the wavevector, k, on the x-axis. The wavevector represents lattice directions of the semiconductor crystal; changing values of k represents a change in momentum of the electron. A comparison of the E-k diagrams for Si or Ge and a compound semiconductor such as GaAs is shown in Figure 4.5. For GaAs, the minimum of the conduction band (CB) is directly above the maximum of the valence band (VB). Accordingly, we refer to such solids as *direct bandgap* materials. In contrast, for Si and Ge, the CB minimum and VB maximum are offset, resulting in an *indirect bandgap*.

$$\lambda = \frac{h}{p} \tag{1}$$

For Si, in order for an electron at the bottom of the CB to recombine with a hole from the top of the VB, the momentum of the electron must shift from k_{cb} to k_{vb} (Figure 4.5b). However, this is not allowed by the Law of Conservation of Momentum. Instead, an indirect recombination mechanism must take place, wherein the electron is captured by an interstitial defect with energy E_r, which facilitates its relaxation to the top of the VB. This process is accompanied by the emission of phonons, or lattice vibrations rather than light emission.

For direct bandgap materials, the electron-hole recombination may occur without any change in momentum, resulting in the emission of photons. We will describe some important applications for direct semiconductors later in this chapter.

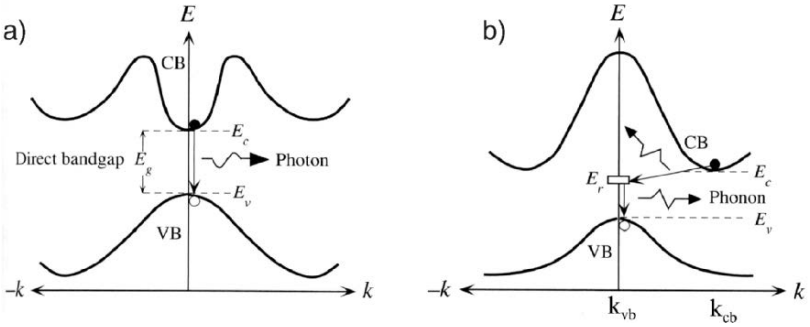


Figure 4.5. Comparison of (a) direct bandgap (*e.g.*, GaAs) and (b) indirect bandgap (*e.g.*, Si, Ge) materials. Reproduced with permission from Kasap, S. O. *Principles of Electronic Materials and Devices*, 2nd ed., McGraw-Hill: New York, 2002.

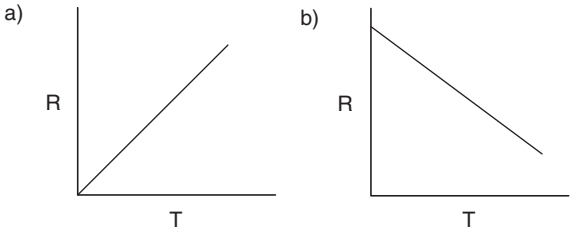


Figure 4.6. Resistivity–temperature relationships for (a) metals and (b) semiconductors.

The temperature dependence for semiconductors is quite different for metals and semiconductors. At a temperature of 0 K, a semiconductor will behave as a perfect insulator. However, metals will exhibit electrical conductivity at absolute zero due to the delocalized electron density and lattices described in Chapter 3. However, as the temperature is increased, the respective conductivities of these materials will be reversed, with metals showing a decrease and semiconductors an increase in conductivity (Figure 4.6). The thermal motion of metal atoms causes less efficient electron mobility through the lattice, whereas a temperature increase causes the bandgap to narrow for semiconductors, resulting in more effective electrical conductivity. As the temperature continues to increase for semiconductors, the linear relationship does not continue to hold, and the resistance begins to increase analogous to metals. Instead, the atomic vibrations caused by the elevated temperatures begin to outweigh the thermally induced decrease in bandgap.

4.2. SILICON-BASED APPLICATIONS

Silicon is the second most abundant element in the earth's crust, next to oxygen. Due to the stability of silicon oxide compounds, elemental silicon does not occur in its free state in nature, but occurs as oxide (*e.g.*, sand, quartz, amethyst, flint, opal, *etc*.)

and silicate (*e.g.*, granite, asbestos, feldspar, clay, mica, *etc.*) minerals. The widespread availability of silicon-containing minerals is responsible for the ubiquitous use of pottery, bricks, glass, and cement since the days of the earliest civilizations.

More recently, the applications for silicon now include electronics. A silicon-based device is found in almost every consumer product available in our world today. Even refrigerators now have extensive microprocessor controls, some fitted with television screens! The popularity of flat-panel televisions has also employed silicon-based technology, especially for thin-film transistor liquid crystal displays (TFT LCDs). In addition to electronics, as the world looks for alternative sources of energy due to dwindling petroleum reserves, silicon-based photovoltaic devices will represent an increasingly important application for our society.

4.2.1. Silicon Wafer Production

The silicon employed for microelectronic and photovoltaic applications must first go through extensive processing to ensure that the material is of utmost purity. This section will describe these steps, with a discussion of perhaps the most intriguing conversion in the realm of materials science: the synthesis of high-purity polished silicon wafers from a naturally occurring form of silicon – sand.

Sea sand is primarily comprised of silicon dioxide (silica), which may be converted to elemental silicon (96–99% purity) through reaction with carbon sources such as charcoal and coal (Eq. 2). Use of a slight excess of SiO_2 prevents silicon carbide (SiC) from forming, which is a stable product at such a high reaction temperature. Scrap iron is often present during this transformation in order to yield silicon-doped steel as a useful by-product.

$$(2) \quad SiO_{2(s)} + 2C_{(s)} \xrightarrow{2{,}000-2{,}500^\circ C} Si_{(s)} + 2CO_{(g)}$$

$$(3) \quad Si_{(s)} + 3HCl_{(g)} \xrightarrow{300^\circ C} SiHCl_{3(g)} + H_{2(g)}$$

$$(4) \quad 2SiHCl_{3(g)} \xrightarrow{1{,}150^\circ C} Si_{(s)} + 2HCl_{(g)} + SiCl_{4(g)}$$

The purity of silicon in this first step is only *ca.* 98%, and is referred to as metallurgical grade silicon (MG-Si). In order for the silicon to be used for electronics applications, additional steps are necessary to decrease the number of impurities. Reaction of MG-Si with hydrogen chloride gas at a moderate temperature converts the silicon to trichlorosilane gas (Eq. 3). When $SiHCl_3$ is heated to a temperature of *ca.* 1,150°C, it decomposes into high-purity silicon and gaseous by-products (Eq. 4). This reaction is typically performed in a bell-shaped Siemens-type reactor, where Si is deposited onto heated electrodes.

Recently, a *fluidized-bed* approach has been developed wherein SiH_4 and H_2 gases are fed into the bottom of a vertical reactor held at a temperature $>600^\circ C$. Silicon seed crystals are suspended in the chamber due to the gas flow, and decomposition of the gaseous precursor causes the nucleation/growth of silicon on the surface of the seeds. When grown to large sizes, the particles no longer remain suspended and are collected at the bottom of the reactor.

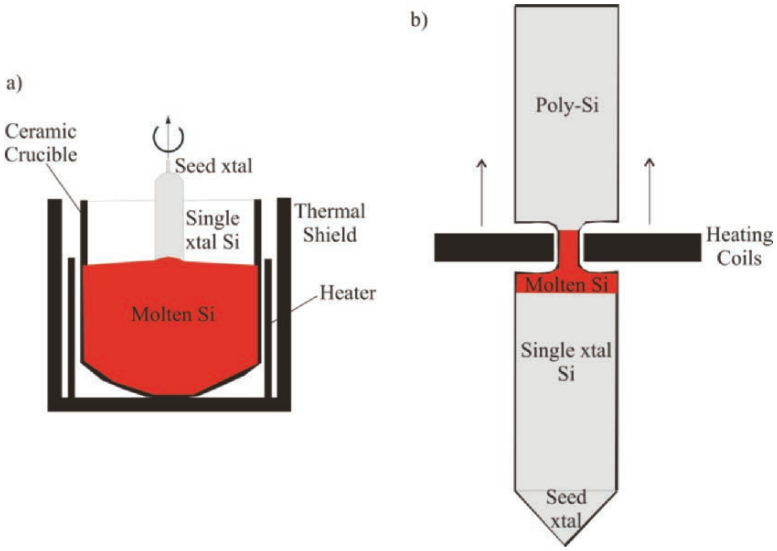


Figure 4.7. Illustrations of the (a) Czochralski (CZ) and (b) float-zone (FZ) methods used to convert polysilicon to single-crystal silicon.

These chemical processes result in electronic-grade silicon (EG-Si), with a purity of 99.999999999%; that is, only one out of every billion atoms in the solid is something other than silicon! To put this into perspective, imagine stacking yellow tennis balls from the earth's surface to the moon; replacing only one of these with a blue ball would represent the level of impurities in EG-Si. Every year, between 15,000 and 20,000 tones of EG-Si is manufactured throughout the world for an ever-increasing number of applications.

Though the purity of EG-Si is suitable for electronics applications, the atomic structure must first be converted from its polycrystalline structure (*polysilicon*) to a single crystal. There are two main methods used for this conversion: Czochralski (CZ) and float-zone (FZ) techniques (Figure 4.7). For CZ growth, silicon and any desired dopants are molten together in a crucible at a temperature above the melting point of silicon, 1,414°C. A rod fastened with a single crystal of silicon is positioned on the surface of the melt, and is pulled upward while rotating (Figure 4.7a). This results in a long cylinder of Si that is referred to as an *ingot*. Due to the surface tension of the liquid, a thin film of silicon first forms on the seed crystal surface. Additional silicon atoms orient themselves to the seed crystal; hence, the final ingot has the same crystal lattice as the original seed. The diameter of the resulting ingot may be finely controlled by manipulating the temperature and rate of pulling/rotation.

In order to ensure high purity of the ingot, this process is normally performed *in vacuo* (*ca.* 10^{-6} Torr), or under an inert atmosphere (*e.g.*, 99.999% Ar) using an inert chamber such as quartz. Even under these reaction conditions, the CZ method technique suffers from O and C impurities that arise predominantly from the crucible walls. It should be noted that small quantities of O impurities are actually

desirable, as they may trap unwanted transition metal impurities, a process referred to as *gettering*. The CZ technique is especially useful to yield doped semiconductors. For example, to yield p-doped Si, the desired concentration of pure Ga metal is added to molten Si within the crucible.

The float-zone technique uses inductive heating to convert poly-Si into single-crystal ingots of EG-Si. As a heating coil is slowly passed from one end to the other, the impurities become concentrated in the moving molten zone, becoming concentrated at the finishing end of the cylinder (Figure 4.7b). The resultant cylinder is typically of greater purity than those using CZ growth, making this technique the most heavily used for modern semiconductor processing. The single-crystal ingot is sliced into thin wafers using a diamond saw, with resultant thicknesses of *ca*. 0.7 mm, and diameters of 300 mm. The surface must be free of topographic defects, microcracks, scratches, and other residual surface imperfections.

To ensure that the wafers remain free from dust and other airborne contaminants, the finished wafers are only handled within a *clean room*. To control the particulate contamination, clean rooms exhibit extensive use of stainless steel, perforated floors and ceiling tiles to promote air circulation, and sloped surfaces to avoid dust accumulation. Yellow lighting is used to prevent light-sensitive material from being exposed to UV light, until desired as a part of chip fabrication (*vide infra*). Prior to entering the clean room, personnel must cover their clothing with a white "bunny suit" made of Tyvek®, a polymeric nonwoven fabric that is nonlinting, antistatic, and contains few surface sites for particulate adhesion (Figure 4.8). En route to the clean room, the worker must also walk over a sticky pad and pass through an air shower to remove dust particles from shoes and clothing. The ratings of clean rooms range from Class 1 to Class 10,000 – an indication of the number of particles per cubic foot. As a familiar reference, in uncontrolled environments such as a typical home or office, the particle count is 5 million per cubic foot!

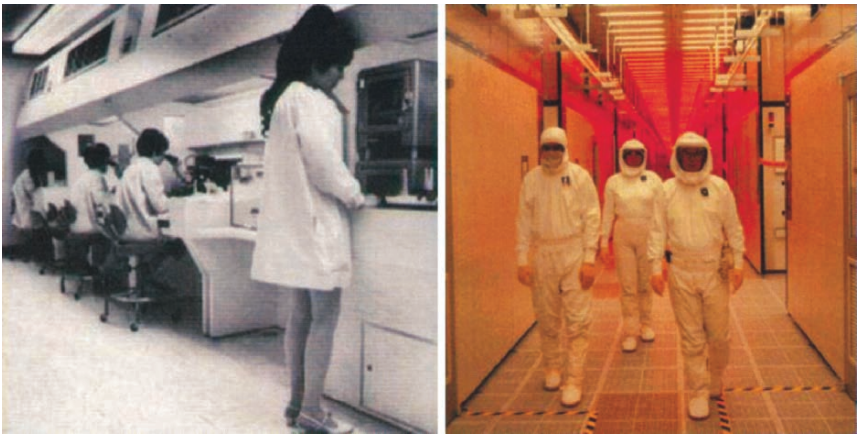


Figure 4.8. The evolution of semiconductor clean rooms from 1968 (left) to the present (right). Photos reproduced with permission from Intel Corporation (http://www.intel.com).

4.2.2. Integrated Circuits

An integrated circuit (IC), or chip, represents the brains behind all electronic devices. An IC is a compilation of billions of complex subunits such as transistors, resistors, capacitors, and diodes that are all interconnected in a specific manner depending on the desired application. Incredibly, this minute world of microcircuitry is fitted onto the surface of a thin silicon substrate about the size of a fingernail! Though ICs have only been used since the early 1970s, consider some of the ways our lives have been changed:

Lightweight laptop computers are easily taken with us while traveling;
Cars monitor their emissions and adjust engine conditions for maximum gasoline efficiency;
Doctors are able to operate on patients from remote locations;
Answers to virtually all questions we may pose are available within minutes via the Internet;
Televisions may be mounted on walls, with a clarity that rivals viewing through a window;
Phones may be used from virtually anywhere on Earth to keep in touch or check email;
GPS systems tell us how to best arrive at our desired destination...
... What a world we live in – what will the next 50 years bring?...

Field-effect transistors: structure and properties

The workhorses of ICs are *transistors*, which act as electronic switches in digital circuitry. Transistors were discovered by Bell Labs in the late 1940s as a replacement for vacuum tubes, which were much larger and consumed significantly more power. The earliest ICs utilized individual transistors; however, these circuits quickly became too large and complex to assemble for all but the simplest applications. In particular, computations were slow due to the long distances traversed by the electrical signals. It was clear that the only way to make ICs behave faster was by increasing the density of transistors. In 1965, Gordon Moore (a co-founder of Intel) predicted that the number of transistors on a chip would double every 2 years. *Moore's Law* has been upheld since this early prediction, with an impressive rise in the number of transistors used in ICs from the first commercial release in the early 1970s, to the latest dual-core processors that run both PC and Mac systems (Figure 4.9).

The first applications for transistors were radios, which utilized a small number (*i.e.*, <10) of transistors. It is hard to believe that current Intel chips, with a package size of 10–20 mm^2, now contain over a billion of individual nanoscopic transistors! Such a technological advancement is only possible through an exponential decrease in the price of single transistors. For instance, the first commercially available transistor was the Raytheon CK703, priced at $18 in 1951. Taking inflation into account, the price of a single transistor would be *ca.* $120 today, resulting in computers that would be worth over $100 billion! Fortunately, developments in chip fabrication processes have made it possible to reduce the production costs for a single transistor to less than $0.000001.

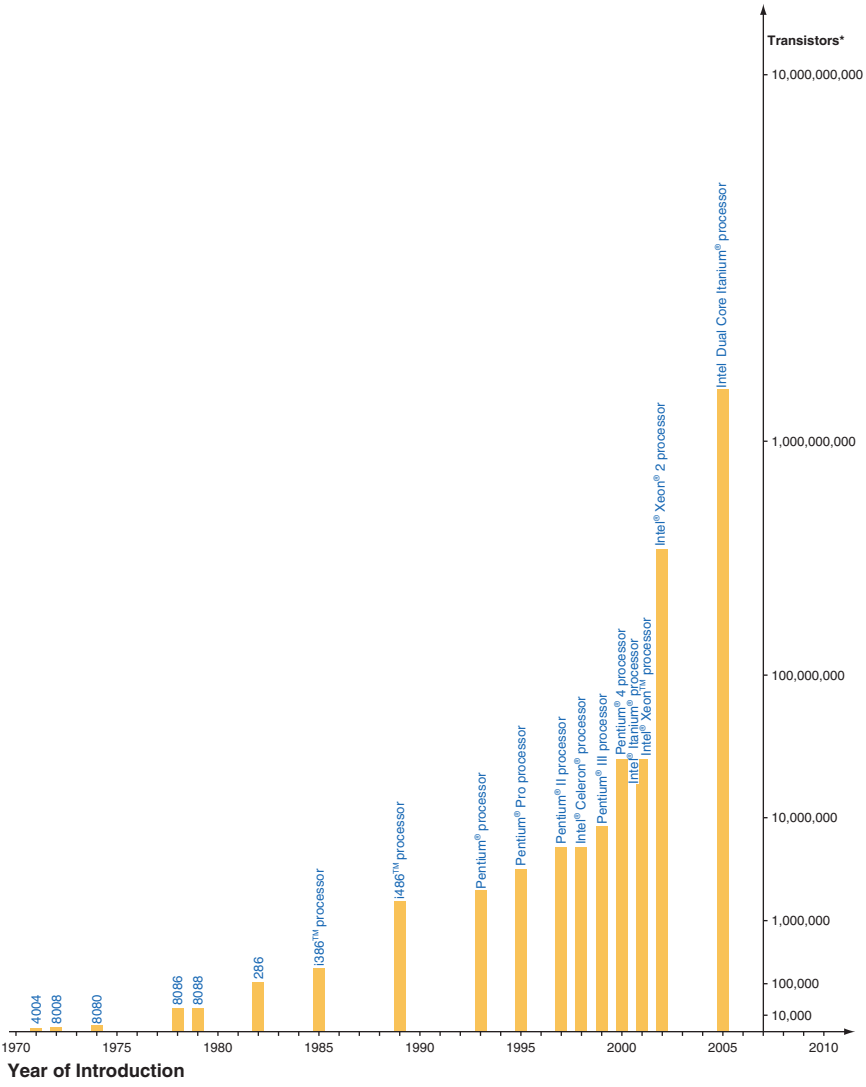


Figure 4.9. The number of transistors employed within Intel chips, as predicted by Moore's Law. Reproduced with permission from Intel Corporation (http://www.intel.com).

The number of transistors within an IC is typically used to classify the complexity of the circuits. For instance, the earliest ICs of the early 1960s were considered small-scale integration (SSI), since they contained fewer than ten transistors. In the late 1960s, ICs contained between 10 and 100 transistors, referred to as medium-scale integration (MSI). Large-scale integrated circuits (LSI), containing 100–1,000 transistors, were prevalent in the early 1970s, responsible for the first handheld calculators. The 1980s brought about very large-scale integration (VLSI), in which

ICs contained more than 1,000 transistors. Although the current architectures with greater than 1 million transistors are often referred to as ultralarge-scale integration (ULSI), the semiconductor industry still considers the latest ICs as being VLSI. Though integrated circuits are our current means for calculations, this paradigm will likely not be the last. As developments in nanotechnology continue, alternative technologies will likely replace transistor-based ICs (*e.g.*, magnetic nanostructures[1]), securing Moore's Law well into the future (Figure 4.10).

The most important component within all transistors is the *p–n junction*, referring to the interface between p- and n-type Si. This contact results in a concentration gradient being established, where the density of electrons on the n-doped side is much greater than in the p-doped region (*vice versa* for the hole concentration). As a result, free electrons from the n-type Si readily diffuse to the p-doped Si nearest the p–n junction. A significant number of free electrons will combine with holes in the p-Si region, leaving behind positively charged n-dopant ions (*e.g.*, As^+, P^+).

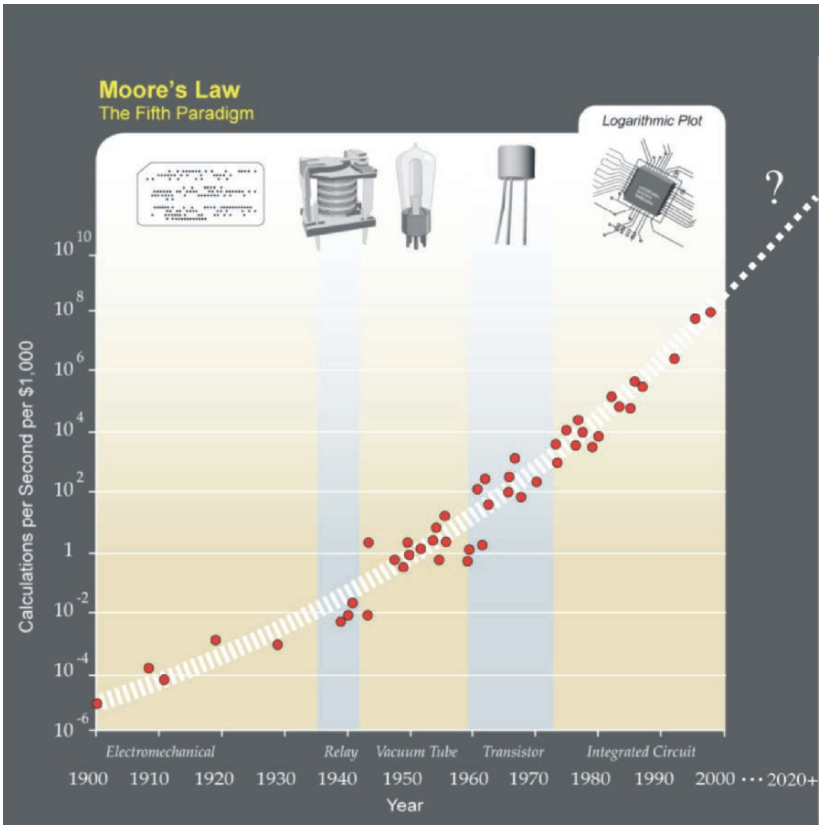


Figure 4.10. Upholding Moore's Law through developments in computational paradigms. Adapted with permission from the Wikipedia Creative Commons, Attribution License v.1.0 (courtesy of Kurzweil Technologies, Inc – http://www.kurzweilai.net/).

Likewise, negatively charged p-dopant ions (*e.g.*, B^-) are also formed on the p-Si side. Since the region in the vicinity of the p–n junction contains ions rather than free electrons/holes, this is referred to as a *depleted region*. The presence of oppositely charged ions within the depletion region forms an electric field (junction potential) that opposes, and eventually ceases, the spontaneous movement of electrons/holes from their respective n- and p-type Si regions. If an external voltage is applied that exceeds the junction potential (forward bias), the electrons in the conduction band of the n-Si side may diffuse through the p–n junction to the p-Si region. Analogously, holes may also migrate freely from p to n regions. This electrically induced flow of electrons in only one direction is referred to as a *diode*.

If a current is introduced to a simple configuration containing islands of n- or p-type Si within a matrix of an oppositely doped Si substrate (*i.e.*, p- or n-type Si, respectively – Figure 4.11a), no electron flow will occur. This is because there is no

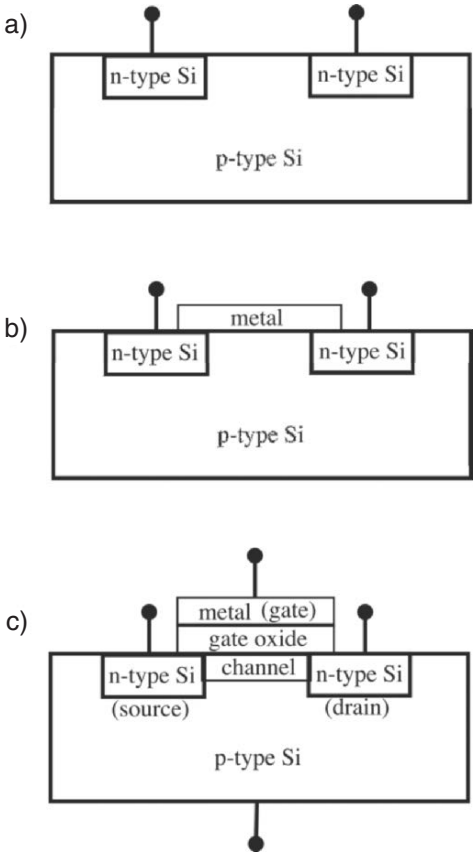


Figure 4.11. Schematic of possible n–p transistor architectures. Shown are: (a) n–p junction not capable of transporting an electrical current, (b) metallic connectivity between electron-rich reservoirs, and (c) an n-channel MOSFET, which features controllable electrical conductivity.

conduit between the electron-rich and electron-deficient regions. Electron flow will occur if a metal is allowed to contact the doped-Si regions (Figure 4.11b); however, there will be no control over the electrical current (analogous to the flow of water through a fully opened water faucet). Hence, the most effective arrangement that will allow for controllable electrical conductivity is shown in Figure 4.11c. This architecture allows one to control the electron flow through a *gate* by varying the magnitude of the applied external voltage on the metal (or polysilicon – the modern composition of complementary metal-oxide-semiconductor (CMOS) gates). Such a configuration is known as a metal oxide semiconductor field-effect transistor (MOSFET), where an insulating layer known as the *gate oxide* is used to prevent uncontrollable tunneling of electrons between the gate and doped-Si substrate. The term "field effect" refers to using an electrical field to control the flow of electrons or positively charged holes.

Let us now consider how an electrical charge may be controlled using a MOSFET. To simplify the discussion, we will first consider a metal-on-semiconductor (nMOS) capacitor – an nMOSFET without source and drain contacts (Figure 4.12a). When a negative bias is applied to the gate (V_{gate}), the positive holes are attracted from the substrate, forming an *accumulation layer* of positive charge on the SiO_2/Si interface (Figure 4.12b). The increase in hole concentration on the Si surface causes the conduction and valence bands to bend upward, increasing the gap between the Fermi level and conduction band.

By comparison, if a positive bias is applied to the gate, the positive holes will be repelled from the SiO_2/Si interface leaving behind a *depleted region* (Figure 4.12c). Since the number of holes and electrons is constant, the electron concentration increases near the silicon surface resulting in a downward bending of the conduction and valence bands (*i.e.*, narrowing the gap between E_F and E_c). However, for p-type silicon, the electron density is negligible for small positive biases. As a result, the gate charges are balanced not by electrons, but by negative acceptor ions (*e.g.*, B^-) in the depletion layer. An increase in the gate bias causes the depletion layer to widen, yielding more acceptor ions in order to balance the gate charge. The degree of E_c and E_v band bending is directly related to the intensity of the positive gate bias. At large positive gate biases, an *inversion layer* will form from the conduction band lying closer to the Fermi level than E_v. That is, at large applied voltages the surface of the Si is inverted from p-type to n-type due to the generation of free electrons that form a channel. In a MOSFET, the inversion-induced channel is located between the *source* and *drain*. Accordingly, there are two types of MOSFETs: n-channel (NMOS or nMOSFET) and p-channel (PMOS or pMOSFET), where the substrate comprises p- and n-type silicon, respectively.

It should be noted that early MOSFETs used metallic gates; however, current chips use polysilicon since it is chemically identical to the underlying substrate. An advantage of polysilicon gates is a much smaller *threshold voltage*, V_t, defined as the gate voltage necessary to induce inversion in the channel region of the substrate. The V_t is related to the difference in the *work function* (*i.e.*, ϕ – the energy required to remove an electron from the surface of a material) between the gate and channel regions. The threshold voltage is especially relevant for current microscopic transistors, which utilize significantly lower applied voltages relative to early MOSFETs.

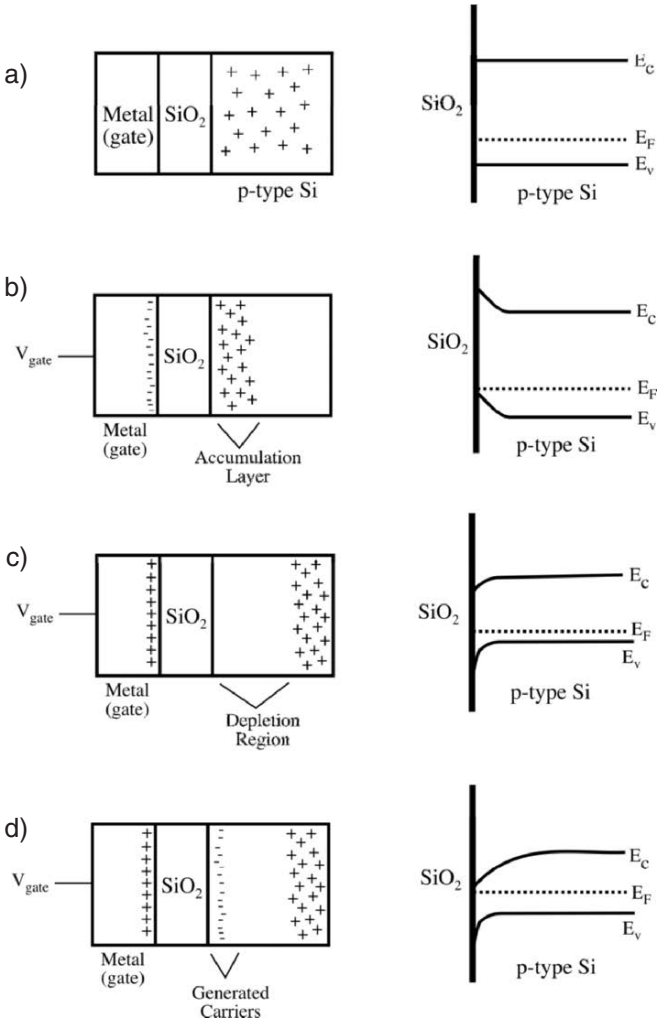


Figure 4.12. Charge distributions (left) and band diagrams (right) for an n-type metal-oxide-semiconductor (nMOS) capacitor, with different gate bias modes. Shown are: (a) $V_{gate} = V_{FB}$, (b) accumulation ($V_{gate} < V_{FB}$), (c) depletion ($V_{FB} < V_{gate} < V_T$), and (d) inversion ($V_T < V_{gate}$). V_T and V_{FB} are threshold and flat-band voltages, respectively. The flat-band voltage is obtained when the applied gate voltage equals the work function difference between the gate and the semiconductor.

That is, as V_t is reduced a weak inversion layer is formed even when $V_{gate} = 0$. This problem is referred to as *subthreshold leakage*; when magnified over a billion transistors, results in a large power dissipation of a modern-day chip. Some of the methods used to address this problem at the nanoscale will be described in Chapter 6.

Nevertheless, the miniaturization of MOSFETs is desirable for a number of reasons. Most obviously, smaller transistors may be more densely packed on ICs.

The shrinking dimensions of individual transistors will shorten the distance from source–drain, resulting in higher computational speeds. The high density of transistors per chip is also essential to maintaining affordable ICs, since the cost of their fabrication is directly related to the number of chips that may be contained on each wafer. In addition to the density advantage, there is a significant operational benefit of smaller MOSFETs. Since transistors may be considered as resistors in the on-state, a smaller transistor will have less electrical resistance as well as lower gate capacitance, allowing more current to pass through in a shorter period of time (*i.e.*, higher processing speeds and less power). Equation 5 shows the dependence of various transistor properties to its observed current.

$$I_{DS} = \frac{W}{2Lt_{ox}}\mu C(V_{GS} - V_T)^2 \tag{5}$$

Where I_{DS} is the drain current; W, the channel width; L, the channel length; t_{ox}, the gate oxide thickness; μ, the channel carrier mobility; C, the capacitance density of the gate oxide; V_{GS}, the gate voltage; and V_T is the threshold voltage.

Perhaps the most crucial component of the transistor structure is the gate oxide, which serves as the insulator between the gate and channel. Ideally, the oxide layer should be as thin as possible to increase the channel conductivity when the transistor is on, and reduce subthreshold leakage when the transistor is off. Since the first MOSFETs were demonstrated, the gate length has steadily decreased to a current size of less than 30 nm (Figure 4.13). For the current "65 nm technology node,"

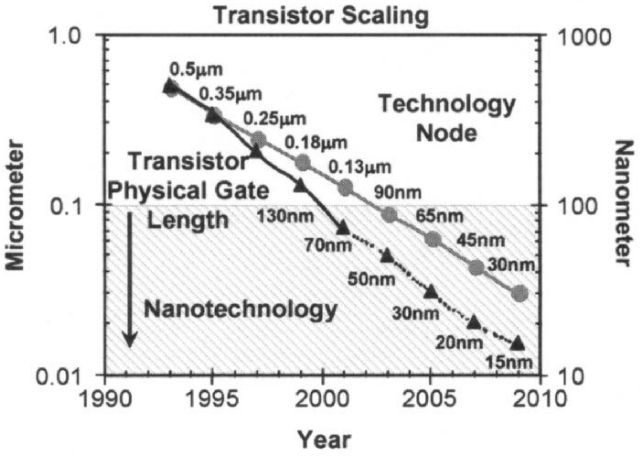


Figure 4.13. Scaling of transistor size. The top line indicates the technology node, whereas the bottom line indicates the physical size of the gate. It should be noted that the "technology node" no longer refers to physical dimensions of the transistor, but is rather an industry term related to a new fabrication process every 2 years, in accord with Moore's Law. By definition, the technology node refers to half the distance (half-pitch) between cells in a DRAM memory chip. It is noteworthy that the gate length, L, officially moved into the nanoregime in the year 2000. Reproduced with permission from Intel Corporation (http://www.intel.com).

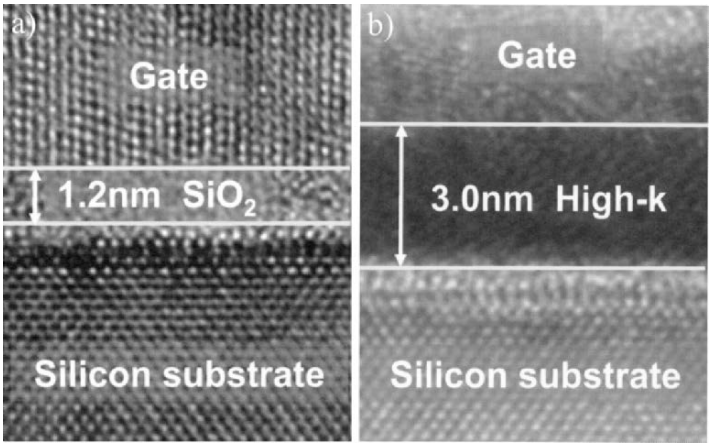


Figure 4.14. Cross-section high-resolution transmission electron microscopy (HRTEM) images of gate oxides for MOSFETs. Provided is an illustration of film thicknesses that result in identical capacitance for SiO_2($\kappa = 3.8$), relative to a high-κ dielectric ($\kappa = 23.9$) gate oxide. Hence, increased gate capacitance will result from thinner films comprising high-κ dielectric materials. Reproduced with permission from Intel Corporation (http://www.intel.com).

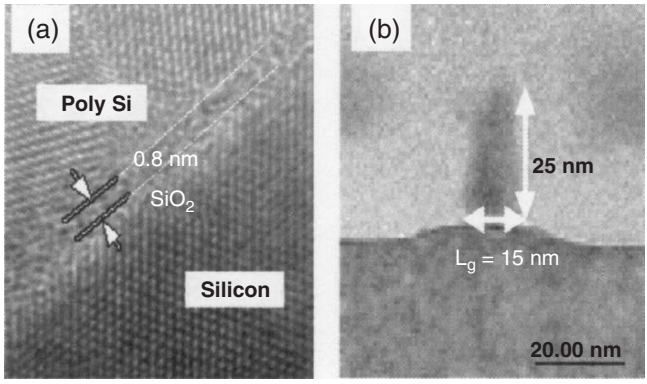


Figure 4.15. Cross-section HRTEM images of next-generation MOSFET gates. Reproduced with permission from Intel Corporation (http://www.intel.com).

this translates to a gate oxide thickness on the order of 1.2 nm, or approximately five atomic layers (Figure 4.14a). Incredibly, transistors with a channel length of 15 nm and SiO_2 thickness of only 8 Å have already been proven in the laboratory (Figure 4.15); channel lengths less than 10 nm may be realized as early as 2011! This represents an entire order of magnitude decrease from only 30 years ago, when the oxide layer thickness was *ca*. 100 nm.

As one can imagine, such a thin insulating layer is prone to electron tunneling between the polysilicon gate and channel, leading to increased power consumption

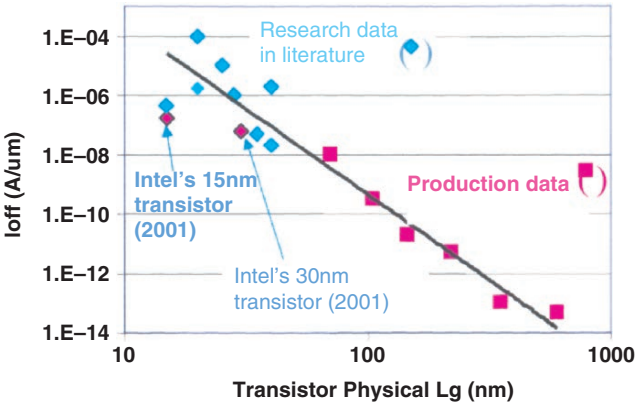


Figure 4.16. The trend of increasing leakage current density with decreasing gate length. Reproduced with permission from Intel Corporation (http://www.intel.com).

and heat buildup. Current transistors exhibit a relatively high *leakage current density*, on the order of 1–10 A cm^{-2} (Figure 4.16). Empirical and theoretical studies have shown that the practical limit for dielectric effectiveness of the gate SiO_2 layer is *ca*. 0.7–1 nm. It should also be noted that as the oxide thickness decreases, other issues arise such as boron diffusion into the channel. This would significantly alter the threshold voltage, and ultimate drain current, of the transistor.

In contrast to the high-performance microprocessor market, the rapidly growing market of low-power applications requires transistors with much lower leakage currents (*ca*. 10^{-3} A cm^{-2}). One intriguing strategy that has been proposed by Intel to minimize off-state leakage current is the design of nonplanar tri-gate transistors (Figure 4.17). Since electrical signals must move up and over the vertical gate, the effective pathway for electrical current transmission is tripled relative to today's planar MOSFETs (Figure 4.17a,b). Further on the horizon, this problem will be circumvented through the use of entirely new paradigms (*e.g.*, ballistic deflection transistors[2]) that employ a variety of nanoarchitectures, as we continue to push Moore's Law to higher limits.

Key features of next-generation designs that will help reduce gate and source/drain leakage include the use of an alternate gate oxide (see below), a silicon-on-insulator (SOI) substrate, and raised source/drain regions (Figure 4.18). The purpose of introducing source and drain cavities with greater thicknesses is rather intuitive. The electrons will have more room to maneuver, resulting in less electron–electron repulsions and lower overall resistance. This translates to a lower operating voltage for the transistor, which results in low power consumption with less heat generation during its operation.

The presence of an insulating layer directly beneath a thin Si channel is reported to reduce off-state leakage from source to drain regions by more than two orders of magnitude. When the top of the buried oxide layer (BO_x) coincides with the bottom of source/drain, the architecture is referred to as a *fully depleted SOI*; a small

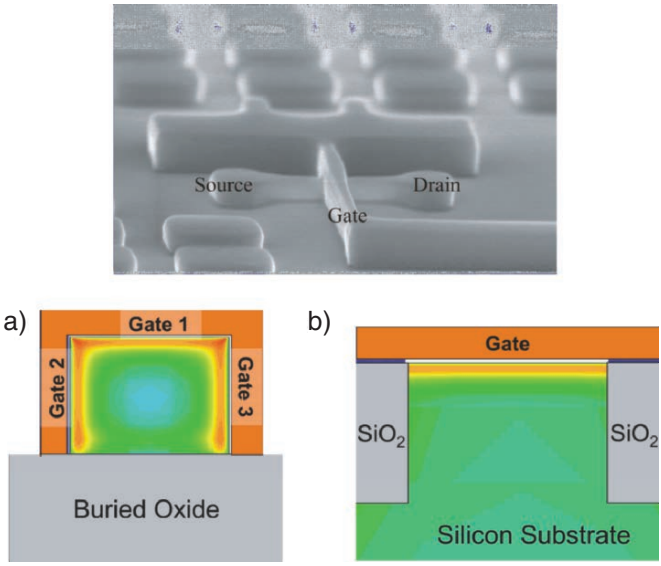


Figure 4.17. SEM image of a nonplanar tri-gate transistors. Also shown is the enhanced electron flow (shown in red) of (a) tri-gate, relative to (b) standard planar MOSFETs. Reproduced with permission from Intel Corporation (http://www.intel.com).

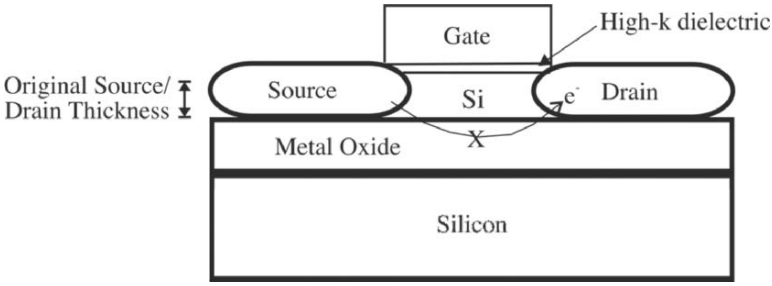


Figure 4.18. Cross-section schematic of an advanced transistor design. Features illustrated are: (a) raised source and drain contacts, (b) an SOI substrate that prevents off-state leakage from source to drain, and (c) a high-κ dielectric gate insulator. These three components employed in tandem are the basis of a faster generation of transistors referred to as "terahertz transistors."[3]

interfacial gap is known as partially depleted SOI, and will result in greater off-state leakage current. Since depleted substrates feature a nonconductive barrier, the transistor is able to switch on/off faster due to more effective drive current while in the on-state.

Another improvement to the transistor design that addresses gate leakage problems is the replacement of the SiO_2 gate oxide with insulating films known as *high-κ dielectrics*, where κ refers to its *dielectric constant* (Figure 4.14b). These materials

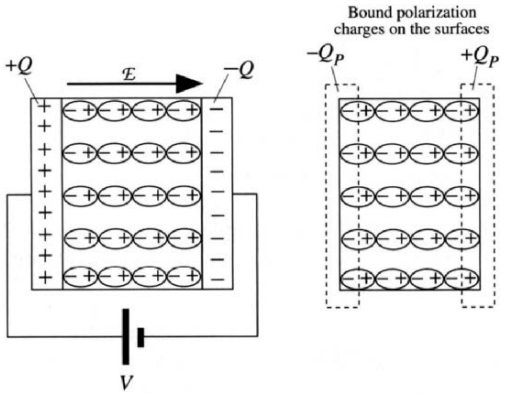


Figure 4.19. Schematic of bound polarization occurring on capacitor plates due to polarization of the dielectric medium. Reproduced with permission from Kasap, S. O. *Principles of Electronic Materials and Devices*, 2nd ed., McGraw-Hill: New York, 2002.

have a larger dielectric constant than commonly used SiO_2 ($\kappa = 3.82$), which allows one to use thicker oxide layers for an identical gate capacitance (Eq. 6).

$$C_{\text{vol}} \propto \frac{\kappa}{t^2} \tag{6}$$

Here C_{vol} is the capacitance per unit volume; κ, the dielectric constant (or, *relative permittivity*); and t is the thickness of the dielectric medium.

Materials that exhibit a high dielectric constant are strongly polarizable. That is, when exposed to an electric field, the dipole moments of the constituent molecules and atoms of the dielectric material will become aligned. This results in the buildup of electrical charges on the surface of the capacitor plates (Figure 4.19). In general, inorganic materials with ionic or partial ionic interactions will possess a relatively large dielectric constant (high-κ), whereas organic molecules with nonpolar bonds are typically low-κ materials. Materials containing relatively large atoms will be most useful as dielectrics, since they exhibit greater polarizabilities due to valence electrons being situated farther from the nucleus.

Analogous to ferromagnetic materials discussed in Chapter 3, *ferroelectric* crystals exhibit a permanent polarization even in the absence of an applied electrical field. As an example, let us consider the pyroelectric crystal $BaTiO_3$. Above the Curie temperature (T_c) of 130°C, $BaTiO_3$ is cubic – refer back to Figure 2.18. Since the positions of the positive and negative charges balance one another, there is no net polarization. However, as the crystal is cooled to temperatures below T_c, the unit cell becomes distorted into a tetragonal array where the Ti^{4+} ion is moved from its central position. This yields a finite polarization vector, resulting in ferroelectricity. It should be noted that all ferroelectric crystals are *piezoelectric* (see Chapter 2), but the reverse is not necessarily true.

In addition to being polarizable, it is also paramount that the dielectric medium be of low conductivity (*i.e.*, have a large bandgap – an offset with Si of >1 eV), to minimize carrier injection into its bands. For example, although the dielectric

constants for Ta_2O_5 and TiO_2 are sufficiently large for high-κ gate oxide applications (see below), both of these materials have small bandgaps (4.5 and 3.5 eV; relative to 8.9 eV for SiO_2), which will result in high leakage currents. On the other hand, more promising candidates include Al_2O_3, HfO_2, and ZrO_2, with bandgaps of 8.7, 5.7, and 7.8 eV, respectively. Other properties that must be inherent in the high-κ dielectric layer include thermodynamic and kinetic stabilities with the Si channel (especially under high-temperature processing conditions), and little/no electrically active defect sites.

The longevity of the dielectric material is dependent on the exposed voltage. When too high a voltage is introduced between the electrodes, dielectric breakdown of the solid may ensue. The solid-state structure, presence of impurities or microstructural defects (*e.g.*, cracks/voids), and ambient conditions (*e.g.*, temperature, humidity) are all paramount in assessing the *dielectric strength* (*i.e.*, resistance to breakdown) of an insulating material. As you might imagine, these factors become even more relevant when considering the 0.8–1 nm SiO_2 layers found in modern MOSFETs.

There are three types of high-κ dielectrics (measured dielectric constants in parentheses):

1. Materials with $4<\kappa<10$ (*i.e.*, relatively weak polarizability) such as: Al_2O_3 (7.5–10), Si_3N_4 (7.9–8.1), sodalime glass (7.2), $Al_6Si_2O_{13}$ (mullite, 6.3), $MgAl_2O_4$ (spinel, 8.6), $MgSiO_4$ (forsterite, 6.2), MgO (9.65), ZnO (8.15), BN (4.15), BeO (6.8), CaF_2 (6.8), $BaCO_3$ (8.5), $SrCO_3$ (8.85), ($KAl[Si_3O_{10}(OH)_2]$) (mica muscovite, 6.5–8.7), and organic polymers (polyurethane: 7.1, polyvinyl fluoride: 7.4, polyvinylidene fluoride: 6.4).
2. Materials with $10<\kappa<100$ such as: MnO_2 (pyrolussite, 12.8), ThO_2 (18.9), Ta_2O_5 (27.6), UO_2 (24), Nb_2O_5 (67), HfO_2 (25), ZrO_2 (12.5–24.7), $Zr(Hf)O_2$ (25), Y_2O_3 (15), La_2O_3 (25), Gd_2O_3 (23), Pr_2O_3 (25), and $BaZrO_3$ (43).
3. Materials with $\kappa > 100$ (*i.e.*, strongly polarizable) such as: TiO_2 (rutile, 85–170), $PbZrO_3$ (200), $Pb(Zr_{0.52}Ti_{0.48})O_3$ ("PZT," 1,800), $SrTiO_3$ (2,080), $Ba(Sr_{0.52}Ti_{0.48})O_3$ (3,000), $Pb(Ni_{0.33}Nb_{0.67})O_3$ (5,500), and $BaTiO_3$ (6,000).

Silicon (oxy)nitride dielectrics are touted as the near-term alternative for SiO_2. Though this would result in only a marginal increase in the dielectric constant relative to other high-κ candidates, the attractiveness of these films is related to their mature deposition techniques, and more desirable electronic barrier properties (*i.e.*, reduced boron diffusion). Not unlike its SiO_2 predecessor, Si_3N_4 films reach a functional limit of 1.2 nm; hence, there is ongoing research for the development of alternative gate oxide materials (and compatible gate stacks – for example, polysilicon gates are not suitable for ZrO_2 and HfO_2) that will be useful in the long-term to keep pace with ensuing transistor improvements.

Two factors are responsible for increasing the speed of transistor switching – the channel length between source and drain, and the speed of electron transport through the channel. Since it is becoming increasingly more difficult to shrink MOSFET dimensions, materials scientists are searching for new ways to address the latter consideration. In 2002, Intel first reported the use of "strained Si" in their 90 nm technology node, in which the silicon lattice in the channel region is strained through compression or expansion. It has been shown that electrons flow through

strained silicon 70% faster than in nonstrained silicon. For ICs, this translates to a 17% increase in speed at the same power, or 35% reduction in power consumption at the same speed.

To understand this effect, we need to consider the Si–Si bonding within the bulk crystalline solid. As we discussed earlier, electrons are promoted from valence to conduction bands due to thermal excitation. The valence band of the extended solid is formed from the overlap of sp^3 hybridized orbitals residing on each Si atom. When an electron migrates from valence to conduction bands under normal circumstances, there is no directional preference. However, when a strain is introduced along a specific direction of the lattice, the energies of the hybrid orbitals along this direction are altered.

A stretching, or tensile, strain will cause the orbitals between neighboring Si atoms to be expanded that weakens Si–Si bonds along the strain direction. Hence, it becomes easier for an electron to be thermally promoted from the valence to conduction band, corresponding to greater electron mobility. This effect is more pronounced for n-channel MOSFETs; the strain-expanded hybrid orbitals between neighboring Si atoms are able to accept electron density with less electron repulsion. By contrast, compressive strain is typically used for p-channel MOSFETs, where the presence of holes is able to offset the electron–electron repulsion within the compressed hybrid orbitals. More simplistically, crystal strain increases the free mean path of electrons (*i.e.*, less scattering through electron–lattice interactions), which increase their velocity under a given field.

Compression strain of p-doped Si is most often accomplished by carving trenches on opposite ends, and filling these with a material with greater lattice spacing than Si. This is typically a material such as $Si_{1-x}Ge_x$ ($x = 0.1$–0.3) solid solutions (Figure 4.20); since the lattice constants of Si and Ge are 5.43 and 5.66 Å, respectively, the lattice constant of the solution is $[5.43(1-x) + 5.66x$ Å$]$. A common method used to induce tensile strain is the deposition of a dense ceramic Si_3N_4 film on top of the channel. Both of these methods result in *uniaxial strain*, which is preferred since localized regions may be altered rather than the entire wafer. In fact, these techniques may be used to modify both nMOSFETs and pMOSFETs on a single chip. By contrast, *biaxial strain* is induced through the deposition of a Si thin film on top of a GeSi substrate that acts as an atomic template for film growth (Figure 4.21). The amount of strain and mobility enhancement depends on the germanium content of the SiGe layer.

Integrated circuit fabrication

From the dawn of civilization, no other field has progressed as rapidly as microelectronics. Gordon Moore, an integrated circuit pioneer and co-founder of Intel stated (in 1998): "If the auto industry advanced as rapidly as the semiconductor industry, a Rolls Royce would get a half a million miles per gallon, and it would be cheaper to throw it away than to park it." The computers we purchase today are almost immediately out-dated, due to constant improvements of microprocessor speeds and memory/storage abilities. A tiny integrated circuit (IC) lies beneath the protective packaging material of virtually any modern electronic device, acting as its nerve cen-

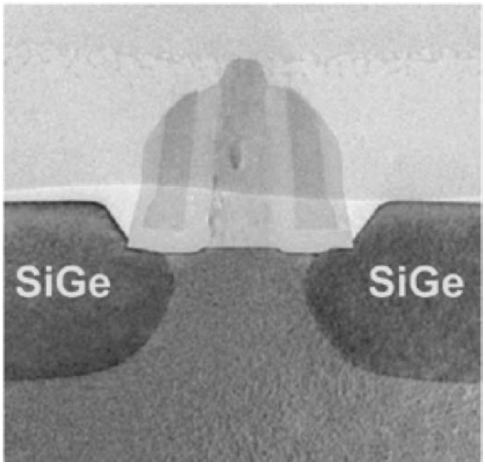


Figure 4.20. Cross-section HRTEM image of a transistor containing a compression-strained channel. Reproduced with permission from Intel Corporation (http://www.intel.com/research/downloads/Strained-Si-paper-IEDM-1203.pdf).

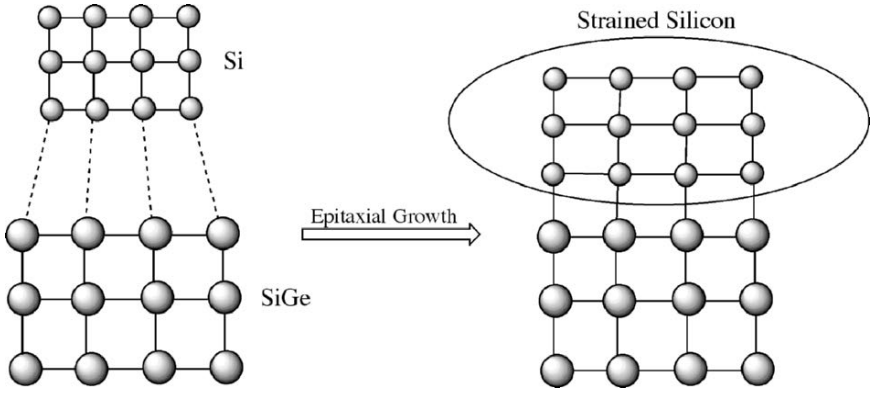


Figure 4.21. Formation of strained silicon from the epitaxial growth of lattice mismatched Si onto a SiGe substrate.

ter. Like any other advanced technology, we take these feats of modern engineering for granted. Beneath the familiar black plastic casing of a computer chip are billions of complex transistor subunits, along with additional components and circuitry, to offer precise control over the flow of electrons.

The major class of digital ICs is the CMOS. The most familiar CMOS applications include microprocessor and RAM chips. As its name implies, CMOS technology involves the complimentary operation of interconnected nMOSFET and pMOSFET pairs (Figure 4.22). Relative to NMOS or PMOS integrated circuitry, CMOS is the

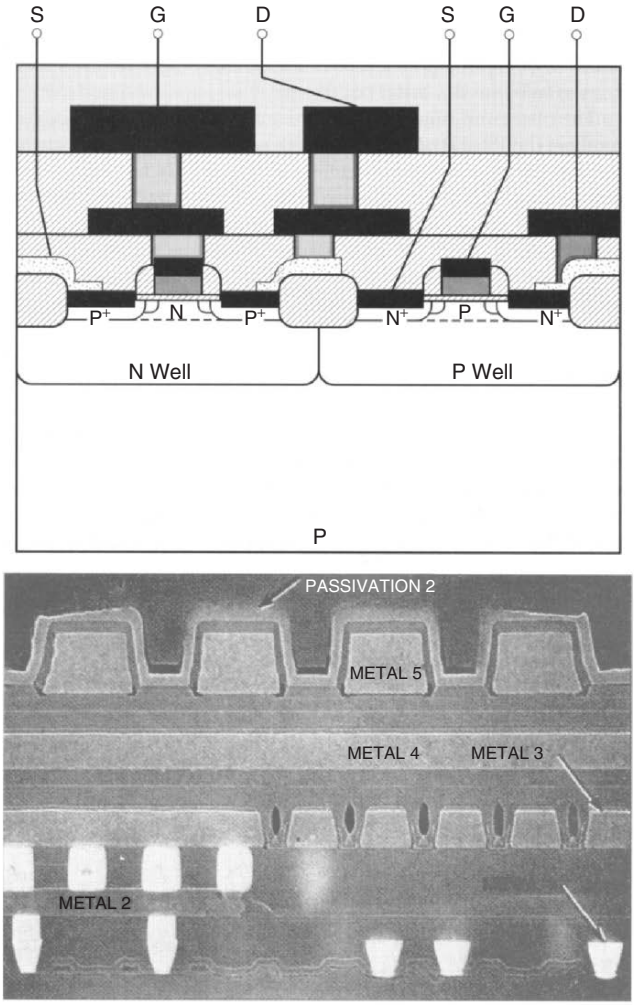


Figure 4.22. Cross-section illustration and SEM image of an integrated circuit. Reproduced with permission from Plummer, J. D.; Deal, M. D.; Griffin, P. B. *Silicon VLSI Technology*, Prentice-Hall: New York, 2000. SEM image Courtesy of Chipworks, Inc., Ottawa, Ontario, Canada.

design-of-choice for the majority of ICs due to distinct advantages in overall design simplicity and reduction in power.

There has also been an increased interest in SOI chips,[4] comprising transistors in which a thin insulating layer lies between a thin layer of Si (in the channel region) and the bulk Si substrate (revisit Figure 4.18). This technology is not new; in fact, SOI ICs have been used since the 1960s for military and space applications. As the relatively high cost of production continues to decrease, SOI substrates will undoubtedly be important for CMOS ICs within future commercial electronic devices. The insulation

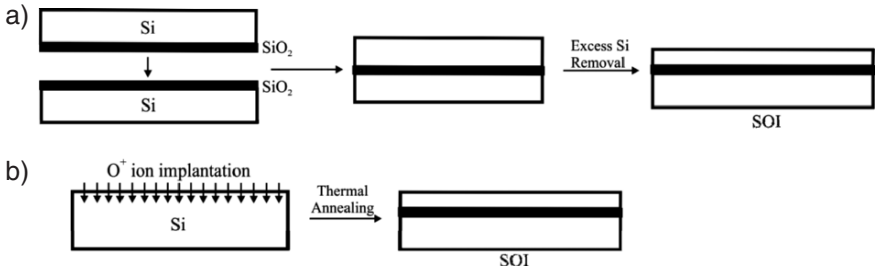


Figure 4.23. Two methods used to fabricate silicon-on-insulator (SOI) wafers. Illustrated are the (a) Smartcut procedure where two oxidized wafers are bonded together and (b) SIMOX procedure where oxygen ions are implanted into a bulk Si wafer.

between source and drain cavities allows the threshold voltage (also decreasing the supply voltage) of the SOI CMOS device to be extremely low, with minimal off-state (*i.e.*, subthreshold) leakage. It is reported that this simple substrate alteration results in chips that operate more than 30% faster, and use *ca.* 20% less power than standard CMOS analogs. The use of less power per chip is important for extending the battery life of portable electronics, as well as dealing with the heat generated by ICs. If one extrapolates the enormous heat generated from high-density bulk Si chips into the near-future, it may be necessary to use liquid coolants within the IC package – clearly not a desirable (or feasible) option.

Without question, the fabrication of modern ICs from grains of sand is one of the greatest feats ever accomplished. More than 300 carefully designed processing steps are used to form a dense packing of nanoscale objects onto a thin wafer of silicon. The silicon wafers used for IC fabrication are obtained through precise slicing of high-purity ingots using a diamond saw. The wafer edges are ground to a round finish polished until they have a defect-free mirrored finish. In order to reduce the cost/transistor, there has been a shift to larger wafers that may hold a greater number of ICs – from diameters of 200 to 300 mm (*i.e.*, 12″). Should a SOI substrate be desired, two methods are typically used for its fabrication (Figure 4.23). A bulk Si wafer may be exposed to high-energy beam of oxygen atoms that diffuse below the surface. Also, two wafers – a bulk Si and one coated with a SiO_2 film – are bonded together using specialized techniques.

Whether the substrate is bulk Si or SOI, the most critical step in IC fabrication is the initial cleaning of the wafer. The RCA clean procedure is the industry standard, consisting of the following complex treatment to remove both organic and inorganic contaminants. It should be noted that these procedures are performed within a clean room, to avoid particulate contamination:

(i) *SC-1 clean*
- The wafer is degreased using acetone, isopropyl alcohol, and ultrapure water (UPW) rinses.
- The native SiO_2 layer is removed from the surface by exposing the wafer to a 50:1 UPW:HF solution for 30 s, followed by a prolonged UPW rinse.

- The wafer is exposed to a 10:2:1 DIW:H_2O_2:NH_4OH solution at 75°C for 10 min, followed by a final UPW rinse. This effectively removes particulate contamination from the surface.
- nitrile gloves and Teflon tweezers are used to handle the wafer during treatments.

(ii) *SC-2 clean*

- Remaining surface oxides and hydroxides are removed by exposing the wafer to a 50:1 UPW:HF etching solution for 15 s, followed by a prolonged UPW rinse.
- The wafer is exposed to a 10:2:1 UPW:H_2O_2:H_2SO_4 solution at 75°C, followed by a UPW rinse. This effectively removes metals/ions from the surface.
- The wafer is rinsed with DIW for *ca*. 20 min., followed by drying under N_2 flow.
- When finished, the polished wafer surface should be reflective with no observable residues.

Since we are considering the starting wafer, it should be noted that often the substrate consists of an epitaxially grown doped Si film on the surface of a more heavily doped bulk Si wafer. If the film is of the same composition of the underlying substrate, this process is referred to as *homoepitaxy*, or simply *epi*. By contrast, *heteroepitaxy* refers to a Si film grown onto a different substrate such as sapphire (α-alumina). Using epi wafer substrates is generally more expensive than bulk analogues, but offers the advantage of fine-tuning the conductivity of the channel region for optimal operation of the CMOS IC. Through careful film deposition methodology, it is also possible to eliminate O and C impurities resulting in further improvements in device performance.

Regardless of bulk, epi, or SOI substrates, a key consideration that is paramount toward subsequent processing steps is the crystal orientation employed. Three important crystal planes for silicon include Si(100), Si(110), and Si(111) – Figure 4.24. The surface atomic densities increase in the order Si(100) < Si(110) < Si(111). Empirically, this translates to available Si–Si bond densities of 6.77×10^{14}, 9.59×10^{14}, and 11.76×10^{14} cm^{-2}, respectively.[5] Hence, the rates required to remove (etch), or react with, surface atoms (*e.g.*, thin-film deposition) should follow the reverse order as above. However, the Si(110) orientation etches fastest due to its more

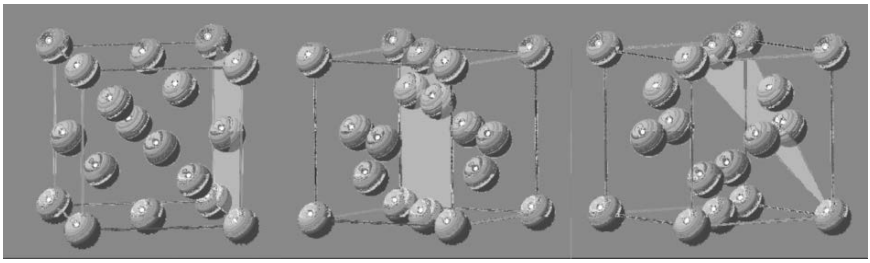


Figure 4.24. Unit cell representations of silicon showing the (100), (110), and (111) planes, respectively.

corrugated structure, relative to the other atomically flat surfaces. To illustrate this effect, exposure of a wafer to a 40% alcoholic KOH solution results in Si etching at rates of 13,000, 6,000, and 90 Å min^{-1}, for the Si(110), Si(100), and Si(111) planes, respectively.

The fabrication of a CMOS IC involves a repeating sequence of film deposition, patterning, and dopant implantation procedures (Figure 4.25). The initial step involves the deposition of SiO_2 onto the wafer surface. You may be wondering why this is necessary, since the cleaning procedures were designed to remove the native silica layer. However, this procedure involves the controlled deposition of an amorphous SiO_2 coating of known thickness onto the wafer surface.[6] By definition, such a coating is referred to as a *thin film*, since its thickness is less than 1 μm. Two reactions are responsible for the formation of SiO_2 (Eqs. 7 and 8; dry and wet oxidation, respectively). A variety of factors govern the resultant thickness of the silica film, such as temperature and pressure/flowrate of O_2.

$$(7) \qquad Si + O_2 \xrightarrow{700-1{,}200^\circ C} SiO_2$$

$$(8) \qquad Si + 2H_2O \xrightarrow{1{,}100^\circ C} SiO_2 + 2H_2$$

It should be noted that SiO_2 deposition does not occur through simple addition to the wafer surface. Rather, as the surface becomes oxidized, oxygen atoms are also able to diffuse from the surface to react with underlying Si atoms. The incoming O_2 or H_2O molecules adsorb dissociatively to the Si surface, resulting in the O/OH groups being either attached to dangling Si bonds, or inserted between Si–Si bonds of the same layer. As the temperature is increased, there is a greater degree of oxygen insertion into Si–Si bonds of underlying layers, which adds to the thickness of the SiO_2 film. A chlorine containing coreactant gas such as HCl or C_2HCl_3 is often present during Si oxidation to stabilize the Si/SiO_2 interface. It has been shown that small amounts (<5%) of chlorine in the gas phase helps to tie up sodium or potassium ions that may be present as contaminants, as well as strengthening the growing oxide film by preventing oxidation-induced stacking faults.[7]

Following oxidation, it has been shown that 0.001% of the silicon atoms in the region immediately surrounding the Si/SiO_2 interface remain unsaturated. This is likely an artifact of incomplete oxidation during the transition from Si to SiO_2. The presence of these defects results in a layer of positive charge that is located within ±2 nm from the interface, which may affect the V_T of the MOSFET, and performance of the IC as a whole. Although these effects are minimized by using Si(100) substrates, additional processing steps also address this issue; for instance, passivation through high-temperature annealing in the presence of hydrogen.

Largely due to the desirable electrical properties of the $Si(100)/SiO_2$ interface, the majority of ICs utilize a Si(100) substrate. That is, since the Si(100) surface is the least dense, there will be a relatively low number of coordinatively unsaturated (*i.e.*, dangling) Si bonds at the surface. As you might imagine, the presence of dangling bonds on a surface greatly affects its reactivity. One way to remove these defects is through *passivation*. For example, a large number of dangling bonds are tied up by H atoms during HF etching to remove native SiO_2 (the RCA clean procedure above).

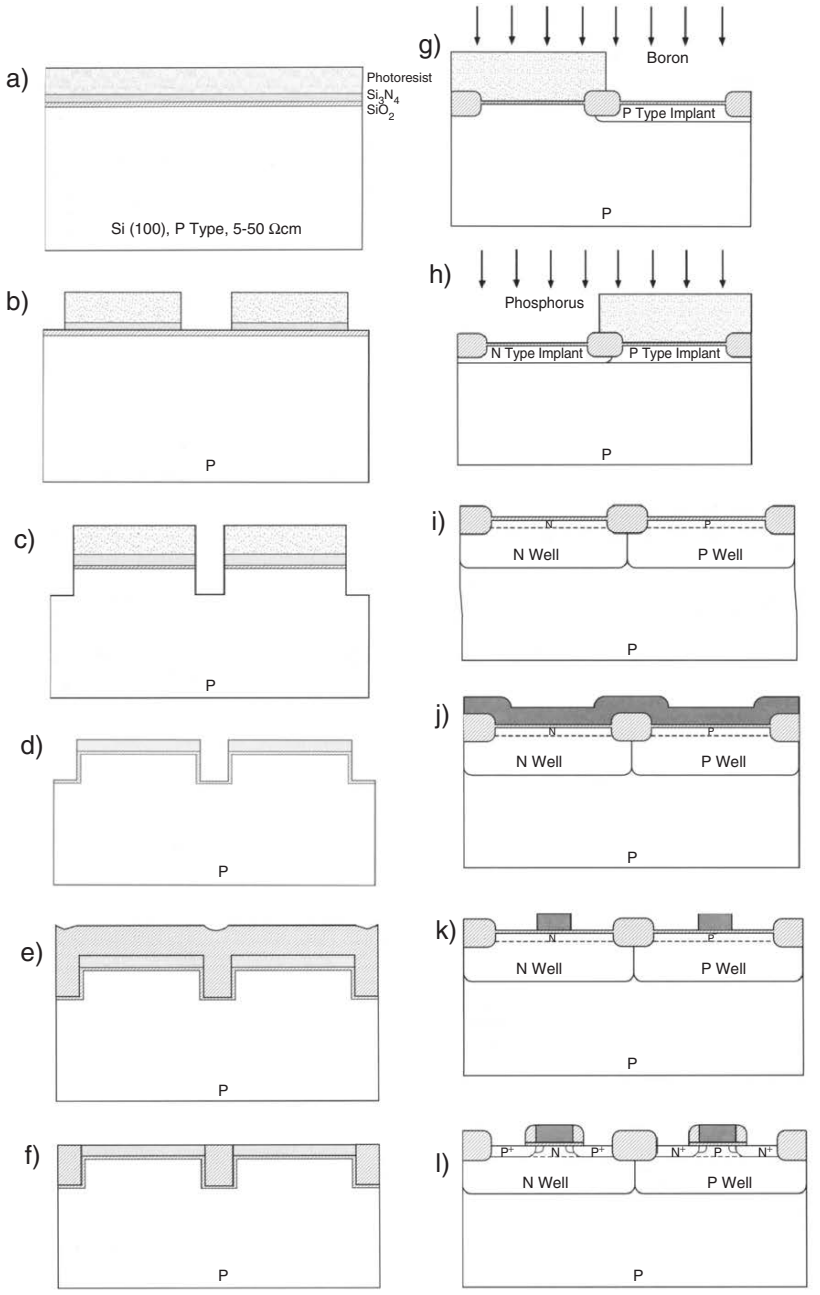


Figure 4.25. The steps involved in fabricating a CMOS IC. Reproduced with permission from Plummer, J. D.; Deal, M. D.; Griffin, P. B. *Silicon VLSI Technology*, Prentice-Hall: New York, 2000.

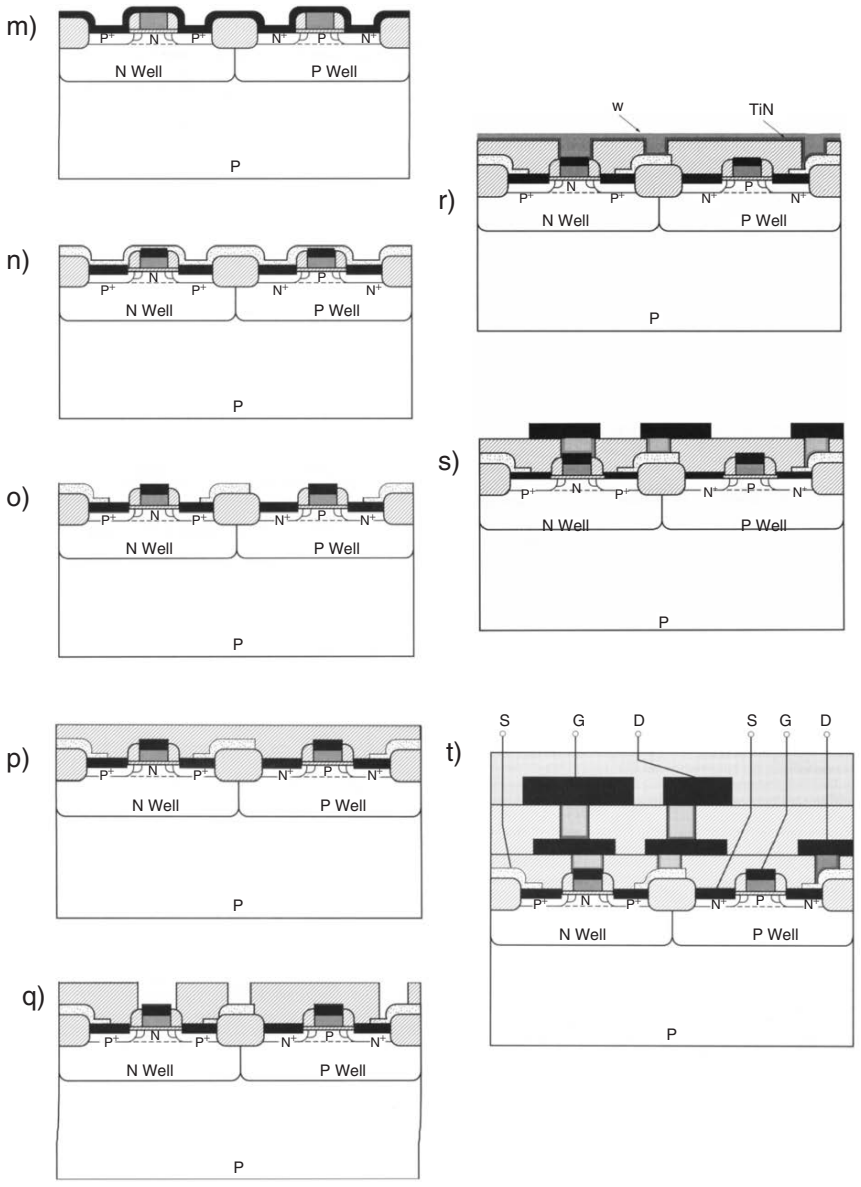


Figure 4.25. Continued

In addition, neighboring unsaturated Si atoms may recombine in a number of ways (*surface reconstruction*), which minimizes the overall energy of the surface.

Also shown in step (a) of Figure 4.25 is the deposition of silicon nitride, Si_3N_4 over the SiO_2 layer. The nitride film is generated through a high-temperature reaction between ammonia and silane gases (Eq. 9). The thicknesses of the SiO_2 and Si_3N_4

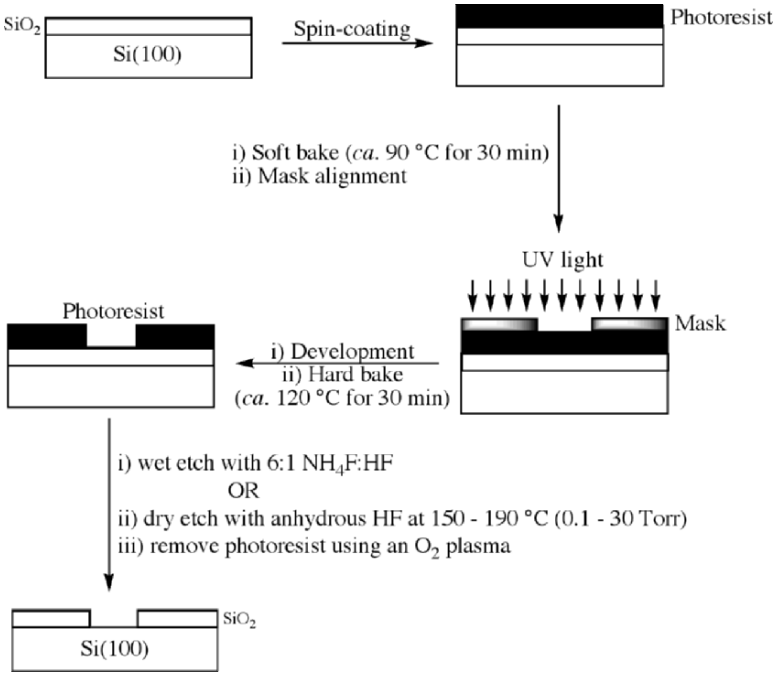


Figure 4.26. Schematic of photolithography and the chemical ("wet") and "dry" methods used to etch patterns in an oxidized Si wafer.

films are typically on the order of 15 and 75 nm, respectively. Since the density of silicon nitride is sufficiently greater than SiO_2 (3.3 g cm^{-3} *vs.* 2.6 g cm^{-3}, respectively), Si_3N_4 is used as an effective passivating layer to prevent oxidation of underlying SiO_2 and Si regions.

$$(9) \qquad 3SiH_4 + 4NH_3 \xrightarrow{800°C} Si_3N_4 + 12H_2$$

The patterning steps are based on *photolithography* (Figure 4.26). In this procedure, a photosensitive compound known as a *photoresist* is first spin-coated onto the surface of the wafer. Frequently, the wafer is pretreated by a "dehydration bake" (to remove adsorbed water) and application of an adhesion promotor (*e.g.*, hexamethyl disilazane, HMDS – used to maintain a strong interaction between the photoresist and SiO_2 surface – Figure 4.27).

A circuit pattern of the microscopic IC known as a *mask* is placed over the wafer, and the uncovered molecules that are exposed to high-energy UV light exhibit a chemical change (Figure 4.28). *Negative tone* photoresists undergo photoinduced crosslinking mechanisms upon UV exposure. The ensuing polymerization reaction is due to the presence of photosensitive groups such as epoxy, vinyl, or aryl halides on the photoresist backbone. Accordingly, they become insoluble during subsequent contact with a basic developing solution (*e.g.*, MOH, tetramethyl

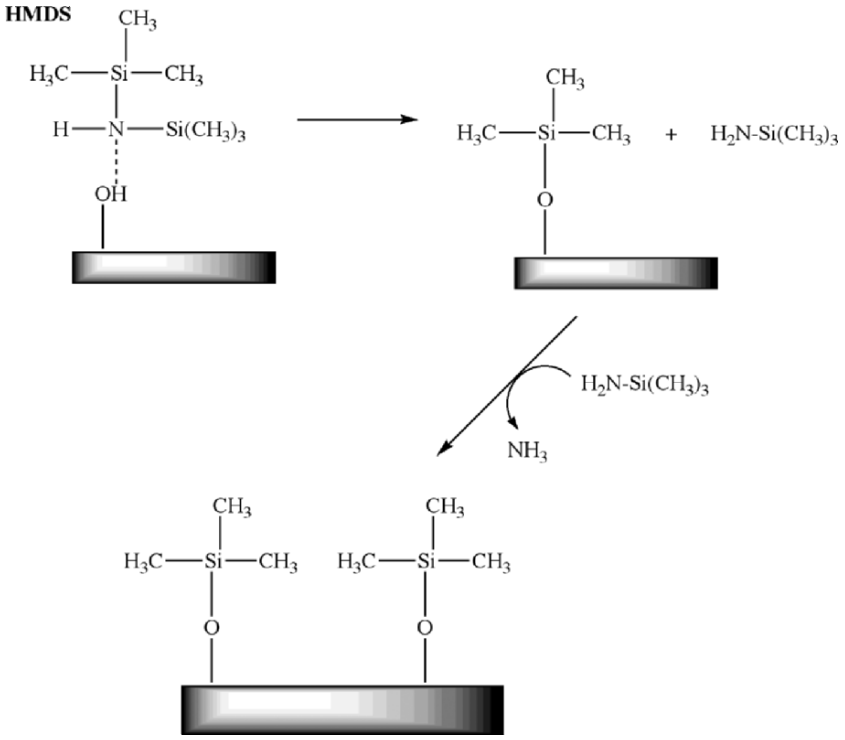


Figure 4.27. The reactions involved in the adhesion of HMDS to a SiO_2 surface. Attachment occurs through available surface hydroxyl groups, with the release of ammonia.

ammonium hydroxide). In contrast, use of *positive tone* photoresists results in the exposed-polymer regions becoming preferentially soluble in the developing solution. Although early photolithographic applications used negative tone photoresists exclusively, the organic matrices of these materials caused swelling, which results in pattern distortion during development. In order to reproduce line features below 3 μm present in today's electronic devices, aqueous base-soluble positive photoresists are most commonly employed. However, aqueous negative tone photoresists have now been developed with line resolutions below 1 μm.

In order to improve the line resolution of photoresists, a technique known as chemical amplification (CAM) is often used for photolithographic applications. The enhancement of quantum efficiency and sensitivity results from a photogenerated acid that will either catalyze a crosslinking (negative photoresists), or deprotection (positive photoresists) mechanism. To facilitate water solubility, modern photoacid generators (PAGs) are typically aryl sulfonium triflate salts (Figure 4.28). Compounds such as *o*-nitrobenzyl tosylate or organic/inorganic onium salts are commonly used as PAGs within organic solvents.

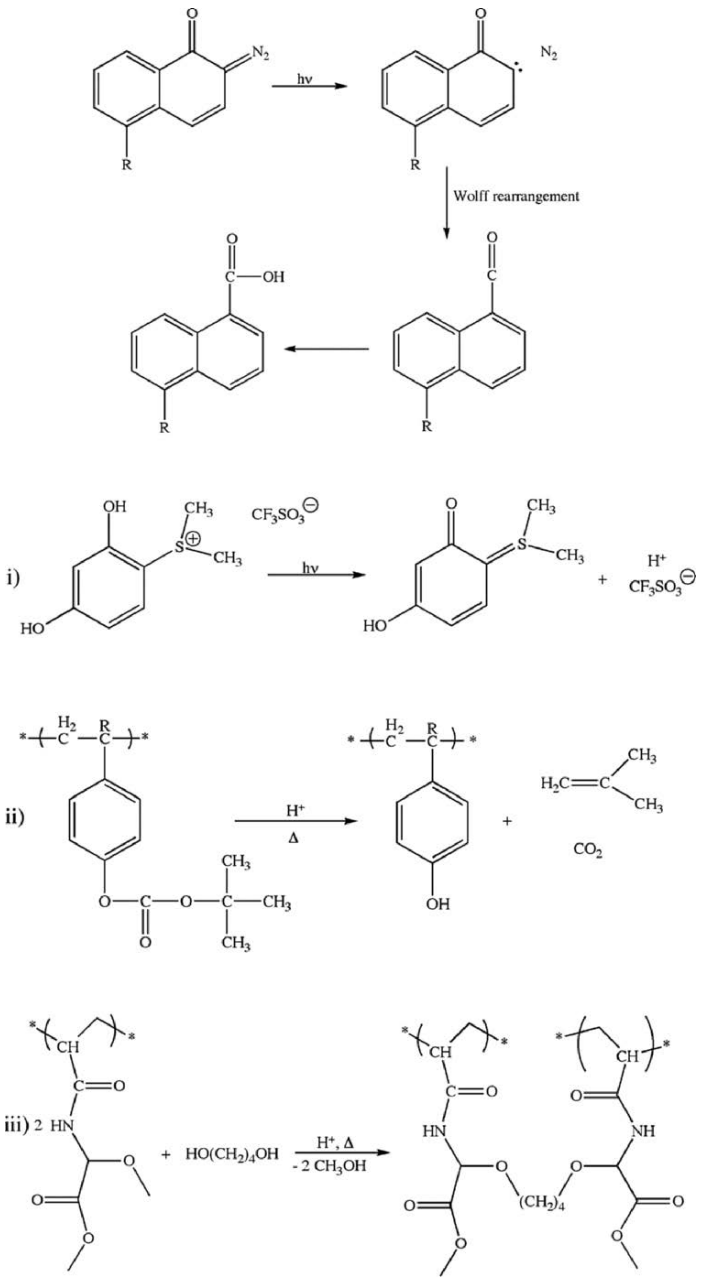


Figure 4.28. Molecular structures and photoinduced reactions of common photoresists. Shown (top) is the positive tone resist containing the active diazonapthoquinone (DNQ) chromophore group. Chemical amplification (CAM) reactions are illustrated in (i)–(iii). Reaction (i) represents photoinduced acid generation; step (ii) is an acid-catalyzed deprotection mechanism (positive tone resist); and step (iii) is an acid-catalyzed crosslinking mechanism (negative tone resist).

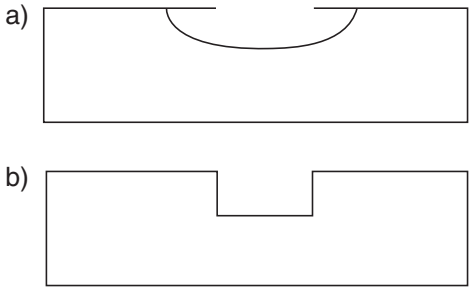


Figure 4.29. Illustration of (a) isotropic etching evidenced by severe undercutting and (b) anisotropic etching with 90° sidewall channels (or *vias*).

Once the pattern in the photoresist layer is formed, a variety of chemical or high-energy plasma techniques may be used to remove exposed underlying layer(s) – a process referred to as *etching*. It should be noted that "wet" etching, or dipping the wafer in an acidic solution, typically results in *isotropic etching* (Figure 4.29a), and a line resolution of $>2\,\mu m$. By contrast, "dry" etching which uses a plasma [8] generally results in *anisotropic* conditions (Figure 4.29b), and line resolutions $<0.25\,\mu m$. It has been shown that the pre-application of an adhesion promotor results in greater anisotropic etching (less undercutting), since the photoresist and wafer surface are in closer contact with one another.

The specific etching conditions used will depend on the surface to be removed. For instance, Figure 4.26 lists the wet and dry etching conditions to preferentially remove SiO_2 from a Si surface. Other plasmas used for selective SiO_2 removal include CF_4/H_2 and C_2F_6/CHF_3. In order to remove Si rather than SiO_2, one would use either a wet etch containing HNO_3, HF, and CH_3COOH (or KOH solutions), or a dry etch using plasmas such as CF_4/O_2, SF_6/O_2, C_2F_6/O_2, or Cl_2/C_2F_6 (for n-type Si; Cl_2/Argon is used for undoped Si). The addition of oxygen to a fluorocarbon plasma increases the amount of free F through the formation of oxyfluorides from fluorocarbons (Eq. 10).[9] However, if more than 10% O_2 is added to the plasma, a decrease in the Si etching rate will result. Most likely, this is due to the competition between O and F atoms for active sites on the silicon surface. Plasma systems used to remove Si_3N_4 layers include CF_4/H_2 or NF_3/Cl_2 combinations (Eq. 11). It should be noted that a O_2/H_2O plasma and/or high-temperature annealing step is typically used to remove the photoresist that remains following the wet/dry etching.

$$(10)\quad CF_4 + O \xrightarrow{\text{plasma}} COF_2 + 2F$$

$$(11)\quad Si_3N_4 + 12F \xrightarrow{\text{plasma}} 3SiF_4 + 2N_2$$

In current semiconductor fabrication, photolithography with 193 nm UV irradiation is able to pattern features that are 37 nm wide. However, the rapid miniaturization of ICs demands improvements in photolithography in order to sharpen the line resolution at size regimes far below 37 nm (Figure 4.30). Although *immersion lithography* has been suggested as a candidate for sub-50-nm line resolution, there

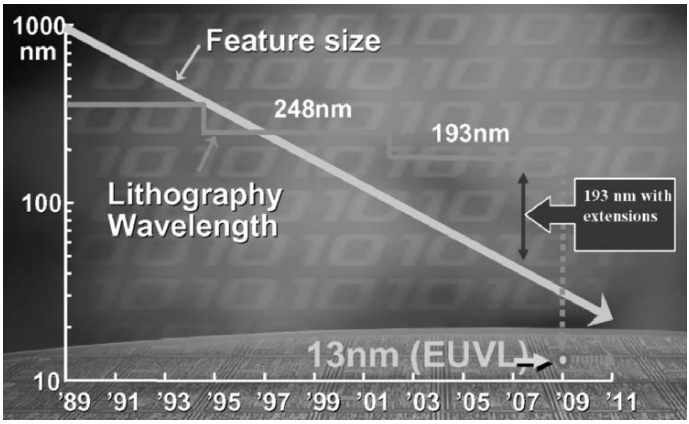


Figure 4.30. The trend of shrinking feature size, which is outpacing the decrease in wavelength currently used for photolithography. Reproduced with permission from Intel Corporation (http://www.intel.com).

are significant technical challenges that must first be addressed. As its name implies, this technique consists of surrounding the lithographic lens with a fluid of greater refractive index (*e.g.*, water: $\eta = 1.47$ at 193 nm; air: $\eta = 1.0$). Since the refractive index of the medium is directly proportional to the aperature (NA), an improvement is made to the imaging resolution (Eq. 12).

$$R = \frac{0.61\lambda}{\mathrm{NA}} \tag{12}$$

Where R is the resolution limit and NA is the numerical aperature of the exposure tool.

Also evident from Eq. 12 is that decreasing the wavelength of exposure will also improve the resulting line resolution. Table 4.1 lists the wavelengths and optimum resolution values for various lithographic techniques. Perhaps the best opportunity to image small features is the use of extreme UV light (EUV), which is actually in the range of soft X-rays (*ca.* $\lambda = 13.5$ nm). A decrease in exposure wavelength translates to new types of photoresists that will be stable upon contact with higher energy radiation. Whereas the DNQ photoresist system is effective in the MUV range, photoresists that contain CAM moieties are used exclusively for DUV and shorter wavelengths.

So far, we have progressed through steps (a)–(c) of the stepwise CMOS IC process (Figure 4.25), with patterning and sequential selective etching to yield a channel (referred to as a *via*) through both the protective oxide/nitride layers and the underlying Si substrate. The resulting trench will be eventually back-filled with SiO_2 (Figure 4.25, step e), which serves to electrically isolate individual devices on the chip from one another. This process, referred to as *shallow trench isolation* (STI), is the method-of-choice for modern CMOS fabrication since it allows the deposition of smaller insulating regions resulting in greater device density, in accord with Moore's Law.

Table 4.1. Resolution Limits for Various Lithographic Techniques

Lithographic technique	Exposure wavelength (nm)	Resolution
Photolithography[a] mid-UV (MUV)	350–450	0.35–3 μm
Photolithography deep-UV (DUV)	248	≥0.25 μm
Photolithography deep-UV (DUV)	193	<70 nm[b]
Photolithography extreme-UV (EUV)	13.5	<65 nm
Electron-beam lithography[c]	*ca.* 1	<50 nm
X-ray lithography[d]	0.4–20	<50 nm

[a]Using a standard chrome-on-glass photomask.

[b]Using advanced photomask techniques such as phase-shift masks (PSMs, Figure 4.32).

[c]Using a photomask of Si_3N_4 membrane and Cr/W patterned regions.

[d]Using a photomask of Si/Si_3N_4/SiC/BN membrane and Au/W patterned regions.

Step (d) of Figure 4.25 illustrates the deposition of a thin film (*ca.* 10–20 nm) of SiO_2 onto the trench sidewalls and bottom. This is used to establish an effective SiO_2/Si interface prior to the deposition of the thick (*ca.* 500 nm–1 μm) back-filled SiO_2 layer. That is, to ensure that the SiO_2 conforms to the entire channel surface, leaving no voidspaces that would decrease the effectiveness of the isolation trench. Once the thick oxide is deposited into the trench, *chemical mechanical polishing* (CMP) is used for *planarization* of the surface prior to subsequent layering steps (Figure 4.25, step f). The silicon nitride layer acts a polishing stop, and is subsequently removed through plasma etching.

The next steps (Figure 4.25g–i) involve the placement of n- and p-type dopants onto the top of the silicon surface. A technique known as *ion implantation* involves the acceleration of B (p-type) or Sb/As/P (n-type) ions to very high kinetic energies (*ca.* 50–200 keV), which then collide with the Si lattice. The chosen energy must be sufficient to penetrate through the SiO_2 layer and a desired depth of Si substrate, but not the exposed photoresist layer that defines the dopant region. For the same accelerating voltage, ions of smaller mass will penetrate to deeper/broader regions of the solid relative to heavier ions (Figure 4.31), due to differences in their kinetic energies and concomitant collisions with Si atoms.

An energy on the order of 15 eV is enough to dislodge a Si atom from its crystalline lattice, forming a Frenkel defect. However, the energies of the impinging dopant ions are an order of magnitude larger, which will cause significant damage to the Si crystal structure. For instance, bombardment of a Si surface with a single arsenic ion with an energy of 30 keV will displace 1,000 Si atoms – all on a timescale of *ca.* 10^{-13} s! Multiply this by thousands of bombarding ions during the implantation process, and one has a picture of a highly perturbed Si crystal. Amazingly, thermal annealing is able to cause the dislodged Si atoms to find vacant sites, removing the Frenckel defects and thus repair the damage. Prolonged high-temperature annealing at temperatures of 1,000–1,100°C is subsequently used to further restore the periodicity of the Si crystal lattice, and diffuse the n- and p-dopant wells (*i.e.*, the substrates for pMOS and nMOS devices, respectively) to their desired Si depths of *ca.* 2 μm. It should be

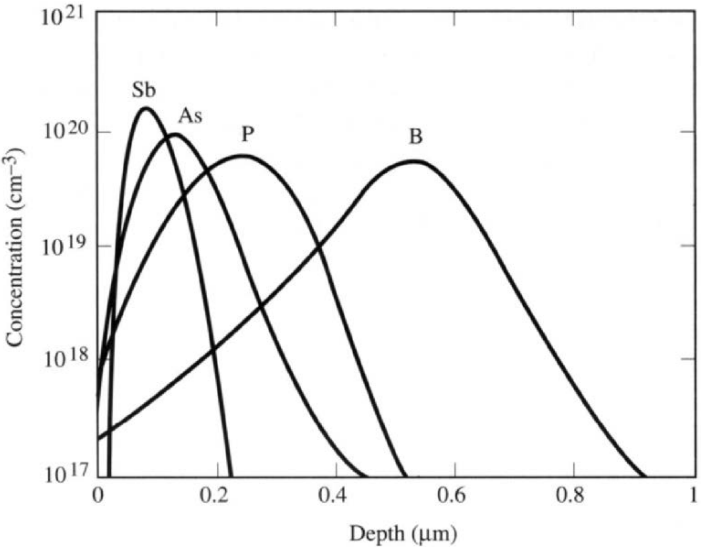


Figure 4.31. Comparison of the Si depth penetration of various n- and p-dopant ions. Reproduced with permission from Plummer, J. D.; Deal, M. D.; Griffin, P. B. *Silicon VLSI Technology*, Prentice-Hall: New York, 2000.

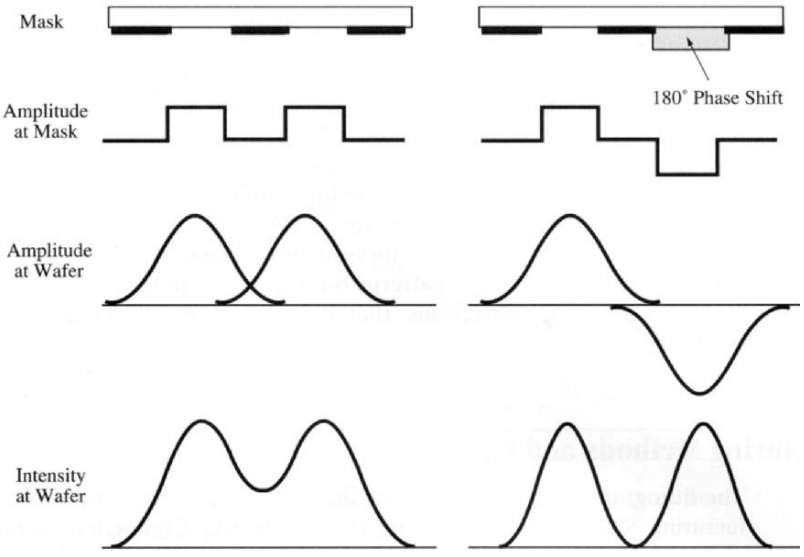


Figure 4.32. The benefits of using phase-shift masks for photolithography. Reproduced with permission from Plummer, J. D.; Deal, M. D.; Griffin, P. B. *Silicon VLSI Technology*, Prentice-Hall: New York, 2000.

noted that the above ion bombardment processes are repeated using an appropriate masking photoresist to introduce a higher concentration of dopants near the surface of the n and p wells. This is done to adjust the threshold voltages (V_T) of the nMOS and pMOS devices (Figure 4.25, step i).

Steps j–k of Figure 4.25 illustrate the formation of the gate electrodes onto the nMOS and pMOS devices. This consists of the deposition of a *ca.* 500 nm thick polysilicon film onto the wafer surface, and subsequent photolithography/etching steps. The specific methodologies used for thin-film deposition will be described in the next section of this chapter. The source and drain of the complimentary MOS devices are shown in step l; this is performed through careful ion bombardment using an appropriate photolithographic mask. Also shown in step l is the deposition of insulating SiO_2 on the sidewalls of the polysilicon gate.

For memory chips and removable flash drives, a gate stack structure known as *SONOS* (polysilicon/SiO_2/Si_3N_4/SiO_2/Si substrate) are typically used. The device is programmed by applying a voltage to the gate to inject charge into the conduction band of the Si_3N_4 layer, resulting in a change in the threshold voltage. When the voltage is removed, the charge remains trapped within the Si_3N_4 layer. High-κ metal oxides such as ZrO_2 and HfO_2 are also capable of trapping injected electrons; if these replace Si_3N_4, the memory device is known as SOMOS (where M refers to metal oxide). The charge-trapping ability is most pronounced for materials that exhibit a high dielectric constant (*i.e.*, are nonconducting and polarizable), and have sufficient density to prevent tunneling to surrounding layers. To erase the device, a reverse bias voltage is applied that removes the trapped charge from the Si_3N_4 region. Typically, the SiO_2 adjacent to the polysilicon gate is thicker than the SiO_2/Si interface, resulting in charge injection occurring through the bottom SiO_2 layer.

Now that the source/drain and gate electrodes have been formed for nMOS and pMOS complimentary devices, the remaining steps involve *metalation* – the selective deposition of metals that form the interconnection between the active devices on the IC. Step m of Figure 4.25 shows the deposition of titanium onto the IC surface, followed by annealing at a temperature of 600°C under a nitrogen environment. Although Ti will readily react with the ambient N_2 to form TiN, titanium will also readily react with Si atoms to form $TiSi_2$ in the regions where a Ti/Si interface exists (at the top of the polysilicon gate, and source/drain cavities). The dark regions illustrated in step n of Figure 4.25 indicate the regions of $TiSi_2$, with TiN formed in all other areas. Metal silicides are used to reduce the resistivity of the polysilicon gate; in addition to $TiSi_2$, other analogues such as WSi_2, and $CoSi_2$ are frequently utilized. The exposed TiN is subsequently patterned using a photolithographic mask, and is etched using a 1:1:5 NH_4OH:H_2O_2 : H_2O wet etching solution (step o).

A thick layer (*ca.* 1 μm) of insulating oxide is then deposited onto the evolving IC, which serves as the interconnect dielectric medium. Although SiO_2 was typically used for this purpose, the latest ICs utilize a high-κ dielectric such as carbon-doped SiO_2 (Figure 4.33) that is reported to decrease the capacitance by up to 20% relative to undoped media. This improvement translates to higher performance at lower power consumption – paramount for high-density modern ICs. Following planarization of the oxide layer (Figure 4.25, step p), selected regions are patterned/etched

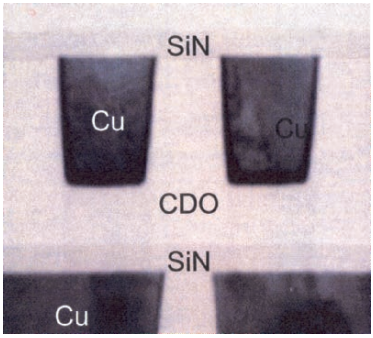


Figure 4.33. Cross-section TEM image of the placement of a carbon-doped oxide (CDO) dielectric between metal interconnects. Reproduced with permission from Intel Corporation (http://www.intel.com).

using our workhorse of photolithography (Figure 4.25, step q). The etched regions are then back-filled with another metal (usually W or Co); however, to improve the adhesion and electrical properties of the interface, a thin buffer layer of TiN is first deposited (Figure 4.25, step r). Subsequent planarization results in a flat surface containing isolated regions of tungsten.

The final steps involve deposition of the interconnect metal (Figure 4.25, step s). Copper is now the metal-of-choice due to its more desirable electrical resistivity, relative to Al (1.7 μΩ cm *vs.* 2.7 μΩ cm, respectively) that was exclusively used in earlier ICs. Due to its low resistivity and high density, titanium nitride is an efficient barrier level that prevents surface oxidation of Cu, as well as the interdiffusion of Cu into adjacent layers. To yield the final multilayer IC shown in step t of Figure 4.25, steps p–s are repeated. Indeed, a long complex process that took weeks in the making.

Perhaps a topic that has been overshadowed by the previous detailed discussion of high-κ dielectrics is the use of materials at the other end of the capacitance spectrum: low-κ dielectrics. In ICs, insulating dielectrics separate the conductive portions (*e.g.*, wire interconnects and individual transistors) from one another. In order to fabricate chips with higher speeds, the transistors must be placed closer and closer together, thus resulting in a thinner insulating layer. This leads to charge buildup and crosstalk, which adversely affects the maximum operating speed and performance of the chip. The use of low-κ dielectrics is important to reduce the parasitic capacitance, hence enabling faster switching speeds and lower heat dissipation. Doping SiO_2 with fluorine reduces the κ from 3.9 to 3.5; other approaches that yield values of $\kappa < 1.5$ involve the use of nanoporous dielectrics based on poly(amidoamine-organosilicon) (PAMAMOS) films (see Chapter 5).

One has to remember that these multistep processes were not performed from start to finish, one IC at a time. If that were the case, each chip would cost millions of dollars! Instead, a large polished wafer that is currently 300 mm in diameter is used as the base to assemble hundreds of ICs (called *dies* when they are perforated from the large wafer). It is easy to see why the semiconductor industry has shifted to larger-diameter wafers; though the fabrication facility (known as a *fab*) costs

billions of dollars to establish, the price/chip is miniscule since one is essentially able to assemble hundreds of chips at once (*ca.* 700 chips/wafer using 300 mm wafers (Eq. 13[10] – an increase of over 200% relative to 200 mm). It has been proposed that the industry will shift to 450 mm wafers in the near future; however, this will dramatically increase the costs associated with single-crystal ingot processing.

$$\text{Dies/wafer} = \frac{\pi(\text{wafer diameter}/2)^2}{\text{die area}} - \frac{\pi(\text{wafer diameter})}{\sqrt{2(\text{die area})}} \quad (13)$$

It is interesting to note that although each wafer goes through the entire CMOS fabrication pathway, all of the completed chips/dies may not be equal. That is, microprocessor chips running at 700 and 900 MHz may have been produced on the same wafer! From our detailed discussion of CMOS fabrication, it is not hard to see how neighboring chips may have slight variations in layer thicknesses/composition, contaminants, *etc*. that will significantly alter their ultimate performance. As you might expect, it is typical for companies to sell the fastest of the fabricated chips at a premium price, and decrease the price accordingly for slower ones.

Thin-film deposition methodologies

The previous section described the deposition of a number of films onto the growing IC – from insulating SiO_2 layers to the copper interconnects. As you might imagine, a number of growth strategies are employed to yield the most desirable films for the particular application. In IC fabrication, individual layers are deposited with varying levels of thicknesses. For instance, relatively thick layers such as photoresist and interconnect dielectrics may be deposited with a higher variability in film thickness. However, for layers such as gate oxide and $TiSi_2$, techniques that are able to deposit films a monolayer at a time are required.

In addition to the application of adhesion promotors and photoresists during IC fabrication, "bulk" deposition techniques such as dip- or spin-coating are commonly used for decorative and/or protective coating applications. Films of organic or inorganic (e.g., sol–gels) materials are also spin-coated onto a desired substrate, air-dried to remove the solvent, and postannealed (if desired) to yield the appropriate morphology/porosity of the final film.[11] It is possible to control the film thickness during spin-coating through varying the solvent, spin rate, drop height, *etc*. However, this technique is not suitable for the growth of thin films where control over film-thickness homogeneity, morphology, composition, conformality, and selectivity are paramount to resultant performance.

As we mentioned in the Introduction, the "bottom-up" approach to materials design, or building the structure one molecule/atom at a time, provides the ultimate in control over the final properties of the material. For thin-film growth, this corresponds to *vapor deposition* techniques, rather than the "top-down" approaches of dip- and spin-coating.

Vapor deposition techniques feature the introduction of gaseous molecular/atomic subunits that self-assemble on the surface of the substrate to yield the desired film. The rate of deposition is on the order of Å min^{-1}, which allows for intimate control over the properties of the growing film. There are two types of vapor

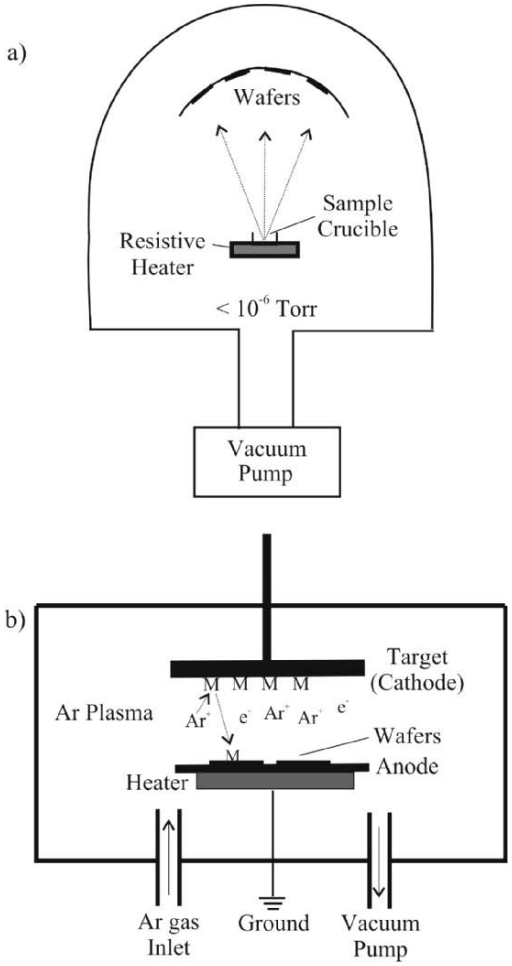


Figure 4.34. Schematic of physical vapor deposition apparati. Shown are: (a) an evaporation system and (b) a sputtering system.

deposition methods: *physical vapor deposition* (PVD) and *chemical vapor deposition* (CVD). Both methods may be used to grow thin films of metals, alloys, oxides, nitrides, carbides, silicon, or amorphous/graphitic carbon. PVD may occur through the evaporation of atoms/molecules from a precursor solid *in vacuo* (*evaporation*, Figure 4.34a), or through use of a high-energy Ar plasma source that causes the vaporization of atoms from a solid target (*sputtering*, Figure 4.34b). Since all atoms of the solid in both techniques will enter the gas phase, the resultant film is only as pure as the solid precursor that is used. As a result, high-purity precursors must be used for PVD; for example, a piece of gold foil of purity $>99.99999\%$ is typically used as a target for Au sputtering.

In general, coatings produced by the PVD process are hard, with a high atomic density due to slow and efficient nucleation/growth. Depending on the exposure time of the substrate with the plasma, the thickness of PVD coatings ranges from a few angstroms to $>30\,\mu m$. Since the substrate is maintained at room temperature, there are no limitations related to the thermal stability of the substrate. This is an important consideration for coating materials that will decompose at high temperatures such as plastics, paper, *etc*.

Two methods for the evaporation of precursors may be employed – resistance heating and electron beam collision. The first method employs a simple alumina crucible that is heated by a W filament. Temperatures as high as 1,800°C may be reached inside the chamber, which is enough for some metals or metal salts to vaporize. Deposition rates for this method are 1–20 Å s^{-1}. The use of an electron beam to assist in the precursor evaporation results in temperatures on the order of 3,000°C, being more suited for the deposition of refractory metals/alloys and metal oxides such as alumina, titania, and zirconia. Since the temperature of the chamber interior is much higher than the walls, the gas-phase ions/atoms/molecules condense on the sidewalls as well as the substrate; this may lead to film contamination as the nonselective coating flakes off the chamber walls.

For high-purity metal or carbon films, an ultrahigh vacuum (UHV) environment (typically $< 1 \times 10^{-6}$ Torr) must be used during PVD. This is necessary in order to prevent the gas-phase reaction of metal/carbon atoms with atmospheric gases (*e.g.*, H_2O, O_2, N_2) that would preferentially form metal oxides, hydroxides, or nitrides rather than the desired film. If mixed phases such as nitrides or oxides are desired, purified nitrogen or oxygen is introduced into the chamber, respectively. For carbide films, targets of both the metal and carbon are placed together within the vacuum chamber. As one would expect, since such a high-purity metal and high vacuum chamber must be used, PVD is relatively quite expensive. Another limitation is the selectivity and conformality of the procedure. While PVD works well to deposit material on surfaces in line-of-sight of the source, nonconformal deposition of complex or rough surfaces (*e.g.*, fibers) is a critical limitation. This issue becomes more important for integrated circuits with feature sizes less than 100 nm, often with high *aspect ratios* (height/width ratio of the surface feature, Figure 4.35).

By contrast, CVD is a process in which gaseous precursors are *reactively* transformed into a thin film, coating or other solid-state material on the surface of a catalyst or substrate. It should be stressed that CVD is no longer limited to thin film growth; this method is now the preferred route to generating fiber-optic preforms,[12] as well as an increasingly diverse nature of nanostructural architectures, especially carbon nanotubes (CNTs) that will be detailed in Chapter 6. The CVD procedure is often denoted as *metal-organic* CVD (MOCVD), which more accurately specifies the use of an organometallic precursor, containing a central metal and ancillary ligands.

The steps involved in the growth of thin films by CVD are shown in Figure 4.36. Once the gas-phase precursor molecules enter the deposition zone (a), they are *physisorbed* to the substrate surface through weak van der Waals interactions (b). The ancillary ligands are removed through thermolysis, leaving the desired residual

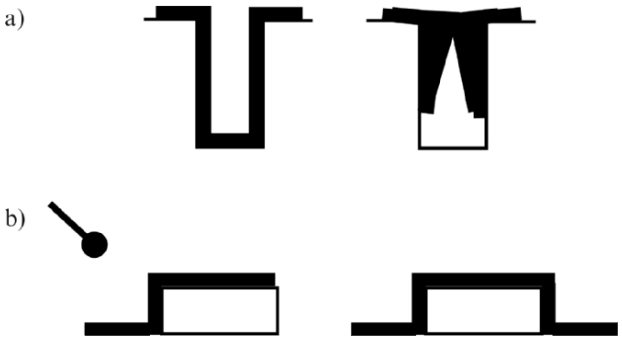


Figure 4.35. Illustration of conformal thin-film growth for (a) trench filling and (b) step coverage. The line-of-sight limitation of PVD, relative to a conformal CVD technique, is shown in (b).

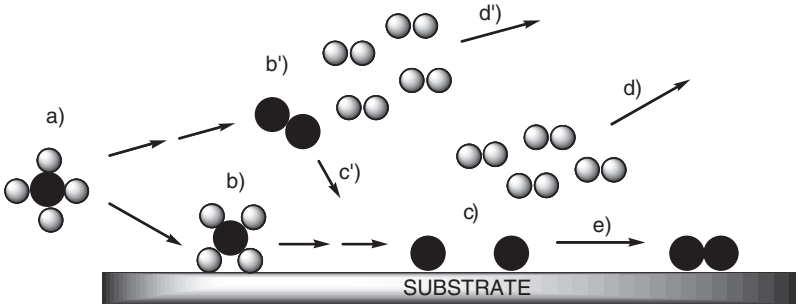


Figure 4.36. The important steps involved in CVD using the thermolysis of precursor molecules.

species on the surface (c), and ligands being removed from the deposition chamber *via* the carrier gas flow (d). Strong covalent interactions are then formed between the surface and adsorbed species, resulting in *chemisorption*. The surface-bound species may migrate along the surface and/or react with other surface species, eventually nucleating on thermodynamically favored positions (e) *en route* toward thin-film growth. Since film growth occurs through surface migration of intermediate species, CVD is the method-of-choice for depositions onto irregular surfaces where conformality is not possible using PVD techniques (Figure 4.37). Though the substrate is often placed horizontally in the CVD chamber, it is often more desirable to tilt the substrate in order to increase the deposition rate and film density. As the substrate is tilted, the precursor vapor is blocked from quickly passing overhead and spends a greater amount of residence time near the substrate surface (Figure 4.38).

In addition to the surface governed reactions described above, there are also gas-phase reactions that may take place between precursor/intermediate molecules. Although these reactions may also be important in the growth mechanism (*e.g.*, for plasma-enhanced CVD), significant gas-phase reactions will result in less desirable impure, granular, nonadhering, and nonconformal films. At relatively high temperatures, the gas-phase precursors may preferentially react with one another rather than

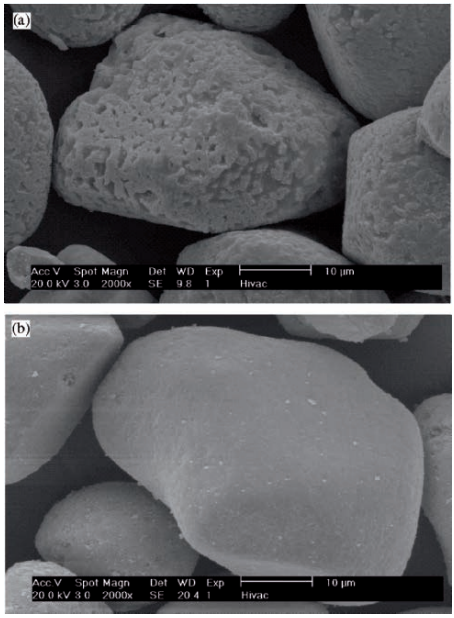


Figure 4.37. SEM images of an electroluminescent phosphor particle, ZnS (used in backlight displays for cell phones, watches, *etc*.), before (a) and after (b) the deposition of an aluminum oxide thin film. This film is a transparent coating that prevents the phosphor particle from undergoing humidity-accelerated decay. A technique known as *fluidized-bed CVD* was used, where a carrier gas both delivered the precursors to a vertically aligned CVD chamber, and dispersed the powdery sample in order to expose all surface regions to the precursor vapors.

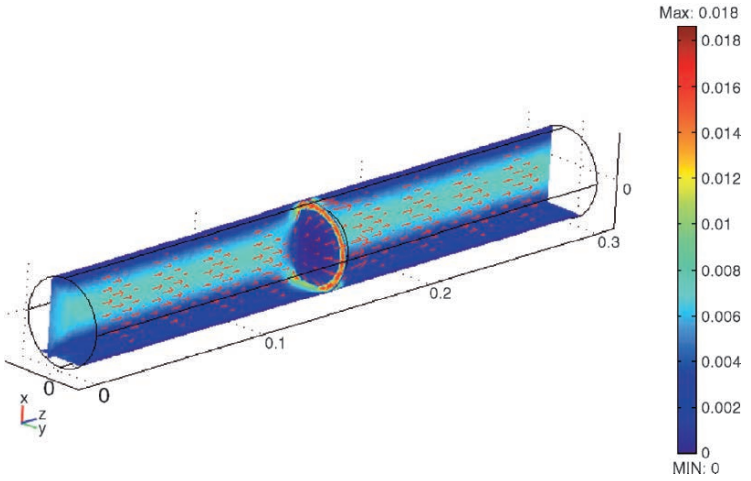


Figure 4.38. Molecular dynamics modeling of the flowrate in the vicinity of a tilted substrate.[13] Shown is the extreme case of a 90° tilt, that is suspended in the chamber to allow for maximum turbulent flow.

adsorbing to the substrate surface (b′, d′). The gas-phase nucleation causes granules to form and fall to the substrate surface due to gravitational forces (c′). Typically, these reactions may be minimized by lowering the deposition temperature, which will facilitate the surface-bound growth route. Though gas-phase growth is generally not desired for thin-film applications, this technique has recently been applied for the synthesis of nanoparticles.[14]

By definition, CVD is a nonequilibrium process. Although thermodynamics may provide useful information about the overall energetics of the growth process, kinetics must be used to provide information regarding reaction pathways or the transformation rates of the gaseous precursors. The kinetic description of CVD is divided into two parts: mass transport and the rates of the specific chemical reactions involving the precursor and intermediate species. Since the growth rate for CVD is relatively slow, the differences in forward and reverse reaction rates for interfacial events are much smaller than the absolute rates themselves.[15] Thus, CVD is frequently treated as a pseudoequilibrium system, considering only the vapor and solid immediately adjacent to the interface, the boundary layer.

At the microscopic level, there are a few key stages of thin-film growth. From adsorbed monomers and surface migration, the first stage is formation of subcritical embryos of varying sizes. These particles will further nucleate while taking on more precursor adsorption, forming a supercritical sized cluster. These clusters will coalesce into growth *islands*, which will expose new regions of the substrate that serve to adsorb additional precursor species. Isolated surface islands eventually grow together, leaving holes and channels that are filled by adsorbing precursor molecules to form a continuous thin film.

Two general types of reactors are used in CVD processes. In a *hot-walled* reactor (Figure 4.39), a tube furnace completely surrounds the deposition chamber containing the substrate. The desired deposition temperature depends on the thermal stabilities of the substrate to be coated, and the reactor (*i.e.*, glass ($T < 600°C$) or quartz ($T < 1{,}100°C$)). Hot-walled reactors are used extensively for laboratory studies,

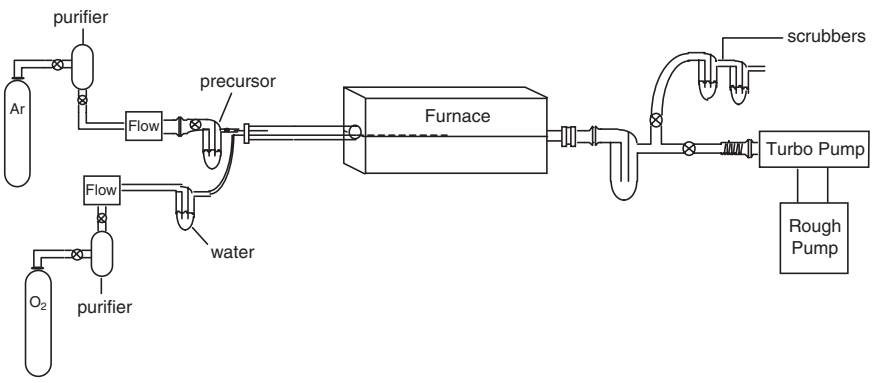


Figure 4.39. Schematic of a horizontal hot-walled CVD reactor. Shown is a two-precursor system, where the water sensitive precursor contacts water vapor directly over the heated substrate.

and also in industry for the CVD of semiconductors and oxides. These reactors are often preferred due to the simplicity of setup, while being able to maintain a uniform temperature over a large number of substrates. However, secondary coating of the reactor walls is unavoidable; frequent cleaning of these reactors is therefore necessary, as deposits may easily flake off these surfaces and contaminate the growing film. To prevent such problems, *cold-walled* reactors may also be used for a CVD process, wherein only the substrate is heated. This focuses the surface and gas-phase reactions to a region immediately surrounding the substrate surface. However, this reactor type generally results in narrower deposition zones and slower growth rates than hot-walled analogues.

Perhaps more than any other materials synthesis technique, semantics becomes a challenge due to the plethora of acronyms that are used to describe a specific CVD process. In particular, it is not sufficient to simply cite "CVD" alone in an article title; a more explicit acronym must be used that states the type of deposition chamber and precursor decomposition methodology employed. If no prefix is affixed to CVD, it usually denotes simple thermolysis of a precursor within a standard cold- or hot-wall reactor. However, if more energetic sources of energy are used to degrade the precursor such as laser, plasma, or microwave plasma, the acronyms laser-assisted CVD (LACVD), plasma-enhanced CVD (PECVD), and microwave plasma CVD (MPCVD) are used, respectively. Rather than using high-energy plasma sources, a relatively new CVD technique uses a heated filament to degrade precursor gases and form reactive intermediates. This technique is referred to as either catalyzed CVD (Cat-CVD) or hot-wire CVD (HWCVD, Figure 4.40).[16]

The pressure of the reactor must also be defined, and acronyms APCVD, LPCVD, and SCF-CVD are used to denote ambient, *in vacuo*, and supercritical pressure conditions within the deposition chamber, respectively. In general, the resultant

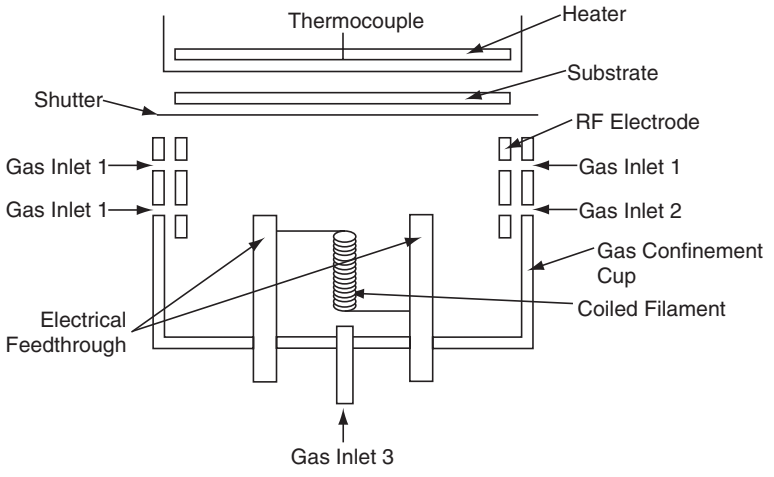


Figure 4.40. Schematic of a hot-wire CVD system. Reprinted from Povolny, H. S.; Deng, X. *Thin Solid Films* **2003**, *430*, 125. Copyright 2003, with permission from Elsevier.

film conformality of a CVD process follows the order LPCVD > APCVD, due to the relative migration rates of intermediate species along the surface of the substrate. However, it has recently been shown that conformal thin films of metal oxides may be deposited on nonplanar substrates such as carbon fibers under APCVD, even at low temperature.[17]

While plasma-enhanced methods are very useful to lower the substrate temperature, the as-deposited films are typically less conformal and often contain more surface impurities than competing methods. In this method, reactive radicals, ions, and atoms/molecules are formed in the gas phase that interact with the relatively low-temperature substrate to generate a film. Some of the more recent applications for plasma CVD include growth of cubic boron nitride (c-BN) thin films.[18]

In recent years, a complementary process known as *atomic layer deposition* (ALD) has been widely utilized.[19] In contrast to CVD, ALD features the sequential exposure of a substrate to two or more precursors. The precursor vapors are pulsed into the reactor one at a time, often being separated by inert-gas purges (Figure 4.41). This establishes user-controlled deposition cycles, each comprising an adsorbed layer of one precursor, followed by the selective surface reaction of a second precursor. Since film growth occurs submonolayer at a time, ALD allows for an unprecedented control over the resultant film thickness, conformality, homogeneity, and stoichiometry. This strategy will become most useful for the deposition of the thin high-κ gate oxide layers (*e.g.*, HfO_2, ZrO_2) required for next-generation CMOS devices.

Oftentimes, the deposition mechanisms between ALD and CVD are quite different, even for identical precursor combinations. In particular, it is well known that thin-film growth by CVD is heavily influenced by side reactions – not as problematic for ALD. A CVD process often generates reaction products such as CO, RH, or HCl that may be preferentially adsorbed onto the substrate surface. This will cause film-growth termination unless the competitive adsorbents are removed through purging with inert gas or introduction of a reducing gas such as H_2.

The chemical nature of the precursor represents the most critical component of a CVD/ALD process. Generally speaking, the choice of a particular precursor is governed by the relative stabilities of the precursor and substrate, as well as the volatility, cost, and hazards of the precursor. The coordination sphere of ligands surrounding the central metal is extremely important; the organic ligands in these precursors may

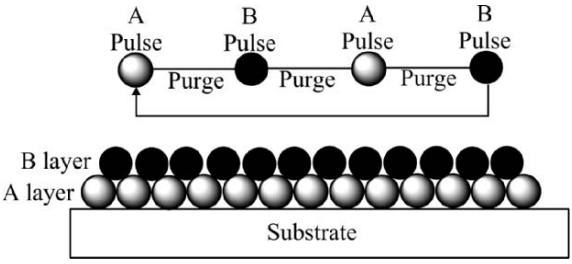


Figure 4.41. Illustration of atomic layer deposition (ALD).

lead to contamination of the films if they are not completely removed through a combination of pyrolysis, reduction, or oxidation processes. Some precursors pose a high risk when being used. For example, $Ni(CO)_4$ has a very high toxicity, $Al(Me)_3$ is pyrophoric, B_2H_6 is explosive, and chloride-containing species are corrosive. In general, organometallic precursors pose lower hazards than hydrides and halides, but are much more costly.

Although it was once essential that volatile precursors be used, this is no longer a synthetic limitation. Within the last decade, the gas/liquid properties of supercritical fluids (*e.g.*, CO_2) have been used to solvate certain precursors, facilitating their use for CVD.[20] Two variations of this technique may be used; supercritical fluid transport (SFT), using the fluid as an aerosol-like delivery vehicle, or *in situ* thin-film growth within a high pressure reactor, known as supercritical fluid deposition (SFD). In these methods the precursor must be soluble in CO_2, which is analogous in solvating ability to hexane with an enhanced fluorophilic character.

A CVD process may involve the use of either single or mixed precursors. The best precursor is a molecule that has sufficient volatility (or CO_2 solubility), and contains labile ligands that will leave no organic residue behind during its surface-catalyzed decomposition. In general, volatile liquids and oils are attractive CVD precursors due to the relative ease of vapor transport through simple carrier-gas bubbling or *direct liquid injection* (DLI) techniques. However, low-melting powders with high volatilities and low decomposition temperatures are regarded as the best candidates for CVD, as they are often less susceptible to premature hydrolysis during transfer to the CVD system. Co-reactant gases such as NH_3, H_2, H_2S, H_2O, *etc.* are often added in order to assist the removal of organic moieties from the organometallic precursor molecule, preventing their incorporation in the growing film.

The use of mixed precursors adds complexity to the control over film stoichiometry. The different volatilities of the precursors often lead to irreproducible vapor-phase concentrations. As a result, flow rates have to be carefully controlled to give films of a desired composition. Another problem that is posed by the use of mixed precursors is the variable rate of hydrolysis/oxidation, resulting in morphologically varied films. Hence, it is most preferable to use a precursor that contains as many desired building blocks of the thin film as possible (Figure 4.42). For the films deposited during IC fabrication, common precursors include silane (SiH_4) with co-reactant gases of H_2 (for polysilicon films), NH_3 (for Si_3N_4), and O_2 (for SiO_2). For tungsten films, a 1:3 WF_6:H_2 precursor mixture is used; TiN is deposited using a $TiCl_4/NH_3$ combination (Eq. 13).

$$(13) \quad 6TiCl_4 + 8NH_3 \xrightarrow{500°C} 6TiN + 24HCl + N_2$$

Prior to CVD studies, it is essential that the potential precursor undergo thermogravimetric analysis (TGA) to determine the most suitable temperature regime for thin-film growth (Figure 4.43). A sharp TGA curve with no remaining residue indicates that the precursor vaporizes without significant ligand decomposition – most desirable for CVD applications. A large mass loss at low temperature also indicates that the compound will be transported into the reactor with minimal gas-phase decomposition. The volatility of the precursor is directly related to its molecular structure;

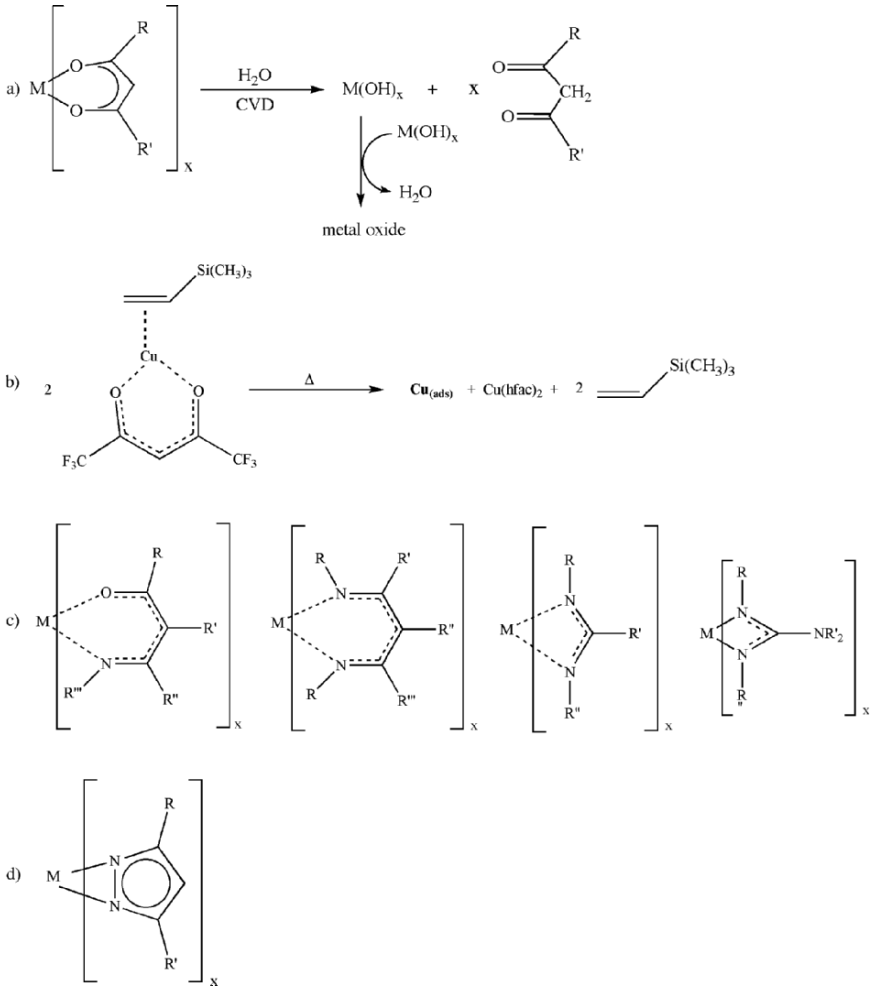


Figure 4.42. Molecular structures of commonly used CVD precursor classes. Shown are: (a) metal β-diketonate (acetylacetonate, acac) complex to grow a metal oxide film (H_2 as the coreactant gas yields a metal film); (b) a heteroleptic (more than one type of ligand bound to the metal) β-diketonate complex to yield a Cu film; the ancillary ligand helps prevent oligomerization, enhancing volatility; (c) various types of complexes to deposit metallic, oxide, nitride, or oxynitride films (depending on coreactant gas(es) used – respective ligands are β-ketoiminato, β-diketiminato, amidinato, and guanidinato; (d) a metal azolato complex commonly used to deposit lanthanide metal thin films.

large molecular weights and/or extensive intermolecular forces will significantly decrease its volatility. An effective method to enhance precursor volatility is to incorporate partially and fully fluorinated ligands. The enhancement may be rationalized either by an increased amount of intermolecular repulsion due to the additional lone pairs or that the reduced polarizability of fluorine (relative to hydrogen) causes fluorinated ligands to have less intermolecular attractive interactions.

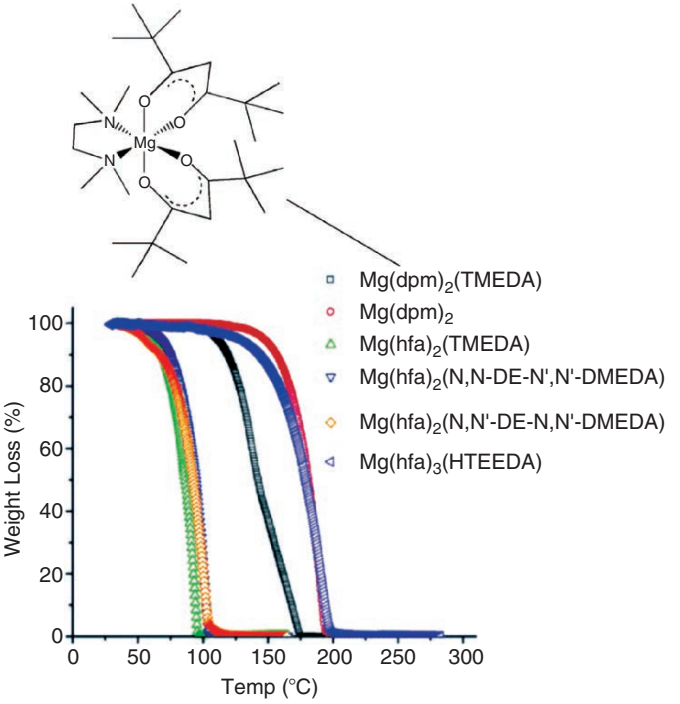


Figure 4.43. Thermogravimetric analysis of organomagnesium CVD precursors, indicating the dependence of molecular structure on its decomposition temperature. The ligand abbreviations are: dpm = 2,2,6,6-tetramethyl-3,5-heptanedionate, TMEDA = N, N, N', N'-tetramethylethylenediamine, hfa = 1,1,1,5,5,5-hexafluoro-2,4-pentanedione, HTEEDA = H(N, N, N', N'-tetraethylethylene diamine). Reproduced with permission from *Chem. Mater.* **2005**, *17*(23), 5697. Copyright 2005 American Chemical Society.

It is often desirable to deposit films that possess more complex stoichiometries. These films may include mixed-metal, and binary/ternary metal oxides and sulfides that are of importance for emerging applications for catalysis and microelectronics. Although these materials may be generated through use of individual species, such as Y, Ba, and Cu β-diketonates for $YBa_2Cu_3O_{7-\delta}$ superconductor films,[21] it is most advantageous to use a single-source precursor for these applications. The fixed ratio of component metals in the individual precursor molecules offers a unique route toward stoichiometric control over thin-film and nanoparticulate growth. Further, lower deposition or postannealing temperatures are often required for single-source precursors relative to their co-reactant analogs.

Perhaps the greatest "Achilles heel" of using single-source precursors is the complex and expensive synthetic procedures that are often required. Although this is usually overcome in laboratory-scale processes, this represents a major limitation for industrial scale-up considerations. Recently, a one-step reaction scheme was designed to yield a host of single-source precursor molecules that were suitable for the CVD growth of I–III–VI semiconductors (Figure 4.44).[22] Subsequent CVD

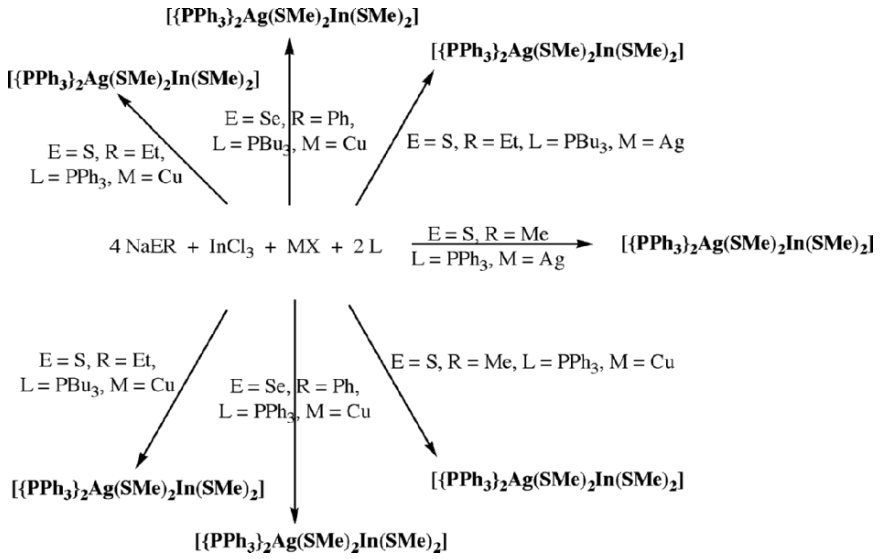


Figure 4.44. Synthetic pathway for single-source precursor design. Reproduced with permission from *Inorg. Chem.* **2003**, *42*, 7713. Copyright 2003 American Chemical Society.

studies using the ternary sulfide precursors were performed using an aerosol delivery methodology, with films of comparable purity to those generated by highly expensive procedures such as PVD from co-evaporation of source metals.

4.3. LIGHT-EMITTING DIODES: THERE IS LIFE OUTSIDE OF SILICON!

Thus far, we have considered the structure and applications of Si – the most heavily employed semiconductor. It should be noted that gallium arsenide (GaAs) is also widely used for FET applications; since GaAs does not form a natural protective layer analogous to SiO_2 onto Si, CVD is used to grow films such as GaS for surface passivation. Many other direct bandgap semiconductors such as II–VI (*e.g.*, ZnS, ZnSe) and III–V (*e.g.*, GaN, GaP) are widely used for optoelectronic (light emission) and photonic (light detection) applications. The most important applications for non-Si semiconductors are light-emitting diodes (LEDs) and solid-state lasers. Unlike Si-based devices, the bandgap of compound semiconductors may be significantly altered by varying the stoichiometry of the composite elements.

At present, the applications for LEDs have been centered on commercial electronic displays such as clock radios, microwave ovens, watches, *etc*. However, the availability of LEDs in a spectrum of colors has opened the floodgates for new applications. The greatest breakthrough was realized in the 1990s with the discovery of wide bandgap blue LEDs, making it possible to create any color of light (Table 4.2). Approximately 10–15% of the traffic lights in the United States have now been replaced with LED-based lamps. New automobiles also utilize this

Table 4.2. Comparison of the Observed Colors of LEDs

Observed color	Wavelength of emission (nm)	Semiconductor
(Infrared)	880	GaAlAs/GaAs
Red	660	GaAlAs/GaAlAs
Red	633	AlGaInP
Orange	612	AlGaInP
Orange	605	GaAsP/GaP
Yellow	585	GaAsP/GaP
Green	555	GaP
Blue	470	GaN/SiC
Ultraviolet	395	InGaN/SiC
White	–	InGaN/SiC

lighting for ultrabright brake and turn-signal lighting. The higher initial cost of the LEDs is quickly recovered due to their greater efficiency in converting electrical current to light emission relative to incandescent lighting. That is, whereas LEDs consume 10–25 W, an incandescent bulb of similar brightness consumes *ca.* 50–150 W. Indeed, once an inexpensive white LED becomes commercially available, our society will forever change as we shift from our longstanding reliance on inefficient and short-lived incandescent bulbs. The light emission from LEDs is classified as a type of *luminescence*, which is different from *incandescence* – the generation of light from a material as a result of its high temperature. Since LEDs glow as a result of an electrical current, this emission is referred to as *electroluminescence*. Two other common types of luminescence include *chemoluminescence* (induced by a chemical reaction(s); *e.g.*, "glow sticks"), and *photoluminescence* (induced through photon excitation; *e.g.*, vaseline glass discussed in Chapter 2). If the emission is prolonged, lasting long after the stimulation source is removed, it is known as *phosphorescence*; otherwise, the short-lived process is termed *fluorescence*.

As we saw earlier in this chapter, the wavelength (and color) of light emitted by a direct bandgap material through electron-hole recombination is influenced by its bandgap. In order to change the wavelength of emitted radiation, the bandgap of the semiconducting material utilized to fabricate the LED must be changed. For instance, gallium arsenide has a bandgap of 1.35 eV (Table 4.3), and emits in the infrared (*ca.* 900 nm). In order to decrease the wavelength of emission into the visible red region (*ca.* 700 nm), the bandgap must be increased to *ca.* 1.9 eV. This may be achieved by mixing GaAs with a material with a larger bandgap, such as GaP ($E_g = 2.35$ eV). Hence, LEDs of the chemical composition $GaAs_xP_{1-x}$ may be used to produce bandgaps from 1.4 to 2.3 eV (and varying colors), through adjustment of the As:P ratio.

The bandgap and concomitant wavelength of light that is emitted from LEDs is related to the bond strength between atoms in the lattice. For these compounds, as the bond strength increases, there is more efficient overlap between molecular orbitals that gives rise to a larger bandgap between bonding and antibonding MOs (*i.e.*, valence and conduction bands of the infinite lattice, respectively). For a particular

Table 4.3. Bandgaps of III–V Semiconductors

Semiconductor	Bandgap (eV)
AlN	6.02
AlP	2.45
GaN	3.50
GaP	2.35
GaAs	1.35
GaSb	0.67
InN	1.95
InP	1.27
InAs	0.36

Group 13 metal, as one moves down the Group 15 Period, the bonding interaction between III–V elements will become weaker through the interaction of more diffuse atomic orbitals. For instance, the bond strengths of Ga–N and Ga–As bonds are 98.8 and 50.1 kcal mol^{-1}, respectively. The larger bandgap for GaN relative to GaAs translates to a short wavelength (blue color) of emitted light that is observed.

Most white LEDs employ a semiconductor chip emitting at a short wavelength (blue), and a wavelength converter that absorbs light from the diode and undergoes secondary emission at a longer wavelength. Such diodes emit light of two or more wavelengths, that when combined, appear as white. The most common wavelength converter materials are termed *phosphors* (*e.g.*, ZnS – Figure 4.38), which exhibit luminescence when they absorb energy from another radiation source. Typical LED phosphors are present as a coating on the outside of the bulb, and are composed of an inorganic host substance (*e.g.*, yttrium aluminum garnet, YAG) containing an optically active dopant (*e.g.*, Ce). Use of such a single-crystal phosphor produces a yellow light, upon combination with blue light gives the appearance of white. A similar result has recently been produced through use of CdSe *nanoparticles* (see Chapter 6).[23] White LEDs may also be made by coating near ultraviolet (NUV) emitting LEDs with a mixture of europium-based red and blue emitting phosphors, plus green emitting copper- and aluminium-doped zinc sulfide (ZnS:Cu,Al). It is also possible for LEDs to emit white light without the use of phosphors. For instance, homoepitaxially grown ZnSe crystals simultaneously emit blue light from the film, and yellow light from the ZnSe substrate.

Recently, there has been much interest in organic light-emitting diodes (OLEDs). Flat-panel televisions, cellular phones, and digital cameras are already beginning to employ this technology; it is only a matter of time before the "holy grail" of *flexible* display screens and luminous fabrics are produced. OLED displays offer many benefits relative to standard CRTs and LCDs such as enhanced brightness, lower power consumption, and wider viewing angles.

The multilayered structure and electroluminescent mechanism of OLEDs is illustrated in Figure 4.45. Depending on whether small organic molecules or long repeating-unit polymers are used (Figure 4.46), the diodes are referred to as OLEDs or PLEDs, respectively. Under positive current, electrons and holes are injected into the emissive layer from opposite directions – from the cathode and anode, respectively. The metal

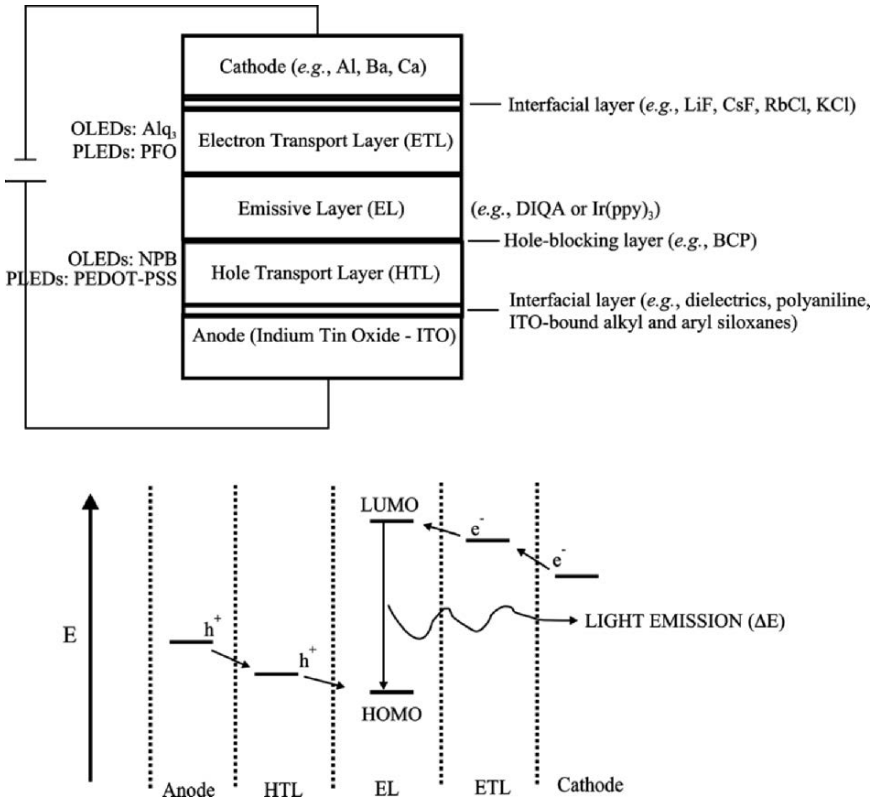


Figure 4.45. Multilayered structure of OLEDs/PLEDs. Also shown are the relative energy levels for individual layers; light is emitted as a result of the radiative recombination of electron-hole pairs.

cathode is usually an alkaline earth or Al, which readily release a valence electron (*i.e.*, possess a low work function). The holes that migrate from the anode are blocked from further transport to the cathode by an organic layer of BCP (Figure 4.46c), which has a relatively low-lying HOMO.

In contrast, the anode is usually tin-doped In_2O_3 (ITO, typically 1:9 Sn:In), since this material is transparent and highly conductive (*ca.* $1 \times 10^4\ W^{-1}\ cm^{-1}$). The substitutional replacement of In^{3+} ions with Sn^{4+} ions results in n-doping of the lattice, injecting electrons into the conduction band. The doping of Sn into the In_2O_3 lattice may also result in SnO sites; this introduces holes in the lattice that reduces its conductivity. The conductivity of ITO is due to both Sn dopants and O vacancies, as represented by the general formula $In_{2-x}Sn_xO_{3-2x}$. Since the organic hole-transport polymer is in direct contact with ITO in an OLED, the surface properties of ITO is important for effective hole injection. It has been shown that surface treatments such as UV ozone cleaning or Ar/O_2 plasma treatments, result in an increase in the work function of the ITO surface as the Sn:In ratio is decreased (and O concentration is increased).

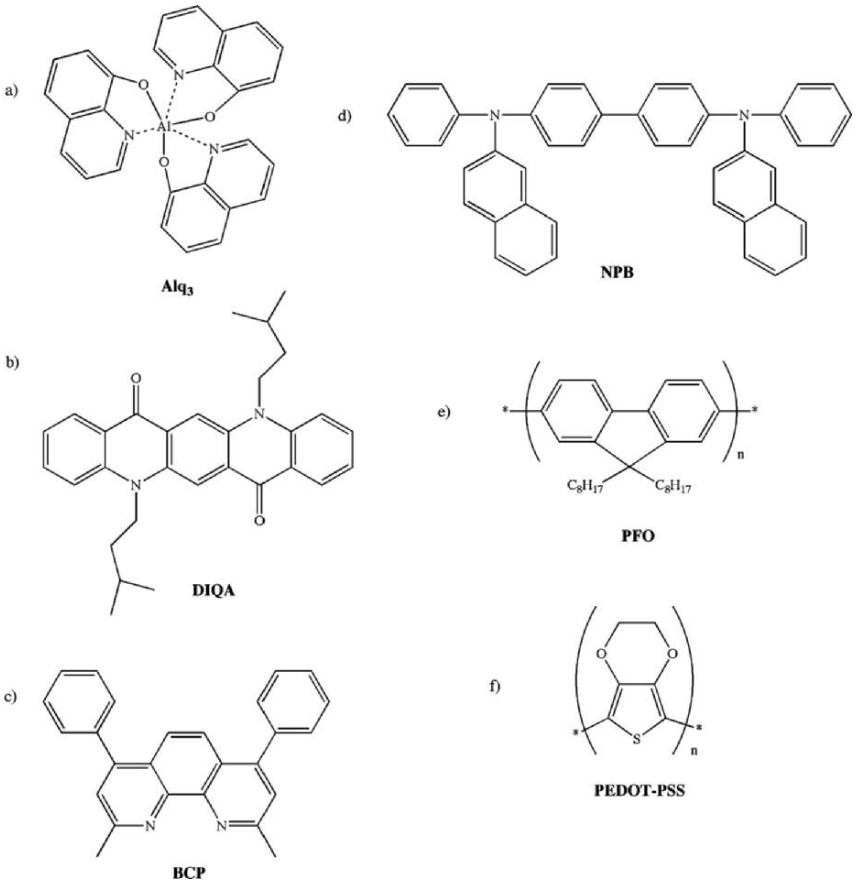


Figure 4.46. Molecular structures of commonly used OLED/PLED materials. Shown are: (a) Alq_3 (tris(quinoxalinato)Al (III)) used as an electron-transport material; (b) DIQA (diisoamylquinacridone) used as an emissive dopant; (c) BCP (2,9-dimethyl-4,7-diphenyl-1,10-phenanthroline) used as an exciton/hole blocking agent; (d) NPB (1,4-bis(1-napthylphenyl amino)biphenyl); (e) PFO (9,9-dioctylfluorene) used as an emissive polymer in PLEDs; (f) PEDOT–PSS (poly-3,4-ethylenedioxythiophene–polystyrene sulfonate) used as a hole transport material in PLEDs.

The recombination of a hole and electron creates a quasiparticle known as an *exciton*, which releases a photon of energy. Organic molecules in the emissive layer facilitate exciton formation – likely through hole-trapping, followed by Coulombic attraction to the free electrons. A challenge in OLED design is to ensure that an equal number of holes and electrons meet in the emissive layer. This is not trivial, since holes migrate much slower than electrons in conductive organic molecules. The organic molecules used in OLEDs/PLEDs are π-conjugated, meaning that p_z orbitals on adjacent $(–C=C–C=)_n$ units overlap, resulting in π valence and π^* conduction bands. The observed color will depend on the HOMO–LUMO gap of the molecule, which may be fine-tuned by altering the

length of the conjugation, or nature of the molecular backbone (*e.g.*, substitution of electron withdrawing/donating groups, *etc.*).

A limitation of traditional OLEDs/PLEDs is their relatively poor quantum efficiency. An exciton may exist in either a singlet (total spin, $S = 0$) or triplet state ($S = 1$), with the triplet outweighing the singlet by a 3:1 ratio. The molecules in the emissive layer are referred to as fluorophores, which yield fluorescent radiation only when a singlet exciton is formed. Hence, the theoretical maximum efficiency that is possible is only 25%. In an effort to improve the efficiency, a number of recent studies have been devoted to using organometallic compounds (transition metals with organic ligands), which allow for rapid intersystem crossing (ISC) from the excited singlet to light-emitting triplet states (Figure 4.47). The process of collecting the excitons in the lowest excited triplet state is referred to as *triplet harvesting*, and results in a 100% quantum efficiency of the OLED.[24] This effect is most pronounced for complexes of 4d and 5d transition metals with well-shielded valence electrons, which exhibit large *spin–orbit coupling*[25] and concomitant mixing of singlet and triplet states.

4.4. THERMOELECTRIC (TE) MATERIALS

As our society attempts to wean itself from a dependence on fossil fuels, many alternative energy sources are being investigated. One interesting potential is from the conversion of waste heat (*e.g.*, vehicle engine heat) into electricity. This conversion is due to the *thermoelectric effect*, first discovered for junctions (or *thermocouple*) of two dissimilar metals. When metals are welded together, free electrons are able to drift across the junction in a preferential direction, based on the different atomic lattice structures of each metal. The migration of electrons effectively leaves one metal electron-deficient (positively charged), and the other metal negatively charged.

If the two metals are held at different temperatures, a voltage will result that is proportional to the ΔT. The ratio ($\Delta V / \Delta T$) is referred to as the *Seebeck coefficient*, α, related to the band structure of the materials involved. The α value for semiconductors is at least two orders of magnitude larger for semiconductors relative to metals, giving rise to much greater voltages. Contrary to the Seebeck effect, when a current is passed through a closed circuit of the two metals, heat is produced at one of the junctions and is absorbed at the other. Since the latter junction is cooled, this *Peltier effect* has been largely exploited for thermoelectric cooling applications (*e.g.*, auto seat coolers, computer component cooling).

A TE device consists of a heat source and sink, joined together *via* n-type and p-type semiconductor materials (Figure 4.48). The *figure of merit*, ZT (Eq. 14), of the device is useful to determine its suitability for power generation or refrigeration – a means to describe the transport properties of the TE material. An effective TE material should have a high Seebeck coefficient (heat conversion efficiency) and electrical conductivity, as well as low thermal conductivity to maintain thermal isolation in the device. The thermal conductivity is related to the transfer of heat through a material through either electron transport or quantized lattice vibrations (phonons). Hence, the ideal TE material has been described as a phonon-glass/electron-crystal,

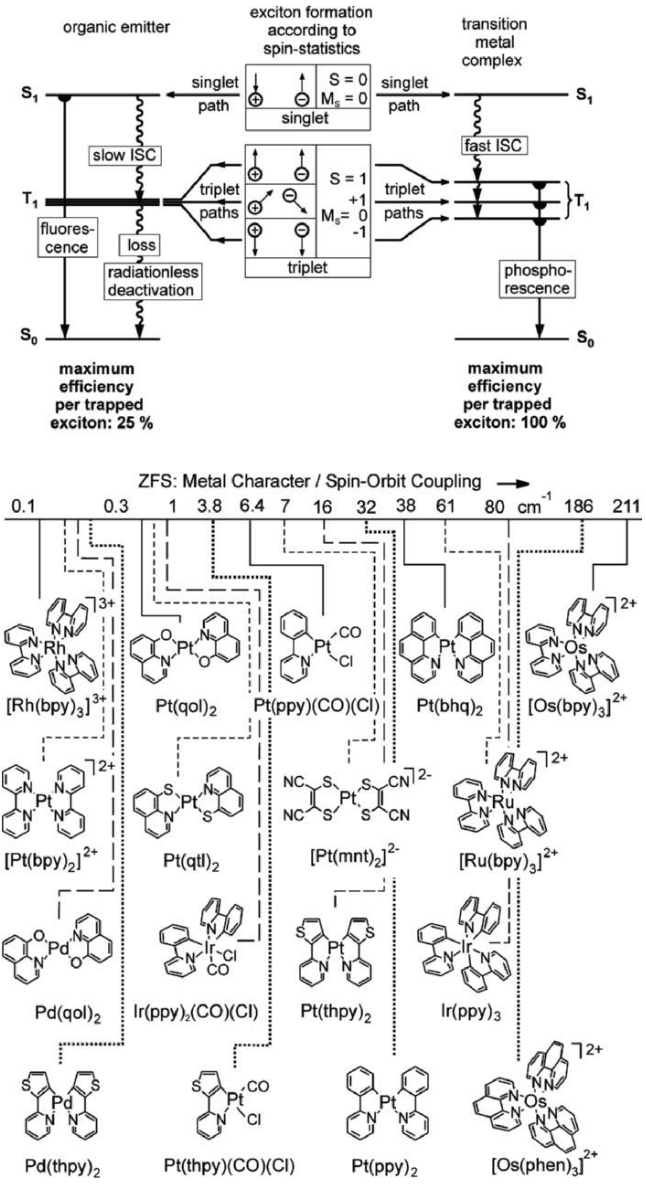


Figure 4.47. Illustration of triplet harvesting. In the absence of a triplet emitter, the triplet excitation energy is converted to heat, losing 75% of the quantum efficiency. Also shown are a variety of organometallic complexes, and their relative spin–orbit coupling values – directly proportional to their use in phosphorescent OLEDs. Reproduced with permission from Yersin, H. *Top. Curr. Chem.* **2004**, *241*, 1. Copyright 2004 Springer Science and Business Media.

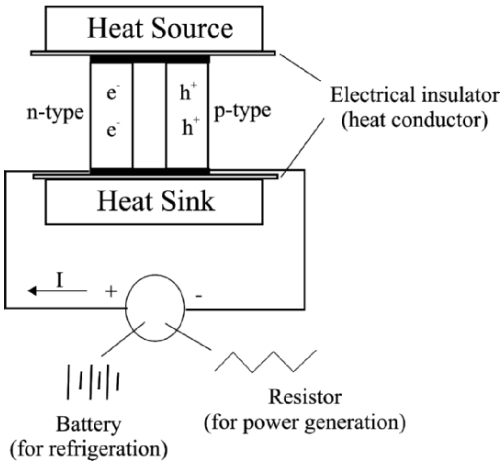


Figure 4.48. Schematic of a thermoelectric device, which may be used for power generation or refrigeration.

having the electrical properties of a crystalline lattice, and thermal properties of an amorphous/glass-like solid.

$$(14) \quad \mathrm{ZT} = \frac{\alpha^2 \sigma}{\lambda}$$

Where α is the Seebeck coefficient ($\mu V/K$; 1–10 for metals, 150–250 for semiconductors); the σ, electrical conductivity; and λ is the thermal conductivity.

Semiconductors are much more effective TE materials (greater ZT values) than metals due to their significantly greater Seebeck coefficients and lower thermal conductivities. A barrier toward higher ZT values for semiconductors is their relatively small electrical conductivity, especially at low temperatures. In order to overcome this limitation, the chemical composition of the semiconductor may be fine-tuned to yield a small bandgap material, or one that is sufficiently doped with a low concentration (*ca.* 10^{18}–10^{19} cm^3) of electronic/thermal carriers exhibiting high mobility through the lattice.

The most widely studied TE material is Bi_2Te_3, consisting of a hexagonal unit cell with repeating…[Te–Bi–Te–Bi–Te…Te–Bi–Te–Bi–Te]… units (Figure 4.49). Whereas Te–Bi layers are bound by strong covalent interactions, the bonding between adjacent Te layers is through weak van der Waals interactions. This results in bulk anisotropic electrical and thermal conductivity, being most pronounced along planes that are perpendicular to the *c*-axis of the unit cell. To further improve ZT values, Bi_2Te_3 crystals may be doped with n- or p-type dopants, with most desired compositions of $Bi_2Te_{2.7}Se_{0.3}$ (n-type) and $Bi_{0.5}Sb_{1.5}Te_3$ (p-type), yielding $\mathrm{ZT} = 1$ at room temperature. This improvement results from a decrease in thermal conductivity of the lattice, brought about by phonon scattering by the dopant atoms (*i.e.*, perturbing the symmetry of the lattice, affecting the organized lattice vibration modes).

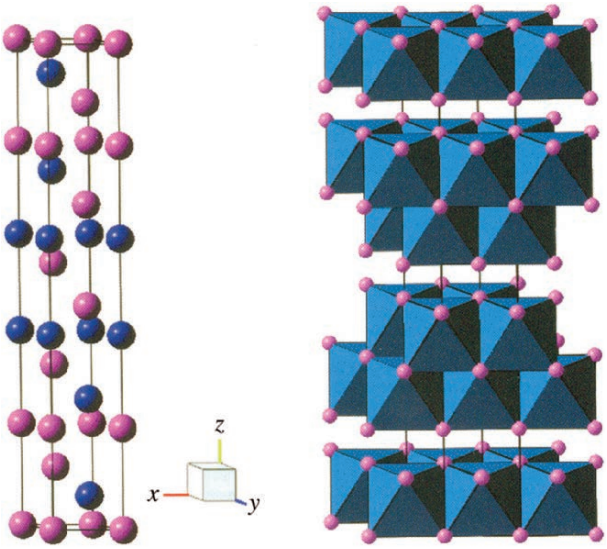


Figure 4.49. Unit cell of Bi_2Te_3 – a widely studied thermoelectric material. The blue atoms are Bi, and the pink atoms are Te. Reproduced with permission from Tritt, T. M.; Subramanian, M. A. *MRS Bull.* **2006**, *31*, 188. Copyright 2006 Materials Research Society.

The following list represents the primary materials classes that have been designed in an effort to optimize ZT for thermoelectric applications.[26] The general strategy is to dope the lattice with sufficient carriers (n- or p-dopants), while also interrupting the phonon transport through the solid through the introduction of large interstitial atoms.

(i) *Complex Solid-State Inorganic Lattices* (*e.g.*, $CsBi_4Te_6$, $Bi_{2-x}Sb_xTe_{3-y}Se_y$, $ZrTe_5$, $Ag_nPb_mM_nTe_{m+2n}$ (M = Sb, Bi), $AgPb_{10}SbTe_{12}$, "Half-Heusler alloys" MNiSn (M = Zr, Hf, Ti), $Zr_{0.5}Hf_{0.5}Ni_{0.5}Pd_{0.5}Sn_{0.99}Sb_{0.01}$). For half-Heusler alloys, the unit cell is a combination of a NaCl lattice of two metals, with the third metal occupying tetrahedral interstitial sites. For instance, for TiNiSn, the Ti and Sn form the NaCl lattice, with Ni occupying 1/2 of the available tetrahedral interstitial sites (*i.e.*, Ti:Ni:Sn = 4:4:4 atoms per unit cell). This combination offers a great deal of control over electronic/thermal conductivity of the solid. While the introduction of Sn (p-doping) increases electrical conductivity, heavy metal atoms such as Ti and Ni cause a decrease in thermal conductivity through phonon scattering.

(ii) *Crystal Structures with "Rattlers"* (*e.g.*, MSb_3-based skutterudites, clathrate architectures – Figure 4.50). In these structures, the dopant atoms are weakly bound to the cage, and "rattle" in response to increasing temperature. As the atom within the cages becomes smaller/heavier, the amount of structural disorder will increase causing a larger decrease in lattice thermal conductivity.

(iii) *Oxides* (*e.g.*, $NaCo_2O_4$, $Ca_3Co_4O_9$, $Al_{0.02}Zn_{0.98}O$). These structures consist of CoO_2 layers, which serves as an effective electronic transport medium within

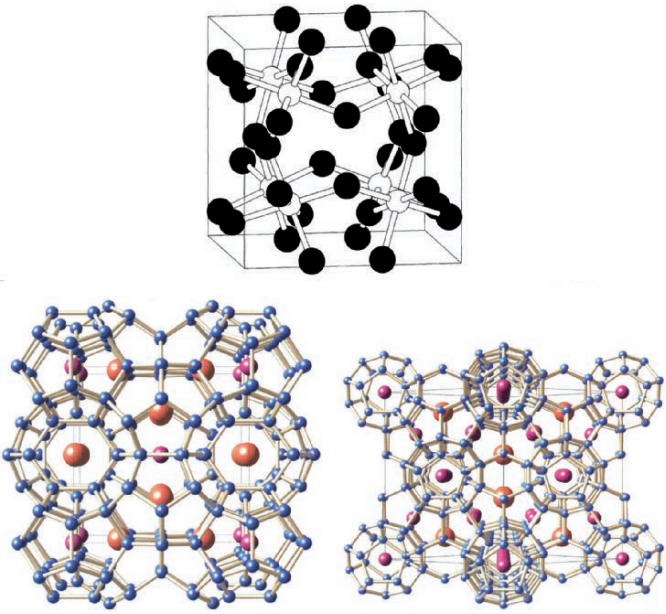


Figure 4.50. Crystal structures of an unfilled $IrSb_3$ skutterudite (top – Sb atoms shown in black) and Type I/Type II clathrates (bottom). In the skutterudite structure, a void is present in the center of the unit cell, surrounded by 12 Sb atoms. For the clathrate structures, tetrahedrally bound framework atoms (*e.g.*, Ge, Sn, Si) are illustrated in blue, and guest atoms within the various cages are shown in orange and pink. Reproduced with permission from (i) Nolas, G. S.; Poon, J.; Kanatzidis, M. *MRS Bull.* **2006**, *31*, 199, Copyright 2006 Materials Research Society and (ii) *Chem. Mater.* **2000**, *12*, 697. Copyright 2000 American Chemical Society.

the lattice. A variety of oxide layers (MO, M = Na, Ca, Sr, Bi) are placed in between the Co oxide layers to serve as phonon scatterers (Figure 4.51).

(iv) *Nanoarchitectures* (*e.g.*, nanotubes and other nanostructures – described in more detail in Chapter 6).

The comparison of ZT values for a host of the above materials is displayed in Figure 4.52, along with the variation of ZT with temperature. The search for new TE materials continues to be an extremely active area of research; recent work has resulted in suitable TE components at virtually all temperature regimes. However, the search continues for materials with ZT values in excess of 2.0; most materials exhibit figure-of-merits ≤ 1.0.

IMPORTANT MATERIALS APPLICATIONS III: PHOTOVOLTAIC (SOLAR) CELLS

The increasingly volatile prices and availability of gasoline and natural gas has brought about a heightened awareness of our dependence on nonrenewable resources. Currently, over 90% of the US electricity supply comes from the combustion of fossil fuels (*i.e.*, coal, oil, and natural gas) and nuclear power generation.

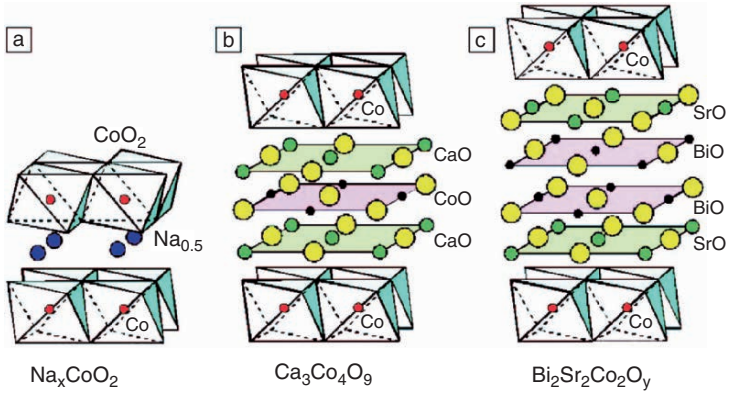


Figure 4.51. Crystal structure schematic of CoO_2-based TE oxides. Shown are (a) Na_xCoO_2, (b) $Ca_3Co_4O_9$, and (c) $Bi_2Sr_2Co_2O_y$. Reproduced with permission from Koumoto, K.; Terasaki, I.; Funahashi, R. *MRS Bull.* **2006**, *31*, 206. Copyright 2006 Materials Research Society.

While the US coal reserve is estimated at 290 billion tons (enough to last another 230 years at current production levels), there is an increased concern about the adverse environmental effects associated with its combustion, namely SO_x, NO_x, and CO_2 emissions. With the energy needs of our world likely to double within our generation, there continues to be active development of alternative renewable energy sources. In an earlier chapter, we dealt with hydrogen-based fuel cells; we now will describe another extremely attractive option – harvesting the unlimited power from the sun.

The majority of commercial photovoltaic cells utilize silicon-based technology. When sunlight comes in contact with a p–n diode, the absorbed energy causes promotion of electrons from valence to conduction bands, generating additional electron-hole pairs. The most noticeable result of this excitation is an effective increase in the number of electrons in the conduction band of p-Si (or holes in n-Si, referred to as the *minority-carrier concentration*). As a result, electrons in the p region will diffuse into the depletion region, where the junction potential propels them back into n-Si (*vice versa* for holes; Figure 4.53a).

In the design of solar cells, a small bandgap semiconductor is desirable since it requires less energy to promote electrons from valence to conduction bands. However, though the resultant photovoltage of the solar cell is directly proportional to the bandgap, higher energy photons would be converted to heat rather than electrical energy. The use of a higher bandgap material would result in a lower photocurrent since only high-energy photons ($E \geq E_g$) would be absorbed. The bandgaps of Si and GaAs (1.1 eV (1,127 nm) and 1.4 eV (886 nm), respectively) are able to absorb a broad range of electromagnetic radiation; however, the efficiencies of Si or GaAs based solar cells are only 20–25%. Thin-film solar cells that utilize amorphous silicon (a-Si) are also currently under investigation. Since Si atoms do not form a continuous crystalline array, surface sites contain significant numbers of dangling bonds, which are easily passivated through reaction with hydrogen (yielding hydrogenated

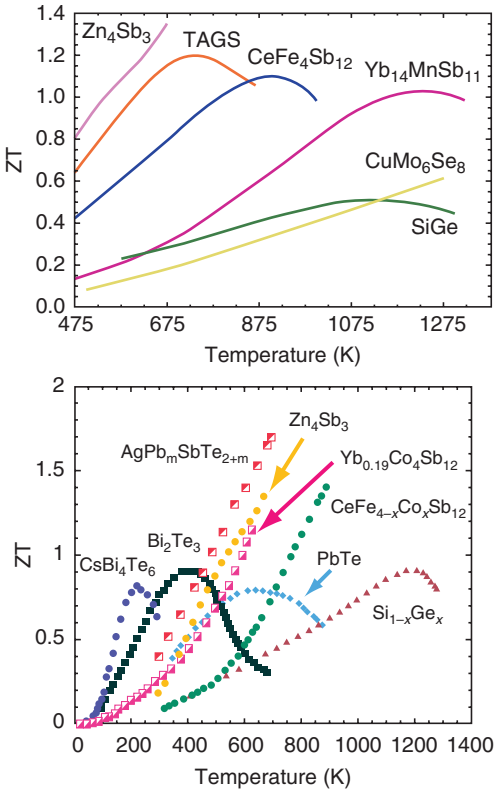


Figure 4.52. The temperature dependence of ZT values for various TE materials. TAGS refers to $(GeTe)_{0.85}(AgSbTe_2)_{0.35}$. Reproduced with permission from Tritt, T. M.; Subramanian, M. A. *MRS Bull.* **2006**, *31*, 188, and *Chem. Mater.* **2006**, *18*, 1873. Copyright 2006 American Chemical Society.

amorphous silicon, a-Si:H). The larger bandgap of a-Si (1.7 eV) is more effective at capturing broadband solar energy; however, the best efficiencies to date are less than 10% for this technology.

In order to capture a larger range of photon energies, multijunction photocells have recently been introduced. This method consists of a stack of semiconductor layers with decreasing bandgaps (Figure 4.53b). Top layers have higher bandgaps, and are able to absorb higher energy photons; lower layers of the cell absorb the transmitted lower energy wavelengths. The multijunction solar cell shown in Figure 4.54b has resulted in an efficiency of 34%, the highest value reported for all photovoltaic cells to date.[27] It has been proposed that replacement of GaAs with a material with an E_g of 1.25 eV would result in a higher overall efficiency, since more photons would be collected by the second layer and fewer would be transmitted to the bottom Ge substrate. With additional tweaking such as altering the number of layers, individual layer thicknesses/stoichiometries, it is expected that multijunction photocells with efficiencies of 45–50% and higher will soon be realized.

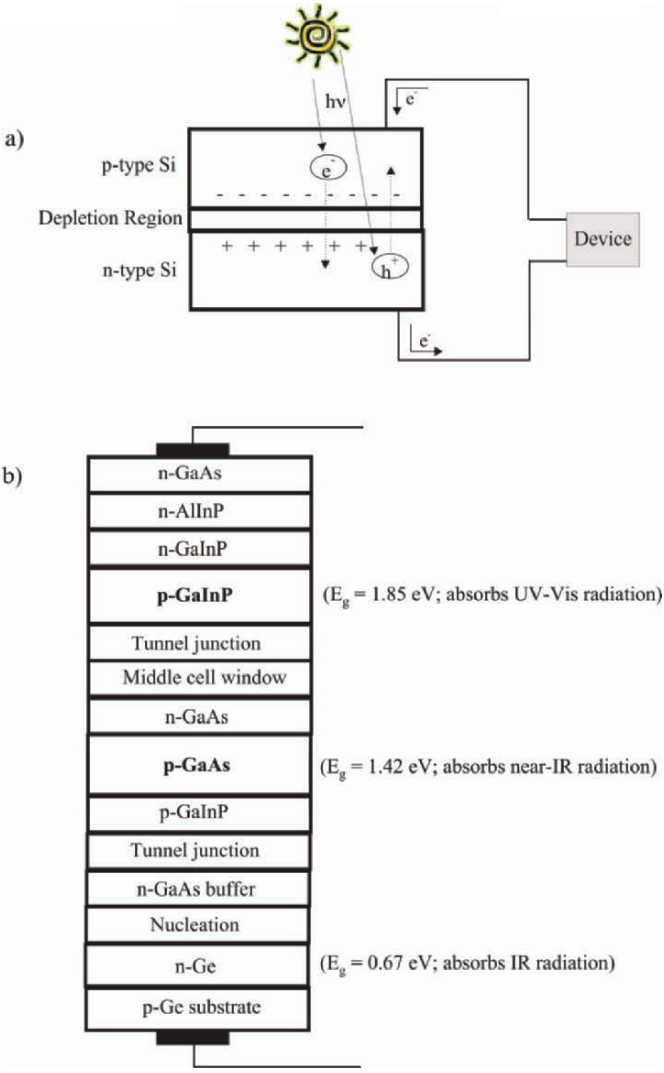


Figure 4.53. Schematic of (a) a single-junction and (b) multijunction photovoltaic cell.

While it is important to control the stoichiometry of each layer to adjust their bandgaps, it is equally important to have as few interfacial mismatches as possible. That is, each layer must be epitaxially grown to ensure that the lattice constants are perfectly matched. It has been shown that a lattice mismatch of only *ca.* 0.01% is enough to cause significant electron-hole recombinations, resulting in lower cell efficiency. CVD is the method of choice for the fabrication of these multilayer devices; as you might expect, cells of this variety are relatively quite expensive.

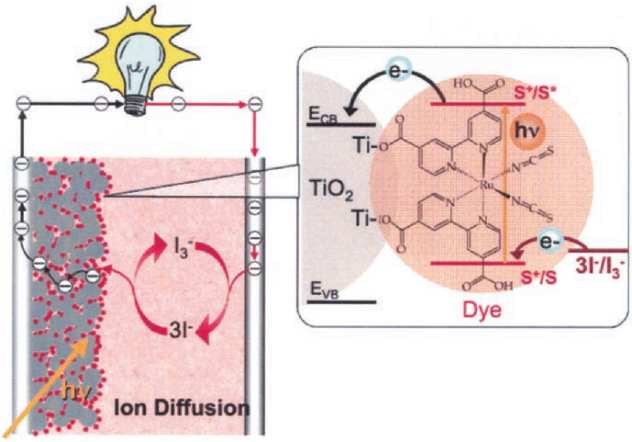


Figure 4.54. Schematic of a dye-sensitized photovoltaic cell. The TiO_2-bound dye molecules act as the light harvester. Electrons are injected into TiO_2, flow to the collector electrode and through the circuit to the counter electrode. The dye is regenerated by electron donation from the $I_3^-/3I^-$ couple (0.536 V). Reproduced with permission from *Inorg. Chem.* **2005**, *44*, 6841. Copyright 2005 American Chemical Society.

In an effort to decrease the cost of production for solar cells, new architectures known as dye-sensitized solar cells (DSC) are of increasing interest. Instead of light being absorbed by inorganic semiconductor layers, absorption occurs in an organic-based film that comprises light-harvesting dye molecules bound to a TiO_2 surface (Figure 4.54). When the dye absorbs a photon, the resultant excitation injects electrons into the titanium, which transports them to the negative electrode, with the positive electrode attached to the electrolyte.[28] To date, the maximum efficiency of DSCs is only *ca.* 11%; however, it is likely that this value will significantly improve as alternate nanomaterials are utilized (see Chapter 6). There are already commercial products that contain DSC technology.[29]

The TiO_2 used in DSCs is not present as a smooth film, but as a porous nanostructural sponge, with a very high surface area. This improves the number of active sites that may absorb light, analogous to the stacked structure of leaves that convert light to energy using the light-harvesting molecule chlorophyll. The oxide nanoparticles are carefully prepared to preferentially expose the (101) face of anatase, which is the lowest energy surface – providing the lowest LUMO level to accept electron density from the sensitizer. It should be noted that other wide bandgap semiconductor oxides such as SnO_2 and ZnO have also been employed within DSCs.

Not unlike their inorganic semiconductor counterparts, the light-harvesting molecule is perhaps the most important component of a DSC. The dyes employed for this application are ruthenium complexes that feature highly conjugated ligands similar to those employed for OLED applications. The absorption of UV–Visible radiation causes $\pi \rightarrow \pi^*$ and metal-to-ligand charge transfer (MLCT) electronic transitions. The latter transition is quite pronounced for ruthenium, since its valence d-electrons

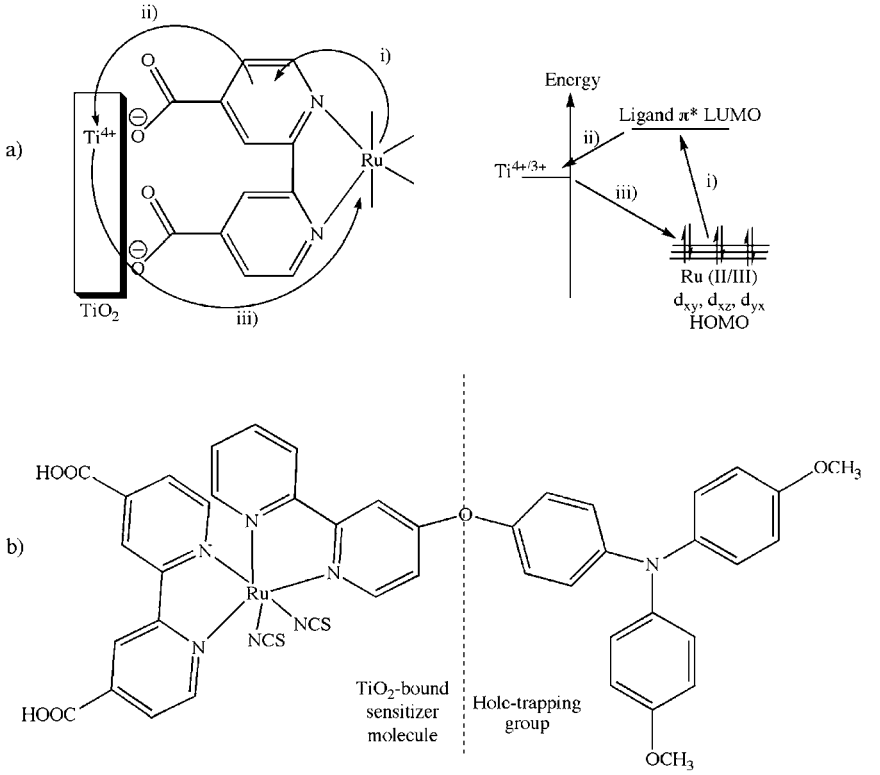


Figure 4.55. (a) Detailed schematic of the forward/backward electron donation between the sensitizer and TiO_2 surface. A carboxylate group is shown on the sensitizer ligand – important for Lewis acid/base binding to the oxide surface. Shown in (b) is an example of a hole-trapping group that prevents back-donation from the TiO_2 surface to the sensitizer (improving the photocurrent generated by the DSC). Adapted with permission from *Inorg. Chem.* **2005**, *44*, 6841. Copyright 2005 American Chemical Society.

are heavily shielded from the nuclear charge and are more easily promoted to ligand unoccupied orbitals. The donation of electron density from the bound sensitizer is due to MLCT which results in the oxidation of the Ru(II) metal center. (Figure 4.55a). The design of sensitizer ligands is not trivial; the π^* LUMO of the ligand must be of appropriate energy to allow electron transport to flow to the TiO_2 surface from the metal ion. However, it is also possible for an electron to be injected back to the sensitizer from the conduction band of the TiO_2 (Figure 4.55a). By designing the sensitizer appropriately, the rate of electron flow from the dye to TiO_2 will be much faster than the reverse reaction. An intriguing way to accomplish this is through the incorporation of a hole-trapping group attached to the ligands (Figure 4.55b). After MLCT takes place and the electron is injected into the TiO_2 layer, the positive charge is transferred to the hole-trapping moiety. This increases the distance of the hole from the surface, which slows the rate of TiO_2-sensitizer back-donation considerably.

References and Notes

1 Imre, A.; Csaba, G.; Ji, L.; Orlov, A.; Bernstein, G. H.; Porod, W. *Science* **2006**, *311*, 205.

2 For a recent press release, see: http://www.rochester.edu/news/show.php?id=2585

3 Many presentations regarding terahertz transistor design may be found at: http://www.intel.com/technology/silicon/micron.htm

4 (a) Sadana, D. K.; Current, M. "Fabrication of Silicon-on-Insulator (SOI) Wafers Using Ion Implantation", in *Ion Implantation Science and Technology*, Ziegler, J. F. ed., Edgewater: Ion Implantation Technology Co., 2000. (b) Colinge, J. P. *Silicon-on-Insulator Technology: Materials to VLSI*, 2nd ed., Dordrecht: Kluwer Academic Publishers, 1997. (c) http://www.ibis.com – Ibis Technology Corporation, Danvers, MA, USA. (d) http://www.chips.ibm.com/bluelogic (e) Marshall, A.; Natarajan, S. *SOI Design: Analog, Memory, and Digital Techniques*, Boston: Kluwer Academic Publishers, 2002.

5 Plummer, J. D.; Deal, M. D.; Griffin, P. B. *Silicon VLSI Technology: Fundamentals, Practice, and Modeling*, Prentice Hall: New Jersey, 2000.

6 Even though single-crystal Si is used as the substrate, the lattice spacings of SiO_2 are much greater than Si (5.431 Å).

7 Kriegler, R. J. *Semiconductor Silicon 1973*, The Electrochemical Society: Princeton, NJ, 1973.

8 A plasma is considered the fourth class of matter, in addition to solids, liquids, and gases. A plasma contains a mixture of ground-state and excited-state atoms, as well as ions.

9 C. J. Mogab, A. C. Adams, and D. L. Flamm, *J. Appl. Phys*. **1978**, *49*(7), 3796.

10 Refer to the following for some recent precedents of sol–gel derived films with controlled porosity. This is typically performed through use of templates that are sacrificially removed following film deposition. (a) Fuertes, M. C.; Soler-Illia, G. J. A. A. *Chem. Mater.* **2006**, *18*, 2109. (b) Velasquez, C.; Rojas, F.; Esparza, J. M.; Ortiz, A.; Campero, A. *J. Phys. Chem. B* **2006**, ASAP article. (c) Xiao, L.; Zhang, H.; Scanlon, E.; Ramanathan, L. S.; Choe, E.-W.; Rogers, D.; Apple, T.; Benicewicz, B. C. *Chem. Mater.* **2005**, *17*, 5328. (d) Kanungo, M.; Deepa, P. N.; Collinson, M. M. *Chem. Mater.* **2004**, *16*, 5535. (e) Li, X. S.; Fryxell, G. E.; Birnbaum, J. C.; Wang, C. *Langmuir* **2004**, *20*, 9095.

11 http://www.ciber-runa.net/guide/ChipCosts.html (a very nice website regarding the cost of production for ICs).

12 For example, http://www.tystar.com/additional_3.htm

13 Courtesy of Leela Rakesh, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859; leela.rakesh@cmich.edu

14 (a) Nasibulin, A. G.; Shurygina, L. I.; Kauppinen, E. I. *Colloid J.* **2005**, *67*, 1, and references therein. (b) Suzuki, K.; Kijima, K. *Jpn. J. Appl. Phys.* **2005**, *44*, 2081. (c) Chen, R. S.; Huang, Y. S.; Liang, Y. M.; Tsai, D. S.; Tiong, K. K. *J. Alloys Compd.* **2004**, *383*, 273. (d) Wang, Y. Q.; Chen, J. H.; Yoo, W. J.; Yeo, Y. -C. *Mat. Res. Soc. Symp. Proc.* **2005**, *830*, 269.

15 Barron, A. R. in *CVD of Nonmetals*, Rees, W. S. ed., Wiley: New York, 1996.

16 (a) Schropp, R. E. *Mater. Res. Soc. Symp. Proc.* **2003**, *762*, 479. (b) Schropp, R. E. *Thin Solid Films* **2004**, *451*, 455. (c) Lau, K. K. S.; Murthy, S. K.; Lewis, H. G.; Pryce, C.; Jeffrey, A.; Gleason, K. K. *J. Fluorine Chem.* **2003**, *122*, 93. (d) Mahan, A. H. *Solar Energy Mater. Solar Cells* **2003**, *78*, 299. (e) Stannowski, B.; Rath, J. K.; Schropp, R. E. *Thin Solid Films* **2003**, *430*, 220. (f) Schroeder, B. *Thin Solid Films* **2003**, *430*, 1. (g) Duan, H. L.; Zaharias, G. A.; Bent, S. E. *Curr. Opin. Solid State Mater. Sci.* **2002**, *6*, 471. (h) Masuda, A.; Izumi, A.; Umemoto, H.; Matsumura, H. *Mater. Res. Soc. Symp. Proc.* **2002**, *715*, 111. (i) Mahan, A. H. *Solar Energy* **2004**, *77*, 931. (j) Matsumura, H.; Umemoto, H.; Masuda, A. *J. Non-Cryst. Solids* **2004**, *338*, 19.

17 (a) Fahlman, B. D.; Barron, A. R. *Adv. Mater. Opt. Electron.* **2000**, *10*, 135. (b) Richards, V. N.; Vohs, J. K.; Williams, G. L.; Fahlman, B. D. *J. Am. Ceram. Soc.* **2005**, *88*, 1973.

18 Zhang, W. J.; Bello, I.; Lifshitz, Y.; Chan, K. M.; Meng, X. M.; Wu, Y.; Chan, C. Y.; Lee, S. T. *Adv. Mater.* **2004**, *16*, 1405.

19 For some recent articles on ALD, see (a) Xiong, G.; Elam, J. W.; Feng, H.; Han, C. Y.; Wang, H.-H.; Iton, L. E.; Curtiss, L. A.; Pellin, M. J.; Kung, M.; Kung, H.; Stair, P. C. *J. Phys. Chem. B.* **2005**, *109*, 14059. (b) Niinisto, J.; Rahtu, A.; Putkonen, M.; Ritala, M.; Leskela, M.; Niinisto, L. *Langmuir* **2005**, *21*, 7321. (c) Sechrist, Z. A.; Fabreguette, F. H.; Heintz, O.; Phung, T. M.; Johnson, D. C.; George, S. M. *Chem. Mater.* **2005**, *17*, 3475. (d) Reijnen, L.; Meester, B.; de Lange, F.; Schoonman, J.; Goossens,

A. *Chem. Mater.* **2005**, *17*, 2724. (e) Matero, R.; Rahtu, A.; Ritala, M. *Langmuir* **2005**, *21*, 3498. (f) Min, Y.-S.; Cho, Y. J.; Hwang, C. S. *Chem. Mater.* **2005**, *17*, 626. (g) Gu, W.; Tripp, C. P. *Langmuir* **2005**, *21*, 211.

[20] (a) Hansen, B. N.; Hybertson, B. M.; Barkley, R. M.; Sievers, R. E. *Chem. Mater.* **1992**, *4*, 749. (b) Lagalante, A. F.; Hansen, B. N.; Bruno, T. J.; Sievers, R. E. *Inorg. Chem.* **1995**, *34*, 5781. (c) Fernandes, N. E.; Fisher, S. M.; Poshusta, J. C.; Vlachos, D. G.; Tsapatsis, M.; Watkins, J. J. *Chem. Mater.* **2001**, *13*, 2023. (d) Cabanas, A.; Long, D. P.; Watkins, J. J. *Chem. Mater.* **2004**, *16*, 2028. (e) Blackburn, J. M.; Long, D. P.; Watkins, J. J. *Chem. Mater.* **2000**, *12*, 2625. (f) Blackburn, J. M.; Long, D. P.; Cabanas, A.; Watkins, J. J. *Science* **2001**, *294*, 141. (g) Cabanas, A.; Blackburn, J. M.; Watkins, J. *J. Microelectron. Eng.* **2002**, *64*, 53. (h) Ohde, H.; Kramer, S.; Moore, S.; Wai, C. M. *Chem. Mater.* **2004**, *16*, 4028.

[21] Hansen, B. N.; Brooks, M. H.; Barkley, R. M.; Sievers, R. E. *Chem. Mater.* **1992**, *4*, 749.

[22] (a) Banger, K. K.; Jin, M. H.-C.; Harris, J. D.; Fanwick, P. E.; Hepp, A. F. *Inorg. Chem.* **2003**, *42*, 7713. (b) Castro, S. L.; Bailey, S. G.; Raffaelle, R. P.; Banger, K. K.; Hepp, A. F. *Chem. Mater.* **2003**, *15*, 3142.

[23] Bowers, M. J.; McBride, J. R.; Rosenthal, S. J. *J. Am. Chem. Soc.* **2005**, *127*, 15378.

[24] For a terrific recent review on exciton formation and OLEDs, see: Yersin, H. *Top. Curr. Chem.* **2004**, *241*, 1.

[25] Interaction of the spin magnetic moment of an electron with the magnetic moment arising from the orbital motion of the electron.

[26] For details regarding all aspects of thermoelectric materials, refer to the March 31, 2006 issue of the MRS Bulletin – devoted entirely to this topic.

[27] http://www.nrel.gov. It should be noted that solar cells are limited by a number of intrinsic and extrinsic losses. Extrinsic sources include reflection, series resistance, absorption within interlayers, nonradiative recombination, and many others. Intrinsic sources include inefficient collection of solar photon energies by each layer, and radiative recombination. For details on these limitations, see: Henry, C. H. *J. Appl. Phys.* **1980**, *51*, 4494.

[28] For a thorough review of DSCs, see: Gratzel, M. *Inorg. Chem.* **2005**, *44*, 6841.

[29] Solar panels using DSC have been produced by Sustainable Technologies, International (www.sta.com.au); other applications are being actively pursued.

Topics for Further Discussion

1. If a circular silicon wafer, with a diameter of 300 mm and thickness of 0.5 mm, is 99.9999999% pure, how many impurity atoms are present in the wafer?
2. What is meant by "density of states," and how does this influence the electrical conductivity of a semiconductor?
3. Your supervisor has asked you to deposit the following thin films; what transition metal/main group complexes would you employ that would impart a high volatility of the overall complex, and low C incorporation in the growing film? Also, think of what coreactant gas(es) would be most appropriate for CVD. (a) ZrO_2 (b) Zr_3N_4 (c) $Zr_xHf_{1-x}O_2$ (d) PZT (e) Mg (f) Alq_3 (*i.e.*, Figure 4.46a).
4. Explain how LEDs/OLEDs operate, and think of some intriguing applications for these types of materials.
5. Why is TiO_2 used in dye-sensitized solar cells (DSCs)? What other oxides would possibly be useful?
6. As discussed in this Chapter, next-generation CMOS ICs will employ high-κ dielectric films as the gate insulator. A potential roadblock to this replacement is the possible incompatibility with current gate stacks – explain this problem, with possible solutions.
7. Explain the differences between ALD, CVD, and PVD. Which one provides the greatest control over the properties of the growing film?
8. In this Chapter, there was a lot of discussion related to silicon-based devices and applications. As we attempt to continue Moore's Law into the future, how likely is it that Si will be replaced entirely? If this paradigm shift *was* proposed, what would be some negative consequences associated with such a move?

9. Why are chlorinated coreactants used during the initial oxidation step of CMOS fabrication, and what are the mechanisms involved with its reactive participation?
10. How are the bandgap (E_g) and dielectric constant (κ) of a thin film determined experimentally?

Further Reading

1. Kasap, S. O. *Principles of Electronic Materials and Devices*, 2nd ed., Prentice Hall: New Jersey, 2002.
2. Plummer, J. D.; Deal, M. D.; Griffin, P. B. *Silicon VLSI Technology: Fundamentals, Practice, and Modeling*, Prentice Hall: New Jersey, 2000.
3. *Physics and Chemistry of III–V Compound Semiconductor Interfaces*, Wilmsen, C. ed., Springer: New York, 1985.
4. Misra, A.; Hogan, J. D.; Chorush, R. A. *Handbook of Chemicals and Gases for the Semiconductor Industry*, Wiley: New York, 2002.
5. Moss, S. J.; Ledwith, A. *The Chemistry of the Semiconductor Industry,* Chapman and Hall: New York, 1987.
6. Turley, J. *The Essential Guide to Semiconductors*, Prentice Hall: New Jersey, 2003.
7. Pierret, R. F. *Semiconductor Fundamentals, Volume I*, 2nd ed., Prentice Hall: New Jersey, 1988.
8. Streetman, B.; Banerjee, S. *Solid State Electronic Devices*, 5th ed., Prentice Hall: New Jersey, 1999.
9. Hamers, R. J.; Wang, Y. "Atomically-Resolved Studies of the Chemistry and Bonding at Silicon Surfaces" *Chem. Rev.* **1996**, *96*, 1261.
10. Waltenburg, H. N.; Yates, J. T. "Surface Chemistry of Silicon" *Chem. Rev.* **1995**, *95*, 1589.
11. Recent reviews of high-κ dielectric materials: (a) Locquet, J. -P.; Marchiori, C.; Sousa, M.; Fompeyrine, J.; Seo, J. W. *J. Appl. Phys.* **2006**, *100*, 051610; (b) Robertson, J. *Rep. Prog. Phys.* **2006**, *69*, 327.
12. Announcement by Intel regarding the 45 nm technology scheduled for 2007: http://download.intel.com/technology/silicon/Press_45nm_106.pdf
13. http://smithsonianchips.si.edu/index2.htm
14. Pierson, H. O. *Handbook of CVD*, 2nd ed., Noyes Publications, 1999.

CHAPTER 5

ORGANIC "SOFT" MATERIALS

Take a minute to look at the room and furnishings around you. Virtually everything you will see is at least partially comprised of organic-based building blocks. From the plastics that surround electronic components to individual carpet fibers, no other type of material is as heavily utilized in our society as organic-based polymers. As first defined by Staudinger in 1920, a *polymer* is any material that is comprised of an extended structure of small chemical repeat units, known as *monomers*. For simplicity, the monomeric unit is almost always clearly identified within the polymer name (*i.e.*, "poly(*monomeric unit*)," Figure 5.1). A polymer is generally comprised of more than 100 repeat units; structures with lower numbers of chemical repeat units are known as *oligomers*. Strictly speaking, all solid-state materials with an infinite structural array are classified as polymers – even inorganic structures such as metals, ceramics, and glasses. However, since we have described inorganic-based materials in previous chapters, we will focus our present discussion on polymeric materials that feature a carbon-containing backbone.

Organic-based materials are generally associated with "soft" characteristics – *i.e.*, relatively low-melting and undergo facile plastic deformation. Though this is generally the case for popular commercial polymers such as plastics and rubbers, there are numerous other polymer classes that exhibit hardness and thermal stabilities that even rival inorganic ceramics. Although polymeric materials have long been used in packaging and other consumer product applications, recent developments have extended the utility range to now include microelectronics, photovoltaics, self-healing materials, and drug-delivery agents that were not envisioned a short time ago. As a testament to the importance of polymers in our world, the current volume of polymers used for commercial and industrial applications already exceeds the volume of steel and aluminum combined. In our era of soaring gas prices, automobile manufacturers are also scrambling to add more plastics to their vehicles. In addition to fabrication cost savings and enhanced design flexibility relative to steel, it is estimated that every 10% of weight reduction will yield a 5% increase in fuel economy.

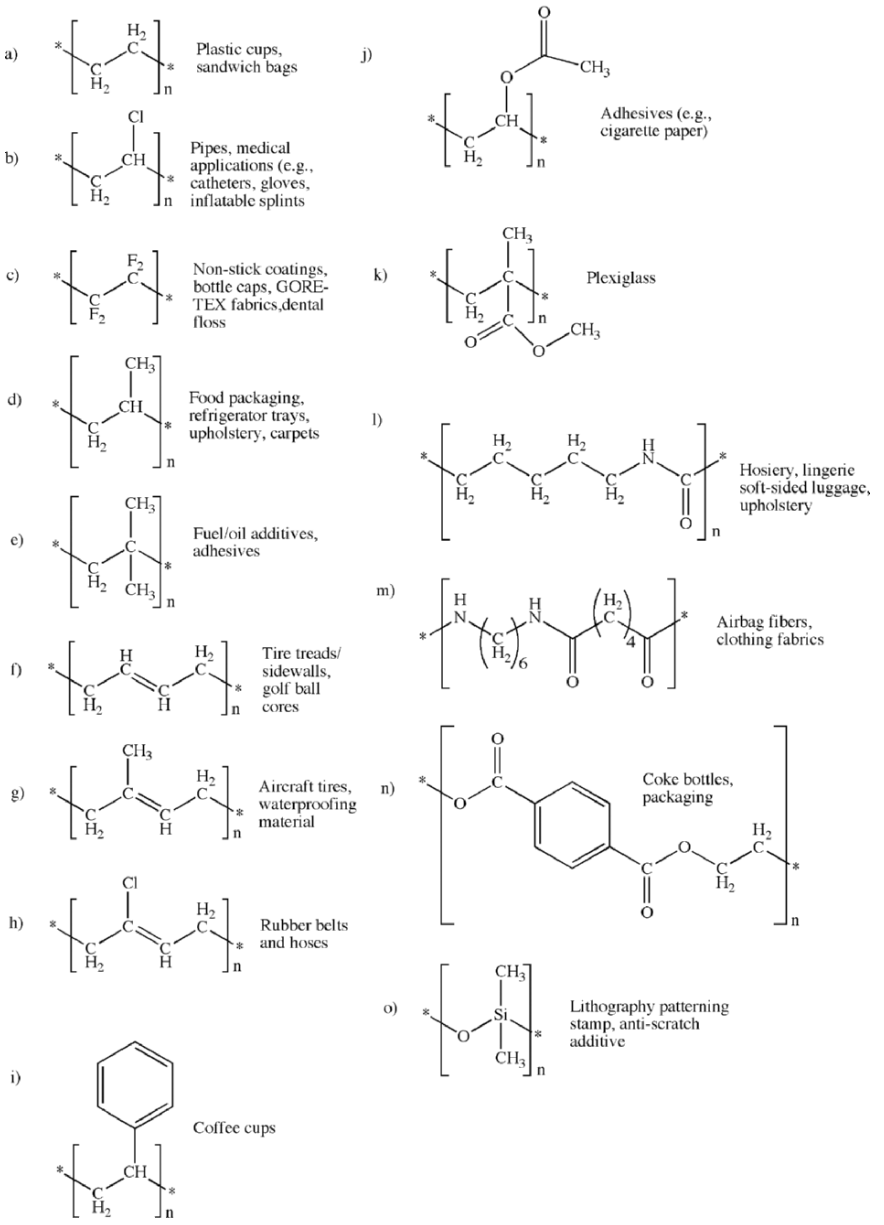


Figure 5.1. Molecular structures of the chemical repeat units for common polymers. Shown are (a) polyethylene (PE), (b) poly(vinyl chloride) (PVC), (c) polytetrafluoroethylene (PTFE), (d) polypropylene (PP), (e) polyisobutylene (PIB), (f) polybutadiene (PBD), (g) *cis*-polyisoprene (natural rubber), (h) *trans*-polychloroprene (Neoprene® rubber), (i) polystyrene (PS), (j) poly(vinyl acetate) (PVAc), (k) poly(methyl methacrylate) (PMMA), (l) polycaprolactam (polyamide – nylon 6), (m) nylon 6,6, (n) poly(ethylene teraphthalate), (o) poly(dimethyl siloxane) (PDMS).

5.1. POLYMER CLASSIFICATIONS AND NOMENCLATURE

The most fundamental classification of polymers is whether they are naturally occurring or synthetic. Common natural polymers (often referred to as *biopolymers*) include macromolecules such as polysaccharides (*e.g.*, starches, sugars, cellulose, gums, *etc.*), proteins (*e.g.*, enzymes), fibers (*e.g.*, wool, silk, cotton), polyisoprenes (*e.g.*, natural rubber), and nucleic acids (*e.g.*, RNA, DNA). The synthesis of biodegradable polymers from natural biopolymer sources is an area of increasing interest, due to dwindling world petroleum supplies and disposal concerns.

Synthetic polymers may be classified under two general umbrellas: *thermoplastics* and *thermosets*. As their names imply, this definition is illustrative of the properties exhibited by these materials under elevated temperatures. For instance, thermoplastics consist of long molecules with side chains or groups that are not connected to neighboring molecules (*i.e.*, not *crosslinked*). Hence, both amorphous and crystalline thermoplastics are glasses at low temperature, and transform to a rubbery *elastomer* or flexible plastic at an elevated temperature known as the *glass-transition temperature* (T_g; Figure 5.2). The T_g is the most important property of polymers, being analogous to the melting point of low molecular weight compounds.[1] In contrast to thermoplastics, thermosets are initially liquids and become hardened by a thermally induced crosslinking process known as *curing*. Also, unlike thermoplastics, since the crosslinking process yields a stable 3D network, thermosets may not be remelted/reprocessed. Thermoset polymers are usually synthesized within a mold to yield a desired shape/part; once the polymer cures, the only way to reshape the material is through machining processes (*e.g.*, drilling, grinding). The most

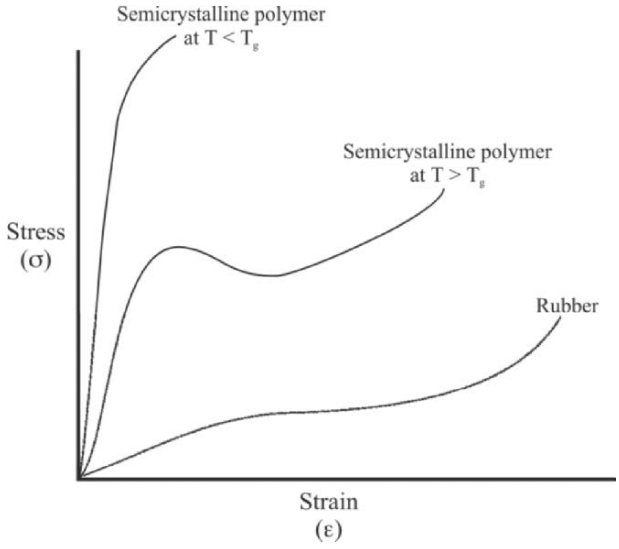


Figure 5.2. Stress *vs.* strain curves for various polymers around its glass-transition temperature. The maximum in the curve that occurs at T_g is referred to as the *yield point* (onset of plastic deformation).

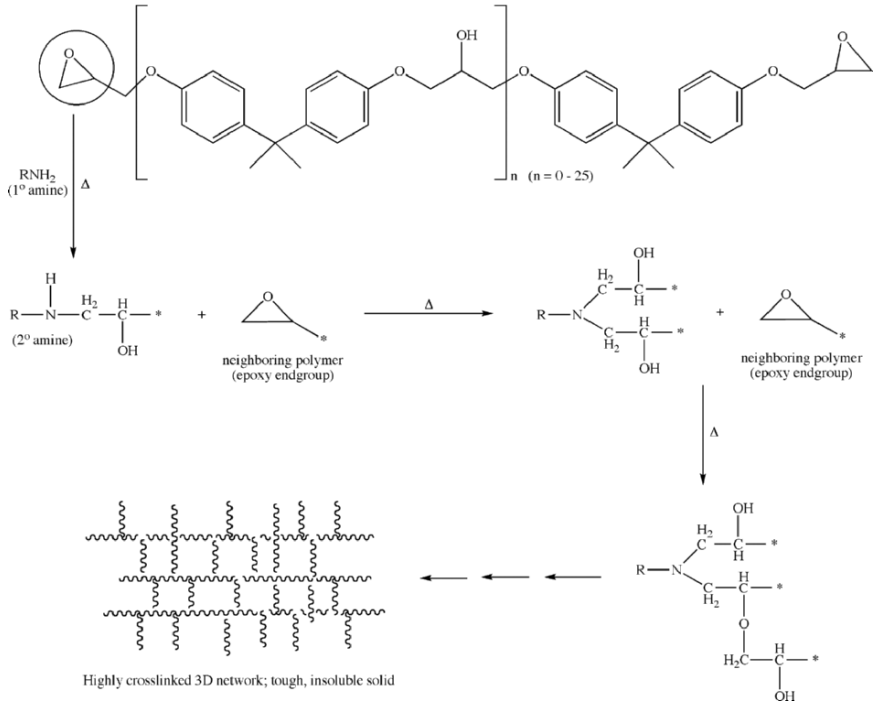


Figure 5.3. Illustration of a hardening mechanism responsible for epoxy resin curing.

common type of thermosetting polymer is epoxy resin, widely used for adhesives and paints/coatings, which harden through crosslinking reactions with a curing agent such as primary and secondary polyamines (*i.e.*, containing reactive $–NH_2$ groups; Figure 5.3. Once the polymer has cured, the only way to reshape the part is by machining (*e.g.*, grinding, drilling, *etc.*).

Regarding the overall structure of polymers, there are five classes of macromolecular architectures, ranging from simple linear arrays to *megamers* – complex structures built from ordered hyperbranched (dendritic) polymers (Figure 5.4). As you might imagine, linear chains are best able to pack into a regular crystalline array; however, as the degree of chain branching increases, only amorphous phases are formed. Due to the pronounced length of polymer chains, it should be noted that regions of crystallinity generally form where only a portion of the polymer chains are regularly organized, while others remain disordered. This structural diversity directly affects physical properties such as tensile strength, flexibility, and opaqueness of the bulk polymer. This chapter will delve into various polymer classes, with a detailed examination of the structural influence on resultant properties.

As a more general structural definition, a polymer synthesized from only one type of monomer is referred to as a *homopolymer*. In contrast, a polymer that is formed from more than one type of monomer is known as a *copolymer*. The terms

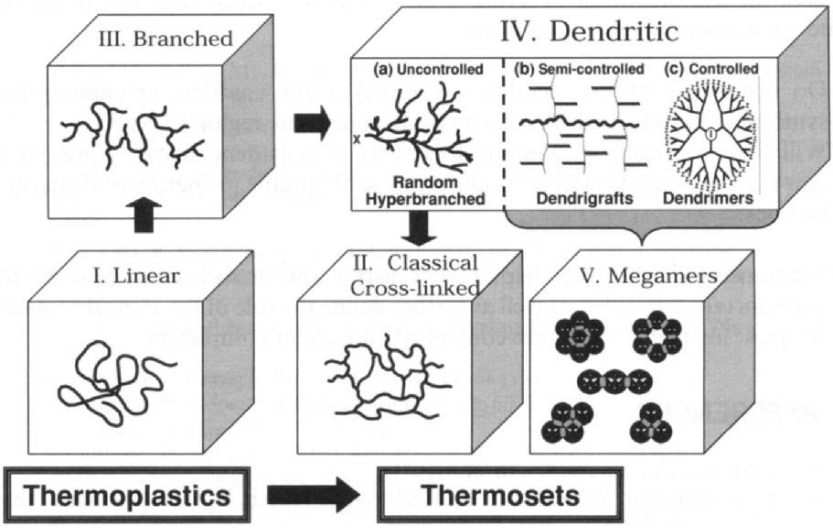


Figure 5.4. The five major structural classes of polymers. Reproduced with permission from Frechet, J. M. J.; Tomalia, D. A. *Dendrimers and Dendritic Polymers*, Wiley: New York, 2001.

terpolymer, tetrapolymer, pentapolymer, *etc.* are used to designate a polymer derived from three, four, five, *etc.* co-monomers. Although homopolymers contain only one repeating chemical unit in their structure, this distinction is not always clear. For instance, polyethylene polymers often contain short-chain branched impurities as a consequence of the polymerization process. However, since this unintentional structural deviation results from a polymerization process involving only one type of monomer, the term *homopolymer* is still most appropriate.

There are four types of copolymers: *random*, *alternating*, *block*, and *graft*, each exhibiting varied physical properties. In contrast to block copolymers, which contain long adjacent sequences of A and B monomers, the chain of an alternating copolymer is comprised of the sequence $(–A–B–)_n$. A random copolymer is formed from the random linkages of A and B monomers along the chain (*e.g.*, –AAABAABBBAAB–). Graft copolymers may be considered as another type of block copolymer, which feature one long chain of one monomer with offshoots of a second monomer emanating along its length (Figure 5.5). The structure, length, and placement of the copolymer units directly affect physical properties such as crystallinity, density, strength, brittleness, melting point, and electrical conductivity (for conductive polymers).

Each carbon within the polymer chain may be considered as a chiral center. As a result, a parameter referred to as the *tacticity* may be defined that describes the stereoregularity of adjacent carbon centers (Figure 5.6). A polymer that contains substituents on the same side of the polymer chain is referred to as *isotactic*, and often exhibits some degree of crystallinity. In contrast, if adjacent substituents are arranged on alternating sides of the polymer chain, a *syndiotactic* polymer is formed. Unless a

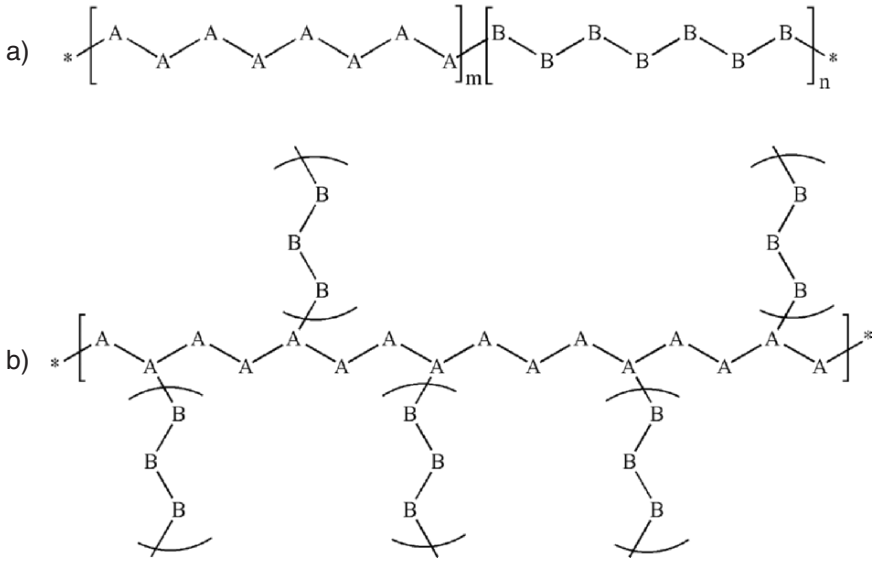


Figure 5.5. Illustration of (a) block copolymers and (b) graft copolymers.

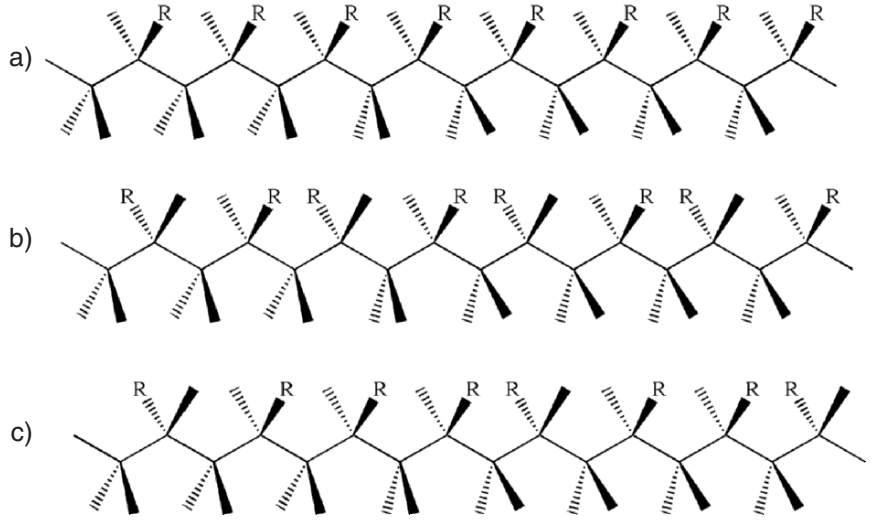


Figure 5.6. The tacticity of polymer chains. Illustrated are (a) isotactic, (b) syndiotactic, and (c) atactic polymers.

polymerization scheme uses a catalyst that confines the nucleation/propagation site, the substituents will be randomly organized – either disordered throughout the entire polymer chain (*atactic*) or along every other repeat unit (*hemiisotactic*). Due to the structural randomness, atactic polymers are almost always entirely amorphous.

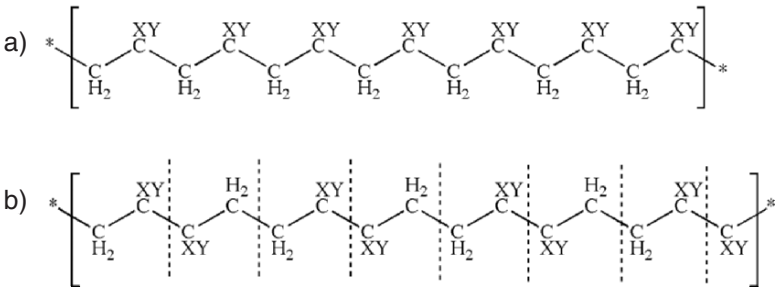


Figure 5.7. Illustration of sequence isomerization exhibited by polymer chains. Shown are (a) head-to-tail and (b) head-to-head/tail-to-tail sequencing.

Another form of isomerism that exists for polymers is the arrangement of adjacent repeat units. For an individual monomeric unit, the subsequent monomer may add in a head-to-tail or head-to-head/tail-to-tail fashion (Figure 5.7). The sequence is often determined by the substituents; for instance, sterically bulky groups (*e.g.*, phenyl) dictate a head-to-tail array for polystyrene. On the other hand, polymers that feature halide substituents (*e.g.*, poly(vinyl fluoride)) contain significant numbers of head-to-head/tail-to-tail sequences.

Thus far, we have considered the organization and composition of the main polymer chain, held together through covalent bonding of neighboring carbons. However, the physical properties of a particular polymer are more strongly affected by the intermolecular forces that exist between individual polymer chains. For instance, crosslinking through covalent bond formation is responsible for vulcanization of rubber (Figure 5.8a). In contrast, the greater flexibility of nylon is due to the relatively weaker interchain hydrogen bonding interactions (Figure 5.8b). Other types of inter- or intrachain interactions include dipole–dipole and van der Waals (induced dipole) forces (Figure 5.8c,d).

5.1.1. Polymerization Mechanisms

Polymers may be synthesized by two general mechanisms: *step-growth* (condensation) and *chain* (addition) polymerizations. Step-growth is the least limiting designation, relative to condensation, since the former is also appropriate for polymerizations that do not eliminate water during the reaction (Eq. 1). Table 5.1 lists the important distinctions between addition and step-growth polymerization schemes. As its name implies, step-growth involves the reactions of functionalized monomers to build up polymer chains through dimers, trimers, *etc.*, *en route* toward oligomers, and polymers. The acid/base catalyzed synthesis of SiO_2 networks by sol-gel (Chapter 2) is a widely used application of step-growth polymerization. In contrast, addition polymerization involves the activation/addition of unsaturated monomers, yielding a

$$n\mathrm{XRX} + n\mathrm{HYR'YH} \rightarrow \mathrm{X(RYR')}_n\mathrm{YH} + n\mathrm{HX} \quad (1)$$

(where X = OH, Cl, *etc.*; Y = NH, O)

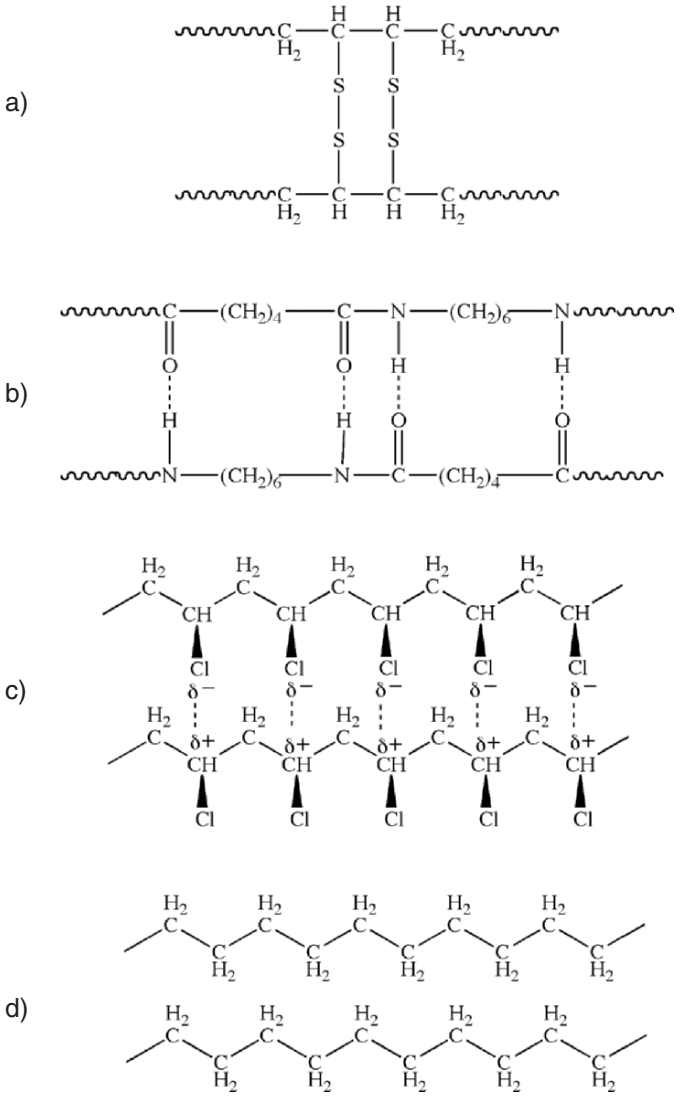


Figure 5.8. The intermolecular forces involved in adjacent polymer chains. Shown are (a) covalent crosslinking (vulcanized rubber), (b) hydrogen bonding (nylon 6,6), (c) dipole–dipole (PVC), and (d) van der Waal interactions (polyethylene).

much higher MW polymer in a relatively short period of time. Since step-growth proceeds through the reaction of neighboring complementary functional groups, small molecular byproducts such as HX, H_2O, *etc.* are generated; addition reactions do not yield such elimination products.

The molecular weight of a polymer may be generally described as the molecular weight of the monomer(s) multiplied by the *degree of polymerization* (DoP) or the

Table 5.1. Features of addition and condensation polymerization schemes

Addition (chain growth)	**Condensation (step-growth)**
1. Unsaturated monomers	Monomers contain ≥ 2 functional groups
2. No products are eliminated	Elimination of H_2O, HCl, *etc.*
3. Only monomer and polymer are present during polymerization	Monomers and polymer are accompanied by dimers, trimers, and oligomeric species
4. Only monomers add to the growing polymer	All intermediate species are reactive, and contribute to the growing polymer
5. Mechanism involves reacting with double bond by active species like free radicals or ions	Involves simple elimination reaction between monomer functional groups
6. Rapidly yields a high MW polymer; crosslinking is achieved through use of monomers with two double bonds (*e.g.*, divinylbenzene)	Molecular weight is typically lower than addition polymerization. The presence of small amounts of multifunctional monomers results in extensive crosslinking (gels)
7. Examples: polyolefins, polydienes, vinyl polymers, acrylic polymers	Examples: polyesters, polyamides, polycarbonates[a], epoxies

[a](fun fact) a polycarbonate layer is used between glass panels to absorb the energy of a bullet blast – "bullet-proof" glass.

number of repeat units. If all of the individual polymer chains were equivalent in length and composition, this would yield an exact molecular weight value analogous to molecular or ionic solids. However, due to the variance in polymer chain lengths, the best we can attain is to assign a molecular weight *distribution*, characterized by using an empirical technique known as *gel-permeation chromatography* (GPC).[2] A broad peak indicates that a greater variance in molecular weight is present, which is not desirable for most applications.

The range of the molecular weight distribution is referred to as the *polydispersity*, given by the polydispersity index (PDI). This value is calculated from the weight average molecular weight, $\overline{x}_w$, divided by the number average molecular weight, $\overline{x}_n$ (Eq. 2). The PDI will always be >1; however, as the polymer chains approach a uniform length, the PDI will approach unity. The types of polymerization used, as well as experimental conditions (*e.g.*, temperature and nature of the catalyst), are paramount in generating a polymer with a narrow polydispersity. For instance, typical addition polymerization, results in PDI values of *ca.* 10–20, compared to 2–3 for step-growth polymerization. However, using precise temperature control to limit termination mechanisms, PDIs of <1.5 may be generated for both techniques.

$$\text{PDI} = \frac{\overline{x}_w}{\overline{x}_n}, \tag{2}$$

where

$$\overline{x}_w = \frac{\sum_i N_i M_i^2}{\sum_i N_i M_i} \quad \text{and} \quad \overline{x}_n = \frac{\sum_i N_i M_i}{\sum_i N_i}$$

(N_i is the number of molecules with molecular weight M_i)

Not surprisingly, nature is far ahead of human ingenuity – biopolymers have PDI values very close or equal to one, indicating that only one length of polymer is present – directly responsible for the high specificity and efficiency of complex living systems.

Addition polymerization

Addition polymerization involves three steps: initiation, propagation, and termination. During initiation, either radicals (Figure 5.9) or ionic species are generated from the controlled decomposition of an initiator molecule. The reactive intermediates are then sequentially added to the C=C bonds of monomers to propagate the growing polymer chain. Free-radical polymerization is the most common method currently used to synthesize polymers from vinyl-based monomers.

Due to the high reactivity of the radical fragments, facile *chain-transfer* may occur (Figure 5.9d i,ii), whereby the radical end of the growing chain abstracts an

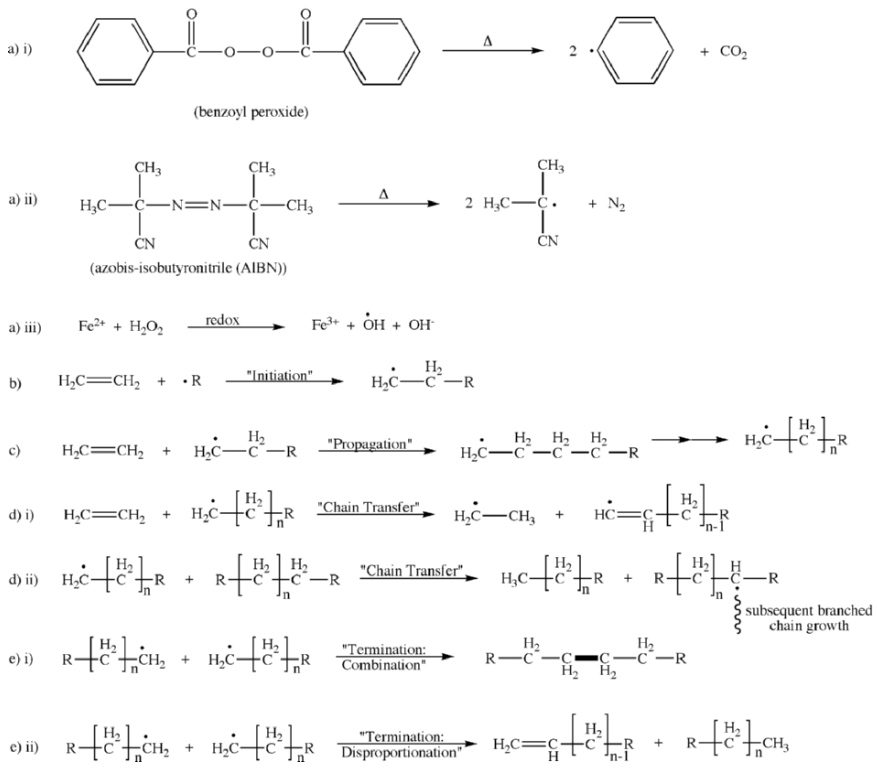
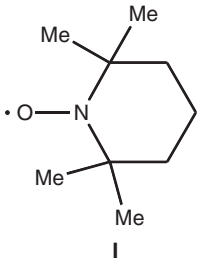


Figure 5.9. Reactions involved in free-radical addition polymerization. Shown are (a) (i)–(iii) generation of free radicals from a variety of initiators, (b) initiation of polymer chain growth through the combination of a free radical and unsaturated monomer, (c) propagation of the polymer chain through the combination of growing radical chains, (d) chain-transfer of free radicals between the primary and neighboring chains, and (e) termination of the polymer growth through either combination (i) or disproportionation (ii) routes.

atom from another molecule/chain. The second molecule may be monomer, solvent, initiator, or other polymers that exist in solution. As a result, the growth of the primary chain is terminated, and a new radical capable of propagation/polymerization is generated. Oftentimes, chain-transfer results in hydrogen abstraction from the second molecule, which causes branching (Figure 5.9d ii). The second molecule may be totally unreactive or moderately reactive with the monomer (*i.e.*, relative to normal propagating radicals). Hence, a deliberate addition of these molecules may be used to inhibit or slow the polymerization process.

Although free-radical polymerization is highly susceptible to chain-transfer, it is possible to suppress these side reactions – resulting in *living polymerization*. By definition, a living process will result in a linear increase in the polymer molecular weight with monomer consumption. In order to gain strict control over the polymerization process, a number of approaches have recently been designed:

(i) *Nitroxide-Mediated Polymerization* (NMP). A stable free radical (*e.g.*, **I**, 2,2,6,6,-tetramethyl-1-piperidinyloxyl (TEMPO)) is added to the solution, acting as a radical scavenger. The TEMPO radical is exceedingly stable due to the nearby methyl groups that supply electron density and help stabilize the unpaired electron that is delocalized over the N–O bond. The coupling of TEMPO with the polymeric radical is reversible, which allows one to control the molecular weight and polydispersity of the resultant polymer by varying the monomer:TEMPO ratio.



I

(ii) *Atom-Transfer Radical Polymerization* (ATRP). A halogenated organic and metal complex are added to the solution, which generate a radical initiator (Figure 5.10). As the monomer reacts with the initiator, the halide moiety preferentially terminates the chain. This process is considered redox-controlled polymer growth, since the polymer chain may grow only as additional atom-transfer reagents are added to the solution.

(iii) *Reversible Addition Fragmentation Chain-Transfer Polymerization* (RAFT). This process features the addition of RAFT agents, containing thiocarbonyl-thio groups. The reaction of radicals with the C═S bond forms a stabilized radical intermediate. In an ideal system, these intermediates do not undergo termination reactions, but rather reintroduce a radical capable of reinitiation or propagation with monomer, while they themselves reform their C═S bond (Figure 5.11).

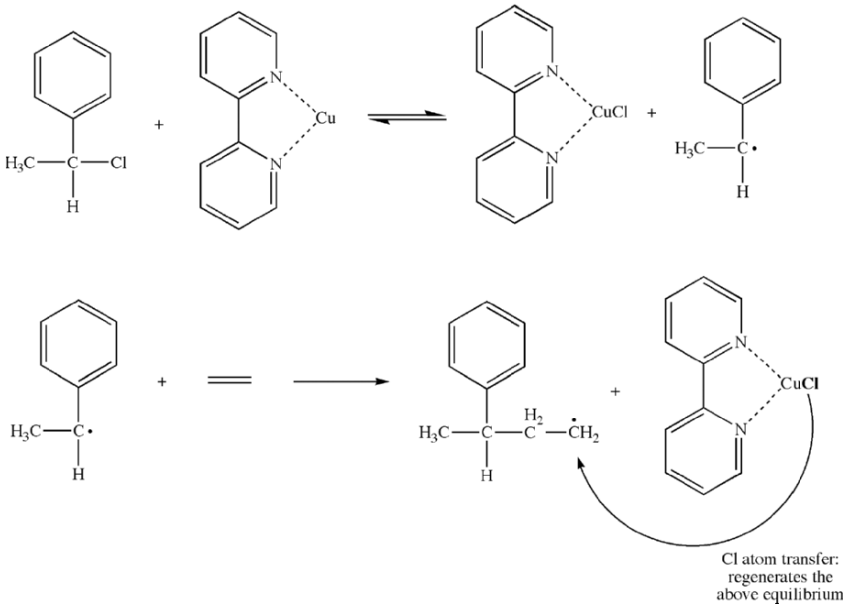


Figure 5.10. Mechanism of atom-transfer radical living polymerization. In this process, addition of atom-transfer agents results in initiating radicals that react with monomers. Rather than terminating polymer growth, halogenated end units are formed that are capable of propagating chain growth when additional monomer is added.

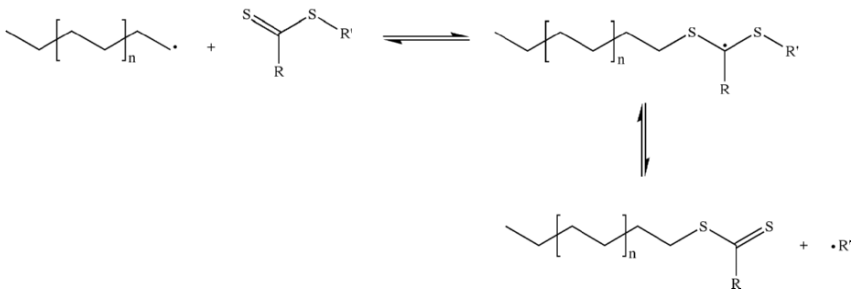


Figure 5.11. Illustration of the general reaction scheme responsible for RAFT polymerization.

The cycle of addition to the C═S bond, followed by fragmentation of a radical, continues until all monomer is consumed. Termination is limited in this system by the low concentration of active radicals.

A recent development in living polymerization are *iniferters* – a single molecule that is capable of *ini*tiating, trans*fer*ring, and *ter*minating the radical polymerization process (Figure 5.12). Upon absorption of UV light, an iniferter generates a carbon radical and sulfur-based dithiocarbamyl radical. Whereas the carbon radicals are extremely reactive toward the monomer, the dithiocarbamyl radical is not

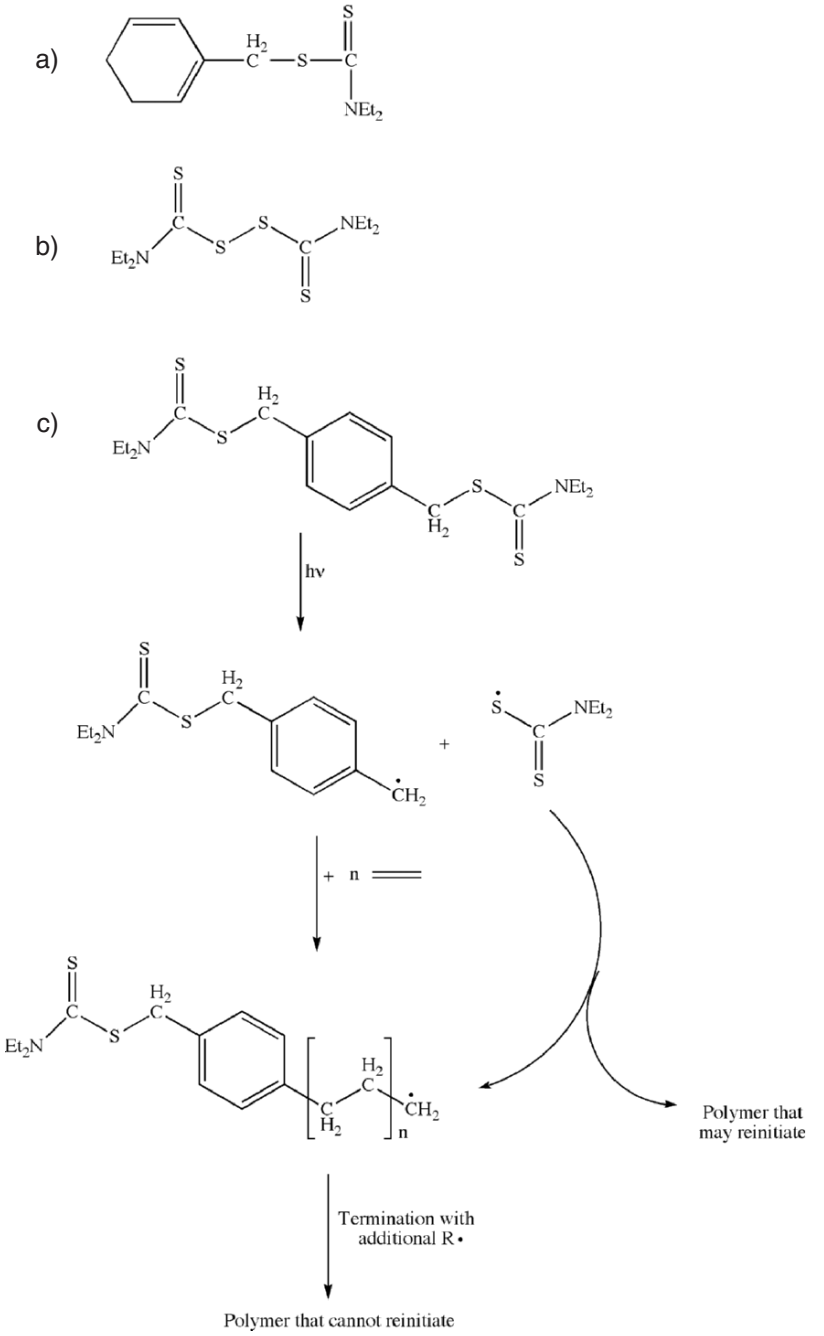


Figure 5.12. Molecular structures of common iniferters (a–c), and illustration of polymerization using an iniferter.

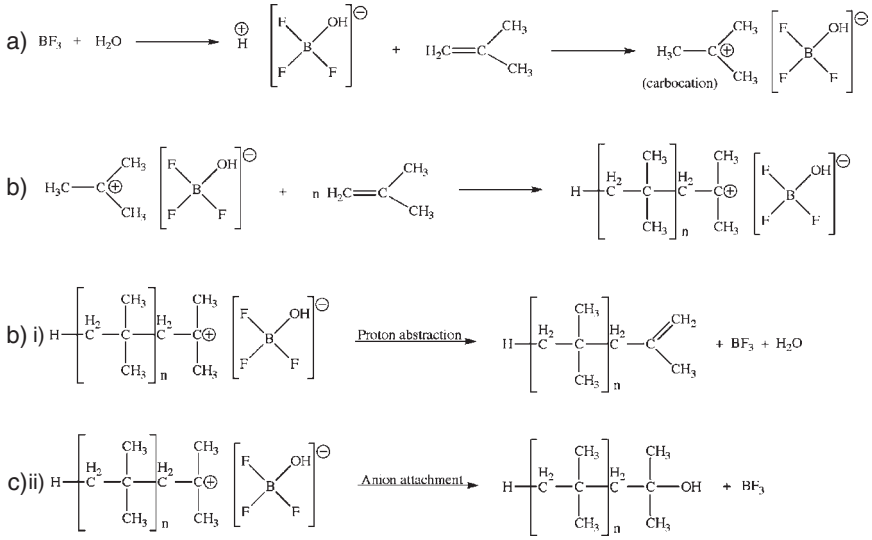


Figure 5.13. Reactions involved in cationic addition polymerization. Shown are (a) generation of a carbocation intermediate from a Lewis acid initiator, (b) propagation of the polymer chain through the combination of the carbocationic polymer chain and additional monomers, and (c) termination of the polymer growth through either proton abstraction (i) or anionic attachment (ii) routes.

sufficiently reactive toward propagation. Termination in iniferter systems may take place through either carbon–carbon or carbon–dithiocarbamyl bimolecular radical termination. The former route results in a dead unreactive polymer, whereas the latter route forms another iniferter species that may reinitiate upon UV light irradiation.

In addition to radicals being generated from initiators, cationic species may also be formed in an addition polymerization process (Figure 5.13). In these systems, a Lewis acid (*e.g.*, $AlCl_3$, BF_3, *etc.*) must also be accompanied by a protic Lewis base (*e.g.*, NH_3, H_2O), which renders the proton as the actual initiator (Figure 5.13a). By contrast, anionic polymerization routes utilize organolithium compounds (*e.g.*, butyllithium) or alkali metal amides (*e.g.*, $NaNH_2$) as initiators (Figure 5.14). In these ionic systems, propagation occurs through the combination of additional monomer with carbocationic/carbanionic intermediate species. For cationic polymerization, termination may occur through either proton or hydroxyl abstraction from the counteranion. By contrast, anionic polymerization is a "living" process; termination does not occur spontaneously, and is controlled through the addition of a Lewis base.

Heterogeneous catalysis

Thus far, we have considered the case for radical polymerization for the conversion of unsaturated monomers into polymer chains. However, none of the aforementioned techniques offer stereoselective control over the growing polymer chain, resulting in purely atactic polymers. In order to gain such control, it is necessary to spatially

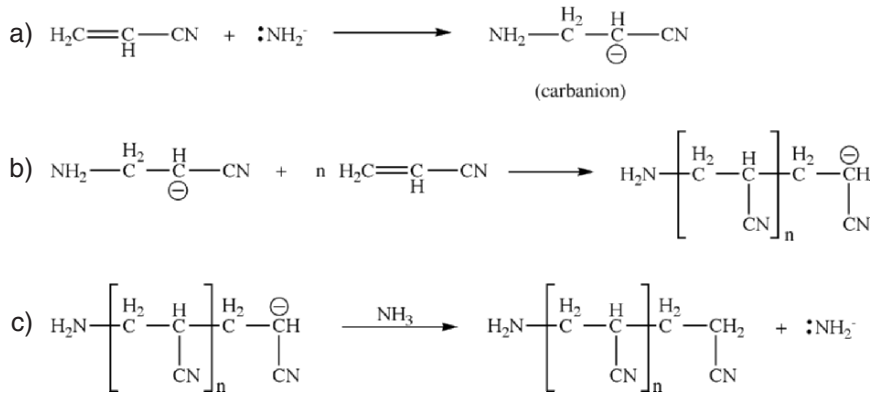


Figure 5.14. Reactions involved in anionic addition polymerization. Shown are (a) generation of a carbanion from a Lewis basic initiator, (b) propagation of the polymer chain through the combination of the carbanionic polymer chain and additional monomers, and (c) termination of the polymer growth through the addition of a Lewis base. Unlike the other addition polymerization schemes, termination does not occur *in situ*, but must be initiated deliberately.

confine the reactive site to control the direction of incoming monomer/growing polymer. The most common method used to control the tacticity of the resulting polymer is *Ziegler–Natta polymerization*.

Ziegler–Natta polymerization is an example of heterogeneous catalysis – a dual-phase system where polymerization occurs on the surface of the catalyst. A crystal of $TiCl_3$ or $TiCl_4$ is used in association with an aluminum alkyl co-catalyst (Figure 5.15). Since the Ti sites on the surface are co-ordinatively unsaturated, monomer may attach along a controlled direction. The aluminum (Lewis acidic) complex acts as an initiator, facilitating monomer coordination through abstraction of a Cl group from the Ti coordination sphere. The Lewis acidic Al site also assists in the intramolecular rearrangements that are essential for chain propagation.

If $TiCl_3$ is used as the catalyst surface, an isotactic polymer is formed. However, changing the surface to VCl_3 yields a syndiotactic product. This difference may be explained by looking at the relative sizes of Ti^{3+} and V^{3+}. Due to an increase in the effective nuclear charge (Z_{eff}) as one moves from left to right of the Periodic Table, the V^{3+} center is smaller which creates more steric hindrance among the coordinated polymer and incoming monomer. As a result, there is less room for the ligands to undergo an equatorial–axial shift for V^{3+}, and the incoming monomer may approach in two directions. Since this spatial ligand shift readily occurs for the larger Ti^{3+}, the monomer approaches along a single direction, resulting in an isotactic polymer. The decrease in reaction rate and polymerization efficiency upon substitution of Ti with V is likely an artifact of electronic differences. Since V^{3+} has an extra d-electron relative to Ti^{3+} (d^2 *vs.* d^1), there is added Coulombic repulsion with the π-electrons of the incoming olefin.

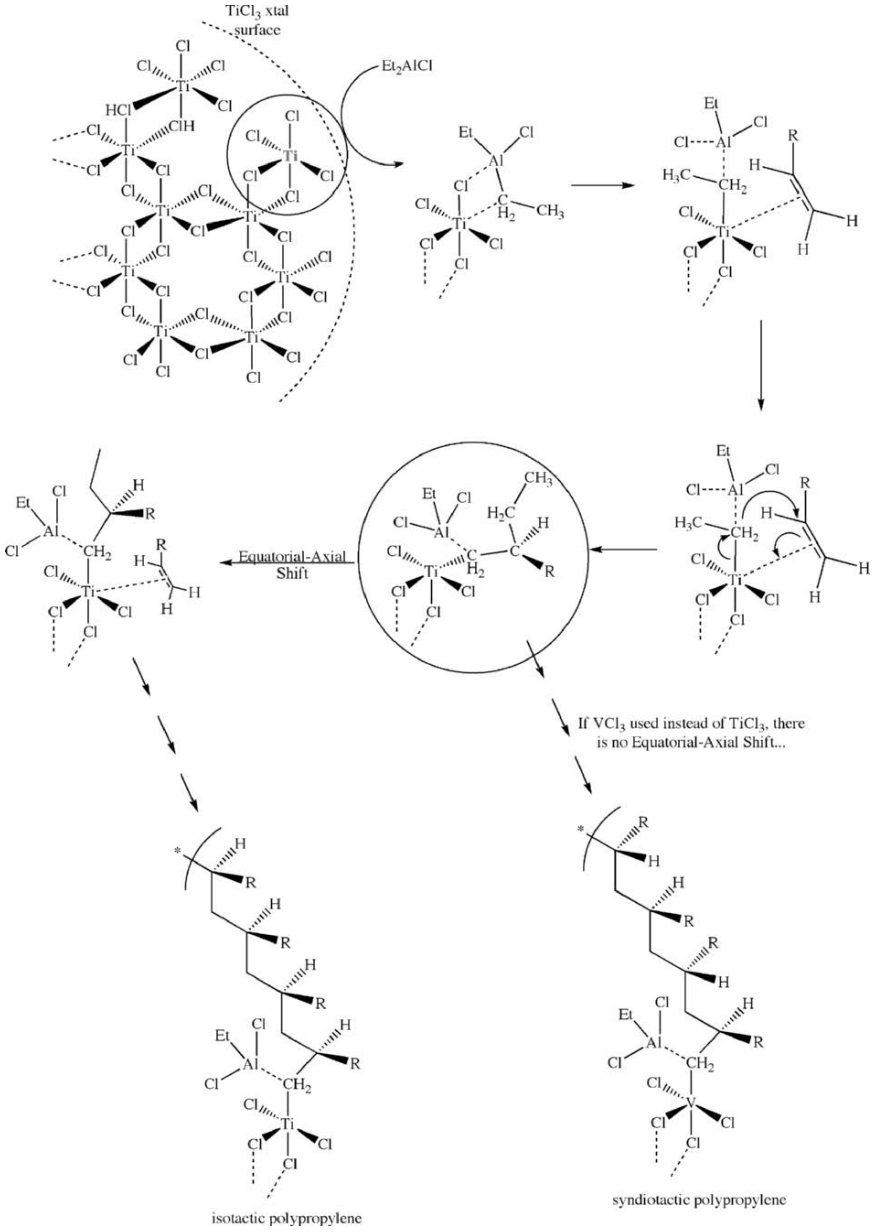


Figure 5.15. Mechanism involved in Ziegler–Natta polymerization for tacticity control over the growing polymer chain.

Homogeneous catalysis

Although separation of products from catalyst is easily accomplished within a heterogeneous system, the polydispersity of the product will be relatively high due to multiple reaction sites on the catalyst surface. In order to improve the overall selectivity, a variety of single-phase *homogeneous* catalytic routes have also been developed. Although the polydispersity of the products are much narrower using homogeneous catalysts, the primary limitation is the difficulty in separating products from reactants/catalyst. However, supercritical fluids (*e.g.*, CO_2 at *ca.* 40°C and 2,000 psi) are now commonly used as the solvent in these systems in order to facilitate product separation. Supercritical fluids have properties intermediate between gases and liquids; minute changes in pressure/temperature of the fluid near its critical point result in dramatic changes in its density and solubility characteristics.

Homogeneous polymerization catalysts ("Kaminsky type") are based on Group 4 metallocene molecules that contain bulky ligands (Figure 5.16). Based on their sterically encumbered metal site, it is easy to see how stereocontrol over the polymer is obtained – through limiting the path of approach for the incoming olefin. A more recent technique to offer further stereoselectivity is "heterogenizing" the catalyst by incorporating the metallocene structure into a solid support such as silica or alumina.

In order to obtain catalytic activity in the system, a co-catalyst featuring an electron-deficient center (*e.g.*, B, Al) must also be present in the solution. The mechanism for activity of aluminum oxide cocatalysts, referred to as *methyl alumoxanes* (MAOs – formed from the controlled hydrolysis of $AlMe_3$), has recently been elucidated – to abstract an alkyl or chloro group from the metallocene structure. For aluminum cocatalysts, this yields a $[(cp)_2MX]^+ \cdots [(^t Bu)_2AlX]^-$ cation/anion pair (X═R, Cl). This activates the metallocene structure toward polymerization, since the M^+ site more readily accepts electron density from the alkene monomer. In order for the ion pair to be active toward polymerization, the anion must be non-coordinating to ensure a high Lewis acidity toward the incoming olefin. Accordingly, one of the most effective cocatalysts is $B(C_6F_5)_3$ – a very strong Lewis

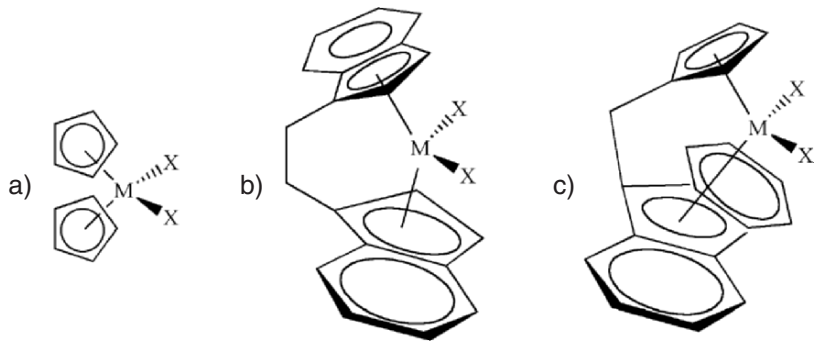
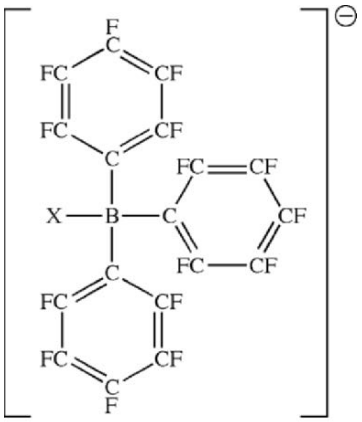


Figure 5.16. Molecular structures of common metallocene homogeneous polymerization catalysts (M═Zr, Hf, X═Me, Cl). The cyclopentadienyl ligands may be abbreviated as "cp" in the molecular formula (*e.g.*, $(cp)_2HfMe_2$).

acid due to highly electronegative fluorinated substituents that pull electron density away from the central B atom. Upon abstraction of Me or Cl groups, the resulting anion $[BX(C_6F_5)_3]^-$ (**II**) is noncoordinating toward the metallocene structure which freely allows the catalyst to accept electron density from the incoming monomer.



II

Since the Lewis acidity of the co-catalyst is most important toward its activity, it was first thought that three-coordinate Al complexes (Figure 5.17) would be most effective for activating the metallocene structure, relative to coordinatively saturated four-coordinate Al compounds. However, it has been shown that the latter structures $\{(CH_3)Al(\mu\text{-}O)\}_n$, $n = 6$–$12\}$, which exist as complex caged structures, are most effective for activating the metallocene catalyst. The term "latent Lewis acids" has been coined for these structures by Barron,[3] indicating that the isolated caged structures do not themselves possess acidity, but become activated due to a cage-opening mechanism upon contact with the metallocene catalyst (Figure 5.18). The magnitude of Lewis acidity is directly related to the cage strain; as one would expect, metallocene activation is more pronounced with increasingly smaller cages.

Three mechanisms have been proposed to explain metallocene-based homogeneous and Ziegler–Natta polymerization schemes. The Cossee–Arlman mechanism (Figure 5.19a) proposes the coordination of the olefin to the vacant metal site, with its subsequent migratory insertion into the M–P (P = growing polymer) bond. The insertion step likely proceeds through a four-membered cyclic transition state. Another mechanism proposes an initial α-hydride shift, resulting in a metal hydride species that may add an olefin and form a growing polymer chain through migratory insertion (Figure 5.19b). The last mechanism that involves α-agostic Ti–H interactions (Figure 5.19c) is generally accepted as being an important part of the Cossee–Arlman mechanism, serving to slow down the insertion reaction and favor the approach of incoming ligand.

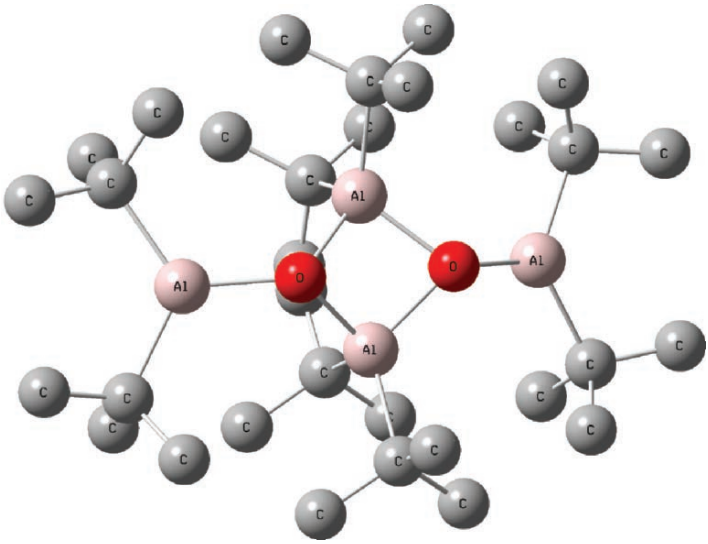


Figure 5.17. Molecular structure of $[(^{t}Bu)_2Al(\mu\text{-}O)Al(^{t}Bu)_2]_2$, featuring three-coordinate, unsaturated Al centers.

It should be noted that early transition metals (Group 4 or 5) are most favorable for transition metal-catalyzed living polymerization. Since these metals have few d-electrons in their valence shell, there is a lesser chance for β-hydride elimination (polymer termination) to occur.[4] In order to purposely terminate chain growth, either the M–C bond may be broken through reaction with hydrogen gas (Figure 5.20a) or thermally induced β-hydride elimination. Whereas the first termination route results in low polydispersity, the latter method results in terminal olefin (Figure 5.20b) that may insert into a neighboring polymer chain yielding a branched, low-density product.

Step-growth polymerization

Thus far, we have considered addition polymerization routes – either catalyzed or uncatalyzed. Although this is sufficient to describe the synthesis of common packaging materials such as polyethylene, polypropylene, polystyrene, *etc.*, other classes of polymers such as nylon, PETE, and polyacrylamide are generated through step-growth mechanisms. Although the synthetic pathway for these polymers is more straightforward than addition polymerization, there are many intricate considerations that affect overall polymer properties.

The general types of step-growth polymerizations are shown in Figure 5.21. Since the resulting polymers contain reactive functional groups on either/both ends, it is referred to as a *telechelic* macromolecule. Although water is typically released as a byproduct of these polymerizations, other small molecules such as alcohols and alkyl halides may also be generated based on the monomers that co-condense. In its most simplest form, condensation reactions result in linear polymers from the reaction of

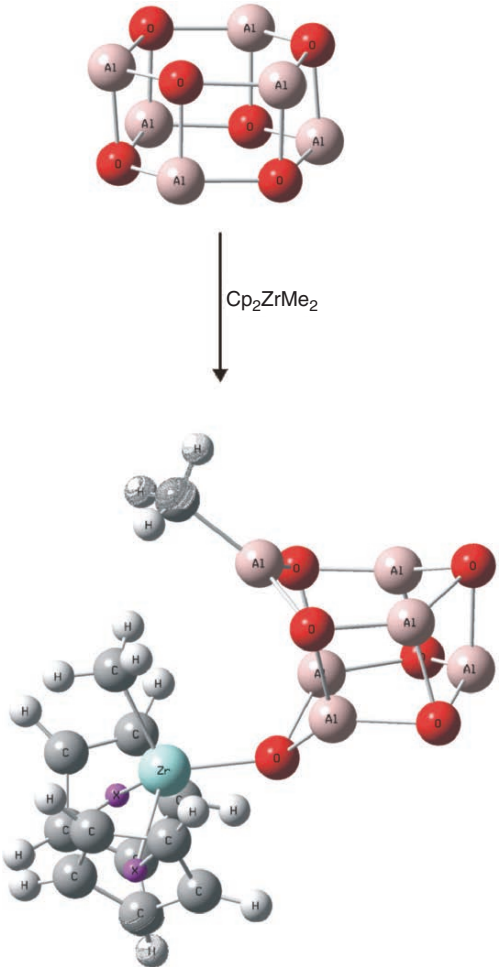


Figure 5.18. Molecular structure of $[Me_2AlO]_6$, and schematic of the cage-opening mechanism of the alumoxane co-catalyst during metallocene-catalyzed polymerization.

bifunctional monomers (Figure 5.22a, where A and B indicate complementary functional groups such as –COOH and –OH). For example, one of the most common linear step-growth polymers is poly(ethylene terephalate) (PET; commonly denoted by the trademarks Dacron or Mylar), used for soft drink bottles due to its impermeability toward liquids and gases. This polymer is obtained through a two-step reaction between ethylene glycol ($HO–C_2H_4–OH$) and the dimethyl ester of terephthalic acid ($CH_3O–C(O)–benz–C(O)–OCH_3$). Both methanol and ethylene glycol byproducts are evolved during the condensation reaction.

In contrast, if the monomers are multifunctional rather than difunctional, a branched polymeric structure will result (Figure 5.22b). For instance, consider either

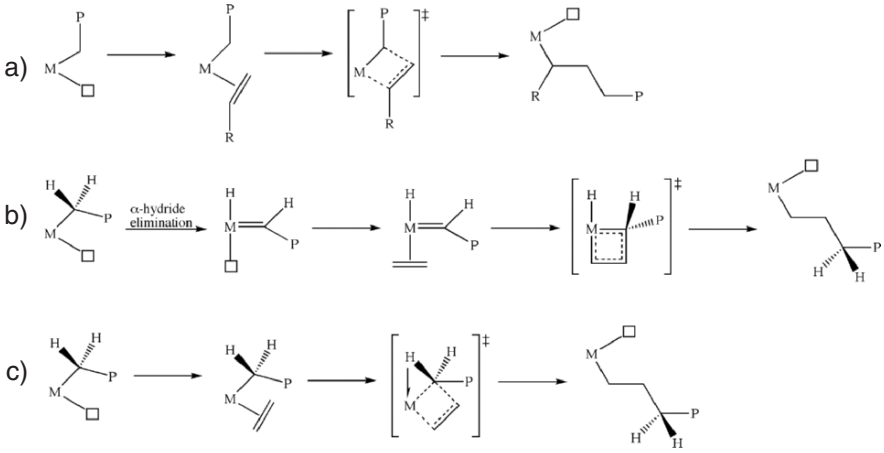


Figure 5.19. Proposed mechanisms for transition metal-catalyzed olefin polymerization.

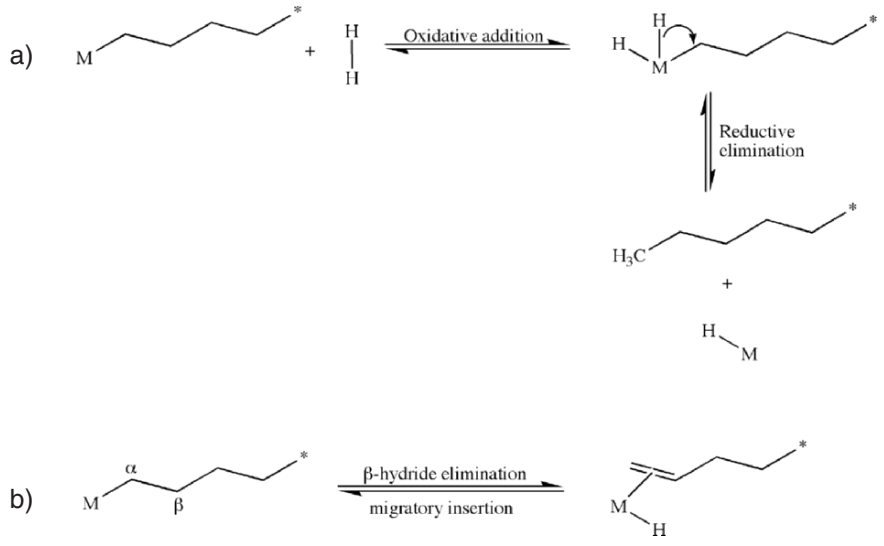


Figure 5.20. Schematic of (a) hydrogenolysis termination of a polymer chain and (b) β-hydride elimination termination, which results in an olefin-terminated polymer chain.

the "$A_x + B_y$" or the "$(AB_2)_n$" systems, where A and B refer to the monomers (subscripts indicate the number of reactive functional groups). In describing branched step-growth polymers, it is generally assumed that:[5]

(i) A and B endgroups react only with each other.
(ii) All endgroups of a given type are equally reactive.
(iii) Intramolecular cyclization reactions do not occur.
(iv) Probability of A–B reactions is independent of molecular size.

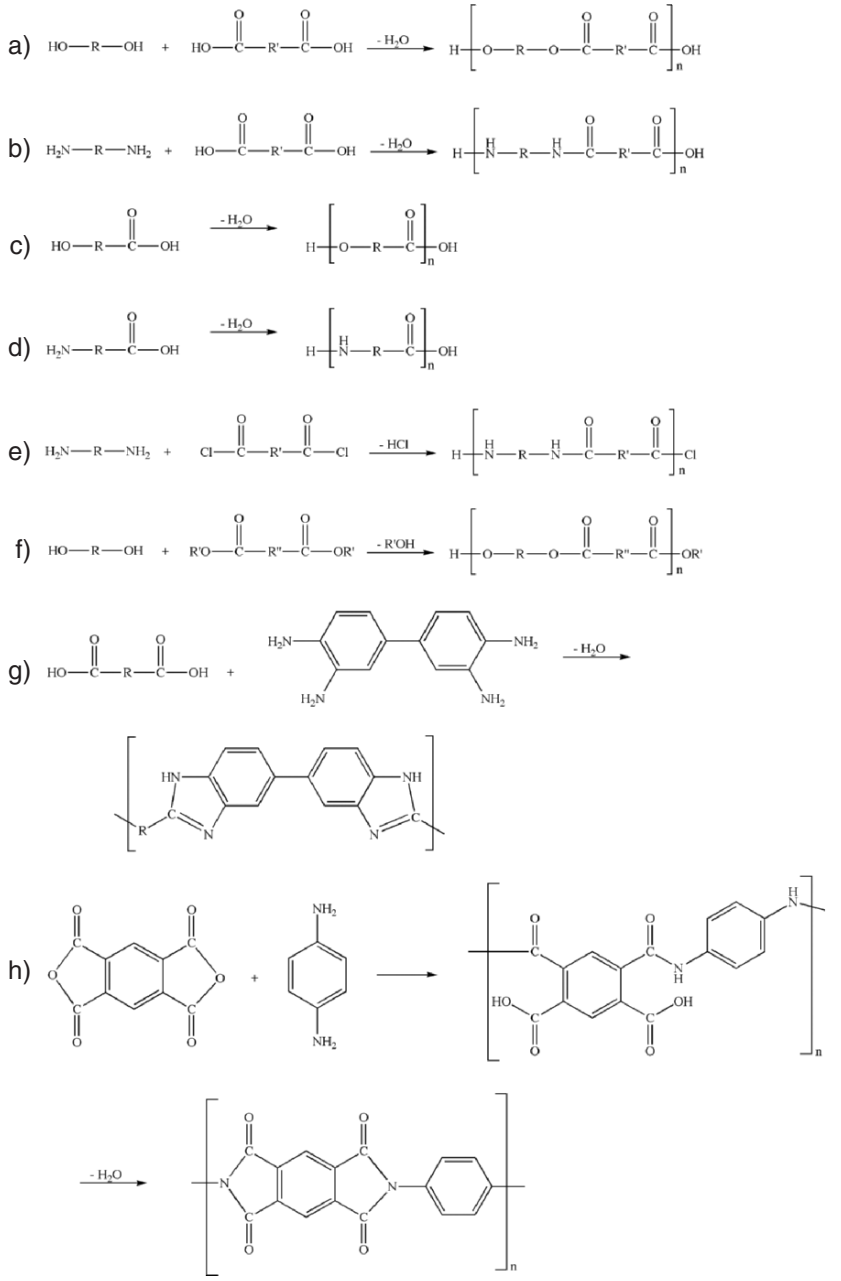


Figure 5.21. Reaction schemes for the most common types of step-growth polymerization. Shown are (a/c) polyester formation, (b/d) polyamide formation, (e) polyamide formation through reaction of an acid chloride with a diamine, (f) transesterification involving a carboxylic acid ester and an alcohol, (g) polybenzimidazole formation through condensation of a dicarboxylic acid and aromatic tetramines, and (h) polyimide formation from the reaction of dianhydrides and diamines.

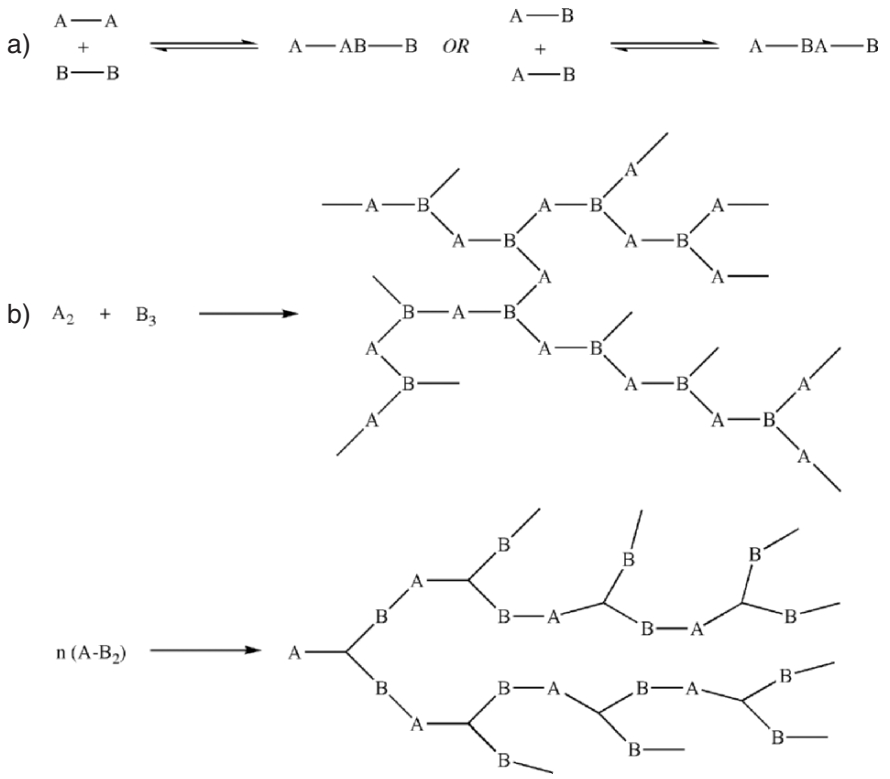


Figure 5.22. Formation of (a) linear and (b) random hyperbranched polymers.

It is interesting to note that whereas $A_x + B_y$ hyperbranched systems have both A and B endgroups, AB_x-based hyperbranched polymers contain only B endgroups. This leads to the discrepancy between the two systems in forming infinite networks (gels). The $A_x + B_y$ system results in gel formation if $x \geq 2$ and $y \geq 3$, whereas it is not possible to form a gel via the AB_x system. Though the polydispersity of hyperbranched polymers is typically quite large, there is increasing interest for a number of blends/coatings applications due to the high density of peripheral groups, and resultant enhanced solubility/surface adhesion, relative to linear analogues. As a more recent extension, network structures that consist of interwoven hyperbranched polymers may also be synthesized (Figure 5.23), which have been proven useful for lithographic applications.

Thus far, we have only considered step-growth polymers obtained through random condensation reactions of multifunctional monomers. The pioneering work of Nobel Laureate Flory[6] in the 1940s helped the polymer community understand the kinetics of branched polymer growth, and suggested that control over sequential step-growth should be possible. However, this was not proven empirically until the work of Vögtle[7] and Tomalia[8] in the late 1970s–early 1980s. Vögtle developed

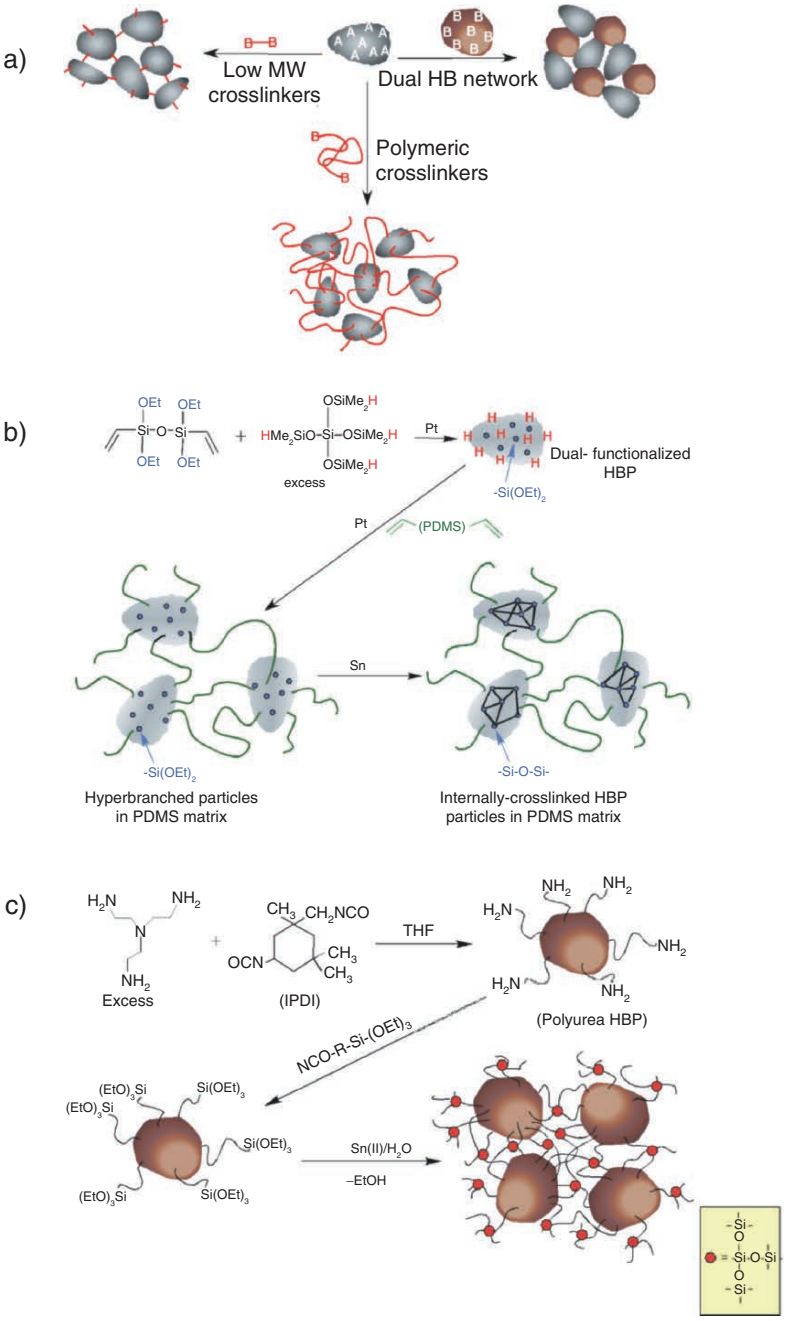


Figure 5.23. The general strategy (a) and examples (b–c) of hyperbranched polymer network structures featuring a siloxane crosslinkage. Reproduced with permission from Meijer, D.; Dvornic, P. R. Fall 2005 ACS meeting, Midland, MI.

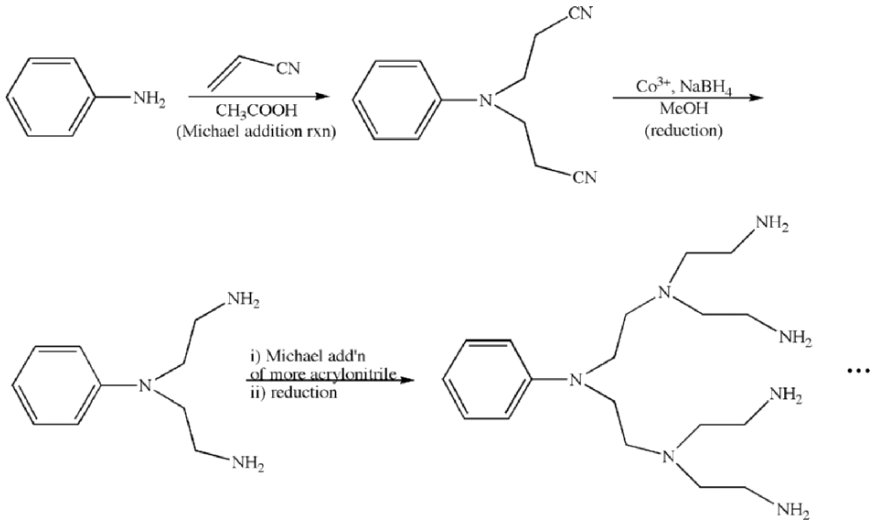


Figure 5.24. The Vögtle approach to yield low molecular weight amines via controlled sequential synthesis.

a repeatable "cascade" route to produce low molecular weight amines (Figure 5.24). However, due to cyclization side reactions, successful polymer growth via the Vögtle approach was not possible – finally being realized in 1993 for the synthesis of poly(propyleneimine) (PPI). The work of the Tomalia group at Dow Chemical was the first to yield a perfectly defined dendritic polymer structure, with an extremely low polydispersity (*ca.* PDI of 1.00–1.05; Figure 5.25).[9] These polymers were coined *starburst dendrimers*, referring to the star-branched architecture and the Greek word "dendra" for tree. Not unlike other major scientific discoveries, the first report of the dendritic architecture was riddled with skepticism by the scientific community. Shortly thereafter, Newkome helped silence the critics with his publication of branched dendrimers that he called an *arborol* (Figure 5.26).[10]

The earliest syntheses of dendritic polymers were divergent in nature; with growth initiating from a core, and outward propagation. The terminal groups of the core are reacted with complementary groups on the monomer, which forms a new branching point for subsequent branching reactions (Figure 5.27a). Most importantly, the terminal functional groups on the monomers are designed to be reactive with only the outwardly growing polymer, which prevents random hyperbranched growth. Such a repetitive procedure results in an exponential increase in reactions that occur on the periphery of the growing polymer, requiring a large excess of reagents. Though this technique is used for the large-scale synthesis of many dendrimers (*e.g.*, poly(amidoamine) – PAMAM), a leading drawback is the relatively high number of defect structures – especially for higher *generations* (a term used to describe the sequential branches emanating from the core of the dendritic structure). With each generation, there is an increased probability for incomplete/side reactions, which are

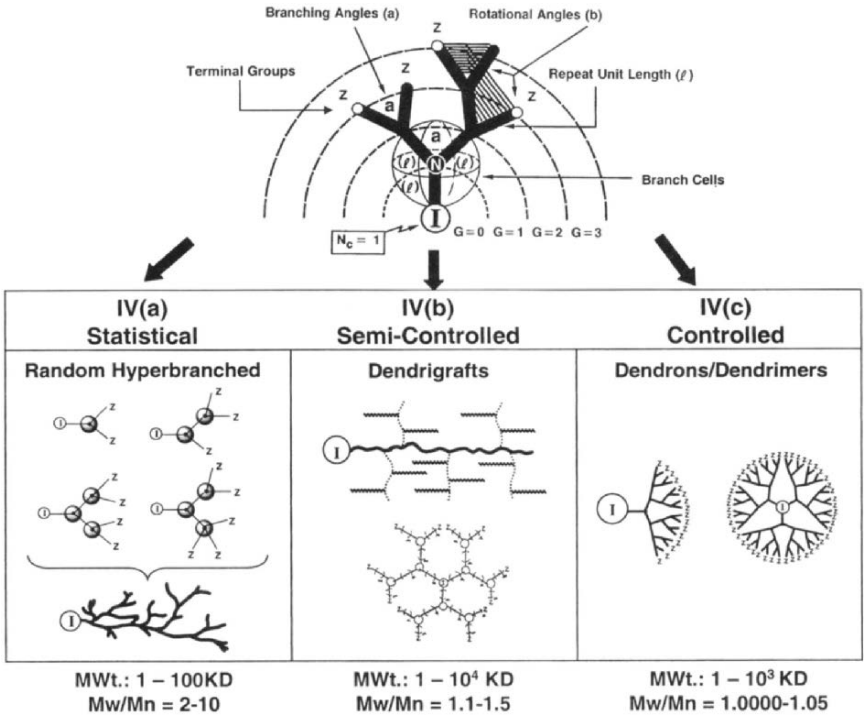


Figure 5.25. Illustration of resultant polymers through varying the degree of control of step-growth polymerization. Each successive growth layer is referred to as a *generation* (G). Reproduced with permission from Frechet, J. M. J.; Tomalia, D. A. *Dendrimers and Dendritic Polymers*, Wiley: New York, 2001.

not easily removed from the solution due to their structural similarity to the final product.

In order to circumvent the purity issues associated with divergent syntheses, Frechet and coworkers designed a convergent approach in the late 1980s.[11] In contrast to the divergent approach, growth initiates from the exterior of the molecule progressing inwardly by coupling endgroups to each branch of the monomer (Figure 5.27b). The functional group at the focal point of the wedge-shaped dendritic fragment (known as a *dendron*) may be reacted with additional monomers to build up higher-generation dendrons. When the desired generation dendron is reached, these units are then attached to a polyfunctional core to form the final dendrimer. This route drastically increases the purity of higher-generation dendrimers relative to divergent syntheses, as there are much fewer reactions per molecule.

In addition, the reactions only require a slight excess of reagent, in contrast to the large excesses that were essential for divergent growth. However, this technique is not useful for commercial large-scale dendrimer synthesis, as the mass of the sample decreases with additional generation growth, and low yields of higher-generation dendrimers due to steric crowding around the focal point of the growing dendron.

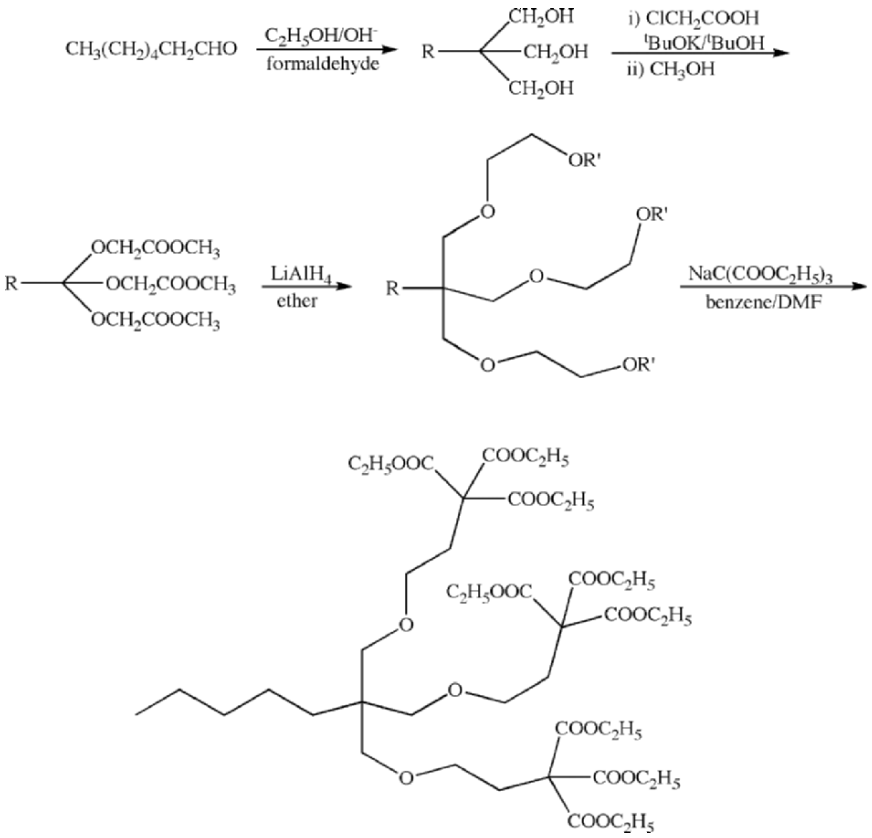


Figure 5.26. The Newkome approach for the sequential step-growth of arborols.

Nevertheless, this is the only route that offers precise structural control over the growing dendrimer, such as being able to modify the focal point/chain ends to yield well-defined unsymmetrical dendrimers. This strategy is being developed to synthesize "bowtie" dendrimers containing both target and drug-delivery agents (Figure 5.28). As an alternative strategy for drug delivery, surface modification of PAMAM with cancer targets and anticancer drugs (Figure 5.29) has also been proven successful in preliminary trials.

To date, the PAMAM dendrimer remains the most heavily utilized for applications, due to its facile scale-up and commercial availability. The first "copolymer dendrimer" was developed by Dvornic and coworkers, which featuring both a hydrophilic PAMAM core and a hydrophobic organosilicon shell (Figure 5.30).[12] Among many other possible applications, this structure proves extremely useful for the encapsulation of air-sensitive polar species within organic solvents (Chapter 6). Due to the water-sensitive alkoxysilyl groups (*e.g.*, Si–OCH_3), facile network formation may also take place (Figure 5.31) via the analogous hydrolysis reactions

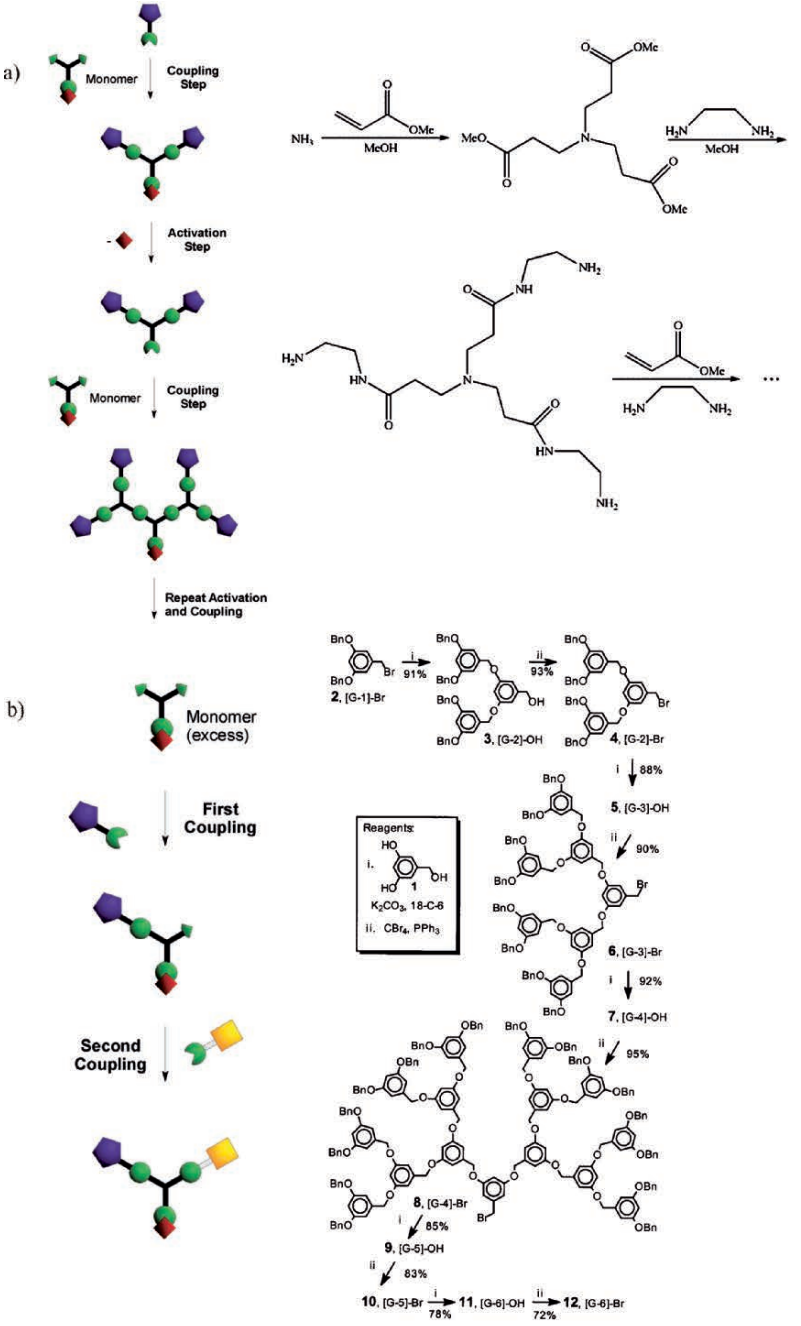


Figure 5.27. Comparison of (a) divergent and (b) convergent dendrimer synthetic routes (adapted with permission from Grayson, S. M.; Frechet, J. M. J. *Chem. Rev.*, **2001**, *101*, 3819. Copyright 2001 American Chemical Society).

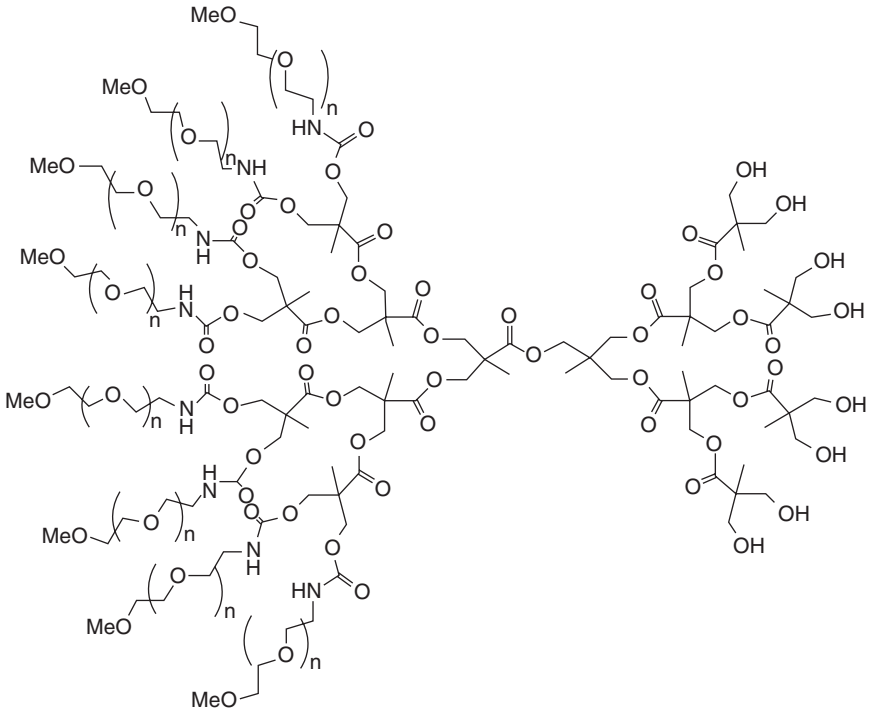


Figure 5.28. Bowtie dendrimer synthesized via the convergent approach, for drug delivery of anticancer drugs to target organs. Reproduced with permission from Gillies, E. R.; Dy, E.; Frechet, J. M. J.; Szoka, F. C. *Mol. Pharm.*, **2005**, *2*, 129. Copyright 2005 American Chemical Society.

that were previously discussed for sol-gel growth of SiO_2 networks (Chapter 2). The crosslinking of dendritic units to form an extended network is referred to as a *megamer* – of increasing interest for functional coatings (*e.g.*, sensors, smart fabrics, *etc.*) applications. A further utility of the PAMAMOS structure is its reactivity toward a glass surface, which contains silanol (Si–OH) reactive groups (Figure 5.32). This results in a permanent coating, with a controllable degree of surface adsorption based on the peripheral groups of the dendrimer.

Although we have described the growth of dendritic polymers as being highly controllable, the resultant size of the polymer is mathematically limited. This is in direct contrast to linear polymers that may increase in size to infinity (as long as they remain soluble within the solvent). As the dendritic structure grows, there become significant steric interactions among the exponentially increasing number of peripheral groups. This phenomenon is known as the *De Gennes dense packing*,[13] and results in a more structurally flawed, globular structure as the dendrimer generation increases.

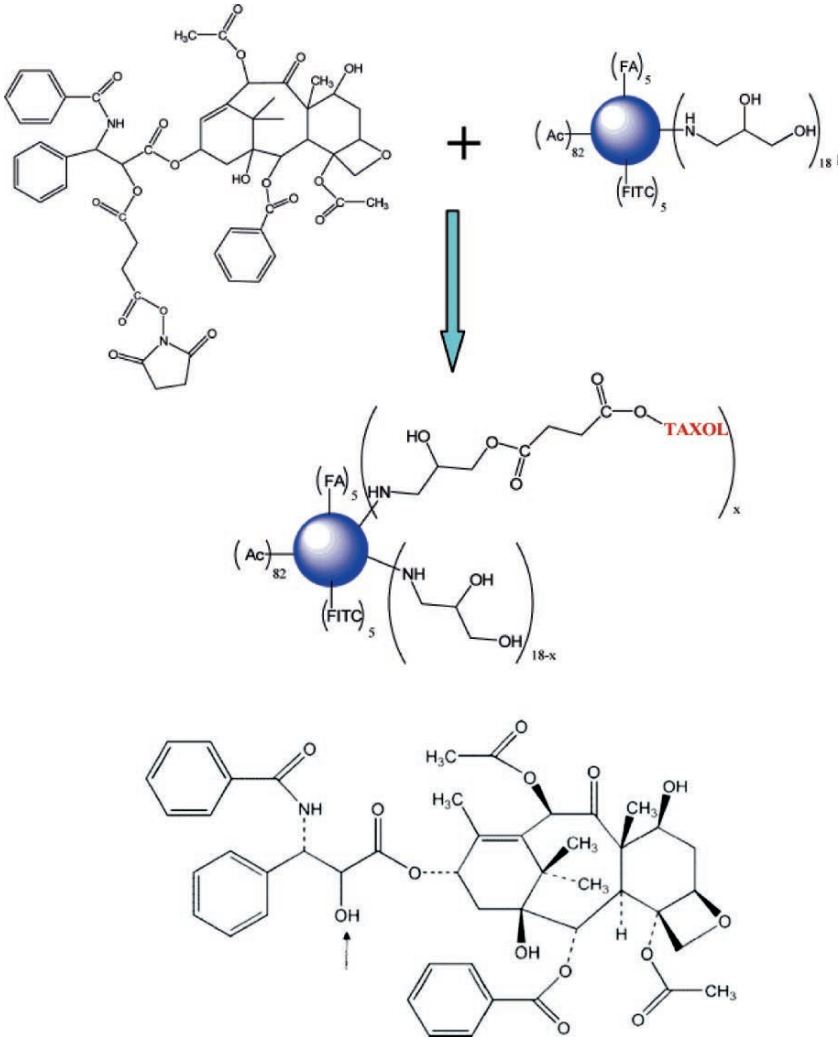


Figure 5.29. PAMAM dendrimer multifunctional conjugates for cancer treatment. The FA group is a folic acid cancer cell target, and FITC is fluorescein isothiocyanate, used as an imaging agent. Also shown (bottom) is the molecular structure for the anticancer drug, taxol, denoting the –OH group that covalently attaches to the dendrimer. Reproduced with permission from Majoros, I. J.; Myc, A.; Thomas, T.; Mehta, C.; Baker, J. R. *Biomacromolecules*, **2006**, 7, 572. Copyright 2006 American Chemical Society.

5.2. "SOFT MATERIALS" APPLICATIONS: STRUCTURE *VS.* PROPERTIES

Thus far, we have considered a large number of polymer types, which are used for a diverse range of applications. In this section, we will now consider how the molecular structure of the polymer drastically affects its properties. For polymers, only slight

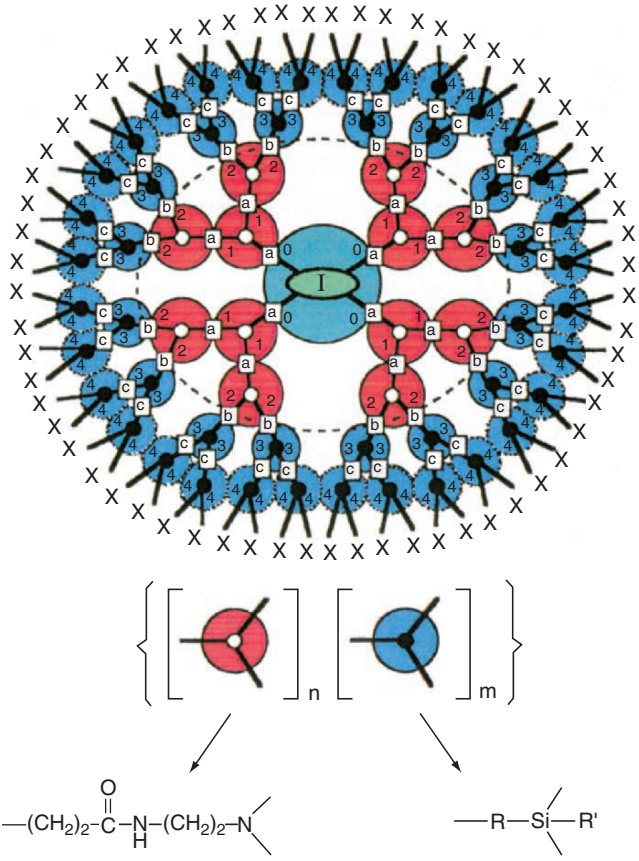


Figure 5.30. Illustration of a poly(amidoamine-organosilicon) (PAMAMOS) dendrimer, with two generations of each PAMAM and organosilicon units. Although the PAMAMOS represents a block copolymer, an unlimited number of other variations that contain a random copolymer array, or varying dendron subunits, may also be synthesized. Reproduced with permission from Dvornic, P. R.; Owen, M. J. *Synthesis and Properties of Silicones and Silicone-Modified Materials*, ACS Symposium Series 838, **2002**, 236.

changes in the polymer backbone, or interaction with neighboring polymer chains, will lead to very different physical properties (and resultant applications). Whereas the backbone structure affects the polymer's general flexibility, the intermolecular interactions between side groups of neighboring polymer chains affect the overall strength, solubility, crystallinity, and glass-transition temperature (T_g).

As you might expect, the influence of the nature and density of side groups *vs.* the backbone structure will strongly depend on the relative composition of each polymer unit, as well as how open the structure is to its environment. Regarding the relative composition of a polymer, even though a polymer may have a polar backbone structure (*e.g.*, ether and/or ester linkages), the structure may not be soluble within polar solvents. That is, if long nonpolar side chains are also present in the polymeric

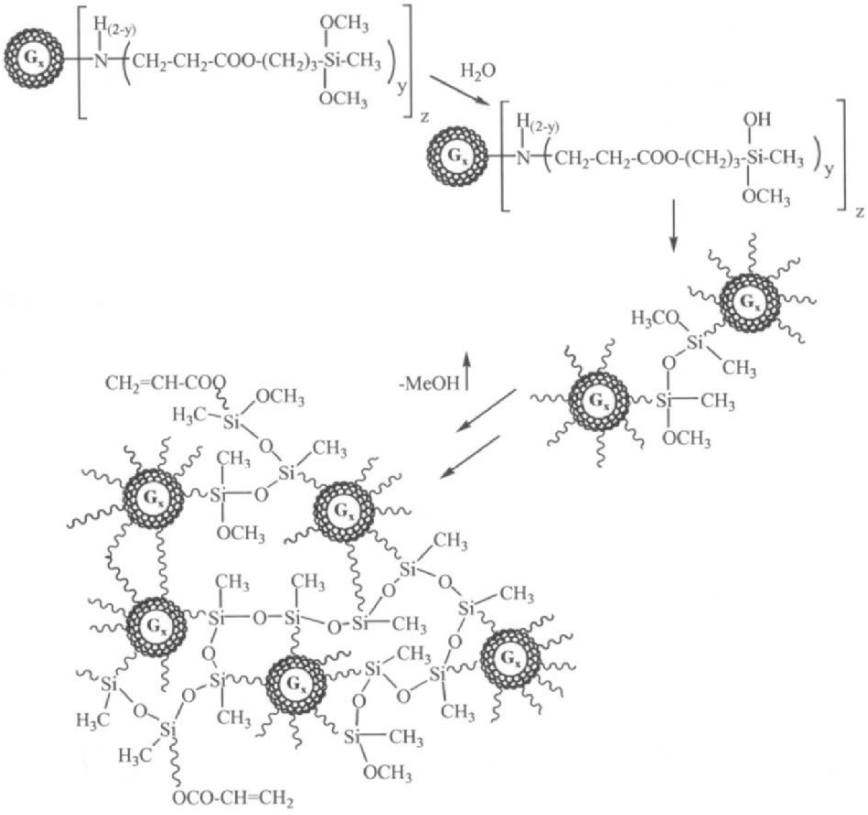


Figure 5.31. Network (megamer) formation through the hydrolysis/crosslinking of neighboring PAMAMOS dendrimer units. Hydrolysis of the C–O–Si bond may also be exploited for the controlled-release of entrained agents (*e.g.*, cancer drugs, *etc.*). It should be noted that subsequent thermal annealing to remove the PAMAM cores results in a nanoporous network that has a dielectric constant (k) of *ca.* 1.5 – of extreme interest for next-generation IC interconnect applications. Reproduced with permission from Dvornic, P. R.; Li, J.; de Leuze-Jallouli, A. M.; Reeves, S. D.; Owen, M. J. *Macromolecules*, **2002**, *35*, 9323. Copyright 2002 American Chemical Society.

structure, its solubility and reactivity will be outweighed by the presence of hydrocarbon chains. A useful analogy to keep in mind is the pronounced decrease in water solubility of simple alcohols, from methanol (completely miscible) to hexanol and higher-hydrocarbon chains (complete immiscibility) as the overall ratio of non-polar:polar functionality increases.

Regarding the interaction of a polymer with its environment, consider the PAMAM dendrimers described in the previous section. For higher generations, the pronounced density of the peripheral functional groups prevents the solvent or other environmental molecules from interacting with the core groups. As a result, the solubility/reactivity of these polymers may be fine-tuned by varying the surface functional groups. As we will see in Chapter 6, the inner-core shielding exhibited

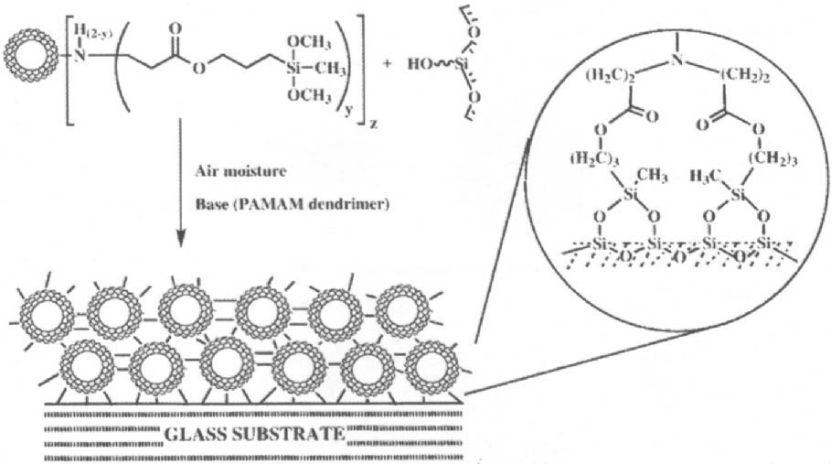


Figure 5.32. The formation of covalently bound coatings of PAMAMOS onto a glass surface. Reproduced with permission from Dvornic, P. R.; Li, J.; de Leuze-Jallouli, A. M.; Reeves, S. D.; Owen, M. J. *Macromolecules*, **2002**, *35*, 9323. Copyright 2002 American Chemical Society.

by dendrimers (*e.g.*, G4–G6 for PAMAM or PPI) allows one to encapsulate a number of species within these "nanocavity reactors" for a number of useful applications for novel nanomaterials synthesis and catalysis.

Table 5.2 lists the various functional groups that may be present in the polymer backbone and as side-groups, with their general effect on overall properties. It should be noted that these comparisons are oversimplications; for example, the presence of crosslinking and/or combinations of different functionalities will yield very different polymer properties. As discussed above, the general trend of chemical inertness and solubility will depend on both the openness of the structure and nature/density of the side groups. For thermal stability, conductivity, and flexibility, the polymer backbone structure is most relevant. Changes from single to multiple bonding, and/or introduction of a delocalized resonance unit in the polymer chain, will have pronounced effects on overall properties. One factor that is not explicitly mentioned in Table 5.2 is the degree of branching. However, it should be intuitive that the greater degree of interaction among adjacent polymer chains will enhance the tensile strength and density of the polymer, as well as increase the T_g. Perhaps the best example of this behavior is for highly branched *low-density polyethylene* (LDPE; used for squeeze bottles, food packaging film, plastic tubing, *etc.*) *vs.* weakly branched *high-density polyethylene* (HDPE; used for tupperware, milk cartons, plastic bags, *etc.*) (Figure 5.33).

Though most of the backbone units appearing in Table 5.2 are carbonaceous, there is one major class of polymers that are comprised of crosslinked $[O–SiR_2–]_n$ units. These polymers are known as *silicones* or *polyorganosiloxanes* – infamously popular due to the bad press over a decade ago concerning silicone breast implants. By varying the –Si–O– chain lengths, Si-alkyl groups, and extent of crosslinking, the resultant polymer may exist as a viscous liquid (*e.g.*, vacuum pump oil), a gel (*e.g.*,

Table 5.2. Influence of polymer structure on resultant properties

[b]Backbone unit/[s]substituent	Induced molecular properties
[b]Saturated carbon (–C–C–)	Chain flexibility, thermal/oxidative reactivity
[b]Unsaturated carbon (–C═C–)	Chain rigidity, high T_g, oxidative reactivity
[b]Aromatic (–C–C═C–)	Chain rigidity, colors, electrical conductivity (if aromatic rings: oxidative resistance, high strength, stacking/self-alignment (liquid crystals))
[b]Ether (–C–O–C–)	Chain flexibility, oxidative/hydrolytic stability (unless in Lewis acidic media), soluble in polar solvents (for small substituents)
[b]Anhydride (–C–C(O)–O–C(O)–C–)	Chain flexibility, water sensitive (especially for short aliphatic chains)
[b]Amide (–NH–C(O)–)	Chain rigidity, crystalline, water sensitive
[b]Siloxane ($–R_3Si–O–$)	Chain flexibility and low T_g (especially for small substituents), stable toward oxidation, acid/base reactive
[b]Phosphazene ($–R_3P═N–$)	Chain flexibility, high chemical inertness
[b]Sulfur (–S–S–)	Chain flexibility, thermal/oxidative reactivity
[s]Hydrogen (–H)	Chain flexibility, low T_g; if bound to non-C atoms: H-bonding (high T_g, m.p., chemical reactivity)
[s]Alkyl ($–CR_3$)	Chain rigidity, hydrophobicity, chemical inertness, noncrystallinity, solubility in nonpolar solvents (properties dependent on size of R groups – most pronounced for aryl groups)
[s]Halogens (–F, –Cl)	Fluoride: extreme chemical inertness, hydrophobicity, insolubility in virtually any solvent Chloride: chemical inertness, chain stiffness, photolytic sensitivity, ↓ solubility, fire retardancy
[s]Hydroxyl (–OH)	Hydrophilic, H-bonded rigid framework (↑ T_g)
[s]Cyano (–CN)	Hydrophilic, dipole–dipole interactions: ↑ T_g and crystalline
[s]Amide ($–C(O)–NH_2$)	Hydrophilic, very high T_g due to strong H-bonding
[s]Ester (–C(O)–O–)	May be hydrophilic or hydrophobic depending on alkyl substituents, T_g also varies with nature of R and tacticity of units
[s]Ether ($–O–CR_3$)	Hydrophilic at low temperatures, solubility and properties vary with R groups and tacticity
[s]Carboxylic acid (–C(O)–OH)	Hydrophilic and hygroscopic in solid-state

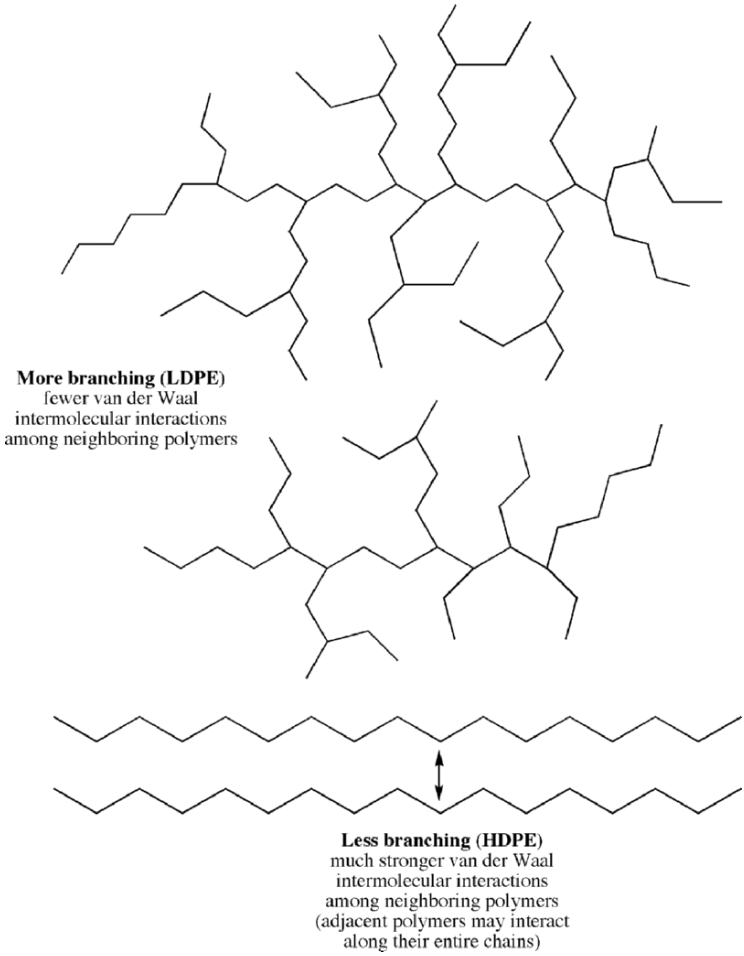


Figure 5.33. Comparison of low-density polyethylene and high-density polyethylene backbone molecular structures. A higher degree of branching (LDPE) results in greater separation of neighboring polymer chains and fewer intermolecular interactions. By comparison, with little/no branching, adjacent chains may closely associate with one another, resulting in strong van der Waal forces along the entire length of the polymer chains.

silicone grease), or rubbery material (*e.g.*, used for remote control keypads). The most widely used silicones are known as *polydimethylsiloxanes* (PDMS), formed from the hydrolysis of chlorosilanes in the presence of water (Eq. 3). In order to increase the polymer size to that required for most applications, the linear and cyclic oligomeric units are subsequently condensed and polymerized, respectively.

$$x\mathrm{Me_2SiCl_2} \xrightarrow{\mathrm{H_2O}} y\underset{\text{(linear)}}{\mathrm{HO(Me_2SiO)}_n\mathrm{H}} + z\underset{\text{(cyclic)}}{\mathrm{(Me_2SiO)}_m} \tag{3}$$

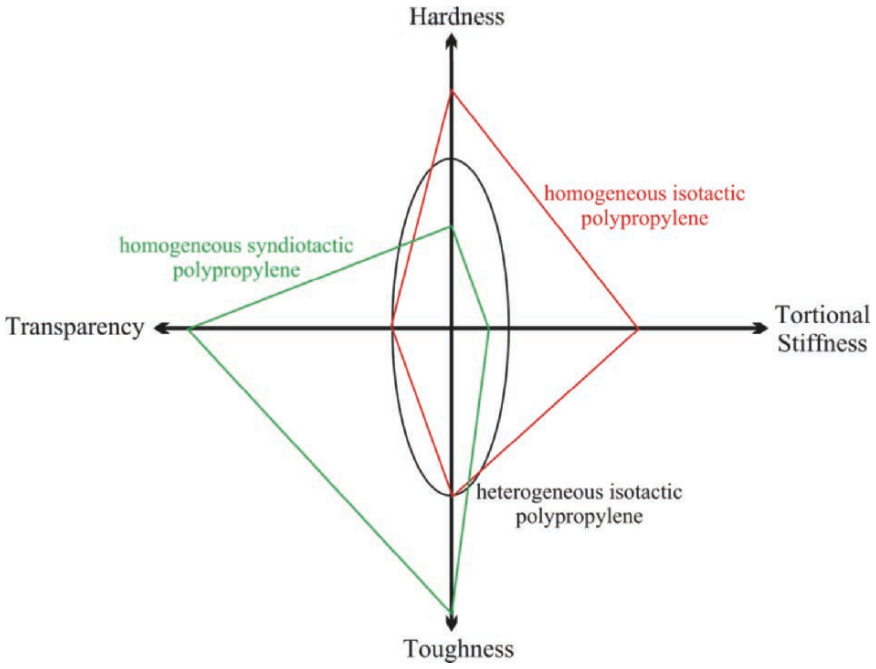


Figure 5.34. Tacticity *vs.* bulk properties for polypropylene polymers. The "homogeneous" and "heterogeneous" notations refer to whether the entire polymer chains, or only regions, are of a certain tacticity.

The crystallinity of a polymer is related to the packing efficiency of individual polymer chains with respect to one another. Not unlike the formation of small molecule single crystals, processing variables such as temperature and time are paramount for influencing the degree of crystallinity. As previously mentioned, non-amorphous polymers may be considered as semicrystalline at best, with regions of crystallinity within the largely disordered polymer matrix. Typical crystalline polymers are polypropylene and polyethylene (especially isotactic and syndiotactic, Figure 5.34), acetals, nylons, and most thermoplastic polyesters. In general, crystalline polymers have high shrinkage, low transparency, a distinct melting point, and possess good chemical and wear resistance. By contrast, polymers that contain bulkier side-groups on their chains, such as polystyrene, polycarbonate, acrylic, ABS, and polysulfone, tend to form amorphous polymers. Based on their disordered structure, amorphous polymers have low shrinkage, a transparent appearance, a broad melting point, and poor chemical and wear resistance.

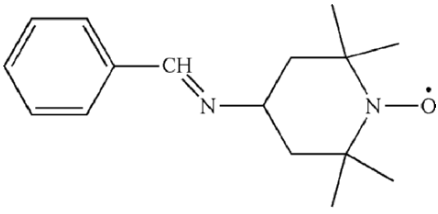
5.2.1. Molecular Magnets

Magnetic materials may date back as far as 2,000 B.C., with the first magnetic mineral, magnetite, discovered in *ca.* 500 B.C.;[14] present-day magnetic materials are typically alloys containing iron, cobalt, or other lanthanide metals. In Chapter 3,

we discussed a variety of magnetic materials and phenomena, which were focused on inorganic alloys and clusters. Since organic compounds are generally diamagnetic (all electrons paired), organic-containing molecules have long been overlooked as candidates for magnetic materials. However, the world of magnetic materials was forever changed in the mid-1980s when Miller and coworkers discovered the first *organic-based* molecular magnet.[15]

In contrast to inorganic compounds that require high temperature sintering processes, organic magnets are synthesized at low temperatures using traditional synthetic techniques. The benefits of organic-based magnets relative to inorganic analogues are their lightweight architecture, tunable conductivity (insulation or semiconductive), tunable solubility, and biocompatibility. With virtually every electronic device employing magnets (*e.g.*, automobiles, computers, audio speakers, televisions, telephones, radios, *etc.*), there will be an increasing market for applications that employ these materials. In particular, the small size of individual molecules relative to inorganic lattices creates the possibility of much greater storage densities for magnetic storage applications (*e.g.*, computer hard drives) – perhaps as high as 200,000 Gb in^{-2}, which is *ca.* three times greater than current alloy thin film materials.[16]

In comparison to inorganic magnets in which electrons reside in d- and/or f-orbitals, organic magnets contain unpaired electrons in p-orbitals. One example of an organic magnet is 4-nitrophenylnitronyl nitroxide (**III**), with a T_c of 0.6 K. In general, the critical temperatures of purely organic magnets ($T_c \leq 1.48$ K) are significantly lower than those that contain metal ions. Hence, for an observable magnetic response at a higher T_c, the p-electron spins are most often coupled with unpaired d-electrons on metal ion(s). The first organometallic (containing both a metal and ligand, with M–C bonds) magnets exploited ligands with conjugated π-systems, capable of metal-like electrical conductivity. To date, the most commonly used ligand systems are reduced forms of TCNQ (7,7,8-tetracyano-*p*-quinodimethane) and TCNE (tetracyanoethylene) (Figure 5.35). Interestingly, if $[TCNQ]^{\bullet -}$ is replaced with $[TCNE]^{\bullet -}$, one unpaired electron is delocalized over a smaller molecular structure that results in a greater spin density and magnetic response. Consequently, the coercivity (1,000 Oe at 2 K) of $[Cp_2Fe]^{\bullet +}[TCNE]^{\bullet -}$ is of the same magnitude as CoPtCr (*ca.* 1,700 Oe), used for magnetic storage. In fact, on a mole-basis, metal TCNE complexes are stronger ferromagnets than iron. However, the low T_c (4.8 K) makes this material impractical for many potential applications.



III

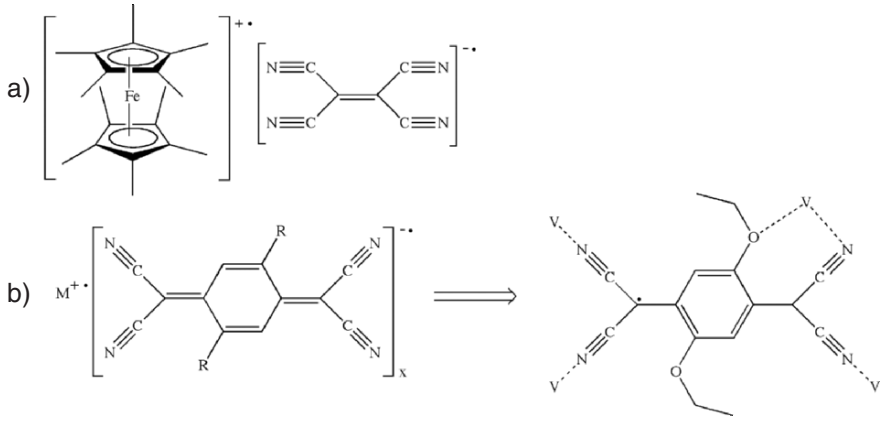


Figure 5.35. Molecular structures of organometallic/transition metal molecular magnets. Shown are (a) $[Fe(Cp^*)_2]^{+\bullet}[TCNE]^{-\bullet}$ and (b) $[M^{+\bullet}][TCNQR_2]^{-\bullet}$, where M═Mn, Fe, Co, Ni, and V; R═H, Br, Me, Et, ^{i}Pr, OMe, OEt, and OPh. For $M[TCNQ]_x$ complexes (b), it was shown that the metal may bind through the terminal N moiety of the cyano group, as well as with the oxygen from a nearby alkoxyl group if the steric bulk is not sufficiently high (*i.e.*, the seven-membered ring will not form with –OPh groups). This chelate formation results in greater stability, which enhances the magnetic coupling and increases the T_c.[17]

The solid-state structure of $[Cp_2Fe]^{\bullet+}[TCNE]^{\bullet-}$ consists of alternating layers of the cationic and anionic species (Figure 5.36). Upon charge transfer between the donor (cation) and acceptor (anion) within a layer, a triplet $S = 1$ state ($\uparrow\uparrow$ *vs.* singlet $\uparrow\downarrow$) is formed through the transfer of a d-electron to the π^*-orbital on TCNE (Figure 5.37). Such intrachain spin alignment serves to stabilize the $[Cp_2Fe]^{\bullet+}[TCNE]^{\bullet-}$ repeat units; however, for bulk macroscopic ferromagnetism, there also has to be interchain spin alignment. This does occur at $T < T_c$, since the distances between $[TCNE]^{\bullet-}$ species and both inter- and intrachain Fe^{3+} ions are roughly equivalent.

The mixture of TCNE and hexacarbonyl vanadium(0) in an organic solvent yields an organic-based magnet with the highest critical temperature to date – *ca.* 400 K (Eq. 4). The detailed solid-state structure of this compound has not been determined due to its air/water reactivity, amorphous morphology, and solvent insolubility. The entrapped solvent is at least partially responsible for its air sensitivity and magnetic susceptibility. That is, CVD of V[TCNE] in the absence of methylene chloride yields air stable films, and acetonitrile-processed V[TCNE] has a significantly lower T_c due to V–NC–CH_3 interactions (less effective ferromagnetic coupling).

$$(4) \qquad x\,\mathrm{TCNE} + \mathrm{V(CO)_6} \xrightarrow{\text{methylene chloride}} \mathrm{V[TCNE]}_x \cdot y\,\mathrm{CH_2Cl_2} + 6\mathrm{CO_{(g)}}$$

The observed magnitude of bulk ferromagnetic ordering depends on both the number and 3D arrangement of unpaired electrons throughout the entire lattice. Since bulk spin coupling is a consequence of the overall solid-state structure, the determination of crystal structures of molecular magnetic materials by X-ray and neutron diffraction remains an active research area. Even at the single-molecule level, the

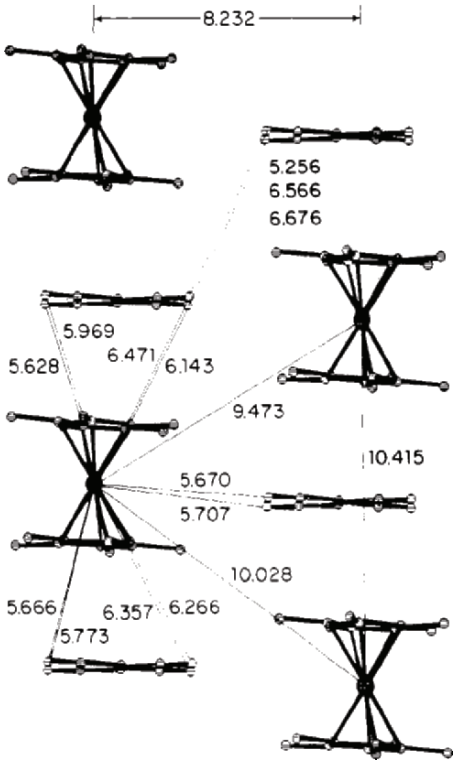


Figure 5.36. Solid-state structure of $[Cp^*_2Fe][TCNE]$, showing relevant intra- and interchain distances. Reproduced with permission from Miller, J. S.; Calabrese, J. C.; Rommelmann, H.; Chittipeddi, S. R.; Zhang, J. H.; Reiff, W. M.; Epstein, A. J. *J. Am. Chem. Soc.*, **1987**, *109*, 769. Copyright 1987 American Chemical Society.

magnetic properties of the material may be complex, and (hopefully) tunable. For example, the antiferromagnetism of a sterically encumbered nickel azide complex may be switched on or off by thermally manipulating the geometry about one of the nickel ions (Figure 5.38). Other examples of light-induced structural transformations represent another emerging area of molecular magnetic research.

Although this chapter is focused on "soft" organic-based materials, we would be remiss if other more recent and active molecular magnetic materials were not mentioned. A number of materials known as *single-molecule* magnets have been identified, with the majority being transition metal oxo clusters that exhibit a 2D honeycomb-layered structure (Figure 5.39a). However, the oxo ligand exhibits multiple bridging possibilities with a wide variability of the M–O–M angles, which greatly affects the M–M magnetic coupling in the solid. Consequently, more recent work has been focused on transition metal cyano (–CN) complexes, which result in a much greater structural control since the ligand may linearly bridge only two metal centers (M–CN–M′). The most widely studied cyano complexes that exhibit either ferri- or ferromagnetic behavior are analogues of Prussian Blue, $Fe^{III}_4[Fe^{II}(CN)_6]_3 \cdot H_2O$ (Figure 5.39b).

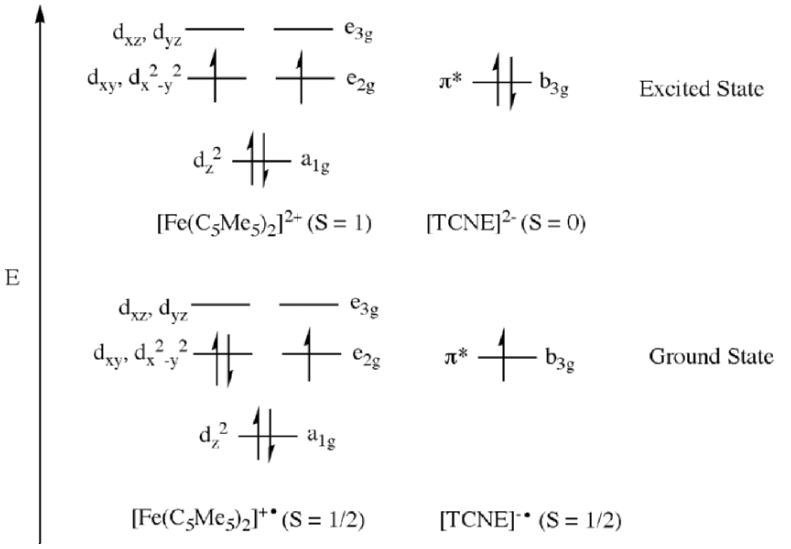


Figure 5.37. Ground- and excited-state electronic configurations of $[Fe(C_5Me_5)_2]$ and TCNE species. Adapted from Miller, J. S.; Calabrese, J. C.; Rommelmann, H.; Chittipeddi, S. R.; Zhang, J. H.; Reiff, W. M.; Epstein, A. J. *J. Am. Chem. Soc.*, **1987**, *109*, 769. Copyright 1987 American Chemical Society.

Before we delve further into these cyano-based structures, let us review a bit of inorganic chemistry, to understand the ligand effect on the electronic configuration of the metal ions in a complex. As you are aware, all ligands behave as a Lewis base toward a metal center, donating an electron pair(s) through filled σ- or π-orbitals. The cyano ligand (and others such as –CO and $–C_2H_4$, *etc.*) represents a special case since it contains an unsaturated (triple) bond. Accordingly, the metal is able to donate electron density back to the ligand into its empty π^*-orbitals, a process known as $M \rightarrow L$ "back-bonding." As you would expect, this phenomenon is most pronounced for metals in low oxidation states, where the metal ion is most apt to reduce its buildup of negative charge. This synergistic electron $L - M\sigma/\pi$ –donation and π^*-back-donation delocalizes the electron density between the ligand and metal, which greatly strengthens the M–L bond. Consequently, a comparative scale known as the *spectrochemical series* has been developed to indicate whether a ligand is active toward back-bonding with the metal. Whereas unsaturated ligands are all considered "strong-field," halides and other σ-donors are all weak-field ligands since they do not have available empty orbitals to accept electron density from the metal.

A number of factors such as nature/oxidation state of metal and nature/charge of ligand will determine the electronic configuration of the metal center in a transition metal complex. For an octahedral complex, the d-orbitals of the transition metal are not degenerate, but are split based on their relative interactions with the ligands. For a metal with $\leq$3 d-electrons, there is no amibiguity with regard to its electronic configuration. However, for d^4 and higher, the fourth electron may be placed either in the lower or higher orbital groups (Figure 5.40). In particular, complexes containing

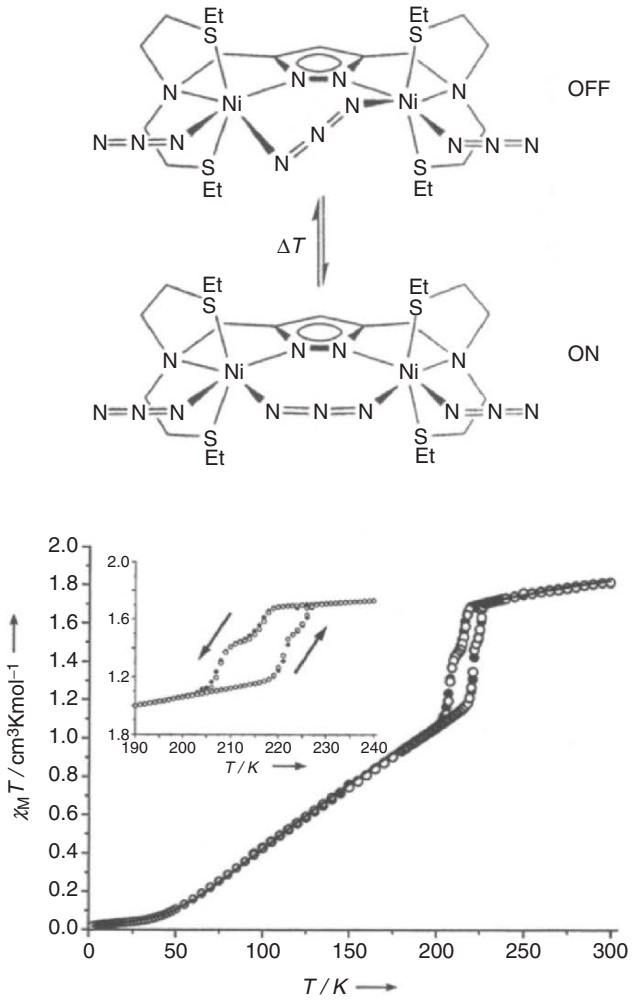


Figure 5.38. Molecular structure (top, showing off/on states for antiferromagnetic coupling of Ni^{2+} ions) and hysteresis curve (bottom) for $[LNi_2(N_3)_3]$ – inset is a magnified view of the 200–225 K region. Reproduced with permission from Leibeling, G.; Demeshko, S.; Dechert, S.; Meyer, F. *Angew. Chem. Int. Ed. Eng.*, **2005**, *44*, 7111. Copyright 2005 Wiley-VCH.

4d/5d metals, low oxidation-state metals (preferring M → L back-bonding), and strong-field ligands will dictate a strong-field (low-spin) configuration. Depending on whether the complex adopts a low-spin or high-spin configuration, the number of unpaired electrons will vary significantly. For example, consider a d^4 complex: low-spin has two unpaired electrons ($S = 1$), whereas high-spin has four unpaired electrons ($S = 2$).

As a terminal ligand, the –CN group always results in a low-spin complex. However, when bridging two metals each end will exhibit different ligand field strengths. That is, the C-end (M ← CN–M) will yield a low-spin configuration

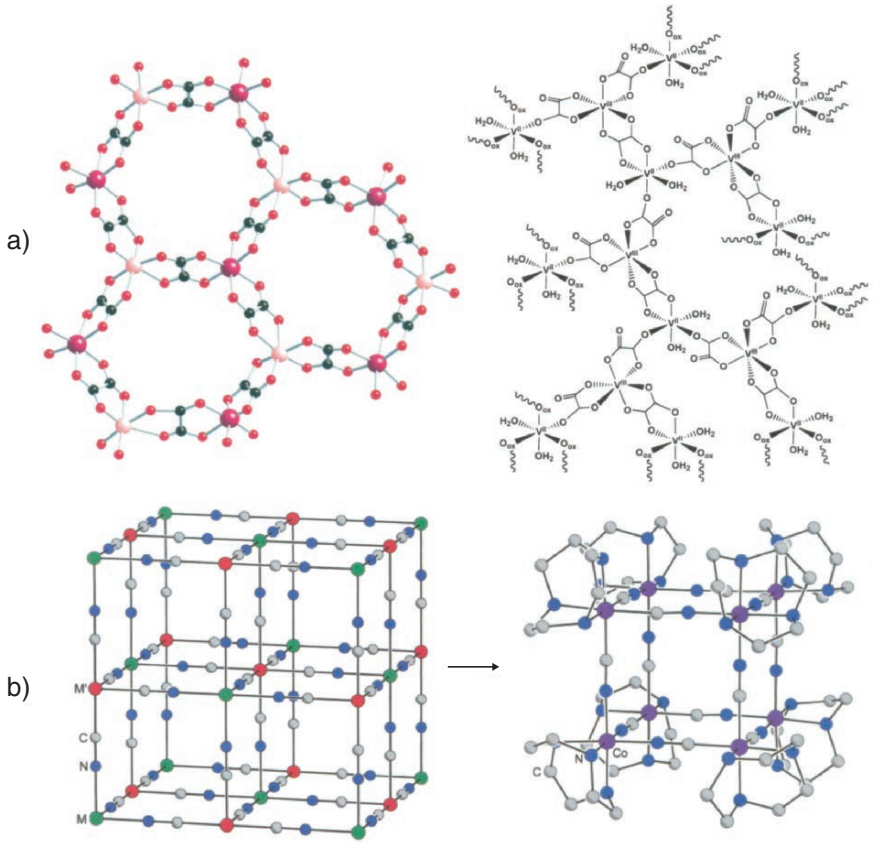


Figure 5.39. Molecular structures of other molecular magnetic materials. Illustrated are (a) tris(oxalato) metalates $[M^{II}M^{III}(ox)_3]$, where M^{II}=Mn, Fe, Ni, Co, Cu, Zn; M^{III}=Cr, Fe, Ru, and (b) the simplified crystal structure of Prussian Blue, with an example of the analogue structure $[(tacn)_8Co_8(CN)_{12}]^{12+}$, where the tacn ligand is 1,4,7-triazacyclononane. Reproduced with permission from (a) Min, K. S.; Rhinegold, A. L.; Miller, J. S. *Inorg. Chem.*, **2005**, *44*, 8433, and (b) Beltran, L. M. C.; Long, J. R. *Acc. Chem. Res.*, **2005**, *38*, 325. Copyright 2005 American Chemical Society.

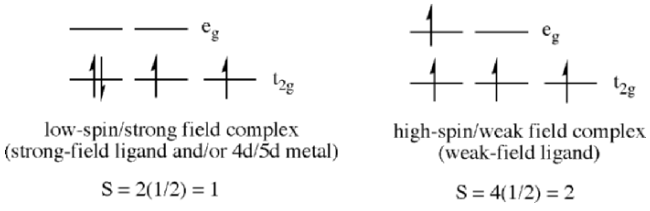


Figure 5.40. Comparative electronic configurations of low-spin (strong-field) and high-spin (weak-field) d^4 metal ions (*e.g.*, Mn^{3+}, Cr^{2+}, V^{+}).

for M, but the N-end (M–CN → M′) will yield a high-spin configuration for M′. Hence, for Prussian Blue, which has a $\rightarrow Fe^{II} \leftarrow CN \rightarrow Fe^{III} \leftarrow NC \rightarrow Fe^{II} \leftarrow$ bonding motif, the d^6 Fe^{II} sites will be low-spin ($S = 0$), whereas the d^5 Fe^{III} sites will be high-spin ($S = 5/2$). This means that ferromagnetic ordering may only occur through distant Fe^{III}–Fe^{III} sites, which occurs at a relatively low T_c (5.6 K). In order to improve the ordering distances with the Prussian Blue array, there has been much interest in replacing Fe with other transition metals (especially early metals in low oxidation states). Hence, the Prussian Blue structure allows for a highly tunable magnetic susceptibility, since a variety of metal ions may be used in association with the cyano ligand. Not only will this affect the number of unpaired electrons that are coupled throughout the lattice, but also the type of ordering (*e.g.*, ferro-, ferri-, antiferromagnetic) based on the metal d-orbitals that house the unpaired electrons. In particular, ferromagnetic ordering occurs when the electrons in the metal ions reside in orthogonal orbitals (*e.g.*, $M(t_{2g})/M'(e_g)$). By contrast, antiferromagnetic coupling occurs when the electrons are housed in orbitals of comparable symmetry (*e.g.*, $M(t_{2g})/M'(t_{2g})$), as illustrated in Figure 5.41.

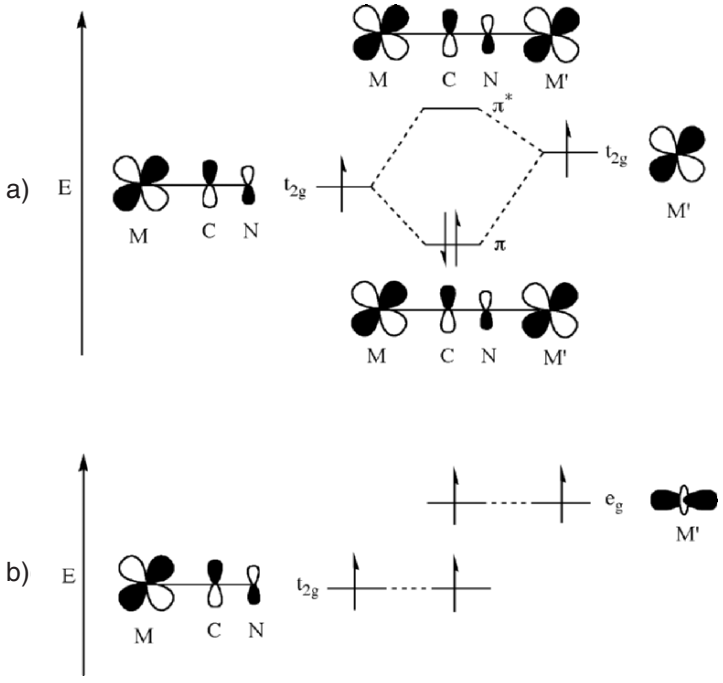


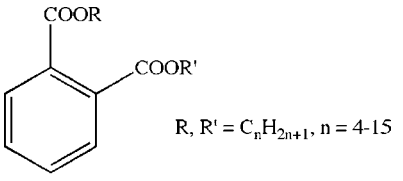
Figure 5.41. Simplified molecular orbital diagrams for an M–CN–M′ unit with octahedrally coordinated metal centers. Shown are (a) antiferromagnetic coupling from overlap of symmetrically aligned orbitals and (b) ferromagnetic ordering from overlap of orthogonal orbitals. Adapted with permission from Beltran, L. M. C.; Long, J. R. *Acc. Chem. Res.*, **2005**, *38*, 325. Copyright 2005 American Chemical Society.

For optical storage applications, it is desirable to have a material that alters its magnetic properties in response to light. This effect is exhibited by Prussian Blue itself, wherein some low-spin Fe sites are converted to a high-spin configuration. More recently, this behavior has been demonstrated for mixed-metal systems such as $K_{0.4}Co_{1.3}[Fe(CN)_6] \cdot 5H_2O$.[18] The light-induced redox reaction may be described by Eq. 5, where diamagnetic Fe^{II} becomes paramagnetic and may couple with other unpaired electrons throughout the solid.[19]

(5) $$Fe^{II} \leftarrow CN \rightarrow Co^{III} \xrightarrow{h\upsilon} Fe^{III} \leftarrow CN \rightarrow Co^{II}$$

5.2.2. Polymer Additives: Plasticizers and Flame Retardants

Although the properties of polymers may be fine-tuned based on the functional groups that are present in their repeat units, a number of components are often added to the consumer formulation to afford desired properties such as color (*e.g.*, inorganic pigments, organic dyes), flexibility, and flame retardancy. Plasticizers are additives that soften a material, enhancing its flexibility. The worldwide market for plasticizers is currently over 5 million metric tons, with over 90% used to soften PVC. The most common plasticizers are phthalates, of the general structure **IV**. Due to the relatively high vapor pressure of these compounds, plasticizers will evaporate from the polymer structure as evidenced by the "new car smell" of new cars, as well as the organic film that becomes deposited on the interior windshield surface.



IV

There are two leading theories concerning the mechanism of activity for plasticizer molecules. The Lubricating Theory suggests that as the polymer is heated, the plasticizer diffuses into the polymer and disrupts the van der Waal interactions among polymer chains. Since network formation is reduced, the T_g is lowered resulting in more flexibility/softness of the bulk polymer. By contrast, the Free Volume Theory suggests that the lowering of T_g is due to the polymer chains being pushed further apart by the interdiffusion of the plasticizer molecules. Since the free volume of the polymer has been increased, the chains are free to move past one another more easily resulting in greater flexibility.

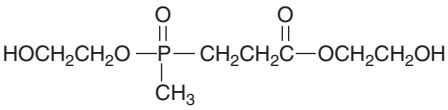
The above mechanistic explanations assume that the plasticizer molecules are not permanently bound to the polymer chains. Since these interactions are relatively weak, there is likely a dynamic adsorption/desorption at various locations among neighboring polymer chains. Accordingly, the plasticizer structure may be fine-tuned to affect its solubility/miscibility with the polymer, as well as its interactions with

polymer chains and with other plasticizer molecules. As the polymer–plasticizer interactions are strengthened, the T_g will increase; at low concentrations, the rigidity of the polymer is increased due to effective rigid-network formation between the plasticizer and polymer. However, as the plasticizer concentration is increased, the additive molecules themselves interact yielding the desired softening characteristics. As illustrated by **IV** above, the molecular structure of a plasticizer contains both polar and nonpolar (hydrocarbon chain) units. It is usually the polar endgroups that bind reversibly with the polymer chains; the length-tunable nonpolar component affords controlled separation of neighboring polymer chains.

Since polymers exhibit a specialized hydrocarbon structure, these materials pose a significant flammability threat. However, if one examines a room following a fire, it is obvious that some polymers withstand ignition much greater than others (Figure 5.42). In fact, these polymers are not naturally fire resistant, but rather contain additives that afford this desirable property. The largest class of flame inhibiting additives is *brominated flame retardants* (BFRs). It is estimated that bromine-containing molecules are added to over 2.5 million tons of polymers each year, with the electronics industry accounting for the greatest consumer market. BFRs are also used in a number of other products such as electronic equipment housing, carpets, paints/stains, fabrics, and kitchen countertops/appliances.

Figure 5.42. Photograph of damage from a candle-ignited fire that destroyed a campus fraternity house in Amherst, MA. Flame retardants likely prevented the television and surface-treated wood cabinet from igniting. Reproduced from http://www.buildinggreen.com.

Due to increasingly stringent environmental regulations, the use of BFRs is being dramatically reduced – especially in Europe. The most widely used alternative is organophosphorus-based (OP) flame retardants (*e.g.*, **V**), which are much more expensive than organohalogen additives. However, these molecules also contribute to environmental hazards, being found in air samples as far away as Antarctica and in rainwater collected across European countries. Part of the problem stems from the fact that OP flame retardants are not as active as BFRs, and must be present in much higher concentrations to be effective – often at the expense of altering the physical properties of the polymer.



V

Not unlike plasticizers, the most common method used to impart flame retardancy is by simple mixing of the additives with the polymer during final processing. However, a growing area of development is the design of functionalized monomers that contain flame retardant groups (*e.g.*, halogens, organophosphorus, Figure 5.43). These monomers are known as *reactive flame retardants*, and yield a polymer that has an inherent flame retardant characteristic – much more controllable than additives that are placed randomly within the polymer. However, this approach is not widely used due to its relatively high production cost. Further, the inclusion of functionalized monomeric units within the polymer chains may alter the physical properties of the bulk polymer.

There are two primary reactive modes for flame retardants: either gas-phase or solid-state. Since a flame consists of a variety of gas-phase radicals and atoms (Eqs. 6 and 7), a gas-phase flame retardant is one that will scavenge the flame-propagating radicals such as $^{\bullet}OH$ and $O^{\bullet}$. In particular, if the hydroxyl radical is suppressed, the exothermic formation of CO_2 (Eq. 8) will be prevented thereby reducing the flame temperature.

(6) $H^{\bullet} + O_2 \Leftrightarrow {}^{\bullet}OH + O^{\bullet}$

(7) $O^{\bullet} + H_2 \Leftrightarrow {}^{\bullet}OH + H^{\bullet}$

(8) ${}^{\bullet}OH + CO \Leftrightarrow CO_2 + H$

Halogenated flame retardants are active in the gas-phase by combining with radicals. The organohalogen compound (MX) first thermally decomposes to form halogen radicals that combine with hydrogen radicals or fuel to form hydrogen halide (HX) (Eqs. 9–10). It should be noted that BFRs all possess aromatic structures since the flame temperature is sufficient to break aromatic C–Br bonds (C–Cl require higher temperatures). By comparison, aliphatic C–Br bonds are broken at temperatures lower than the flame, resulting in predecomposition of the organobromine compound and negligent flame retardancy. Often, antimony oxide (Sb_2O_3) is used in

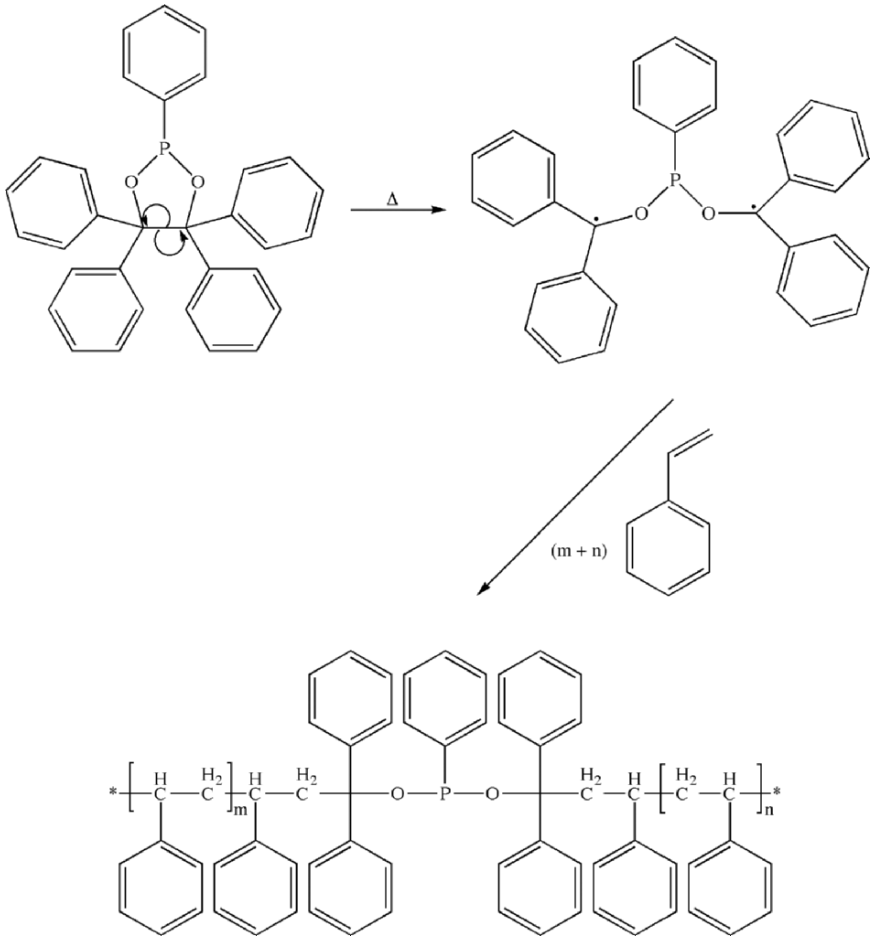


Figure 5.43. Example of the reactive flame retardant approach through incorporation of an organophosphorus unit in the polymer main chain.[20]

combination with BFRs since the side-product of $SbBr_3$ is sufficiently volatile, and carries a high concentration of bromine into the gas-phase.

(9) $MX \Leftrightarrow M^{\bullet} + X^{\bullet}$

(10) $X^{\bullet} + H^{\bullet} \Leftrightarrow HX$

The hydrogen halide species behave as flame inhibitors by scavenging hydrogen, hydroxyl and oxygen radicals (Eqs. 11 and 12).

(11) $^{\bullet}H + HX \Leftrightarrow H_2 + X^{\bullet}$

(12) $^{\bullet}OH + HX \Leftrightarrow H_2O + X^{\bullet}$

By contrast, solid-state flame retardants such as organophosphorus agents act by passivating the surface of the polymer toward heat, flame, and oxygen through formation of a carbonaceous coating known as a *char*. When thermally decomposed, some phosphorus compounds generate acids that promote surface crosslinking reactions and resultant char formation. Quite often, phosphorus flame retardants also have gas-phase reactivity, whereby $PO^{\bullet}$ radicals help to scavenge flame-propagating species in the flame. Though the mechanism is not currently known, the use of nitrogen compounds (or N-containing polymers such as poly(amide)s) enhances char formation. This is most likely due to a lower number of reactive C/H/O-containing units that volatilize to fuel the flame.

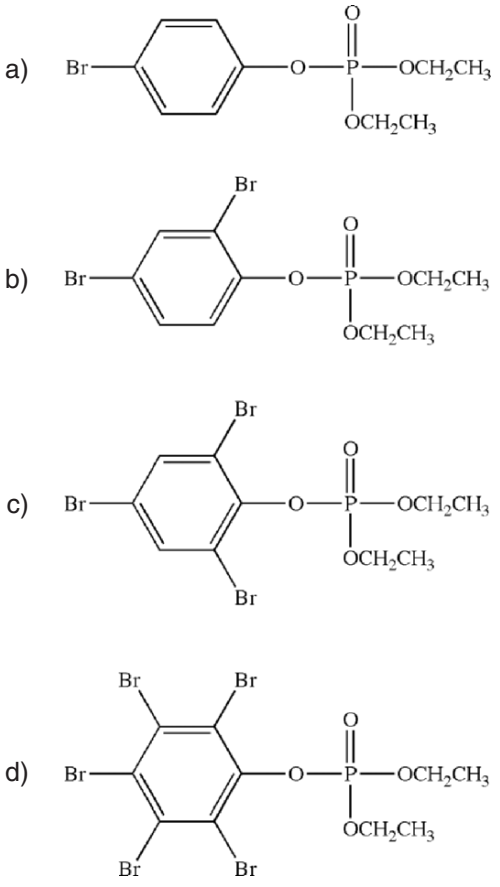


Figure 5.44. Molecular structures of dual-action brominated organophosphorus flame retardants. Shown are (a) (4-bromophenyl)diethylphosphate, (b) (2,4-dibromophenyl)diethylphosphate, (c) (2,4,6-tribromo phenyl)diethylphosphate, and (d) (2,3,4,5,6-pentabromophenyl)diethylphosphate.

Another interesting area of recent development is the synthesis of dual-action flame retardants that contain both organophosphorus and organobromine units (Figure 5.44). Since these agents are active in both gas- and condensed phases, the activity is proposed to be much higher than current additives at much lower concentrations – in accord with environmental regulations. In addition, the synthesis of these agents is relatively inexpensive, which is also a stringent guideline in the development of alternative additives for such a large market sector.

IMPORTANT MATERIALS APPLICATIONS IV: SELF-HEALING POLYMERS

Imagine getting into an accident only to have your vehicle revert back to its original shape before your eyes! Though this certainly sounds like something out of a science fiction novel, these materials are fast becoming a reality. The general strategy for this activity is to have the healing agents as a part of the polymer, which become activated upon crack formation. This was first demonstrated with microcapsules of a urea-formaldehyde shell that contained a dicyclopentadiene monomer, suspended within an epoxy polymer matrix. Of course, without a polymerization catalyst, nothing would happen; hence, crystals of Grubbs' catalyst were also dispersed in the polymer (Figure 5.45). The encapsulation of the Grubbs' catalyst within a wax microsphere has been reported to improve the dispersity in the epoxy matrix and prevent side reactions between the catalyst and the amine-based epoxy curing agents.[21]

A technique referred to as *ring-opening metathesis polymerization* (ROMP) using Grubbs' catalyst (Figure 5.46) was chosen due to rapid polymerization under ambient conditions, in the presence of oxygen and water. Further, these conditions were also suitable for low shrinkage upon polymerization – extremely important for self-healing materials. When a crack ruptured a microcapsule, the monomer became exposed to the catalyst and a strong, highly crosslinked system was formed that bonded together adjacent crack faces within minutes. This same technology is being developed for other materials such as ceramics and glasses for an unlimited number of applications.

An alternative technique that does not involve microcapsules or transition metal catalyst has recently been developed. This method features a reversible polymerization using Diels-Alder reactions of furan and imide-based monomers (Figure 5.47). When the monomers are combined, a highly crosslinked polymer is formed with properties similar to epoxy or polyester composites. However, when a crack propagates through the material, the reactions are reversed generating the monomers in the vicinity of the fracture. If this region is heated and clamped together, the monomers will polymerize to generate a self-healed material. The benefit of this approach is the ability of this material to undergo multiple fracture-healing cycles without a loss of activity. However, the area must be heated in order to under regeneration. Further developments in this exciting area are surfacing all the time – including an interesting proposal for self-healing knee implants using encapsulated poly(methylmethacrylate) and peroxide initiators.[22]

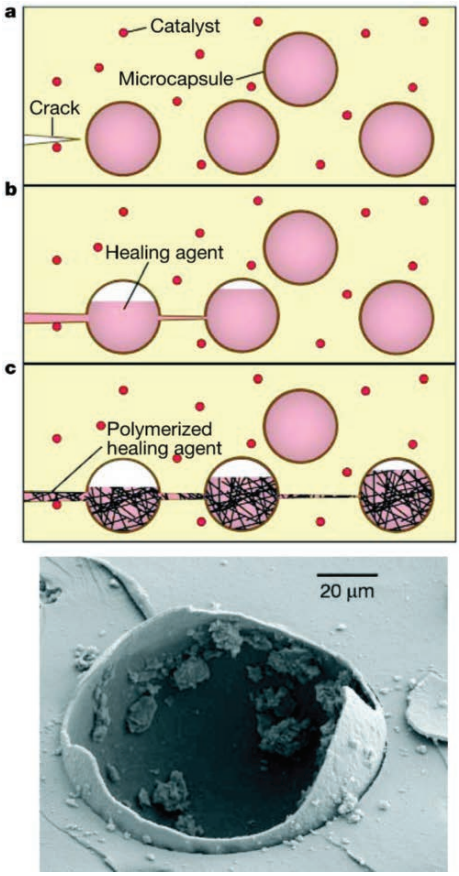


Figure 5.45. Schematic illustrating the mode of action of self-healing polymers through embedded healing-agent microcapsules that are activated by a propagating crack. Also shown is a polymeric microcapsule following rupture. Reproduced with permission from White, S. R.; Sottos, N. R.; Geubelle, P. H.; Moore, J. S.; Kessler, M. R.; Sriram, S. R.; Brown, E. N.; Viswanathan, S. *Nature*, **2001**, *409*, 794. Copyright 2001 Macmillan Magazines.

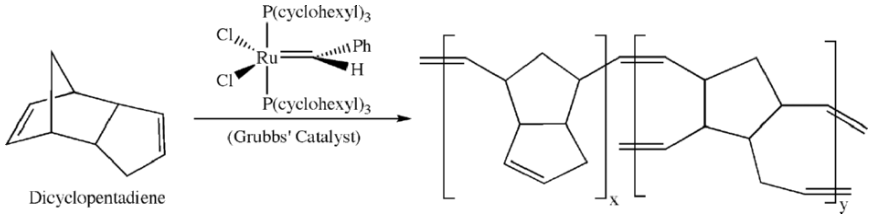


Figure 5.46. Ring-opening metathesis polymerization (ROMP) of dicyclopentadiene catalyzed by Grubbs' catalyst.

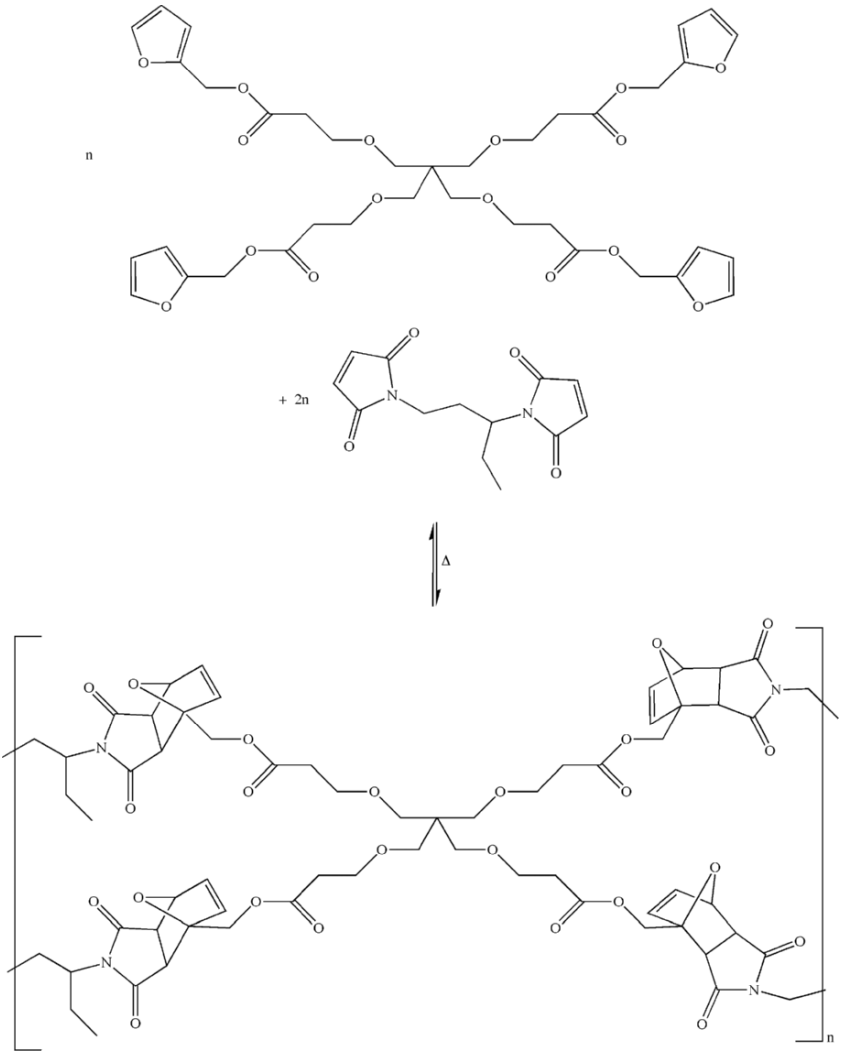


Figure 5.47. Thermally induced self-healing through a Diels-Alder reaction.[23]

References and Notes

1 It should be noted that melting points refer to the temperature required to separate molecules from one another. By contrast, the glass-transition temperature refers to the temperature required to perturb the bonds of the polymer backbone.

2 Gel-permeation chromatography (GPC) is a subdivision of size-exclusion chromatography (SEC), in which macromolecular species are separated from one another based on their size. As its name implies, GPC employs a gel (usually cross-linked polystyrene) as the stationary phase, with detection through either light-scattering or refractive index.

3 Harlan, C. J.; Bott, S. G.; Barron, A. R. *J. Am. Chem. Soc.*, **1995**, *117*, 6465.

4 In general, high molecular weight polymers are typically formed from catalysts derived from early transition metals (Groups 4–6). For late transition metals, the β-hydride elimination mechanism is more preferred, leading to greater numbers of oligomers and dimers.

5 Meijer, D. Fall 2006 ACS meeting, Midland, MI.

6 Flory, P. J. *J. Am. Chem. Soc*., **1952**, *74*, 2718.

7 (a) Buhleier, E.; Wehner, W.; Vögtle, F. *Synthesis*, **1978**, 155. (b) Moors, R.; Vögtle, F. *Chem. Ber.*, **1993**, *126*, 2133. (c) Vögtle, F.; Weber, E. *Angew. Chem. Int. Ed. Engl.*, **1974**, *13*, 814.

8 (a) Tomalia, D. A.; Baker, H.; Dewald, J.; Hall, M.; Kallos, G.; Martin, S.; Roeck, J.; Ryder, J.; Smith, P. *Polym. J.*, **1985**, *17*, 117. (b) Tomalia, D. A.; Baker, H.; Dewald, J.; Hall, M.; Kallos, G.; Martin, S.; Roeck, J.; Ryder, J.; Smith, P. *Macromolecules*, **1986**, *19*, 2466. (c) Tomalia, D. A.; Hall, M.; Hedstrand, D. *J. Am. Chem. Soc.*, **1987**, *109*, 1601. (d) Tomalia, D. A.; Berry, V.; Hall, M.; Hedstrand, D. M. *Macromolecules*, **1987**, *20*, 1164. (e) Tomalia, D. A. *et al.* US Patents 4,587,392 (1986); 4,558,120 (1985); 4,568,737 (1986); 4,599,400 (1986); 4,631,337 (1986); 4,507,466 (1985). (f) For an extensive review, see: Tomalia, D. A.; Naylor, A. M.; Goddard III, W. A. *Angew. Chem. Int. Ed. Engl.*, **1990**, *29*, 138.

9 It should be noted that Robert Denkwalter and coworkers from Allied Corporation were granted the first patent for dendrimers (US Patent 4,410,688 – filed 11 December 1981 and published 18 October 1982: http://www.hairlosshelp.com/html/researchframepatentus.htm), which represents the first dendrimer-related publication. However, the term "dendrimer" can be traced back to A. J. Vogel. The generally accepted definition of dendrimers is highly branched and monodisperse polymers (particularly for convergent growth), with a degree of branching of 1.0.

10 (a) Newkome, G. R.; Yao, Z.; Baker, Z. R.; Gupta, V. K.;. *Org. Chem.*, **1985**, *50*, 2004. (b) Newkome, G. R.; Yao, Z.; Baker, G. R.; Gupta, V. K.; Russo, P. S.; Saunders, M. J. *J. Am. Chem. Soc.*, **1986**, *108*, 849. (c) Newkome, G. R.; Baker, G. R.; Saunders, M. J.; Russo, P. S.; Gupta, V. K.; Yao, Z.; Miller, J. E.; Bouillion, K. *J. Chem. Soc., Chem. Commun.*, **1986**, 752.

11 (a) Hawker, C. J.; Frechet, J. M. J. *J. Am. Chem. Soc.*, **1990**, *112*, 7638. (b) Hawker, C. J.; Frechet, J. M. J. *Macromolecules*, **1990**, *23*, 4726. (c) Wooley, K. L.; Hawker, C. J.; Frechet, J. M. J. *J. Am. Chem. Soc.*, **1991**, *113*, 4252.

12 (a) Dvornic, P. R.; de Leuze-Jallouli, A. M.; Owen, M. J.; Perz, S. V. *Macromolecules*, **2000**, *33*, 5366. (b) Dvornic, P. R.; Li, J.; de Leuze-Jallouli, A. M.; Reeves, S. D.; Owen, M. J. *Macromolecules*, **2002**, *35*, 9323.

13 De Gennes, P. G.; Hervet, H. *J. Phys. Lett.*, **1983**, *44*, 351–360.

14 In 2,000 B.C., magnetic stones are mentioned in the oldest medical textbook ever discovered – the Yellow Emperor's Classic of Internal Medicine. Over 2,500 years ago, a shepherd named Magnes found mineral stones sticking to the metal nails in his sandals; he called the mineral magnetite.

15 (a) Miller, J. S.; Calabrese, J. C.; Epstein, A. J.; Bigelow, R. W.; Zhang, J. H.; Reiff, W. M. *J. Chem. Soc., Chem. Commun.*, **1986**, 1026. (b) Miller, J. S.; Epstein, A. J. *Angew. Chem. Int. Ed. Eng.*, **1994**, *33*, 385. (c) Miller, J. S.; Epstein, A. J. *Chem. Eng. News*, **1995**, *73*(40), 30. (d) Manriquez, J. M.; Yee, G. T.; McLean, R. S.; Epstein, A. J.; Miller, J. S. *Science*, **1991**, *252*, 1415. (e) Miller, J. S.; Epstein, A. J. *J. Chem. Soc., Chem. Commun.*, **1998**, 1319. (f) Pokhodnya, K. I.; Epstein, A. J.; Miller, J. S. *Adv. Mater.*, **2000**, *12*, 410. (g) Hatlevik, O.; Buschmann, W. E.; Zhang, J.; Manson, J. L.; Miller, J. S. *Adv. Mater.*, **1999**, *11*, 914. (h) Miller, J. S. *Inorg. Chem.*, **2000**, *39*, 4392.

16 Beltran, L. M. C.; Long, J. R. *Acc. Chem. Res.*, **2005**, *38*, 325.

17 Vickers, E. B.; Selby, T. D.; Thorum, M. S.; Taliaferro, M. L.; Miller, J. S. *Inorg. Chem.*, **2004**, *43*, 6414.

18 (a) Verdaguer, M.; Bluezen, A.; Marvaud, V.; Waissermann, J.; Seuleimann, M.; Desplanches, C.; Scuiller, A.; Train, C.; Garde, R.; Gelly, G.; Lomenech, C.; Rosenman, I.; Veillet, P.; Cartier, C.; Sato, O.; Einaga, Y.; Iyoda, T.; Fujishima, A.; Hashimoto, K. *J. Electrochem. Soc.*, **1997**, *144*, L11. (b) Einaga, Y.; Sato, O.; Ohkoshi, S.; Fujishima, A.; Hashimoto, K. *Chem. Lett.*, **1998**, 585.

19 For a very nice review on 3D cyanide-based magnets see: Miller, J. S. *MRS Bull.*, **2000**, *11*, 60.

20 Howell, B. A. Presented in part at the 15th International Conference on Advances in Additives and Modifiers for Polymers and Blends, Las Vegas, NV, February, 2006.

21 Rule, J. D.; Brown, E. N.; Sottos, N. R.; White, S. R.; Moore, J. S. *Adv. Mater.*, **2005**, *17*, 205.

22 Hasenwinkel, J. M.; Lautenschlager, E. P.; Wixson, R. L.; Gilbert, J. L. *J. Biomed. Mater. Res.*, **2002**, *59*, 411.

[23] Chen, X.; Dam, M. A.; Ono, K.; Mal, A.; Shen, H.; Nutt, S. R.; Sheran, K.; Wudl, F. *Science*, **2002**, *295*, 1698.

Topics for Further Discussion

1. Explain the differences between a thermoplastic and thermoset.
2. What is meant by "tacticity," and how does this influence overall polymer properties?
3. Explain the effects of the initiator concentration on the degree of polymerization (free-radical route) of styrene.
4. Why was the reported discovery of dendrimers fought with so much controversy (*i.e.*, what about the structure/route did the leading scientists not believe)?
5. What is the chemical origin of the organic film that forms on the windshield of a new car? How about the "new car smell"?
6. What is meant by the term "iniferter," and what are some recent candidates that have been synthesized in the laboratory?
7. What is the purpose of MAO in olefin polymerizations?
8. Explain why polymers crystallize in a chain array rather than the thermodynamically favored extended chain form.
9. Examine the polymeric materials in your immediate vicinity and determine whether they were formed via condensation or addition polymerization routes.
10. What are the structures of conductive polymers, and what is the mechanism for electronic transport in the solid-state?
11. Your supervisor has asked you to make a polymeric material with the following properties – what are some possible backbone and side-group structures that you would try to incorporate in the polymer?
    (a) High-melting point, flexible
    (b) Low-melting point, conductive
    (c) Able to withstand strong acids or bases
    (d) Flexible at low temperatures and becoming more rigid at elevated temperatures

Further Reading

1. http://www.socplas.org/industry/defs.htm – great webpage that provides definitions and historical anecdotes of all of the major classes of polymer resins.
2. Al-Malaika, S.; Golovoy, A.; Wilkie, C. A. ed. *Chemistry and Technology of Polymer Additives*, Blackwell Science: Malden, MA, 1999.
3. Allcock, H. R.; Lampe, F. W.; Mark, J. E. *Contemporary Polymer Chemistry*, 3rd ed., Prentice-Hall: New Jersey, 2003.
4. Painter, P. C.; Coleman, M. M. *Fundamentals of Polymer Science*, 2nd ed., CRC: New York, 1997.
5. Flory, P. J. *Principles of Polymer Chemistry*, Cornell University Press: Ithaca, NY, 1953.
6. Odian, G. *Principles of Polymerization*, 3rd ed., Wiley: New York, 1991.
7. Young, R. J.; Lovell, P. A. *Introduction to Polymers*, 2nd ed., CRC: New York, 2000.
8. Stevens, M. P. *Polymer Chemistry: An Introduction*, 3rd ed., Oxford University Press: Oxford, 1998.
9. Fortin, J. B.; Lu, T. -M. *Chemical Vapor Deposition Polymerization: The Growth and Properties of Parylene Thin Films*, Springer: Berlin Heidelberg New York, 2003.
10. Hsieh, H.; Quirk, R. P. *Anionic Polymerization*, CRC: New York, 1996.

CHAPTER 6

NANOMATERIALS

Imagine how much control over resultant properties you would have if you were able to deposit and maneuver individual atoms into predefined arrangements, *en route* toward a new material. This is fast becoming a reality, and is the realization of the ultimate in "bottom-up" materials design. Thus far, one is able to easily fabricate materials comprised of a small number of atoms, with features on the nanometer scale (10^{-9} m) – one-billionth of a meter. To put this into perspective, think of a material with dimensions approximately 1,000 times smaller than the diameter of a human hair follicle! As we will see, it is now even possible to push individual atoms around a surface using specialized techniques.

We are at the crossroads of unprecedented applications that will only be possible using nanoscale building blocks. More effective devices will be constructed to remove pollutants from the environment and detect/deactivate chemical and biological warfare agents. Integrated circuitry with the capabilities of current workstations will be the size of a grain of sand and will be able to operate for decades with the equivalent of a single wristwatch battery. Robotic spacecrafts that weigh only a few pounds will be sent out to explore the solar system, and widespread space travel will be possible for the masses. Oh, yes – one that is near to us all – inexpensive alternative energy sources will power our vehicles, rather than depending on dwindling oil reserves and the daily fluctuations of (soaring) gas prices![1]

In order to gain rapid progress toward these intriguing goals, the level of government and private funding in the nanosciences continues to soar. This has spawned a number of institutes in recent years, focused on research, development, and commercialization of nanoscale discoveries, as well as public education/outreach. Some recent examples are[2]:

- *National Nanotechnology Initiative* (Federal R&D Program, Washington, DC); http://www.nano.gov
- *Richard E. Smalley Institute for Nanoscale Science and Technology* (Rice University, Houston, TX); http://cohesion.rice.edu/centersandinst/cnst/index.cfm
- *Institute for Nanotechnology* (Northwestern University, Evanston, IL); http://www.nanotechnology.northwestern.edu/index.html
- *Nano Science and Technology Institute* (Cambridge, MA); http://www.nsti.org/about/

- *National Cancer Institute Alliance for Nanotechnology in Cancer* (Bethesda, MD); http://nano.cancer.gov/
- *ASME Nanotechnology Institute* (New York, NY); http://www.nanotechnology-institute.org/about.html
- *Nanotechnology Institute* (Philadelphia, PA); http://www.nanotechinstitute.org/nti/index.jsp)
- *Center for Nanoscale Chemical–Electrical–Mechanical Manufacturing Systems* (Urbana, IL); http://www.nano-cemms.uiuc.edu/
- *Nano/Bio Interface Center* (University of Pennsylvania); http://www.nanotech.upenn.edu/
- *Center on Nanotechnology and Society* (Chicago-Kent College of Law, Illinois Institute of Technology); http://www.nano-and-society.org/
- *The NanoTechnology Group, Inc.* (Nanoscience educational outreach, Wells, TX); http://www.thenanotechnologygroup.org/
- *The Foresight Institute* (Palo Alto, CA); http://www.foresight.org
- *The Institute for Soldier Nanotechnologies* (Massachusetts Institute of Technology); http://web.mit.edu/isn/aboutisn/index.html

The first national network focused on the design/fabrication/testing of nanomaterials was instituted in 2004 through funding from the National Science Foundation. The National Nanotechnology Infrastructure Network (NNIN) consists of a conglomerate of 13 sites across the country (Figure 6.1) that are focused on all aspects of nanomaterials. Since "nanotechnology" is such an interdisciplinary field, many more nanorelated research centers will likely be instituted in the near future. However, as with all scientific disciplines, a major roadblock toward research progress in the US is domestic student recruitment. There are a declining number of degrees awarded in the sciences within recent years (from B.S. to Ph.D. levels) in the US, relative to other foreign countries (*e.g.*, China, India). This represents an ominous forecast

Figure 6.1. The 13 sites of the National Nanotechnology Infrastructure Network (FY2004-FY2009). Reproduced from http://www.nnin.org.

that our advanced technology and warfare capabilities will greatly lag behind other countries, threatening our everyday way of life and the "superpower" status long enjoyed by the US.

As exciting as the futuristic applications may sound, is it possible that such technological growth may be associated with dire societal consequences? In Eric Drexler's book "Engines of Creation,"[3] it was forecasted that self-replicating nanomachines would take over all life on Earth! Although that notion is far from reality, there may be more serious issues that arise through introduction of nanomaterials into the biosphere. History has revealed that chemistry is a two-edged sword, with benefits that greatly improve our lives but also negative consequences for human health and our environment. For example, think of the chlorofluorocarbons (CFCs) that were once used for refrigerants. Their discovery was heralded as one of the greatest triumphs of modern science. Alas, many years after their worldwide adoption, it was realized that they contributed to the ozone hole and likely increase in the incidences of skin cancer. Likewise, emissions from factories and automobiles were not considered as problematic many years ago, but we are now aware of the destructive consequences of these sources (*e.g.*, acid rain and global warming – the leading cause of catastrophic changes in our climate/weather patterns).

What will arise from the widescale introduction of nanoscale materials into our world? Are we on the verge of upsetting the natural balance in ways that cannot be overturned? These are serious questions that may only be answered through far-reaching research endeavors, many of which are currently being investigated. In particular, we must first determine the toxicological/environmental consequences of nanoscale materials with comparisons to known contaminants such as colloids, aerosols/smoke particulates, and asbestos. Some important questions that must be addressed[4]:

(i) Do nanomaterials bioaccumulate?
(ii) What is the danger of exposure *via* skin absorption, ingestion, or inhalation routes?
(iii) What is the fate, transport, and transformation of nanosized materials after they enter the environment?

In addition to these toxicological-based questions, a number of ethical considerations must also be addressed[5]:

(i) Equity issues – will nanotechnology be utilized to solve third-world problems, or will it primarily be used to increase the prowess of industrially advanced countries?
(ii) Privacy issues – imagine a world where you have invisible sensors/microphones – need we say more…
(iii) Security – how will our country and others defend itself against invisible nanoweaponry?
(iv) Human–machine interactions – there are many religious and philosophical issues associated with embedded nanodevices within the human body.

The first news report on the potential damaging effects of nanoscale materials surfaced about a decade ago, when TiO_2/ZnO nanoparticles from sunscreen were found to cause free radicals in skin cells, damaging DNA. Since then, there have

been an increasing number of such reports suggesting that nanostructures are able to traverse across membranes in the body, with an increasing toxicity with decreasing nanoparticulate dimensions. Perhaps the most widely reported study surfaced in mid-2004, where it was shown that fullerenes, a nanoscale allotrope of carbon, cause brain damage in aquatic species. Many additional studies are needed to determine the full impact of nanomaterials before their full worldwide adoption.

The introduction of a new architecture such as nanomaterials necessitates the need for new terminology and methods of classification and characterization. We must also understand the mechanisms by which individual nanostructures may assemble into larger materials, as this will greatly affect the properties of the bulk device for a particular application. This chapter will focus on all of these important issues, with an introduction to the various types of nanomaterials, laboratory techniques used for their synthesis, and (perhaps most importantly) their role in current/future applications.

6.1. WHAT IS "NANOTECHNOLOGY"?

Although there is much current excitement about nanomaterials, there is really nothing new about nanoscience. In fact, the earliest civilizations used nanoscale materials for a variety of applications. For example, the Mayans used a magnesium aluminum silicate clay called palygorskite, which contained nanosized channels that were filled with water. The Mesopotamian civilizations used colored glass for decorative applications that contained embedded metallic nanoparticles.

Physics Nobel Laureate Richard Feynman gave the first lecture regarding the applications for nanoscale materials. His talk, entitled "There's Plenty of Room at the Bottom," was delivered on 29 December 1959 at the annual American Physical Society meeting on the campus of Caltech. Appendix 2 contains a transcript of his entire talk, which contains references to a future world that was never before imagined. Feynman pointed out that designing materials atom-by-atom is a real possibility, as it would not violate any physical laws. He also predicted such sci-fi accomplishments as writing 24 volumes of the Encyclopedia Brittanica on the head of a pin, and even more amazingly, the complete reproduction of every book ever produced to fit within a small handheld pamphlet of less than 40 pages! To put these prophetic statements into context, at the time he delivered this speech, computers such as UNIVAC 1 filled an entire room (Figure 6.2) and carried a price tag of over $1 million.

The first use of the term "nanotechnology" was by Norio Taniguchi in 1974 at the International Conference on Precision Engineering (ICPE). His definition referred to "production technology to get extra high accuracy and ultra fine dimensions, *i.e.*, the preciseness and fineness on the order of 1 nm (nanometer), 10^{-9} m in length."[6] Although many definitions for nanotechnology have been suggested, NASA recently suggested the most thorough description:

> The creation of functional materials, devices and systems through control of matter on the nanometer length scale (1–100 nm), and exploitation of novel phenomena and properties (physical, chemical, biological) at that length scale.[7]

Figure 6.2. Photo of the room-sized UNIVAC 1 computer system that was introduced in the late 1950s.[8]

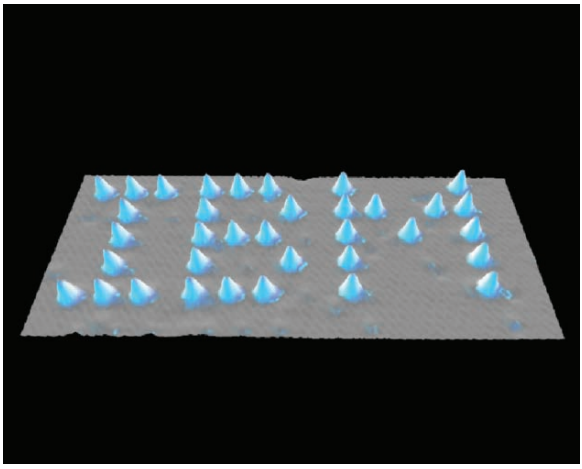


Figure 6.3. Scanning tunneling microscope image of the placement of individual Xe atoms on a Ni(110) surface – no surprise, by researchers at IBM. Reproduced with permission from Eigler, D. M.; Schweizer, E. K. *Nature* **1990**, *344*, 524. Copyright 1990 Macmillan Publishers Ltd.

Although Feynman and Drexler certainly popularized nanotechnology, their influence did not directly lead to the design of nanoscale materials. Rapid progress in nanotechnology could only take place after the arrival of sophisticated instrumentation, capable of viewing and manipulating materials on the nanoscale. In the 1980s, scanning probe microscopy (SPM) was developed which allowed scientists to fulfill Feynman's vision of pushing individual atoms around a surface (Figure 6.3).

This technique was co-invented by Calvin Quate and Hemantha Kumar Wickramasinghe. Interestingly, when Quate and Binnig first submitted their work to the peer-reviewed journal *Physical Review Letters*, it was rejected due to such far-fetched claims as being able to measure forces on individual atoms. However, these results were eventually published, directly influencing the future of molecular nanotechnology. The 1986 Nobel Prize in Physics was awarded to Gerd Binnig and Heinrich Rohrer to honor their design of the scanning tunneling microscope (STM). They shared the Prize with Ernst Ruska, the inventor of the first electron microscope, another essential tool for the modern nanomaterials scientist. In fact, the resolution of modern electron microscopes are now high enough to provide images of individual atoms, and are often fitted with detectors that are capable of determining the chemical composition and/or oxidation state of the surface atoms. Chapter 7 will describe these and other instruments that are commonly used for materials-related research and development.

6.2. NANOSCALE BUILDING BLOCKS AND APPLICATIONS

The first question almost everyone new to the nanoregime asks is "why are nanomaterials so special?" The leading advantage of this size regime is the large surface area/volume ratio exhibited by nanomaterials (Figure 6.4). Accordingly, this translates to a very high surface reactivity with the surrounding surface, ideal for catalysis or sensor applications. Further, since biological systems feature the systematic organization of nanoscale materials (*e.g.*, proteins are 1–20 nm in size, the diameter of DNA is *ca*. 2.5 nm), being able to fabricate materials in this size regime holds promise for artificial components within cells (that have *ca*. 10,000–20,000 nm

Figure 6.4. Comparison of the surface area/volume ratio of macroscopic particles (marbles) and nanoscopic aluminum oxide particles. Since nanoparticules contain a proportionately large number of surface atoms, there are a significantly greater number of adsorption/reaction sites that are available to interact with the surrounding environment. Further, whereas bending of a bulk metal occurs *via* movement of grains in the >100 nm size regime, metallic nanostructures will have extreme hardness, with significantly different malleability/ductility relative to the bulk material.

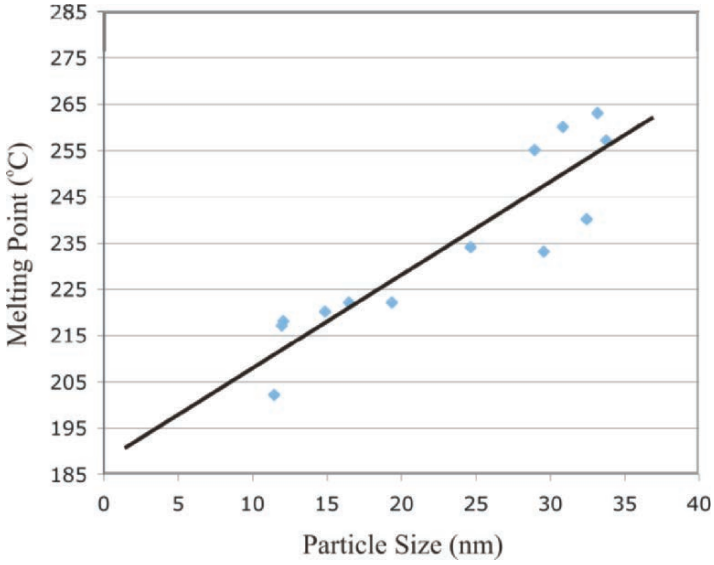


Figure 6.5. Decrease in the melting point of gold nanoparticles with decreasing diameter. It should be noted that the melting point of bulk gold is 1,064°C! Adapted with permission from Unruh, K. M. et al. "Melting Behavior in Granular Metal Thin Films," *Materials Research Society Symposium Proceedings*, vol. 195. Materials Research Society, Apr 16–20, 1990. Copyright 1990 Materials Research Society.

diameters) to diagnose/fight diseases, ilnesses, viruses, and other superficial weaknesses (*e.g.*, artificial muscles).

Another key benefit for nanomaterials is the ability of varying their fundamental properties (*e.g.*, magnetization,[9] optical properties (color), melting point (Figure 6.5), hardness, *etc*.), relative to bulk materials without a change in chemical composition. Although bulk properties such as melting point and hardness are related to the enhanced surface interactions among nanoparticulates, the size-tunable electronic properties are due to quantum confinement effects, as discussed in later in this chapter.

Since we live in a macroscopic world, the next generation of materials will be of similar physical dimensions as today's consumer products. That is, we have shrunk down the size of cell phones and computers to almost their useful limits – any further, and one would inconveniently need to use a sharp stylus to dial a phone number! However, as articulated in Chapter 4, although the size of electronic devices will remain somewhat constant, the speed and computational ability of these devices must continue to increase. This translates to materials that are built from the ground up, one nanoscale building block at a time. However, it is synthetically too expensive (and not industrially scaleable) to arrange such small units into their desired positions by hand. Consequently, materials chemists are largely focused on "bottom-up" techniques that afford the *self-assembly* of nanoscale species. As we will see later in this chapter, parallel efforts in "top-down" processing are being developed

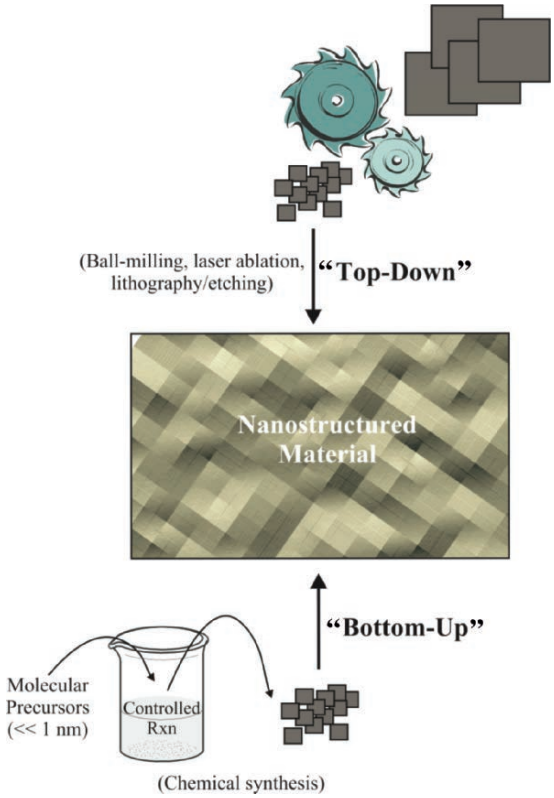


Figure 6.6. Comparison of the "top-down" and "bottom-up" approach to nanomaterials synthesis.

by materials engineers to yield nanoscale building blocks and devices through advanced lithographic, ablation, and etching techniques (Figure 6.6). In this respect, one can consider a nanoscale object as being "mesomolecular" or "mesoatomic" – an aggregate of smaller molecular/atomic subunits.

There are two primary types of nanoscale building blocks that may be used for further device fabrication and applications:

(i) 0D (*e.g.*, nanoparticles, nanoclusters, nanocrystals)
(ii) 1D (*e.g.*, nanotubes, nanofibers, nanowires)

The direct incorporation of these nanoarchitectures in existing materials to improve their properties is often referred to as *incremental nanotechnology*. However, as we will see later in this chapter, the self-assembly of these nanosized building blocks into 2D and 3D architectures may yield entirely new devices and functionalities – referred to as *evolutionary nanotechnology*.

6.2.1. Zero-Dimensional Nanomaterials

Analogous to the period in this sentence, a "zero-dimensional" structure is the simplest building block that may be used for nanomaterials design. These materials have

diameters <100 nm, and are denoted by nanoparticles, nanoclusters, or nanocrystals, which are used synonymously in the literature. However, in order to continue our rapid nanoscience developments, the active participants must share a common language. Since there has been no broad adoption of such terminology, a goal of this chapter is to provide explicit definitions and examples (Figure 6.7), in order to avoid the current nomenclature ambiguities.

The term *nanoparticle* is generally used to encompass all 0D nanosized building blocks (regardless of size and morphology), or those that are amorphous and possess a relatively irregular shape. Herein, we will define nanoparticles as amorphous or semicrystalline 0D nanostructures with dimensions larger than 10 nm, and a relatively large (≥15%) size dispersion. For amorphous/semicrystalline nanostructures smaller in size (*i.e.*, 1–10 nm), with a narrow size distribution, the term *nanocluster* is more appropriate. This distinction is a simple extension of the term "cluster," which is typically used in inorganic/organometallic chemistry to indicate small molecular cages of fixed sizes (Figure 6.7). Analogous to bulk materials, the agglomeration of noncrystalline nanostructural subunits should best be termed a *nanopowder* (Figure 6.7).

It is also important here to note the difference between nanoparticles/nanoclusters and traditional *colloids*, which date back to the early 1860s (Table 6.1). We are all familiar with the term colloid, which is used to describe solid/liquid and solid/gas suspensions such as milk, paints, butter, smoke, and smog. Although both types of materials have sizes within the nanoregime, the leading difference is the control one has over composition and morphology. As we will see shortly, in order to stabilize metal nanostructures, a stabilizing agent must be used to prevent agglomeration into a larger powder. This is also the case for colloids, which generally employ polydispersed organic polymers and other ionic species that may adsorb to the colloid surface. Such a variation in the nature of the encapsulating environment leads to a large dispersity in overall morphology and properties of colloids. By contrast, in order for nanomaterials to be used for "bottom-up" design, their synthesis and resultant properties must be reproducible. This is easily accomplished through the use of stabilizing agents with well-defined structures that do not react with/surface deactivate the entrained nanostructures (*e.g.*, dendrimers, polyoxoanions, *etc*.).

Thus far, we have defined nomenclature for amorphous 0D nanostructures. Analogous to bulk materials, any nanomaterial that is crystalline should be referred to as a *nanocrystal*. This term should be reserved for those materials that are single-crystalline; if a particle exhibits only regions of crystallinity, it is better termed a nanoparticle or nanocluster depending on its dimensions. Transmission electron microscopy, especially in tandem with electron diffraction is most useful in determining the crystallinity of any nanostructure (Figure 6.8).

A special case of nanocrystal that is comprised of a semiconductor is known as a *quantum dot*.[12] Typically, the dimensions of these nanostructures lie in the range 1–30 nm, based on its composition (see below). Quantum dots currently find applications as sensors, lasers, and LEDs. In fact, new high-density disks (*e.g.*, HD-DVD and Blu-ray high-definition DVD formats) may only be read *via* blue lasers, which

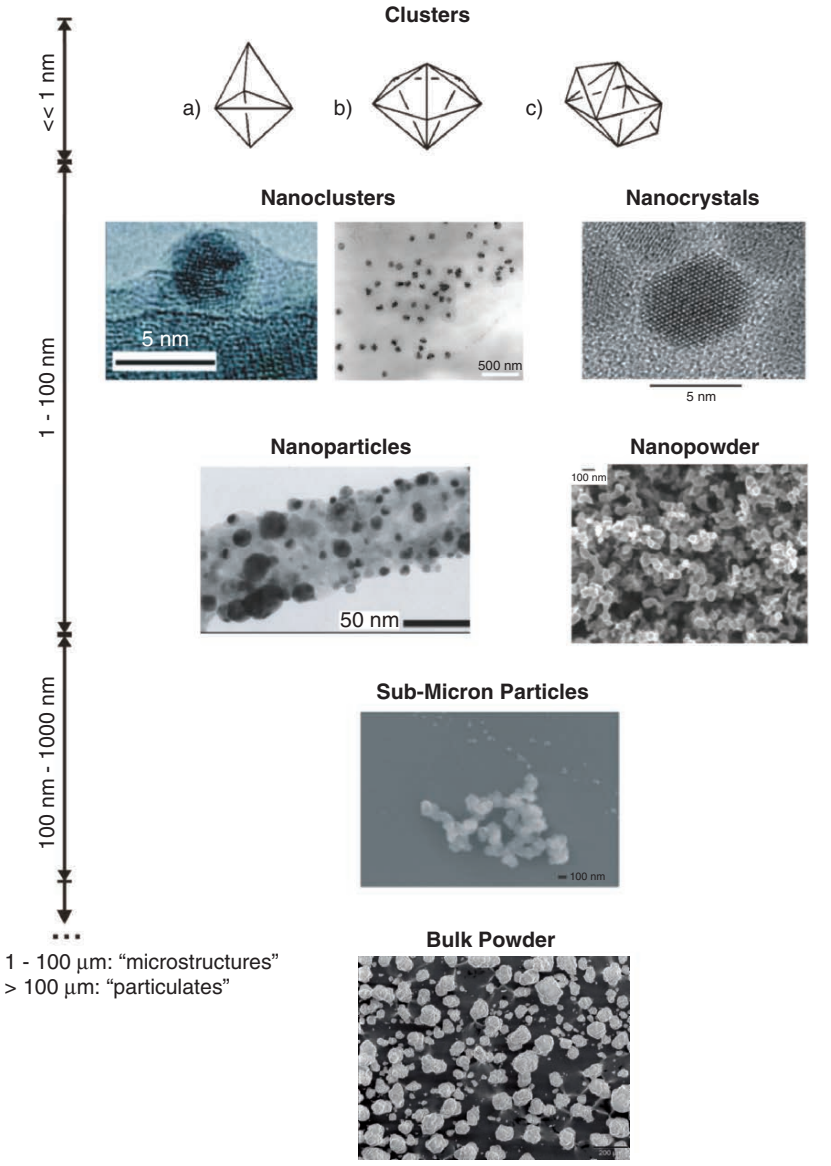


Figure 6.7. 0D nanostructure nomenclature. Shown are the well-defined cage sizes of molecular *clusters* $(Os_5(CO)_{16})$, $(Os_6(CO)_{18})$, $[Os_8(CO)_{22}]^{2-}$.[13] Comparatively, *nanoclusters* should be used to describe 0D nanostructures of a homogeneous size distribution.[14] By contrast, *nanoparticles* exhibit a greater range of sizes/shapes.[15] *Nanocrystals* are characterized by the presence of an ordered lattice array of the constituent subunits, as illustrated by a single nanocrystal of CdSe.[16] In stark contrast to a nanocrystal, an example of a *nanopowder* is shown that consists of microscopic grains, each comprised of nanoscale amorphous units.[17] The size regime that is intermediate between the nano- and microregimes is best referred to as *submicron.*[18] The bulk powder scale bar is 200 µm.

Table 6.1. Comparison of 0D Nanoarchitectures with Traditional Colloids[10]

Nanoparticles/nanoclusters	Colloids
Size: 1–100 nm (nanoclusters: 1–10 nm)	Typically >10 nm
Homogeneous molecular composition ≤15% Size dispersion (less polydispersity for nanoclusters relative to nanoparticles)	Poorly defined compositions >15% Size dispersion
Reproducible synthesis (control over size, shape, and composition)	Nonreproducible, uncontrollable morphology/composition
Reducible physical properties and catalytic activity	Nonreproducible properties (*esp.* irreproducible catalytic activities[11])
Soluble in polar/nonpolar organic solvents (depending on stabilizing agent)	Typically only soluble in polar solvents
Contain clean surfaces	Contain surface-adsorbed species such as –OH, –X, $–OH_2$, *etc*.

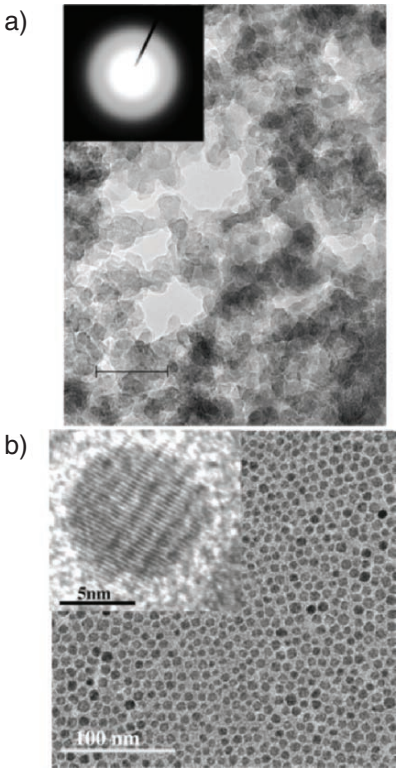


Figure 6.8. TEM images of amorphous nanoclusters (a), and nanocrystals, (b). The inset of (a) shows selected area electron diffraction (SAED); the absence of a pattern indicates an amorphous structure. The scale bar is 20 nm. The high-resolution TEM image inset in (b) shows lattice spacings of an individual nanocrystal. Reproduced with permission from (a) Liu, S.; Fooken, U.; Burba, C. M.; Eastman, M. A.; Wehmschulte, R. J. *Chem. Mater.* **2003**, *15*, 2803. (b) Tirosh, E.; Shemer, G.; Markovich, G. *Chem. Mater.* **2006**, *18*, 465. Copyright 2003 & 2005 American Chemical Society.

Table 6.2. Calculated Exciton Bohr Radii for Various Semiconductors

Material	r_B (Å)
Si	55
CdS	315
CdSe	61
CdTe	100
ZnO	18
ZnS	50
PbS	204
PbSe	460
InAs	340
InSb	540

are fabricated from quantum dots. Long-term applications for these structures will likely include optical computing and high-efficiency solar cells.

As you may recall from Chapter 4, when an electron is promoted from the valence to conduction bands, an electron–hole pair known as an *exciton* is created in the bulk lattice. The physical separation between the electron and hole is referred to as the *exciton Bohr radius* (r_B) that varies depending on the semiconductor composition. In a bulk semiconductor crystal, r_B is significantly smaller than the overall size of the crystal; hence, the exciton is free to migrate throughout the lattice. However, in a quantum dot, r_B is of the same order of magnitude as the diameter (D) of the nanocrystal, giving rise to *quantum confinement* of the exciton. Empirically, this translates to the strongest exciton confinement when $D \leq 2r_B$.

Analogous to the "particle-in-a-box" model from introductory physical chemistry, exciton quantum confinement results in discrete energy levels rather than the continuous bands of a bulk semiconductor crystal. Since the gap between adjacent energy levels is infinitesimally small for a bulk semiconductor, the bandgap is considered as a fixed value. In contrast, since the dimensions of a quantum dot are extremely small, the addition/subtraction of a single atom will significantly change the nanocrystal dimensions and bandgap. Table 6.2 lists the r_B values for common semiconductor crystals whose bandgap may be easily fine-tuned by simply changing the diameter of the quantum dot, as long as the dimensions are smaller than r_B.

In the early 1980s, Efros described the size-dependent electronic properties of quantum dots, first delineating that the bandgap, E_n, will increase from the bulk value based on a $1/R^2$ confinement energy term (Eq. 1). According to quantum confinement theory, electrons in the conduction band and holes in the valence band are spatially confined by the potential barrier of the surface. Due to confinement of both electrons and holes, the lowest energy optical transition from the valence to conduction band will increase in energy, effectively increasing the bandgap.

$$(1) \qquad E_n = E_g + \frac{h^2\pi^2}{2\mu R^2},$$

where E_g is the bandgap of the bulk semiconductor; h, Planck's constant; R, the radius of the quantum dot; and μ is the reduced mass of the exciton given by $m_e m_h/(m_e + m_h)$. Here m_e and m_h are masses of the electron and hole, respectively.

This model was expanded by Brus and coworkers to include Coulombic interaction of excitons and the correlation energy (Eq. 2).

$$E_n = E_g + \frac{h^2\pi^2}{2\mu R^2} - \frac{1.786\,e^2}{\varepsilon R} + 0.284E_R, \quad (2)$$

where E_R is the Rydberg (spatial correlation) energy of bulk semiconductor:

$$E_R = \frac{\mu\, e^4}{2\varepsilon^2\varepsilon_0^2 h^2} = 13.56\frac{\mu}{\varepsilon^2 m_e},$$

where ε_0 is the permittivity of free space; ε, the dielectric constant of the bulk semiconductor; and m_e is the mass of the electron.

Hence, the absorption energy of quantum dots will shift to higher frequency with decreasing diameter of the dots, with a dependence of $1/R^2$. This is readily observed from the reflected colors of quantum dots with varying diameters, shifting from blue to red with increasing size (Figure 6.9).

For anyone who has admired the bright red colors of stained glass windows, it may be surprising to note that this too is a nanoparticle phenomenon. In fact, the reds and yellows observed in these antiquated glasses are typically caused by the presence of

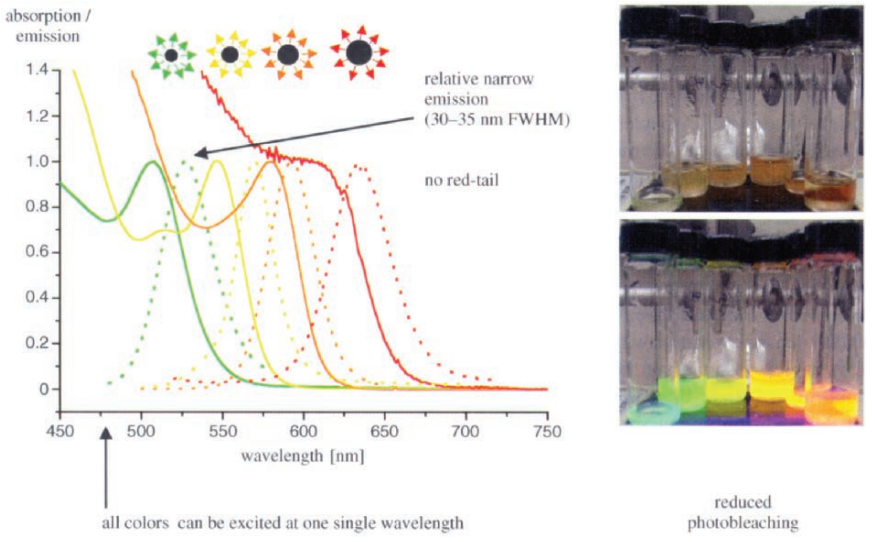


Figure 6.9. The size-dependency on the optical properties of CdSe nanocrystals. With decreasing size, the fluorescence peak is shifted to shorter wavelengths. Also shown is the reduced photobleaching exhibited by CdSe nanocrystals – (top-bottom photos): suspension of nanocrystals in solution, before and after irradiation with UV light, respectively. Reproduced with permission from Pellegrino, T.; Kudera, S.; Liedl, T.; Javier, A. M.; Manna, L.; Parak, W. J. *Small*, **2005**, *1*, 48. Copyright 2005 Wiley-VCH.

gold and silver nanoparticles, respectively. However, for metallic nanoclusters/nanoparticles with diameters *ca.* > 2 nm, the operating principle is different from semiconductor quantum dots, since there is no bandgap between valence and conduction bands, and the energy states form a continuum analogous to bulk metal.

For these metallic nanostructures, another phenomenon known as *surface plasmon resonance* (often denoted as localized surface plasmon resonance, LSPR) is active for these structures, involving specific scattering interactions between the impinging light and the nanostructures. In particular, the oscillating electric field of the incoming light causes the coherent oscillation of the conduction electrons, resulting in a concomitant oscillation of the electron cloud surrounding the metal nuclei (Figure 6.10).[19]

The leading theory that describes the scattering behavior of small spherical particles with light dates back to the work of Mie in the early 1900s (Eq. 3).[20]

$$(3) \qquad E(\lambda) = \frac{24\pi N_A r^3 {\varepsilon_m}^{3/2}}{\lambda \ln(10)} \left[\frac{\varepsilon_i(\lambda)}{(\varepsilon_r(\lambda) + 2\varepsilon_m)^2 + \varepsilon_i^2(\lambda)} \right],$$

where $E(\lambda)$ is the extinction (sum of absorption and scattering); N_A, the density of the nanostructures; r, the radius of the nanostructure; ε_m, the dielectric constant of the metallic nanostructure; λ, the wavelength of absorbing radiation; and ε_i, ε_r are the imaginary and real (respectively) portions of the λ-dependent dielectric function of the nanostructure.

This theory is still relevant today, alongside Raleigh scattering theory, to describe the colors of our sky, as well as the appearance of suspensions such as milk and latex paints. It should be noted that for nonspherical nanostructures, the denominator of the bracketed term above is replaced with:

$$(\varepsilon_r(\lambda) + \chi\varepsilon_m)^2,$$

where χ may range from a value of 2 (perfect sphere) to 17 (*e.g.*, for a 5:1 aspect ratio nanostructure) and beyond.[21]

Based on Eq. 3, the factors that govern the oscillation frequency (and the observed color) are: electron density (size/shape of the nanostructures, Figure 6.11), the effective nuclear charge of the nuclei, and the size/shape of the charge distribution (polarization effects, strongly affected by the dielectric constant of the metal). As you might expect, further effects toward the resonance frequency/intensity are

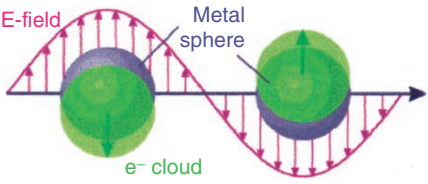


Figure 6.10. Schematic of LSPR for a nanosphere, showing the induced oscillation of the electron cloud relative to the nuclei. Reproduced with permission from Kelly, K. L.; Coronado, E.; Zhao, L. L.; Schatz, G. C. *J. Phys. Chem. B* **2003**, *107*, 668. Copyright 2003 American Chemical Society.

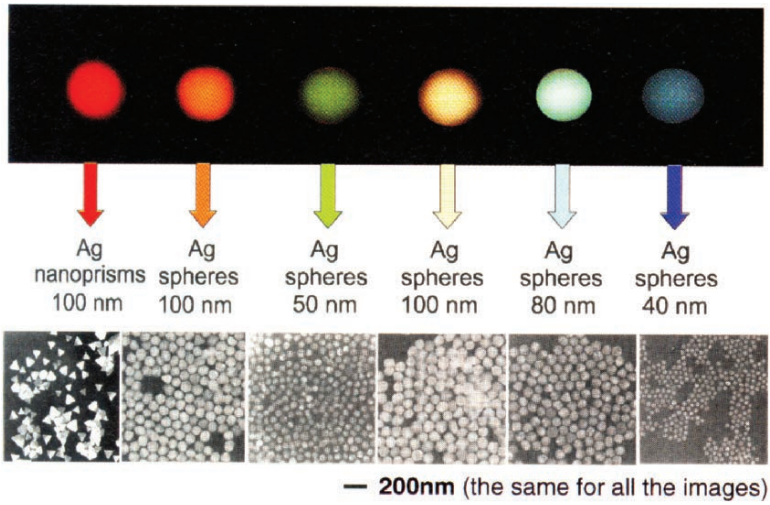


Figure 6.11. The influence of size and shape on the light-scattering, and resultant colors, of silver nanoparticles. Reproduced with permission from Mirkin, C. A. *Small* **2005**, 1, 14. Copyright 2005 Wiley-VCH.

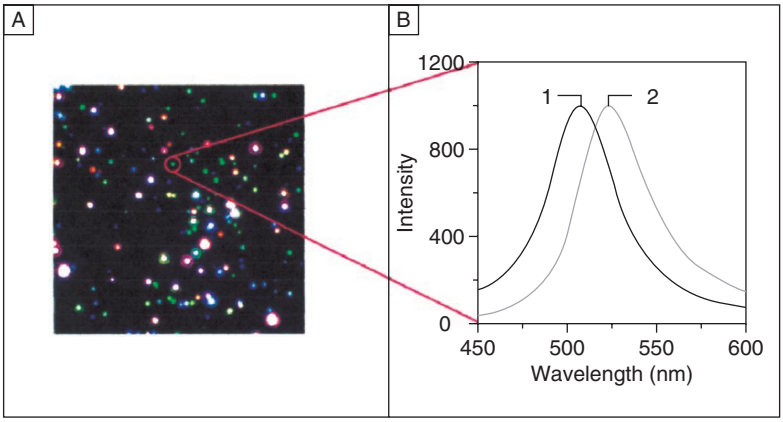


Figure 6.12. Biological sensing using Ag nanoparticles. Shown is (a) a dark-field optical image of surface-functionalized (biotinylated) Ag nanoparticles, and (b) the shift in wavelength before (1) and after (2) exposure to 10 nM streptavidin. Reproduced with permission from Haes, A. J.; Stuart, D. A.; Nie, S.; Duyne, R. P. V. *J. Fluoresc.* **2004**, *14*, 355. Copyright 2004 Springer Science and Business Media.

observed from the solvent and substrate that surrounds the nanostructures. As such, noble metal nanoparticles have been used as tunable platforms for biological sensing (Figure 6.12).

For the smallest of metallic nanoclusters with dimensions *ca.* < 2 nm, the surface plasmon absorption disappears. Since so few atoms comprise discrete nanoclusters of this size, the spacings between adjacent energy levels (referred to as the *Kubo*

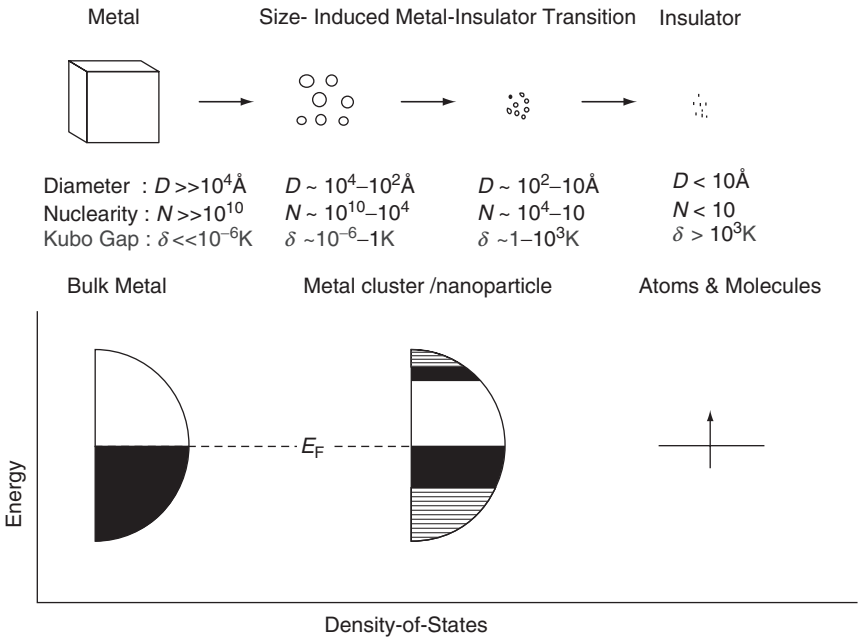


Figure 6.13. Schematic of the density of states exhibited by bulk metal relative to increasingly smaller nanoclusters. The approximate diameter, nuclearity, and Kubo gap for each size regime are indicated. As the nanocluster size decreases, the energy continuum of the bulk metal is transformed into discrete energy levels, especially at band edges. Reproduced with permission from Rao, C. N. R.; Kulkarni, G. U.; Thomas, P. J.; Edwards, P. P. *Chem. Soc. Rev.* **2000**, *29*, 27. Copyright 2000 Royal Society of Chemistry.

gap, δ, Eq. 4) become comparable to the thermal energy, kT – especially at lower temperatures and smaller nanocluster diameters. This results in a shift in conductive properties of the nanocluster, from metallic to semiconducting and insulating, with decreasing size (Figure 6.13). In band theory, the breadth of a band is directly related to the strength of interactions among nearest neighbors. For atoms/molecules (Figure 6.13, far right) this interaction is weak, which corresponds to a narrow band. However, as additional atoms are added to the solid, stronger interactions will ensue, resulting in a greater density of states near the Fermi level – additional states near the band edges develop last. Hence, the insulating properties of very small nanoclusters are due to the small number of neighboring atoms, which are held together primarily by nonmetallic interactions (*e.g.*, van der Waal forces).[22a] As nanocluster size increases, the atomic s/p and d orbitals from a larger number of constituent atoms will broaden into bands, forming an energy continuum.

Quantum confinement effects also cause a change in the optical properties of metallic nanoclusters. For instance, since the spacing between intraband energy levels increases with decreasing nanocluster size, the 4s/p to 5d fluorescent emission for gold (*e.g.*, Au_{40} – *ca.* 1.1 nm diameter, Au_8, *etc.*) will become increasingly blue-shifted with decreasing dimensions.[22]

$$\delta = \frac{3E_f}{2N}, \quad (4)$$

where δ is the Kubo gap, energy spacing between adjacent E levels; N, the nuclearity, the number of atoms in the nanocluster; and E_f is the energy of the Fermi level.

Mechanism for the nucleation/growth/agglomeration of metal nanoclusters

In order to maintain control over the composition and morphology of a 0D nanostructure, it is essential that we understand the self-assembly mechanism for these structures. Only within the last decade have we figured out ways to repeatably control the morphology and composition of nanoparticles/nanoclusters. This explains why the mechanistic details of nanocluster growth have not surfaced until recently. The first system to be investigated were metallic iridium nanoclusters, formed through the hydrogenation of a polyoxoanion-supported Ir complex (Figure 6.14).[23] Since Ir does not have an observable surface plasmon resonance profile, the rate of nanocluster growth was determined by following the hydrogenation of alkenes over time. This is possible in this system since the initial complex is not an active toward alkene hydrogenation; catalytic activity arises only from the reduced metal (Ir^0).

The overall four-step mechanism for nanocluster growth/agglomeration is shown in Figure 6.15.[24] Although this pathway is based on metallic nanocluster studies, other types of nanoclusters/nanoparticles would likely follow a similar route.[25] The first step involves the slow, continuous nucleation of clusters that are much smaller than 1 nm. When an energetically favored critical nucleus size is reached (*ca*. 15 atoms for Ir^0; varies depending on the metal), additional atoms rapidly aggregate to the surface – a process referred to as *autocatalytic surface growth*. The term "autocatalytic" is used since the nanoclusters formed from the nucleation step are also reactants for subsequent surface growth. Accordingly, the autocatalytic step will proceed faster as the reaction progresses, effectively shutting down the nucleation step – essential to achieve monodisperse nanocluster growth. This is a crucial finding, since this explains the strict size control that is possible for metal nanocluster growth. That is, the size of the growing nanocluster may be controlled by varying the relative values of k_1 and k_2, as well as the availability of additional precursor molecules. In this fashion, metal nanoclusters have been designated as "living metal polymers," analogous to organic living polymers discussed in Chapter 5.

Interestingly, it has been shown that the growth of nanoclusters proceeds through the formation of "magic number" (or closed shell) clusters that exhibit unusual electronic stability. For fcc or hcp transition metals, stable clusters contain 13, 55, 147, 309, 561, 923, 1,415,… metal atoms, where the number of surface atoms in the nth shell is given by $10n^2 + 2 (n = 1, 2, 3, 4, 5, \ldots)$.[26] It should be noted that intermediate magic-number nanoclusters represent only local minima in the potential energy surface, relative to the global minimum of a bulk metal with the lowest possible surface area (Figure 6.16).[27] The high yield of magic number nanoclusters is a consequence of kinetically controlled surface growth. That is, once these favored intermediate structures are formed, they are less reactive toward autocatalytic surface growth relative to nonmagic number clusters.

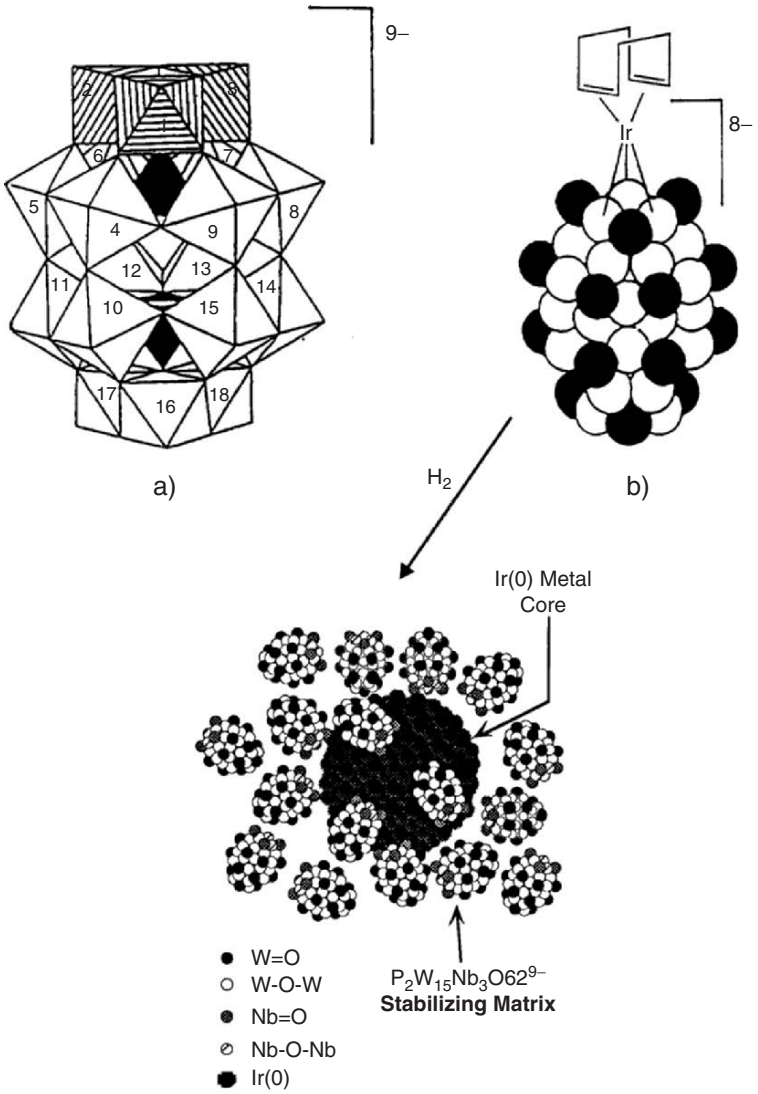


Figure 6.14. Polyoxoanion-stabilized Ir^0 nanocluster formation. Shown is (a) Polyhedral representation of the $\alpha-1, 2, 3-P_2W_{15}Nb_3O_{62}^{9-}$ stabilizing anion, and (b) a space-filling representation of the $[(1, 5-COD)Ir(P_2W_{15}Nb_3O_{62})]^{8-}$ complex. In (a), the three Nb atoms are indicated by the striped octahedra in positions 1–3. The WO_6 polyhedra occupy the 4–18 positions, and the internal PO_4 tetrahedral units are illustrated in black. In (b), the black spheres represent M–O terminal groups, and the white spheres represent M–O–M bridging groups. The Bu_4N^+ and Na^+ counterions are omitted for clarity. Images (a) and (b) reproduced with permission from Finke, R. G.; Lyon, D. K.; Nomiya, K.; Sur, S.; Mizuno, N. *Inorg. Chem.* **1990**, *29*, 1784. The bottom image of the stabilized nanocluster is reproduced from Watzky, M. A.; Finke, R. G. *Chem. Mater.* **1997**, 9, 3083. Copyright 1997 American Chemical Society.

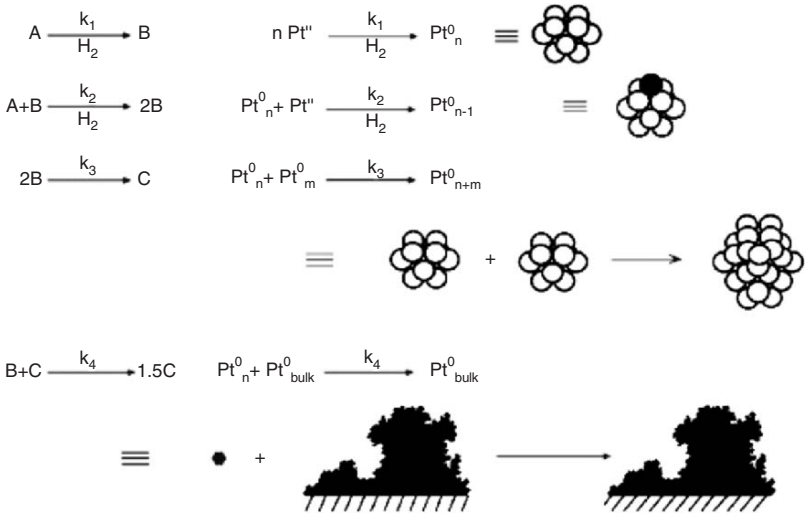


Figure 6.15. Schematic of the four-step mechanism for transition metal (*e.g.*, Pt) nanocluster formation. Shown are (i) nucleation to a desired cluster size; (ii) autocatalytic growth onto the cluster surface; (iii) diffusive agglomerative growth of two nanoclusters; and (iv) autocatalytic agglomeration into bulk metal particulates. Reproduced with permission from Besson, C.; Finney, E. E.; Finke, R. G. *J. Am. Chem. Soc.* **2005**, *127*, 8179. Copyright 2005 American Chemical Society.

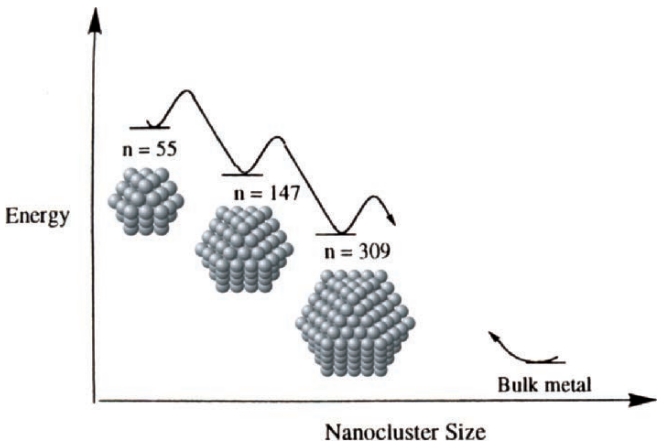


Figure 6.16. Comparative energy levels for "magic number" noble metal nanoclusters relative to bulk metal. Also shown are the perfect fcc arrays of the nanoclusters, in which more than 75% of the atoms are located on the surface. Adapted with permission from Finke, R. G. in *Metal Nanoparticles: Synthesis, Characterization, and Applications* Feldheim, D. L.; Foss, C. A. eds., Dekker: New York, 2002. Copyright 2002 Taylor & Francis.

The last two steps of the nanocluster growth mechanism involve agglomeration. This phenomenon whereby larger (nano+) particulates grow at the expense of smaller nanoclusters is often referred to as coarsening or *Ostwald ripening*. Whereas step 3 illustrates a bimolecular aggregation of two nanoclusters, the last step involves the formation of bulk metal particles through a second autocatalytic surface-growth process. These two steps are obviously undesired pathways during nanocluster growth. However, it is essential to have these mechanistic steps (finally) stated, since this will lead to future quantitative studies that will identify the best conditions to prevent agglomeration (varying ratios of k_1–k_4, stabilizing agents, metals, *etc*.). This data is currently lacking, and one simply "hopes for the best" regarding the long-term stability of the grown nanoclusters based on a chosen stabilizing agent. A general prediction of the overall mechanism shown in Figure 6.15 is that lower concentrations and higher temperatures are most conducive to yield nanoclusters rather than bulk metal (*i.e.*, $k_1 \gg k_3$ and k_4). Indeed, this prediction is backed up by experimental data.

The first 0D nanoarchitecture: the fullerenes

Without question, our modern "nanotechnology revolution" was catalyzed by the mid-1980s discovery of carbon nanoclusters known as *fullerenes* (*e.g.*, C_{60}, Figure 6.17).[28] The 1996 Nobel Prize in Chemistry was awarded to Richard Smalley, Robert Curl, and Sir Harold Kroto for this discovery, which focused the worldwide spotlight on unique nanoscale materials and their possible applications. To date, the most common applications for fullerenes include MRI contrast agents (exploiting its container properties), drug-delivery agents (through surface functionalization), fulleride-based superconductors, and light-activated antimicrobial agents[29] (Figure 6.18). It should be noted that the term fullerene does not simply

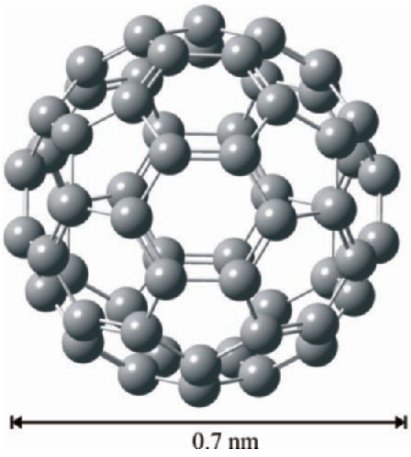


Figure 6.17. Molecular structure of Buckminsterfullerene, C_{60}, containing alternating six- and five-membered rings of sp^2 hybridized carbon atoms. This is only one isomer for C_{60}, out of a staggering total of 1,812 possible structures.[34]

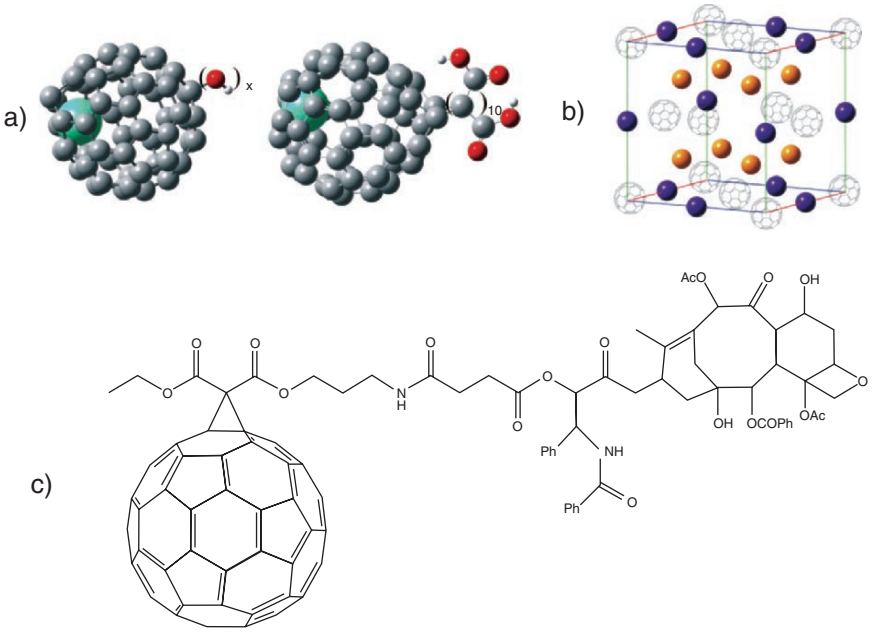


Figure 6.18. Examples of fullerene applications. Shown are a) $Gd^{3+}@C_{60}–[OH]_x$ ($x \approx 27$) and $Gd^{3+}@C_{60}–[CCOOH]_{10}$ used as MRI contrast agents,[35] (b) the ionic unit cell for the superconductive alkali metal fulleride $Cs_xRb_yC_{60}$,[36] and (c) the conjugate structure of C_{60} covalently bound to the lung cancer drug Paclitaxel.[37]

mean the C_{60} carbon allotrope termed *Buckminsterfullerene* (or "Buckyballs"),[30] but rather the entire class of closed-cage carbon clusters that are comprised of exactly 12 pentagons, and a varying number of hexagons (*e.g.*, 20 hexagons for C_{60}).

As with other major discoveries such as Teflon and nylon,[31] the synthesis of C_{60} was serendipitous. In fact, the experiments that led to the discovery of fullerenes were aimed at simulating the environment of a carbon-rich red giant star. As such, the discovery of C_{60} has been dubbed "the celestial sphere that fell to earth."[32] The system featured the laser vaporization of a graphite target into a helium carrier gas wherein the atoms nucleated into clusters. The gas was then cooled using supersonic expansion, and injected into a time-of-flight mass spectrometer for analysis (Figure 6.19a). Reactive gases such as hydrogen or nitrogen could also be added to the carrier gas, for the synthesis of other reaction products. It should be noted that this technique is now used to generate nanoclusters of a wide variety of metals, semiconductors (*e.g.*, Si), and insulators (*e.g.*, Al_2O_3) depending on the nature of the target and co-reactant gases employed.

The first large-scale synthesis of fullerenes was discovered in 1989 by Huffman and Kratschmer; this provided a macroscopic quantity of C_{60} in order to confirm the proposed icosohedral structure. Their technique consisted of the arc-evaporation of graphite electrodes *via* resistive heating within an atmosphere of *ca*. 100 atm. helium

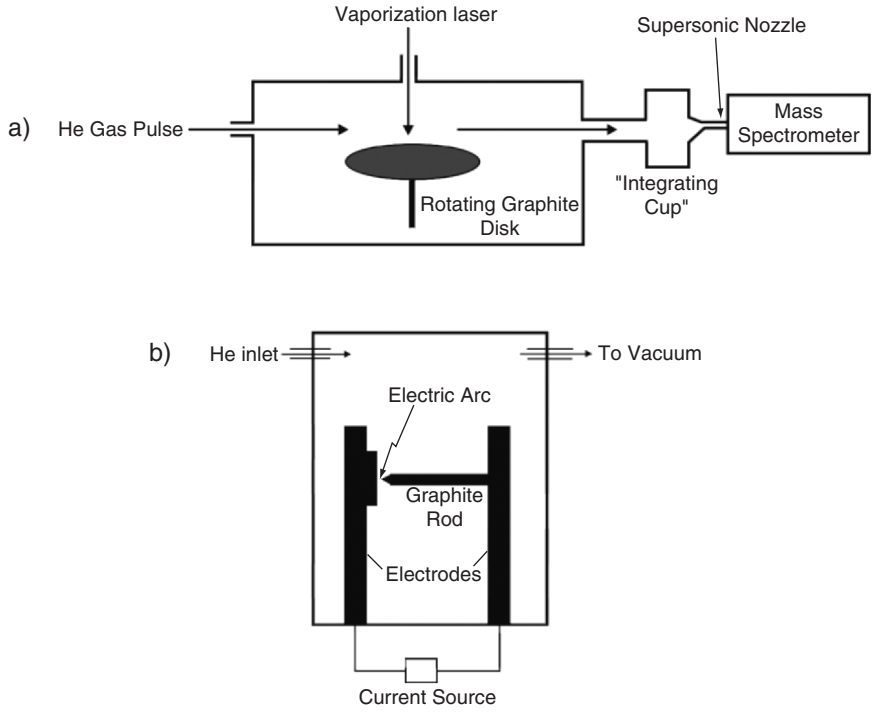


Figure 6.19. Schematics of apparati first used to synthesize fullerenes. Illustrated are (a) the Smalley/Curl supersonic laser evaporation system, and (b) the Huffman/Kratschmer electric arc apparatus.

(Figure 6.19b).[33] Other more recent techniques such as high-temperature combustion of benzene and a benchtop graphite arc process (Figure 6.20) were developed in order to reduce the cost and complexity associated with fullerene synthesis. Not only will such improvements allow for more widespread fabrication of fullerenes for research/applications (*i.e.*, synthesis no longer limited to groups with supersonic laser and arc plasma systems), but may also result in lowering the cost due to a more straightforward industrial scale-up.

Buckyballs represent the smallest fullerene that obeys the *Isolated Pentagon Rule* (IPR) – *i.e.*, an energetic requirement that pentagons be surrounded by hexagons, so that adjacent pentagons do not share an edge. Calculations show that π bonds shared between six-membered rings have large positive bond resonance energies (BREs) and bond orders, indicating a high degree of aromaticity and stability/unreactivity. However, π bonds between adjacent five-membered rings have large negative BREs with very small bond orders, indicating a much lower thermodynamic stability. Most likely, this difference is due to the increased ring strain that would be imposed in the fullerene structure as a result of two smaller rings directly adjacent to one another. Theoretical calculations indicate that the strain energy of the icosohedral Buckminsterfullerene structure is at least 2 eV lower than any other non-IPR isomer, of which there are over 1,800 possibilities.

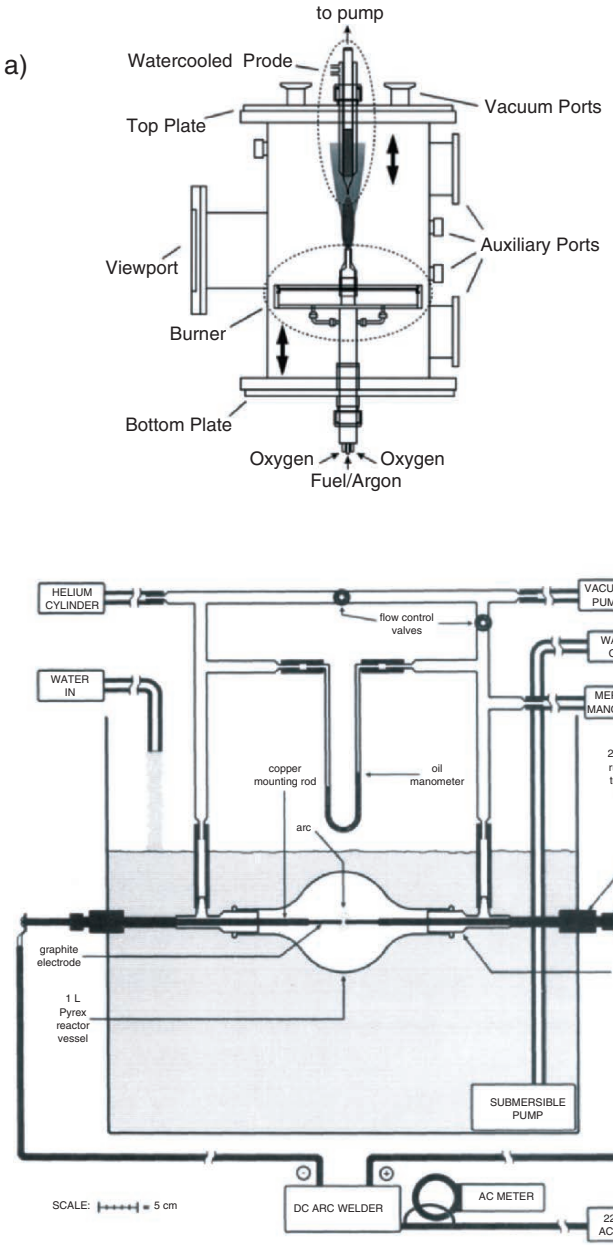


Figure 6.20. Cross-section schematics of reactors used for fullerene synthesis. Shown are (a) a reduced-pressure fuel-rich pyrolytic chamber, and (b) a benchtop modified arc evaporation system. Reproduced with permission from (a) Hebgen, P.; Goel, A.; Howard, J. B.; Rainey, L. C.; Vander Sande, J. B. *Proc. Combust. Inst.* **2000**, *28*, 1397, Copyright 2000 Elsevier, and (b) Scrivens, W. A.; Tour, J. A. *J. Org. Chem.* **1992**, *57*, 6932. Copyright 1992 American Chemical Society.

Interestingly, it has recently been reported that adjacent pentagons containing at least one N atom instead of C (*e.g.*, $C_{58}N_2$ rather than C_{60}), may actually be more stable than C_{60} (Figure 6.21).[38] This apparent anomaly is a direct contradiction of the IPR. The most plausible explanation is the reduction of ring strain due to sp^3 hybridization of N, as well as the addition of π-electron density (from the N lone-pair electrons) to the pentagons, resulting in an enhanced aromaticity/stability. To date, only short-lived azafullerenes $C_{59}N$ and $C_{58}N_2^-$ have been identified experimentally; the search continues for stable structures, since these will likely result in dramatically different properties and associated applications relative to their C-only analogues.

Although fullerenes have been actively investigated for more than two decades, there is an ongoing debate regarding the growth mechanism of these nanoclusters. Since the formation of fullerenes *via* laser/arc/combustion techniques occurs too rapidly to isolate intermediate species, most of the mechanistic proposals are based on theoretical techniques (quantum mechanical and molecular dynamics). It was once thought that fullerenes were formed from the folding of preformed graphitic sheets that emanated from the target following laser ablation. However, a variety of experiments have shown that the growth process initiates from small linear chains of carbon atoms. As the number of carbon atoms increases, the chains preferentially connect into ring structures due to their greater stabilities. In particular, for C_n where $n < 10$ (with the exception of C_6 as discussed below), linear species are the preferred morphology rather than rings (Figure 6.22). The preference for ring formation for $n = 6$, and $n \geq 10$ (especially for C_{10}, C_{14}, C_{18}, *etc.*) is due to the enhanced aromaticity/stability of planar rings when there are $4n + 2\pi$ electrons (where $n = 1, 2, 3, \ldots$ – the *Hückel rule*).

Figure 6.23 illustrates a proposed mechanism for the subsequent steps of fullerene growth, involving the self-assembly of carbon rings. When $n \geq 30$ or so, the monocyclic rings are proposed to form graphitic sheets. The "pentagon road" mechanism proposed by Smalley[39] assumes that the graphitic sheets contain both hexagon and pentagon units. Closure of the sheets to form Buckyballs effectively results in growth

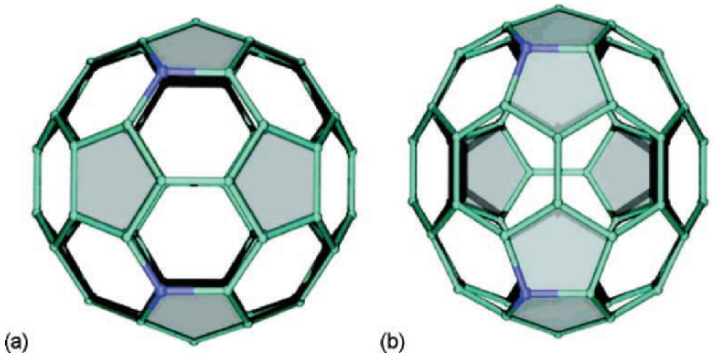


Figure 6.21. Illustration of $C_{58}N_2$ that (a) satisfies the IPR and (b) violates the IPR with adjacent pentagons. The structure with adjacent pentagons, (b), is more stable than (a) by 12.5 kcal mol^{-1}. Reproduced with permission from Ewels, C. P. *Nano Lett.* **2006**, 6, 890. Copyright 2006 American Chemical Society.

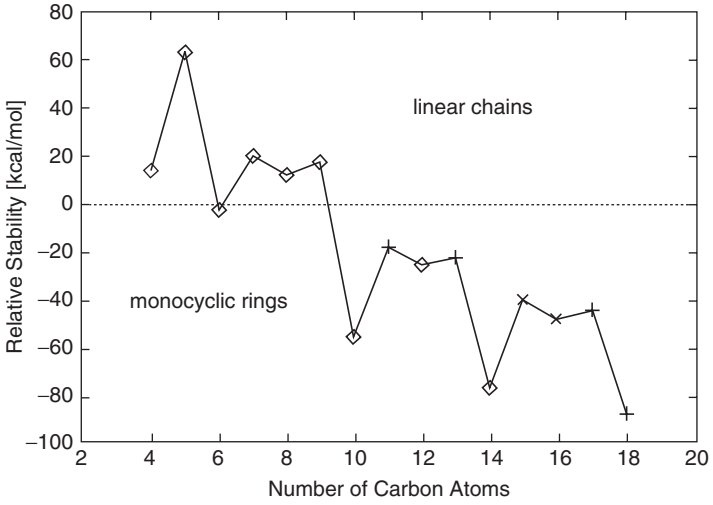


Figure 6.22. The relative stabilities of linear-chain carbon clusters vs. monocyclic rings with changing cluster size. Reproduced with permission from Hutter, J.; Luthi, H. P.; Diederich, F. *J. Am. Chem. Soc.* **1994**, *116*, 750. Copyright 1994 American Chemical Society.

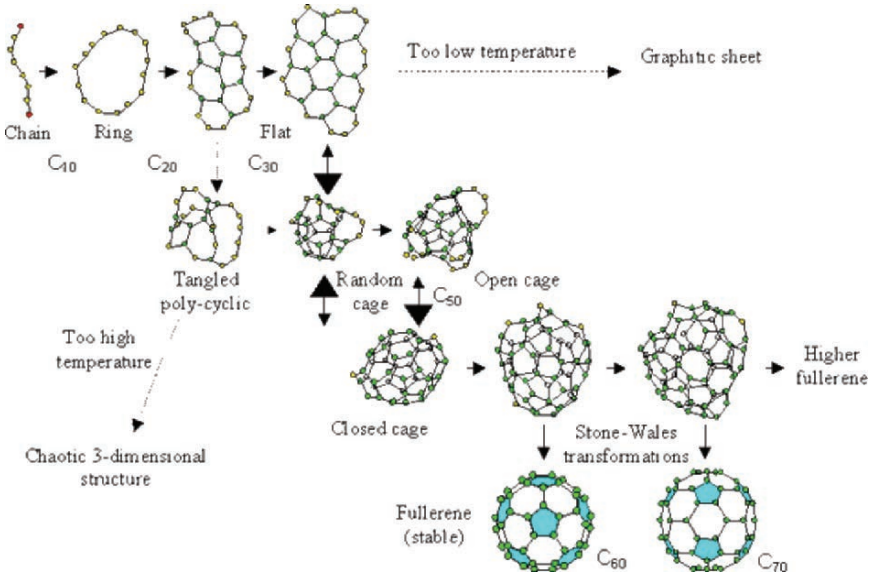


Figure 6.23. Proposed mechanism for fullerene growth. Reprinted from Yamaguchi, Y.; Maruyama, S. *Chem. Phys. Lett.* **1998**, *286*, 343. Copyright 1998, with permission from Elsevier. 87985734.

termination. In contrast, the "fullerene road" model assumes the initial formation of smaller non-IPR fullerenes, which undergo thermal rearrangement to yield C_{60} and higher fullerenes.[40]

As discussed earlier, pentagon units are essential to the fullerene structure, since they allow the planar graphitic sheet to curl. The driving force for this rearrangement is likely the C–C bonding of edge carbons that satisfies their unfilled valences. As the prevailing mechanism points out, adequate annealing is required in order to incorporate a sufficient number of pentagons. For instance, if the cooling rate is too high, amorphous soot particles will be preferentially formed rather than fullerenes. In addition, an overall low growth temperature will not be sufficient to cause cage formation, yielding planar graphitic fragments instead of fullerenes.

Regardless of the proposed mechanism, a final thermal annealing step is likely required to organize the hexagon and pentagon subunits into the lowest-energy IPR arrangement. This rearrangement step is known as the *Stone–Wales (SW) transformation*, and involves a concerted reorganization of the hexagon/pentagon units. We already saw an example of a rare SW transformation where the non-IPR N-containing species was actually lowest in energy (Figure 6.21). However, most often this rearrangement occurs in the opposite direction – transforming adjacent pentagons into a hexagon-isolated structure. It should be noted that the Stone–Wales transformation is actually thermally forbidden *via* the Woodward–Hoffman rules; calculations show an energy barrier of at least 5 eV for this pathway. However, it has been shown that this rearrangement may likely be catalyzed by additional carbon and/or hydrogen atoms that are present during laser/arc or thermal combustion syntheses (Figure 6.24).

Interestingly, a fullerene structure may serve as a nucleation site for additional layers of graphite en route toward multishell fullerenes (Figure 6.25). These are denoted as "$C_{60}@C_{240}$" where the @ symbol represents the encapsulated species. There are even triple-layered structures such as "$C_{60}@C_{240}@C_{560}$."[41] Though very small quantities ($<0.01\%$) of multilayered fullerenes are found in the soot resulting from laser vaporization, the yield may be improved by *in vacuo* sublimation of the vapor phase at a high temperature (*ca.* 1,200°C).

Although the "brute force" methods of laser/arc and high-temperature pyrolysis represent the most common techniques for generating fullerenes, a goal of the synthetic organic chemist has long been the solution-phase, stepwise synthesis of C_{60}. In 1999, a promising step in that direction was accomplished with the first nonpyrolytic synthesis of "buckybowls."[42] These structures had a bowl-shaped structure, and consisted of the hexagon-isolated pentagon backbone exhibited by fullerenes (Figure 6.26). In early 2002, a chlorinated C_{60} precursor was reported using a traditional 12-step organic synthesis. This compound was subsequently converted to Buckyballs using high-temperature vacuum pyrolysis (Figure 6.27). The yield of C_{60} was $<1\%$ – certainly not useful for commercial production of Buckyballs! However, the novelty of this approach was that pyrolysis did not decompose the precursor into smaller units, but rather served to stitch together adjacent arms of the molecular precursor. Hence, this method provides a targeted route toward individual fullerenes

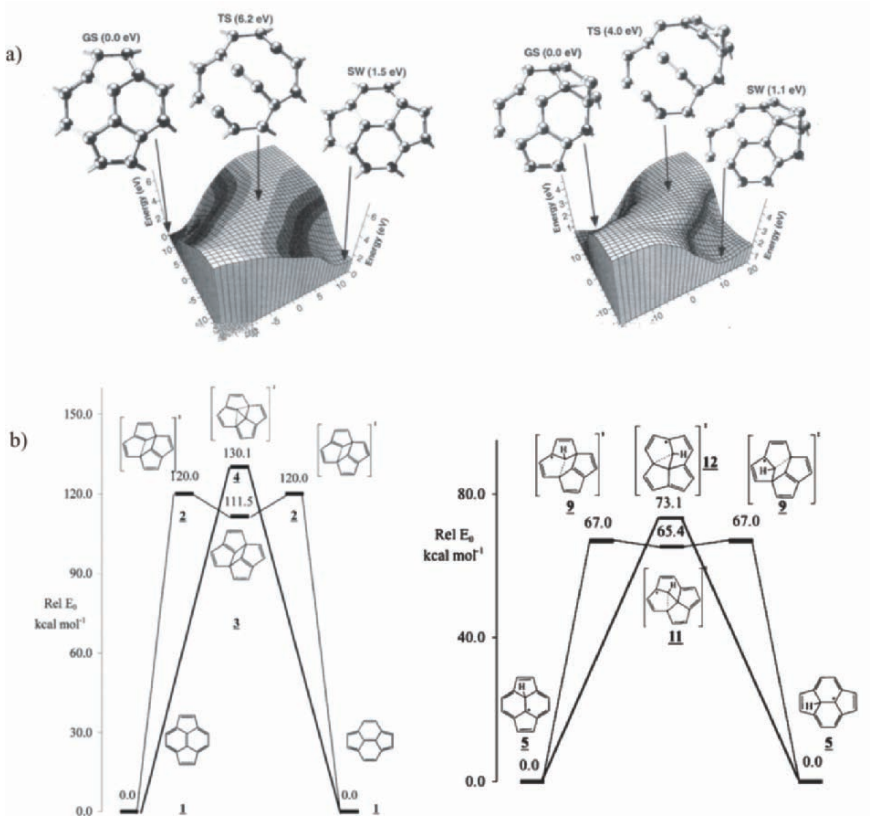


Figure 6.24. The potential energy surface of the Stone–Wales transformation before/after the addition of catalyzing moieties such as (a) carbon, and (b) hydrogen atoms. Reproduced with permission from (a) Eggen, B. R.; Heggie, M. I.; Jungnickel, G.; Latham, C. D.; Jones, R.; Briddon, P. R. *Science* **1996**, *272*, 87, Copyright 1996 AAAS; and (b) Nimlos, M. R.; Filley, J.; McKinnon, J. T. *J. Phys. Chem. A* **2005**, *109*, 9896. Copyright 2005 American Chemical Society.

based on the structure of the precursor, rather than high-energy methods that always result in a mixture of fullerene products.

In addition to pristine fullerene structures, it has been discovered that various metal ions may be encapsulated inside the caged structure to yield *endohedral fullerenes*. Thus far, a variety of alkali and lanthanide metals, Group V atoms, noble gases, and neutral molecules such as CO and H_2O have been sequestered inside the C_{60} structure. Calculations have shown that the encapsulation of noble gas atoms and small ions (*e.g.*, Li^+, F^-, Cl^-) actually stabilize the fullerene cage, whereas larger species (*e.g.*, Rb^+, Br^-, I^-) destabilize the cage.[43] *Metallofullerenes* ($M@C_x$) are typically grown by either laser ablation of metal-doped graphite disks at high temperature (*ca.* 1,200°C), or carbon arc techniques with metal-doped graphite rods. An example of a

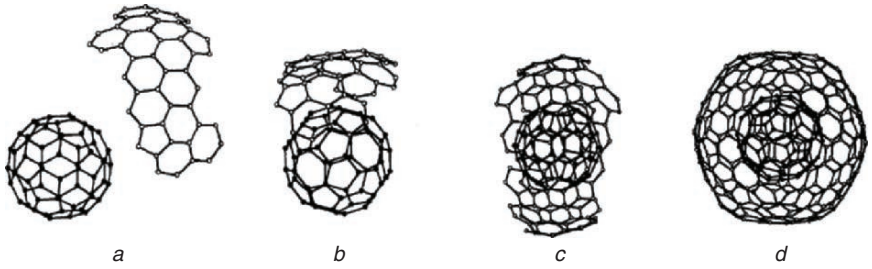


Figure 6.25. Proposed scheme for the formation of the multishell fullerene $C_{60}@C_{240}$. Reproduced from Mordkovich, V. Z.; Shiratori, Y.; Hiraoka, H.; Takeuchi, Y. *Synthesis of Multishell Fullerenes by Laser Vaporization of Composite Carbon Targets*, found online at http://www.ioffe.rssi.ru/journals/ftt/2002/04/p581–584.pdf.

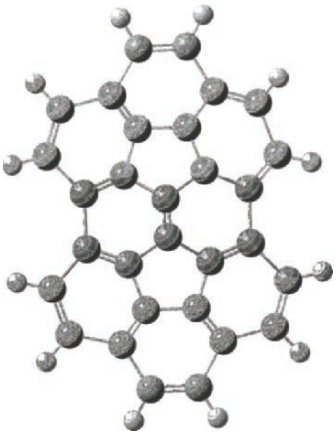


Figure 6.26. Backbone structure of semibuckminsterfullerene, $C_{30}H_{12}$.

metallofullerene ($Gd^{3+}@C_{60}$) was shown in Figure 6.18a; these structures are commonly employed as MRI contrast agents.

You may be wondering "*how does the ion get inside the cage?*" That is, does this occur during the growth of the fullerene structure itself, or does the metal ion insert after the cage is already formed? It has been shown that the latter occurs, with the exact entrance pathway dependent on the size of the dopant species. Small dopants such as He or H^+ may directly pass through either hexagon or pentagon units of the cage toward the vacant core. However, for larger ions/atoms, some framework C–C bonds must be reversibly broken in order to accommodate the incoming species – aptly referred to as a *window mechanism* (Figure 6.28). Since non-IPR fullerenes have relatively large strain energies due to fused pentagons, this process should occur readily for these structures. Indeed, there has been much recent interest in synthesizing "unconventional" metallofullerenes such as $Sc_2@C_{66}$.[44] Unlike the empty C_{66} counterpart, these structures are stable since the incoming metal atom donates

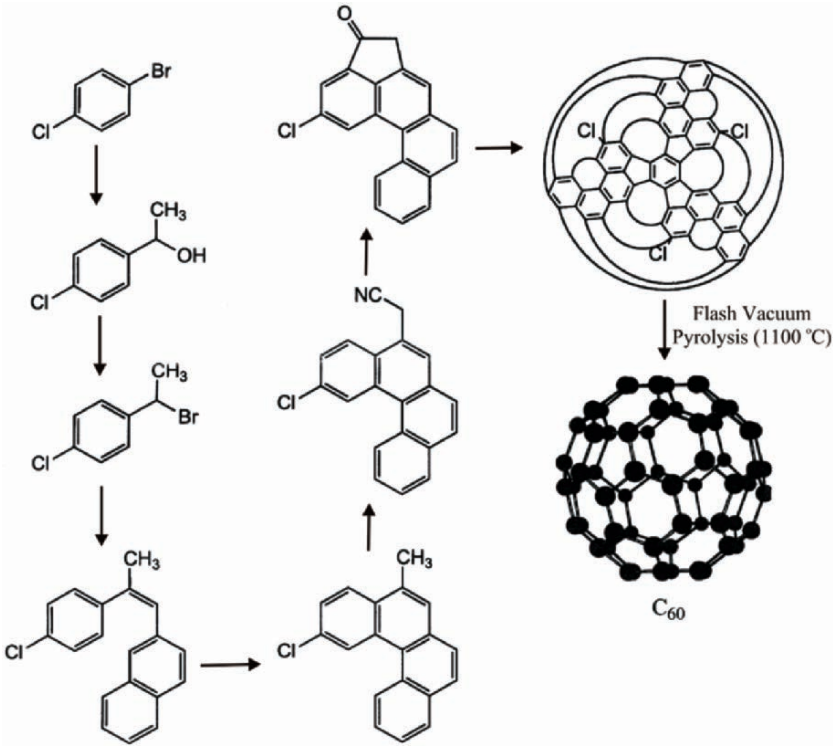


Figure 6.27. Synthesis of a C_{60} precursor. Reproduced with permission from Scott, L. T.; Boorum, M. M.; McMahon, B. J.; Hagen, S.; Mack, J.; Blank, J.; Wegner, H.; de Meijere, A. *Science* **2002**, *295*, 1500. Copyright 2002 AAAS.

electron density to the C–C bond between adjacent pentagons, causing a decrease in the local bond strain.

As you might imagine, relatively large ions such as Cs^+, Y^{3+}, or Sc^{3+} are likely not encapsulated through a simple reversible windowing mechanism. In order for this to occur, more than one C–C bond would need to be broken (Figure 6.28b), which increases the energetic barrier for this to occur. Recently, a "hole-repairing mechanism" was proposed for $Y@C_{82}$ metallofullerenes, in which calculations predict the combination of a large C_{76} open-cage fullerene and a smaller C_6Y fragment that effectively repairs the framework hole (Figure 6.29).[45]

The solution-phase synthesis of 0D nanostructures

Most of our discussion thus far has involved some rather extreme synthetic environments of laser, arc, or pyrolysis. However, a preferred route toward nanoclusters/nanoparticles of metals and their compounds is through use of relatively mild conditions – often taking place at room temperature on the benchtop. This is not possible for carbon nanoallotropes, since the precursor (*e.g.*, graphite) contains

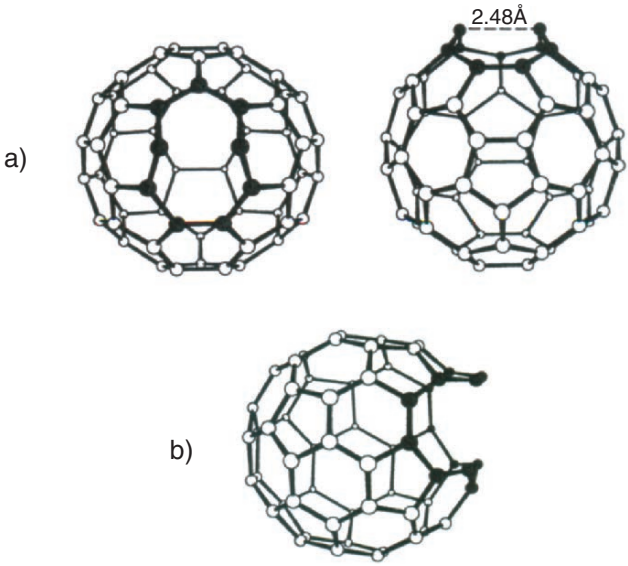


Figure 6.28. Theoretical intermediates during endohedral fullerene formation. Image (a) shows the nine-membered ring formed by a single pentagon–hexagon bond. By comparison, image (b) illustrates the formation of a larger 13-membered ring by breaking two framework bonds. Reproduced with permission from Murry, R. L.; Scuseria, G. E. *Science* **1994**, *263*, 791. Copyright 1994 AAAS.

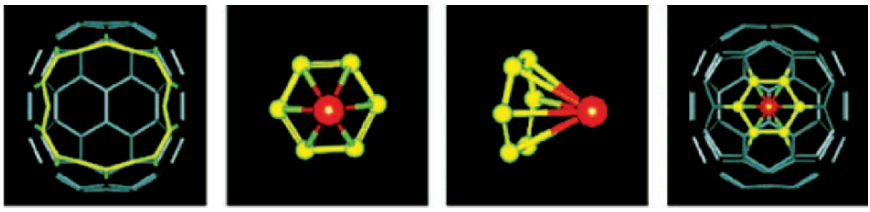


Figure 6.29. Images of the calculated structures involved in the "hole-repairing" mechanism for endohedral fullerene growth. Shown from left to right are: the C_{76} open cage, top and side views of the C_6Y fragment, and the final $Y@C_{82}$ metallofullerene. Reproduced with permission from Gan, L.-H.; Wang, C.-R. *J. Phys. Chem. A* **2005**, *109*, 3980. Copyright 2005 American Chemical Society.

strong covalent bonding, which requires a significant amount of energy for dissociation prior to self-assembly. However, for metal nanostructural growth, the simple reduction of metal salts (usually *via* $NaBH_4$, H_2, or hydrazine as reducing agents) is amenable for mild, solution-phase growth (for example, Eq. 5 for Cu metal). In theory, any metal with a larger standard reduction potential (E^0) than the reducing agent (*e.g.*, -0.481 V for borohydride ion) is a candidate for reduction to its metallic form. This includes most of the first-row transition metal ions, and many others from the main group/transition metal series. However, it should be noted that solution pH

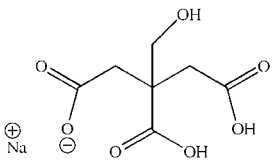
and side-reactions (*e.g.*, metal ions being converted to borides by BH_4^- rather than reduction) often provide a barrier toward successful metal ion reduction.

$$(5)\quad Cu^{2+} + 2Cl^- + 6H_2O + 2NaBH_4 \rightarrow Cu^0 + 2NaCl + 7H_2 + 2B(OH)_3$$

If the above reaction were to be carried out as-is, a metallic film or bulk powder would be formed rather than nanostructures. That is, as the metal ions are reduced, they would instantly agglomerate with one another to form larger particulates. Hence, the most crucial component of nanostructure synthesis is the stabilizing agent that isolates the growing nuclei from one another. We saw an example of this earlier, with polyoxoanions being used to stabilize Ir nanoclusters (Figure 6.14). Some desirable traits of a stabilizing (or entraining) agent are:

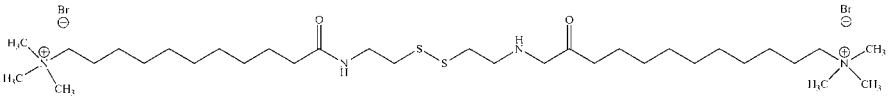
(i) Chemically unreactive toward the growing nanocluster, rendering an unpassivated nanocluster surface
(ii) Structurally well-defined (size/shape), which allows for the controlled growth of the encapsulated nanocluster
(iii) Comprised of light elements (organic-based), so its structure does not interfere with the characterization of the entrained nanocluster. This will also facilitate its sacrificial removal from the nanocluster by pyrolysis at relatively low temperatures, if desired
(iv) Surface-modifiable, to allow for tunable solubility and selective interactions with external stimuli. In addition, to afford controllable self-assembly of entrained nanoclusters on a variety of surfaces through chemisorption, if desired

In aqueous solutions, the most common method used to stabilize nanostructures is the use of organic "capping" ligands. For instance, the *Turkevich process,* which dates back to early colloidal growth of the 1950s, uses sodium citrate (**I**) to entrain the reduced gold nuclei.



I

Particle diameters on the order of 10–20 nm may be synthesized using this method. In this case, since gold is easily reduced ($E^\circ = 1.00\,V$ for $AuCl_4^-$), the citrate reagent acts as both the reducing and stabilizing agent. More recently, a cationic stabilizing agent bis(11-trimethylammoniumdecanoylaminoethyl)disulfide dibromide (TADDD, **II**) has been utilized to grow nanoclusters with diameters $<10\,nm$. If $NaBH_4$ is used as the reducing agent, the sulfide bond is cleaved resulting in a –SH capping group. The thiol is chemisorbed to the surface of the growing nanostructure surface to prevent agglomeration (*esp*. effective for "thiol-philic" noble metals such as Pt, Ag, and Au).[46]



II

Recently, there has been much interest in the use of structurally perfect *dendrimers* such as poly(amidoamine) (PAMAM, Figure 6.30) as stabilizing templates for nanocluster growth. By varying the peripheral groups and number of repeat branching units (known as "Generations"), one is able to fine-tune the size of the

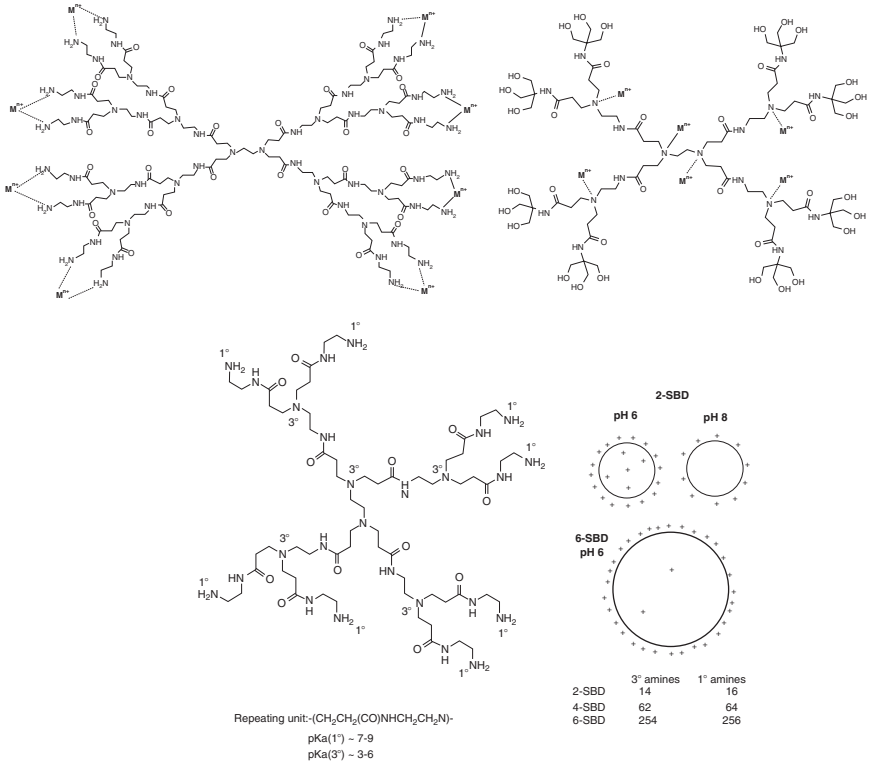


Figure 6.30. Molecular structure of a second Generation (G2) amine-terminated PAMAM dendrimer, illustrating the positions of the metal ions chelated to the primary amine groups (prereduction). In comparison, a G2 hydroxyl-terminated PAMAM dendrimer is shown, with the metal ions now preferring to chelate to the interior tertiary amine groups. Shown on the bottom is the effect of protonation on G2/G6 amine-terminated PAMAM dendrimers. A schematic on the lower right illustrates the positions of the protonated amines at varying pH values. As the generation size increases, the surface density also increases which limits the access of protons (or chelating metal ions) to interact with the interior tertiary amine groups. Reproduced with permission from Kleinman, M. H.; Flory, J. H.; Tomalia, D. A.; Turro, N. J. *J. Phys. Chem. B* **2000**, *104*, 11472. Copyright 2000 American Chemical Society.

entrained nanocluster. Though amine-terminated dendrimers and hyperbranched polymers may be used as a template for M^{n+} chelation and subsequent chemical reduction, the size of the resultant nanoparticle is relatively large, with a greater degree of agglomeration possible. This is especially the case for hyperbranched polymers that exhibit a random structure, which results in a much greater nanoparticle polydispersity. On the other hand, if the primary surface amines ($–NH_2$) are either protonated ($–NH_3^+$), or replaced with hydroxyl groups (–OH), the prereduced metal ions are forced further into the interior of the dendritic structure (Figure 6.30). This results in much smaller diameters and narrow polydispersities for the reduced metal nanoclusters.[47]

In addition to controlling the surface moieties and solution pH, the generation size of the dendrimer is also paramount for successful nanocluster growth. As the degree of branching increases, so does the surface density, which prevents the incoming metal ion (or H^+ during protonation) from entering the interior of the dendritic architecture. Alternatively, for smaller generations, the entrained species becomes easily dislodged from the interior due to its open structure. Hence, the most effective PAMAM size range for nanocluster growth is between the fourth and sixth generations (G4–G6), which exhibit strong container properties (Figure 6.31). As an illustration of the extreme flexibility of the dendritic architecture, the core may also be altered to change its solubility characteristics, or allow the penetration of species through the periphery at high generations (Figure 6.32). Since dendrimers containing an almost unlimited range of cores and peripheral groups have been synthesized, it is now possible to easily control nanocluster properties such as composition, size, morphology, solubility, and encapsulation (*e.g.*, control the release rate of entrained medicinal agents/sensors based on structural or environmental changes, for targeted drug delivery or *in situ* monitoring).

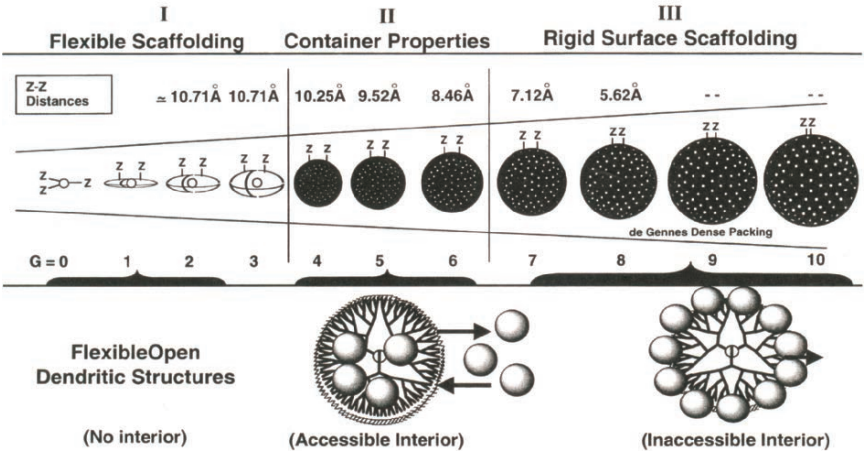


Figure 6.31. Relative sizes and surface densities of PAMAM dendrimers, showing the most suitable range for nanocluster growth as Generation 4 to Generation 6. Reproduced with permission from *Dendrimers and other Dendritic Polymers*, Frechet, J. M. J.; Tomalia, D. A. eds., Wiley: New York, 2001.

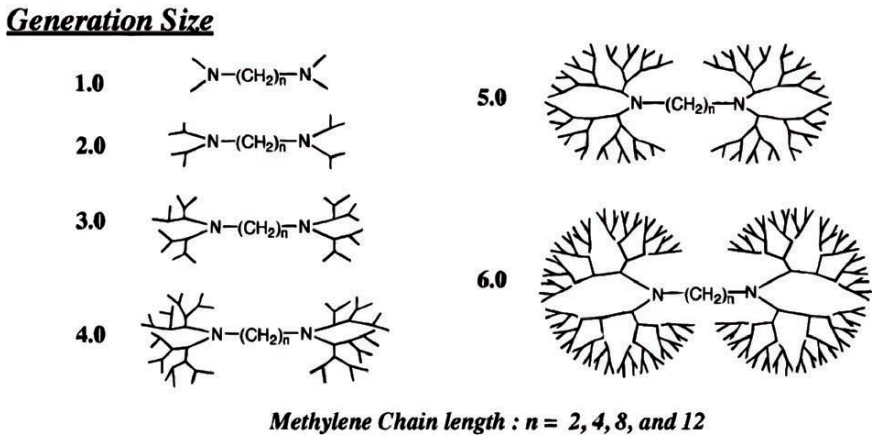


Figure 6.32. Molecular structures of dendrimers modified with long-chain aliphatic cores. Unlike traditional dendritic structures with smaller cores, as the generation size increases, there is an available channel for external species to enter the dendrimer interior. Reproduced with permission from Watkins, D. M.; Sayed-Sweet, Y.; Klimash, J. W.; Turro, N. J.; Tomalia, D. A. *Langmuir* **1997**, *13*, 3136. Copyright 1997 American Chemical Society.

It should be noted that metal nanocluster growth using dendritic templates is strongly governed by the degree of complexation of the precursor metal ions. For instance, silver nanoclusters are not possible using hydroxyl-terminated PAMAM dendrimers since Ag^+ is not strongly chelated to tertiary amine groups. However, if Cu^0 nanoclusters are first generated within the structure, followed by Ag^+, a redox reaction will facilitate the displacement of Cu^0 with Ag^0 within the dendrimer interior (Eq. 7).

$$(7) \qquad Cu + 2Ag^+ \rightarrow Cu^{2+} + 2Ag$$

In addition to simple metallic nanostructures, more complex intermetallic species have also been synthesized through the introduction of more than one metal. For instance, bimetallic nanoclusters may be generated *via* three routes within a dendritic host (Figure 6.33). In addition to already being proven for core–shell nanoclusters, this route should also be amenable for the growth of trimetallic nanostructures for interesting catalytic applications.[48]

Other polymers may also serve as effective stabilizing agents for nanostructural growth. For instance, poly(vinylpyrrolidone) (PVP, **III**), poly(styrenesulfonic acid) sodium salt (PSS, **IV**), and poly(2-ethyl-2-oxazoline) (PEO, **V**) were recently used to generate a number of intermetallic nanoalloys *via* a mild *metallurgy in a beaker* approach developed by Schaak and coworkers.[49] Since individual metal nanoparticles are in intimate contact with high surface reactivity and low melting points, the use of high-temperature annealing is not required, unlike bulk-scale alloy synthesis. The PVP architecture has also been shown to facilitate the growth of Au@Ag core–shell nanostructures, as well as Ag nanowires and nanocubes.

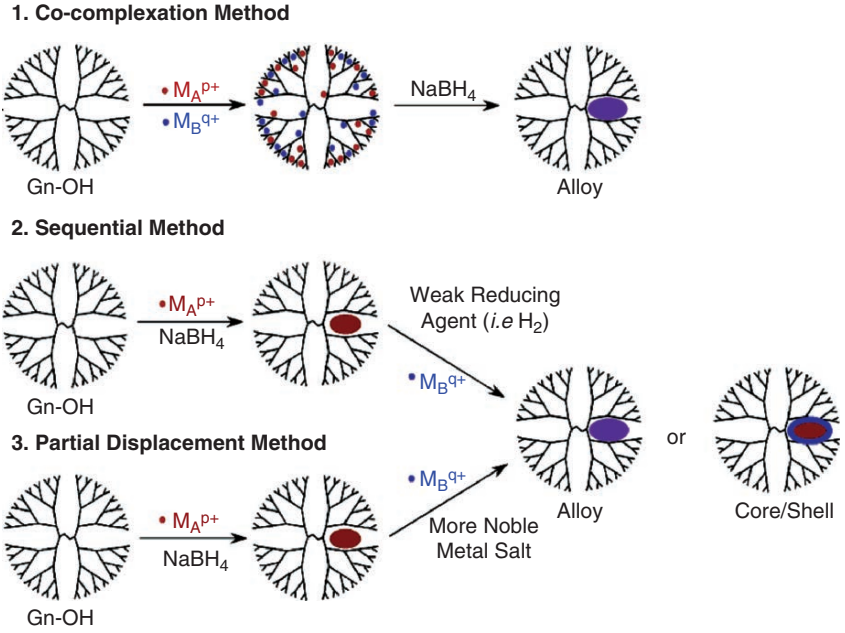
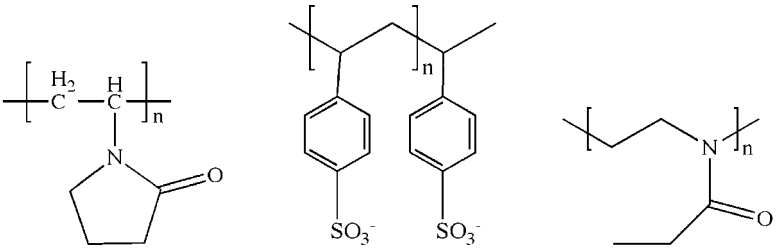
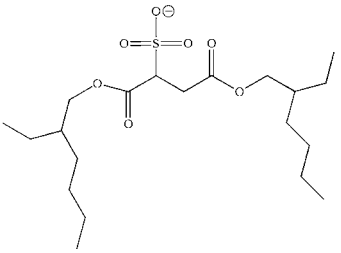


Figure 6.33. Schematic of the three methods used to generate bimetallic nanoclusters within a dendritic host. Reproduced with permission from Scott, R. W. J.; Wilson, O. M.; Crooks, R. M. *J. Phys. Chem. B* **2004**, *109*, 692. Copyright 2004 American Chemical Society.

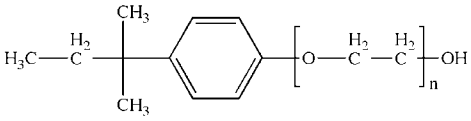


For the synthesis of nanostructures within nonpolar solvents, one uses stabilizing agents that contain alkyl chains rather than –OH endgroups. One of the first capping agents to be used for noble metal colloidal growth was alkylthiols (CH_3–$[CH_2]_X$–SH). In this system, the –SH end is bound to the surface of the nanostructure, and the long organic tail is responsible for dispersion within the organic solvent. Though these capping stabilizers worked well for colloidal growth to prevent agglomeration, even allowing solvent removal/redispersion into organic solvents, it was relatively difficult to control the size dispersity of the nanoparticles.

Accordingly, systems that contain a nanoreactor template have been used most recently for controlled nanostructural growth. In fact, everyone who has washed dishes or laundry already has some experience with these types of stabilizing agents, known as *micelles*. These compounds contain both polar (–OH, cationic/anionic) and nonpolar (aliphatic) ends. Soaps and surfactants work by surrounding the dirt particle with the nonpolar ends, leaving the hydrophilic polar groups exposed to the surrounding water molecules. This results in pulling the dirt particle from the surface of the clothing fiber, forming an aqueous suspension. In a similar fashion, an oil-in-water microemulsion may be set up using common surfactants such as sodium bis(2-ethylhexyl)sulfosuccinate (also referred to as Aerosol OT or AOT, **VI**), or the nonionic surfactant Triton-X (**VII**) in the presence of the biphasic oil/water mixture.



VI



VII

Since most precursors for solution-phase nanostructural growth are ionic metal salts, a typical micelle would not be effective since the precursor would not be confined to the interior of the microemulsion. Hence, *reverse micelles* (or inverse micelles, Figure 6.34) are used to confine the precursor ions to the aqueous interior, which effectively serves as a nanoreactor for subsequent reduction, oxidation, *etc. en route* to the final nanostructure. Not surprisingly, either PAMAMOS dendrimers (Chapter 5) or dodecyl-terminated (hydrophobic) PAMAM dendrimers (Figure 6.35) have been recently employed for this application.

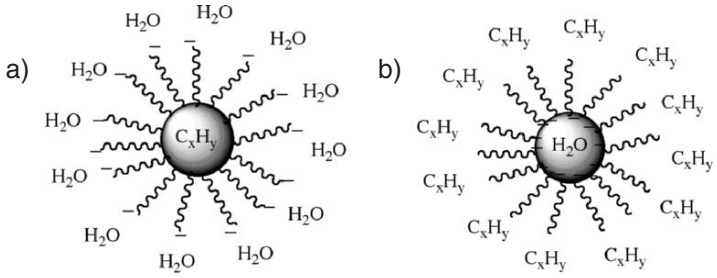


Figure 6.34. Comparison of a traditional micelle used to entrain organic oils/dirt using an anionic surfactant, (a), and a reverse micelle used to stabilize aqueous nanoreactors within a nonpolar solvent, (b).

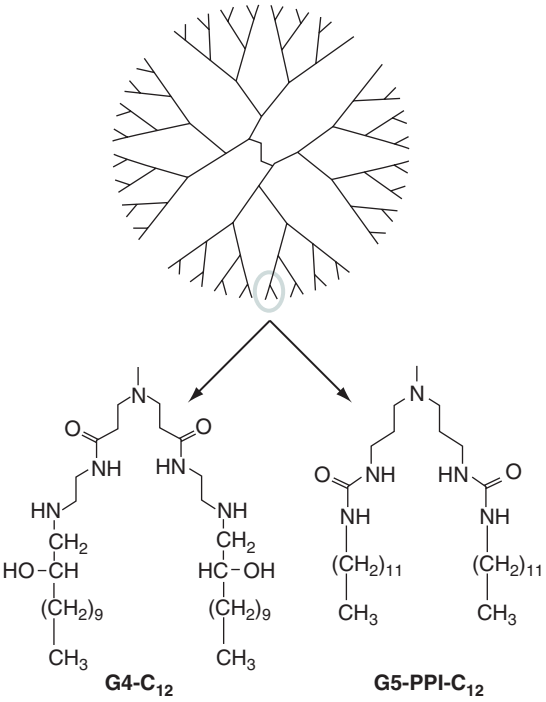


Figure 6.35. Hydrophobic-functionalized PAMAM (G4-C_{12}) and poly(propyleneimine) (PPI) dendrimers, which serve as templates for Au nanocluster growth. Reproduced with permission from Knecht, M. R.; Garcia-Martinez, J. C.; Crooks, R. M. *Langmuir* **2005**, *21*, 11981. Copyright 2005 American Chemical Society.

It should be noted that dendrimer-entrained nanoclusters synthesized within aqueous solutions may also be phase-transferred into an organic solvent by mixing with alkylthiols dissolved in a nonpolar solvent.[50] This also results in monodisperse nanoclusters, with much less polydispersity than early colloidal syntheses that employed thiol-based entraining agents. That is, the nanocluster size has already been controlled *via* intradendrimer stabilization. In contrast, the use of alkylthiols from the initial stages of growth is not as effective toward preventing agglomeration during the nucleation step.

It is not always necessary for the metal ions to be encapsulated within a stabilizing polymer during chemical reduction. For instance, the reduced metal may be entrained by polymerization precursors, such as postreduction living radical polymerization that takes place on the surface of gold nanoparticles (Figure 6.36). This results in a dense "polymer brush" that encapsulates the metallic nanoparticle, effectively stabilizing the structure against agglomeration. Subsequent alignment and surface reactivity of the resultant nanostructures may be fine-tuned by varying the nature of the polymer coating.

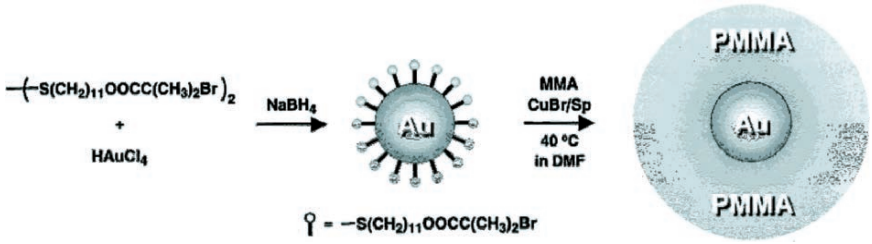


Figure 6.36. Schematic of the formation of gold nanoparticles coated with free-radical polymerization initiators that subsequently yield Au@polymer nanostructures through a surface-controlled living polymerization process. Reproduced with permission from Ohno, K.; Koh, K.-M.; Tsujii, Y.; Fukuda, T. *Macromolecules* **2002**, *35*, 8989. Copyright 2002 American Chemical Society.

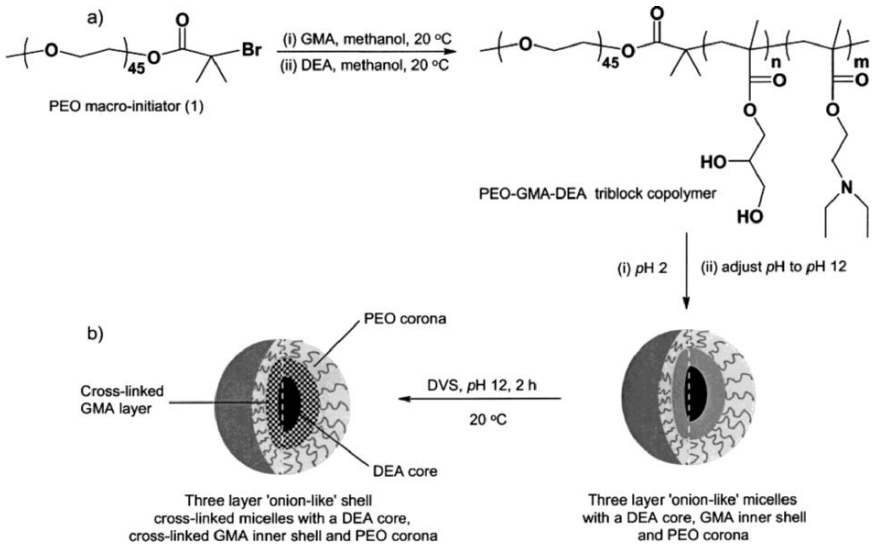


Figure 6.37. Synthetic scheme for three-layer cross-linked micelles. Shown is the micellation of poly[(ethylene oxide)-block-glycerol monomethacrylate-block-2-(diethylamino)ethyl methacrylate] (PEO-GMA-DEA) triblock copolymers to the final "onion-like" layered nanostructure. Reproduced with permission from Liu, S.; Weaver, J. V. M.; Save, M.; Armes, S. P. *Langmuir* **2002**, *18*, 8350. Copyright 2002 American Chemical Society.

Increasingly complex stabilizing agents such as *cross-linked micelles* have been developed in recent years. These nanostructures consist of a polymeric core that is cross-linked with surrounding polymer layer(s) (Figure 6.37). In contrast to solid nanoparticles, a hollow *nanoshell* may be synthesized by sacrificially removing the core material by chemical or thermal decomposition. Such intriguing nanobuilding blocks will likely be of extreme importance for next-generation medical treatment/sensing, hydrogen storage, ion-exchange, and microelectronics applications.

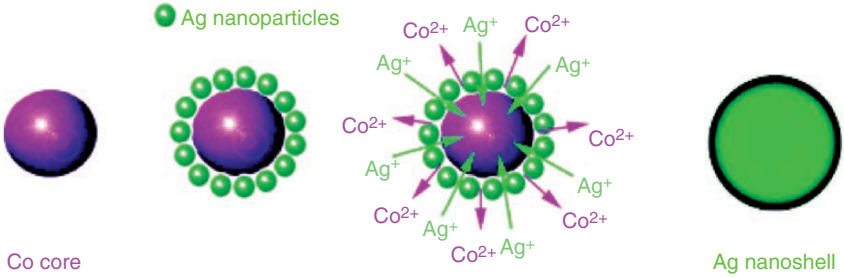


Figure 6.38. Schematic of Ag nanoshell formation from a nanostructural Co core. Due to favorable redox couples between Ag and Co, a nanoshell of metallic silver forms at the expense of the inner Co core. Reproduced with permission from Chen, M.; Gao, L. *Inorg. Chem.* **2006**, *45*, 5145. Copyright 2006 American Chemical Society.

Another strategy for nanoshell growth consists of applying a thin metallic coating onto silica or polystyrene templating spheres, with subsequent sacrificial removal of the template by hydrofluoric acid (HF) or toluene, respectively. The reverse case of a polymer/ceramic coating onto removable metallic nanoparticles has also been exploited to yield nonmetallic nanoshells.[51] However, the removal of a relatively large core (*i.e.*, typically >200 nm) from a nanoscale coating generally results in significant deformation of the resultant nanoshell.

The structural robustness of nanoshells has been recently improved through the use of sacrificial cores of smaller diameters. An example of this strategy is the growth of silver nanoshells from nanostructural Co templates (Figure 6.38). It should be noted that $NaBH_4$ reduction of ligand-capped Co^{2+} ions preferentially yields Co_2B rather than metallic cobalt. However, a postexposure of oxygen converts the boride into metallic cobalt nanostructures (Eq. 8).[52] Once the Co core was formed, silver ions were introduced into the system through addition of $AgNO_3$. Since $Ag^+ (E_{red}^\circ = 0.799\ V)$ is preferentially reduced relative to $Co^{2+} (E_{red} = -0.377\ V)$, the exchange of Ag for Co occurs spontaneously as Ag^+/Co^0 boundaries are formed. This process continues as Ag^+ ions diffuse through the growing Ag^0 layer, until all of the metallic Co is consumed from the core.

$$(8) \qquad 4Co_2B + 3O_2 \rightarrow 8Co + 2B_2O_3$$

Rather than redox-governed nanoshell growth, diffusion-governed routes are also possible. For example, the exposure of cobalt nanoparticles to oxygen, sulfur, or selenium does not simply form a Co-chalcogenide coating, but rather hollow nanoshells of the cobalt chalcogenide (Figure 6.39).[53] This interesting result is a modern extension of an effect that has been studied since the late 1940s – the *Kirkendall Effect*. This phenomenon describes the differential diffusion rates of two species in direct contact with one another at elevated temperatures (*e.g.*, Cu and Zn in brass). During the growth of oxide or sulfide films, pores are formed in the solid due to a vacancy-exchange mechanism. That is, the outward movement of metal ions through the oxide layer is balanced with an inward migration of lattice vacancies

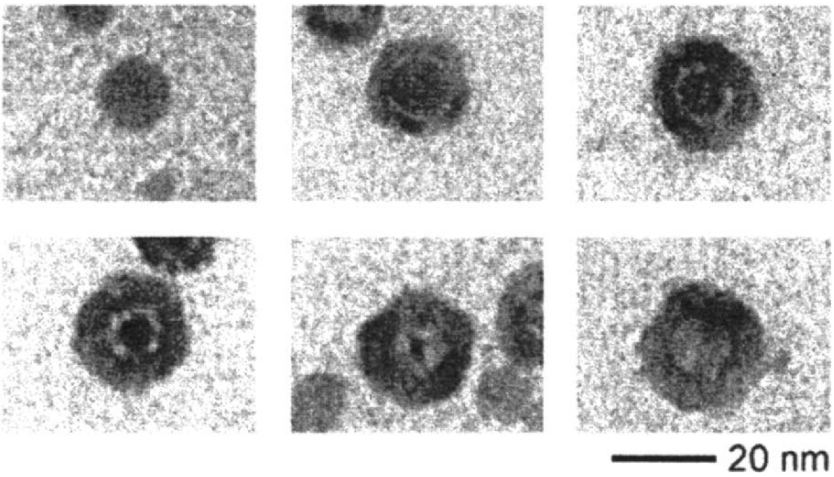


Figure 6.39. Hollow nanoshell formation *via* the Kirkendall Effect. Shown is the evolution of the hollow nanoshell after reaction times of (left–right): 0 s, 10 s, 20 s, 1 min, 2 min, 30 min. An interesting feature is the formation of "bridges" that connect the core to the sulfide shell, facilitating the fast outward migration of Co. Reproduced with permission from Yin, Y.; Rioux, R. M.; Erdonmez, C. K.; Hughes, S.; Somorjai, G. A.; Alivisatos, A. P. *Science* **2004**, *304*, 711. Copyright 2004 AAAS.

that settle near the metal/oxide boundary. Due to the large number of defects and volume of bulk solids, the pores at the metal–oxide interface do not coalesce into an ordered array. However, for a nanostructural system that is relatively free of defects and exhibits a large surface/volume ratio, the pores readily aggregate into a single hollow core.

Not surprisingly, 0D nanostructural growth need not be limited to metallic structures, but may also include other compounds such as metal oxides, sulfides, *etc*. There are a copious number of applications for nanostructural oxides such as high-density magnetic storage, heterogeneous catalysis, gas sensors, electrolytes for lithium batteries, and fuel cells. Sun and coworkers recently described a novel method for the synthesis of hexane-suspended Fe_3O_4 (magnetite) nanoclusters through the reaction between $Fe(acac)_3$, oleylamine, oleic acid, and 1,2-hexadecanediol at *ca*. 200°C within phenyl ether.[54] This one-pot synthesis results in the thermal replacement of the acetylacetonate ligands (recall Chapter 4 – CVD precursors) from the iron center, followed by oxide formation from reaction with the alcohol. The oleylamine and oleic acid condense *in situ* to form a long-chain inverse micelle that is capable of suspending the nanoclusters in nonpolar media. For purification, one simply adds an excess of polar solvent (methanol or ethanol) to precipitate the nanostructures. Following centrifugation, the nanoclusters are resuspended in hexane, with this precipitation/resuspension procedure repeated 3–4 times to afford monodisperse, chemically pure Fe_3O_4 nanoclusters.

Another more recent strategy that we developed is the synthesis of metal oxide nanoparticles from the simple reaction of interdendritic stabilized basic metalate salts

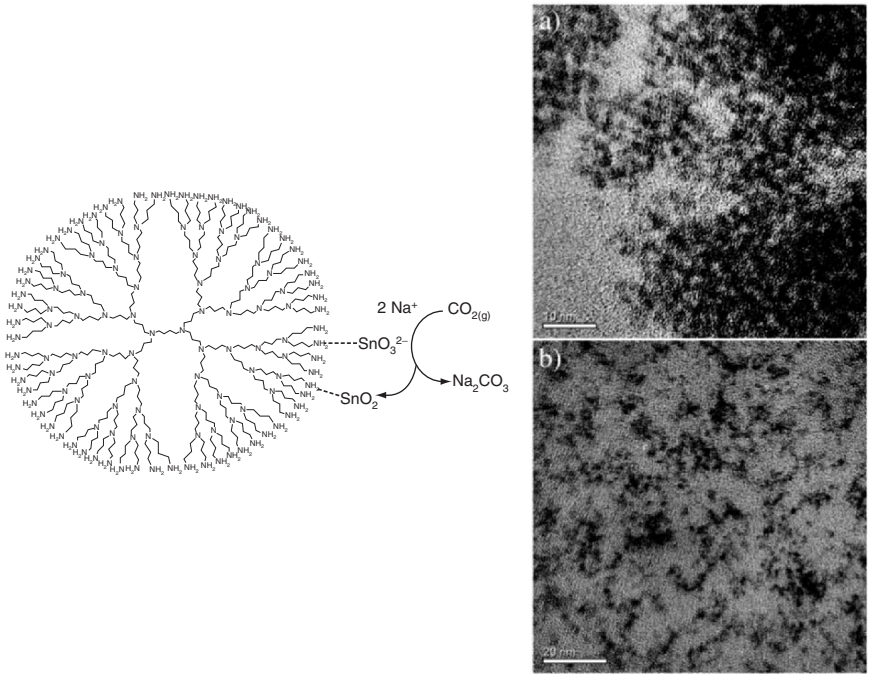


Figure 6.40. Schematic of the growth of tin oxide nanoclusters at room temperature. The TEM images on the right illustrate interdendritic stabilized nanoclusters using (a) PAMAM and (b) amine-terminated poly(ethyleneimine) hyperbranched polymer hosts. Reproduced with permission from Juttukonda, V.; Paddock, R. L.; Raymond, J. E.; Denomme, D.; Richardson, A. E.; Slusher, L. E.; Fahlman, B. D. *J. Am. Chem. Soc.* **2006**, *128*, 420. Copyright 2006 American Chemical Society.

with CO_2 (Figure 6.40).[55] This strategy should work for any metalate salt that is typically formed from the reaction of the metal hydroxide with a strong base (Eq. 9, for sodium hafnate).

(9) $2NaOH + Hf(OH)_4 \rightarrow Na_2HfO_3 + 3H_2O$

In addition to oxides, a number of other compositions may be synthesized. In general, any stoichiometry is "fair game" through the same reactions that one does in bulk scale. For example, sulfides through reaction of precursors with H_2S, nitrides through NH_3 exposure, *etc*.[56] As long as a suitable entraining agent is used, one is able to control the resultant size/morphology of the nanostructures. With current advances in compound and metallic nanoclusters, with the properties of each being fine-tuned *via* quantum confinement effects, one can begin to imagine intriguing designs for future microelectronic devices.

Self-assembly of 0D nanostructures into arrays

Now that you understand how 0D nanostructures are synthesized and stabilized, it is worthwhile noting how these structures are aligned into more complex arrays.

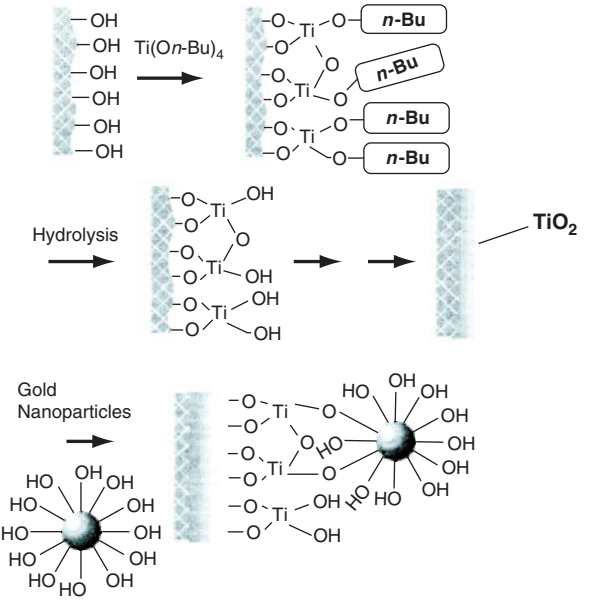


Figure 6.41. Illustration of the chemisorption of a nanostructure onto a hydroxylated TiO_2 surface. The hydroxyl groups on the periphery of the nanostructure stabilizing agent provide the handle for surface adsorption. Reproduced with permission from Liu, S.; Zhu, T.; Hu, R.; Liu, Z. *Phys. Chem. Chem. Phys.* **2002**, 4, 6059. Copyright 2002 PCCP Owners Societies.

As you recall, the "bottom-up" approach to materials design features the purposeful placement of individual nanoarchitectures, in order to build specific functional devices one unit at a time. Though one would have the greatest control over the device properties through the growth/arrangement of individual atoms, current synthetic methodology most easily yields nanoclusters/nanoparticles that consist of small groups of atoms ($N_{\text{atoms}} \sim 50+$).

As we have seen, a stabilizing agent is used to prevent agglomeration of growing 0D nanostructures. In fact, this component also provides an effective "handle" to bind the nanostructure to a particular surface. Once appropriate reactive groups (*e.g.*, –OH, NH_2) are placed on a surface through monolayer formation, the stabilizing group surrounding the nanostructure spontaneously becomes chemisorbed (Figure 6.41). If the nanostructures are encapsulated with stabilizing agents of well-defined sizes, then the spacing between adjacent nanostructures will also be highly ordered and predictable (Figure 6.42). Though this most often results in a 2D matrix of nanostructures, it is also possible to create alignment into 1D chains. For instance, Au nanostructures that are stabilized by a long-chain thiolated poly(ethylene oxide) polymer form a linear array *via* interactions with sulfate/acid groups of a polysaccharide compound (chondroitin sulfate c, **VIII**) (Figure 6.43).

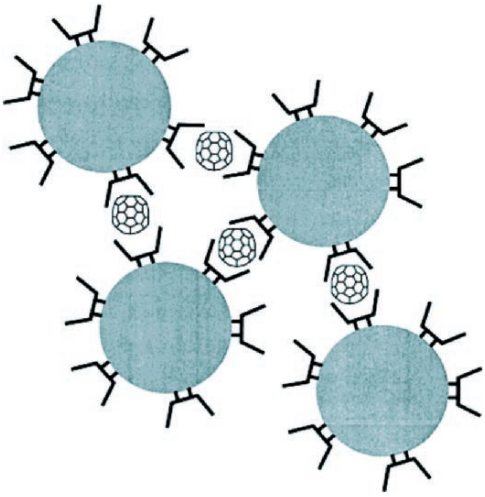


Figure 6.42. Schematic of the controlled spacing between individually stabilized gold nanoparticles by the formation of fullerene inclusion complexes. Reproduced with permission from Yonezawa, T.; Matsune, H.; Kunitake, T. *Chem. Mater.* **1999**, *11*, 33. Copyright 1999 American Chemical Society.

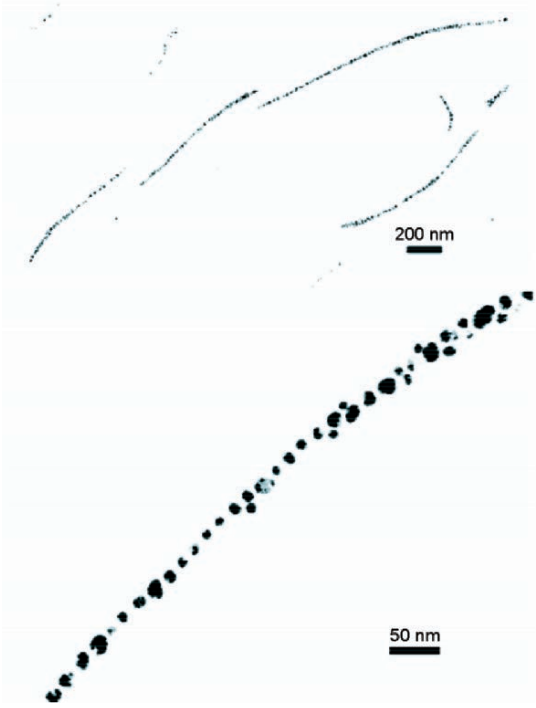
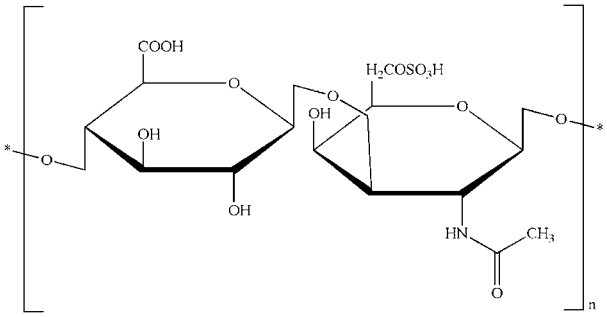


Figure 6.43. TEM images of a 1D array of Au nanoparticles formed through the interaction of PEO and polysaccharide chains. Reproduced with permission from Takagi, K.; Ishiwatari, T. *Chem. Lett.* **2002**, 990. Copyright 2002 Chemical Society of Japan.



VIII

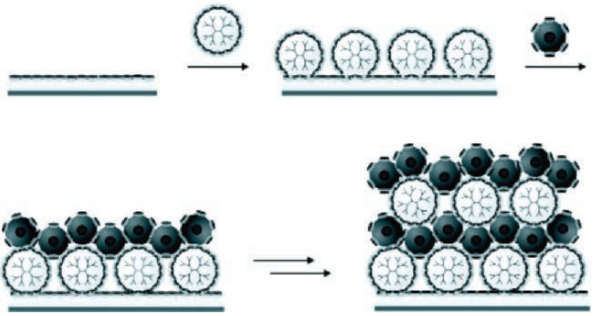


Figure 6.44. Illustration of LbL growth of cyclodextrin-stabilized Au nanoparticles (dark spheres) and adamantyl-terminated PPI dendrimers. Reproduced with permission from Crespo-Biel, O.; Dordi, B.; Reinhoudt, D. N.; Huskens, J. *J. Am. Chem. Soc.* **2005**, *127*, 7594. Copyright 2005 American Chemical Society.

The most popular method used to achieve 3D arrays such as layered nanostructural thin films, is the *layer-by-layer* (LbL) self-assembly pioneered by Decher in the early 1990s.[57] This technique is based on the sequential adsorption of species with complimentary functional groups (*e.g.*, ionic charges) on a variety of substrates. There are many device applications for LbL thin films, such as self-cleaning surfaces, surface deactivation of warfare agents, solar energy, drug delivery, and optoelectronics. An interesting recent precedent for LbL nanostructural film growth uses sequential layering of cyclodextrin-stabilized Au nanoparticles and adamantyl-terminated dendrimers (Figure 6.44).[58] The film thickness was reported as 2 nm per bilayer, allowing for strict control over the resultant film thickness.

On the topic of self-assembly, we would be remiss if we did not mention a recent precedent that scores high in the "cool" category. We are referring to the development of surface-rolling molecules that are aptly termed *nanocars*/*nanotrucks*.[59] Rather than stabilized nanoparticles, these nanovehicles are organic molecules that contain fullerenes as wheels (Figure 6.45). The placement of the nanocars onto a gold surface

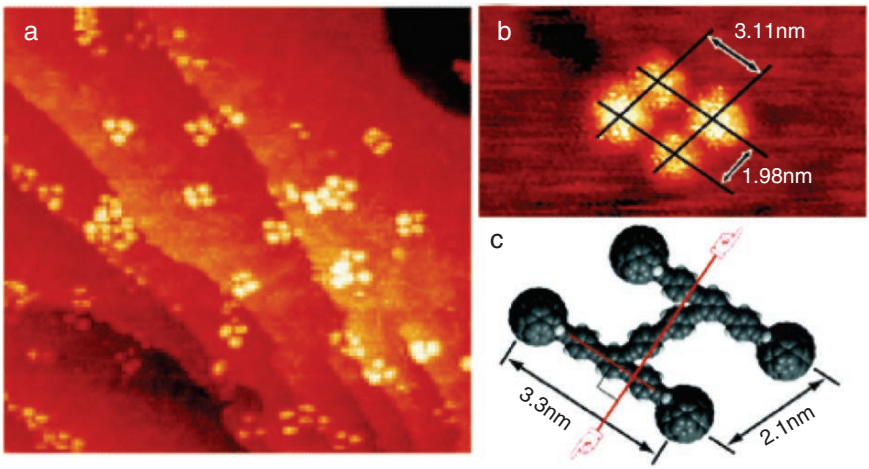


Figure 6.45. Images and dimensions of a nanocar on a Au(111) surface. The bright features are the fullerene wheels. Reproduced with permission from *Nano Lett.* **2005**, 5, 2330. Copyright 2005 American Chemical Society.

is as easy as spin-casting from a toluene suspension. Due to strong adhesion forces between fullerenes and metal surfaces, the nanocars spontaneously deposit with all four "nanowheels" on the substrate. The axles consist of triple-bonded alkyne groups that allow rotation of the fullerene wheels, transporting the nanocar across a gold surface in response to changes in temperature.[60] Though only pivoting and translational motion are currently possible, this work sets an important step toward the realization of nanomachines for an endless number of possible applications.

6.2.2. One-Dimensional Nanostructures

The second class of nanoscale building blocks, referred to as 1D nanostructures, is reserved for those materials that have nanoscale dimensions that are equivalent in all but one direction. For instance, let us consider the 2D architecture of this page. Recall that a 0D nanostructure is analogous to the period following this sentence (length = width); a 1D nanostructure is analogous to the number "1" (length > width).

Since we began the discussion of 0D nanostructures with nomenclature, we will follow suit in this section. Once again, it is easy to be confused by the common synonymous use of the terms nanotube, nanofiber, nanowire, and nanorod. However, if you think of the analogous bulk materials without the prefix "nano," there should be no ambiguity regarding the proper use of these descriptors (Figure 6.46). The common thread among all of these structures is that their diameters must be within the 1–100 nm range; typically, their lengths are within the micron (or larger) regime. A *nanotube* is a 1D structure that contains a hollow core, whereas the other three nanoarchitectures are solid throughout. The term *nanofiber* should be

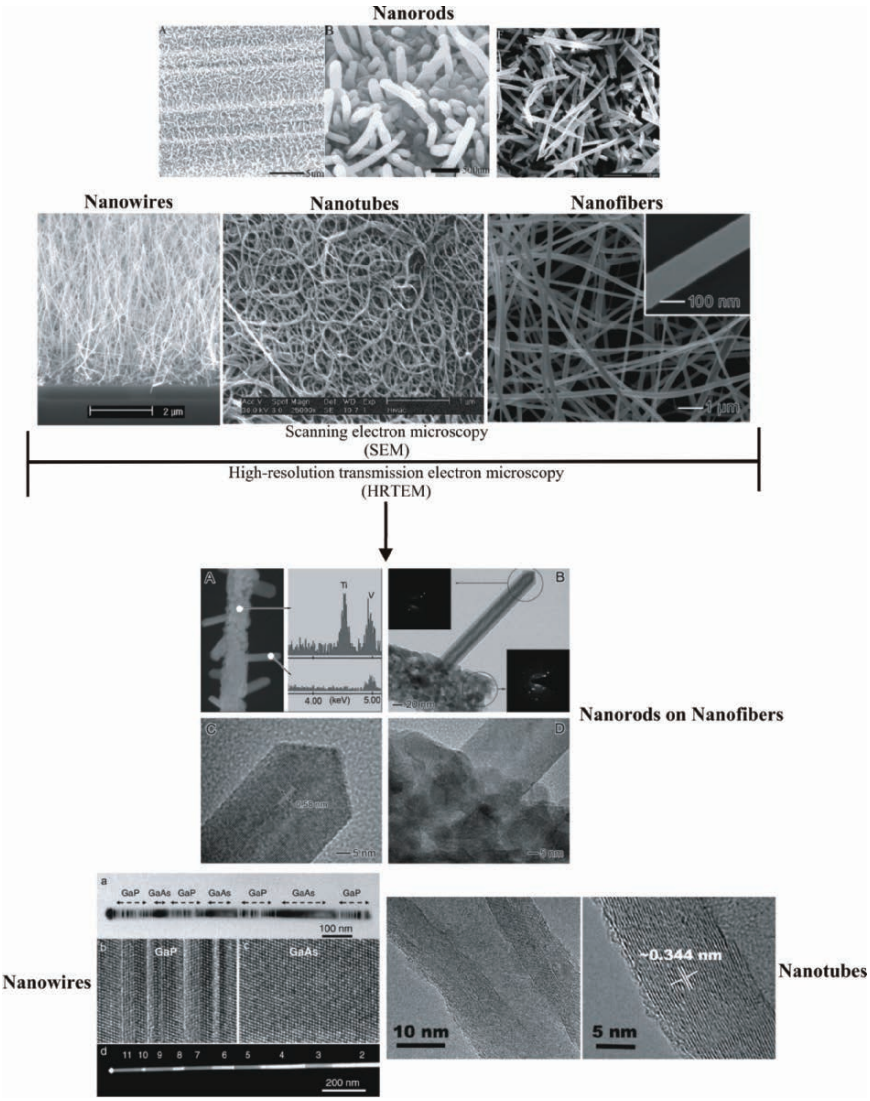


Figure 6.46. Classifications of 1D nanostructures. The top portion shows SEM images of various nanostructures.[63] Whereas the morphology of nanowires, nanotubes, and nanofibers look identical by SEM, nanorods are notably different, with much shorter lengths and straight sidewalls. The bottom portion shows high-resolution TEM images, which provide morphological details of the nanostructures.[64] The top TEM image shows crystalline nanorods grown on the surface of amorphous nanofibers. The bottom two TEM images illustrate the difference between nanowires/nanotubes – the latter contains a hollow core. It should be noted that crystalline nanorods may also be termed "nanocrystals," as their morphology resembles that of needle-like bulk crystallites.

reserved for 1D nanostructures that are amorphous (and usually nonconductive) such as polymers and other nongraphitized carbonaceous structures. By contrast, a *nanowire* designates a structure that is crystalline, with either metallic or semiconductive electrical properties.

A *nanorod* is typically a crystalline 1D nanostructure, with an overall length comparable to its width (*i.e.*, both dimensions are <100 nm). As their name implies, another feature of nanorods is their rigid sidewall structures. However, since crystalline nanorods exhibit the same overall shape as needle-like bulk crystals, the term "nanocrystal" is probably more appropriate for these structures (or, more explicitly: "rod-like nanocrystals"). Whereas nanowires, nanofibers, and nanotubes exhibit an interwoven array, nanorods are completely linear in morphology. As such, nanorods are capable of stacking onto each other to yield interesting 2D and 3D arrays – not usually as easy to perform with the "spaghetti-like" morphology of the other 1D nanostructures.

Carbon nanotubes

Without question, the most widely studied 1D nanomaterial is the *carbon nanotube* (CNT). These structures were first discovered by Iijima in 1991,[61] and consist of a graphitic sheet(s) of sp^2 hybridized carbon atoms (*i.e.*, graphene[62]) rolled into a tubular array. Based on the layers of graphene sheets that comprise the CNT, the structures are designated as single-walled, double-walled, or multiwalled nanotubes (SWNTs, DWNTs, or MWNTs, respectively – Figure 6.47). The diameters of CNTs range from 1 nm (SWNTs) to >30 nm (MWNTs), with aspect ratios (length:width) ranging from 100 to greater than 1×10^6. Even though the diameters of CNTs are

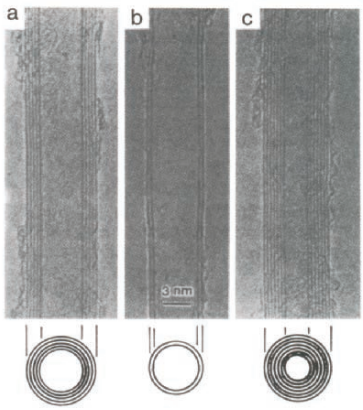


Figure 6.47. TEM images of MWNTs, formed from the folding (a) 5-, (b) 2- (*i.e.*, DWNT), and (c) 7-stacked graphene sheets. The spacing between individual concentric cylinders is 0.34 nm – the distance between adjacent planes in graphite. Reprinted from Dresselhaus, M. S.; Dresselhaus, G.; Eklund, P. C. *Science of Fullerenes and Carbon Nanotubes*. Copyright 1996, with permission from Elsevier.

orders of magnitude smaller than a human hair, their tensile strength is *ca.* 20 times greater than steel – a property attributed to extremely strong sp^2 bonding between neighboring hexagonal units.

The electrical conductivity of SWNTs may vary from metallic to semiconducting, depending on the way a graphene sheet is folded (Figure 6.48). In particular, the diameter and helicity of a SWNT are uniquely characterized by the *chirality vector* (or *Hamada vector*), $\vec{C}$, which connects crystallographically equivalent graphene lattice sites. Vector indices designated by (n, m) are used to indicate the direction and length of the chirality vector (Eq. 10). When $n = 0$, the SWNT is denoted as the *zigzag* conformation; when $n = m$, the SWNT is in its *armchair* form. For all values in between these extremes, the nanotubes are designated simply as *chiral*.

$$\vec{C} = n\vec{a}_1 + m\vec{a}_2, \quad (10)$$

where n, m are the integers denoting the number of unit cell vectors along two directions in the crystal structure of graphene and $\vec{a}_1, \vec{a}_2$ are the graphene unit cell vectors.

For metallic SWNTs, the electrical conductance may exceed silver or copper by three orders of magnitude. Calculations have shown that $(n, 0)$ or zigzag SWNTs exhibit metallic conductivity when $n/3$ is an integer, and semiconducting properties for all other values of n. Similarly for chiral SWNTs, when $(2n + m)/3$ is an integer, the tubes are metallic (otherwise are semiconducting). Finally, armchair SWNTs, with $n = m$, exhibit metallic conductivity (Figure 6.49). Most importantly, electronic band structure calculations show that metallic and semiconducting CNTs are dependent only on (n, m) – that is, slight structural variations result in dramatic changes in their electronic properties. For example, the bandgap of semiconducting SWNTs may be fine-tuned from *ca.* 10 meV to 1 eV – with no required addition of dopants, unlike bulk Si counterparts. If semiconducting nanotubes exhibit the same chirality, the bandgap is inversely proportional to the diameter (*e.g.*, $E_g(7, 0) > E_g$ (10,0)).

The tunable electronic properties of CNTs are being explored for next-generation IC architectures. As you may recall from Chapter 4, traditional Si-based microelectronic devices will likely reach a fundamental limit within the next decade or so, necessitating the active search for replacement materials. Accordingly, an area of intense investigation is molecular electronics – in which the electronic device is built from the placement of individual molecules.[65] Not surprisingly, the interconnects of these devices will likely be comprised of CNTs and other (semi)conductive 1D nanostructures such as nanowires.

Since CNTs have a high electrical conductivity and contain sharp tips, these nanomaterials are the best-known field emitter[66] of any material to date. In general, the smaller the radius of curvature of the tip, the more concentrated the electric field will be, which corresponds to increased field emission at low required voltages. This property is currently being exploited for the design of flat-panel field emission displays. Though plasma and LCD displays are hot ticket items at electronics stores, they both possess inherent disadvantages. Plasma displays are extremely heavy, consume a significant amount of energy, and are prone to "burn-in," which permanently degrades the screen. On the other hand, LCD screens are expensive to produce and often lack the response time required to view fast-paced sporting events/movies

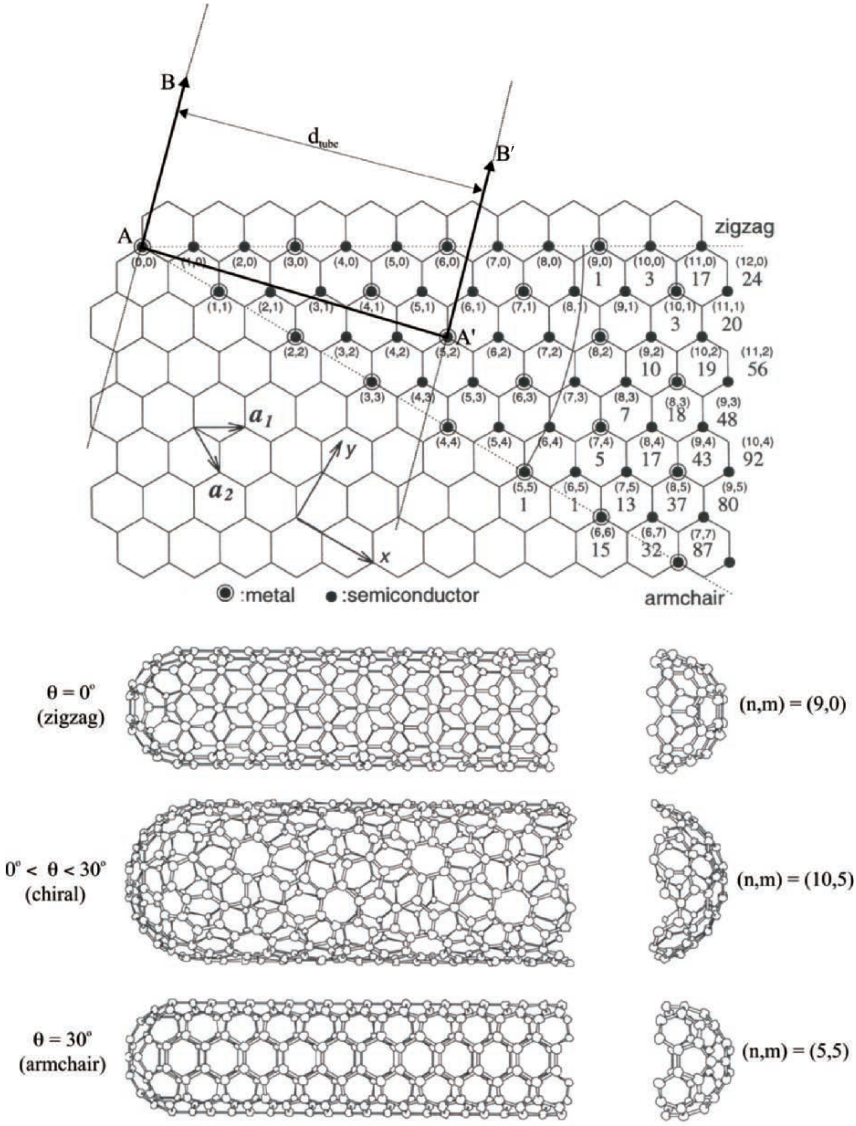


Figure 6.48. Illustration of the honeycomb 2D graphene network, with possible unit cell vector indices (n,m). The dotted lines indicate the chirality range of tubules, from $\theta = 0°$ (zigzag) to $\theta = 30°$ (armchair). For θ values between 0 and 30°, the formed tubules are designated as chiral SWNTs. The electrical conductivities (metallic or semiconducting) are also indicated for each chiral vector. The number appearing below some of the vector indices are the number of distinct caps that may be joined to the (n,m) SWNT. Also shown is an example of how a (5,2) SWNT is formed. The vectors AB and A′B′, which are perpendicular to the chiral vector (AA′) are superimposed by folding the graphene sheet. Hence, the diameter of the SWNT becomes the distance between AB and A′B′ axes. Reprinted from Dresselhaus, M. S.; Dresselhaus, G.; Eklund, P. C. *Science of Fullerenes and Carbon Nanotubes*. Copyright 1996, with permission from Elsevier.

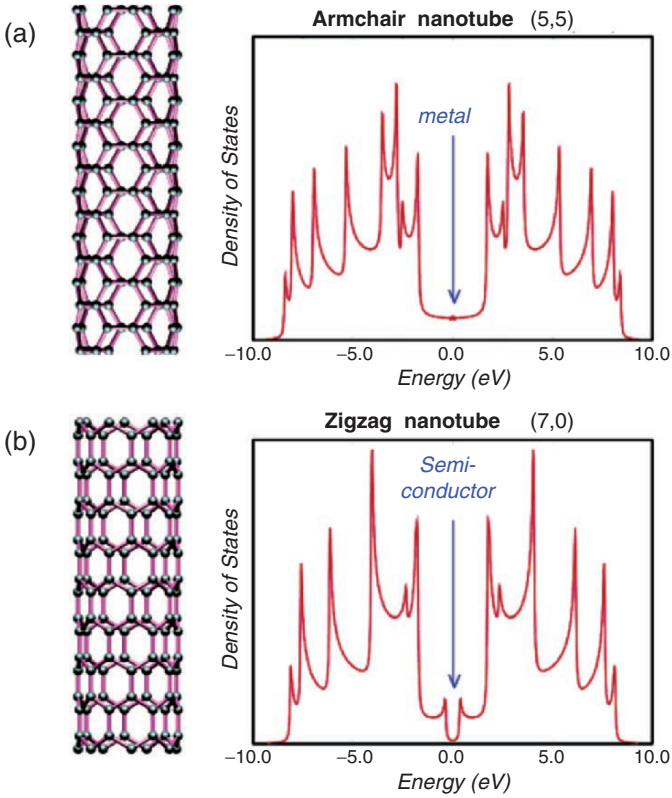


Figure 6.49. Comparison of the density of states, and resultant electronic properties of armchair and zigzag SWNTs. Reproduced with permission from Charlier, J.-C. *Acc. Chem. Res.* **2002**, *35*, 1063. Copyright 2002 American Chemical Society.

without blur. Hence, as far as picture quality is concerned, traditional CRT displays are still among the finest quality. The replacement of this technology with CNTs is a logical step in the evolution of display panels. Rather than a single electron gun, CNT-based screens will contain a separate nanotube electron gun for each individual pixel in the display – dramatically enhancing the resolution and clarity of the picture. Further, in contrast to current large flat-panel televisions, the overall weight of CNT-based analogues will be significantly lower, and they will consume far less power. This concept has already been proven in prototypes and is scheduled to reach the commercial market within 1–2 years.

Another electronic application for CNTs is for next-generation field-effect transistor (FET) design. The "proof-of-concept" for CNTFETs was demonstrated in the late 1990s, with a simple bridging of two noble metal electrodes with a SWNT (Figure 6.50 – top). However, the electrical characteristics of this new FET design were less than desirable, with high contact resistance ($>1\,\mathrm{M\Omega}$) and low drive currents. Researchers at IBM have since modified the original design wherein the

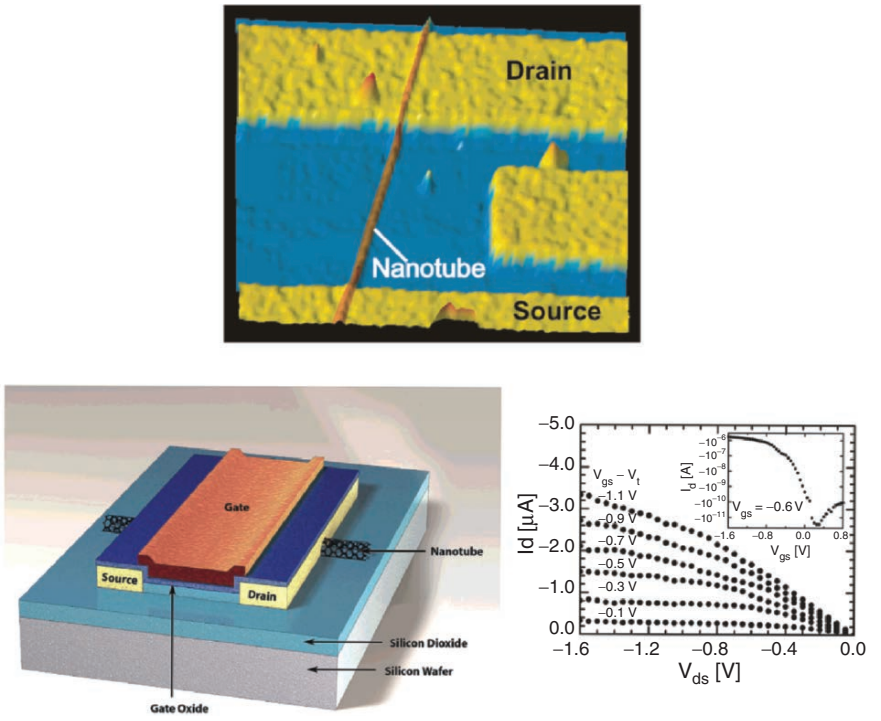


Figure 6.50. Atomic force microscope (AFM) topographical image of an original CNTFET (top). The source/drain electrodes were fabricated on a SiO_2 thin film grown on a Si wafer. The wafer itself served as the gate electrode. Also shown (bottom) is an illustration of a top-gated CNTFET design, with an oxide thickness of 15 nm and CNT length of 300 nm. The device electrical characteristics at room temperature is also provided. The inset shows the transfer characteristic of the FET (output voltage of the device as a function of the input voltage). Reproduced with permission from Avouris, P. *Acc. Chem. Res.* **2002**, *35*, 1026. Copyright 2002 American Chemical Society.

semiconducting SWNTs are placed directly onto an oxidized Si wafer, followed by the deposition of the source and drain (Co or Ti) electrodes (Figure 6.50 – bottom). Through subsequent annealing, a stronger interaction is afforded between the electrodes and CNT channel, which reduces the contact resistance.

Though CNTFETs are in a relatively early stage of development, Avouris at IBM provided a recent comparison of the output from a top-gate CNTFET (Figure 6.50 – bottom) and existing Si-based FETs.[67] For CNTs with an average diameter of 1.4 nm, the ON current for a top-gated CNTFET is on the order of $2{,}100\,\mu A\mu m^{-1}$ at V_{DS} (drain voltage) = V_{GS} (gate voltage)–V_T (threshold voltage) = 1.3 V. In comparison, the highest drive current in a p-CMOS under the same conditions is $650\,\mu A\mu m^{-1}$ for a gate length of 50 nm. The *transconductance* (ratio of the output current variation to the input voltage variation) of the CNTFET is $2{,}300\,\mu S\mu m^{-1}$; the value for an analogous Si p-CMOS is $650\,\mu S\mu m^{-1}$. Hence, the values obtained thus far show that CNTFETs outperform Si-based FETs. As improvements continue

to be made to the design of nanotube-based transistors, this technology should be ready for the market just as the "glass ceiling" of Si CMOS devices is reached. It should be noted that the extremely efficient thermal conductivity of CNTs will also be exploited for cooling applications for future computers – of increasing concern as the chip density continues to soar.

In addition to the above tunable conductive properties, CNTs are the strongest and stiffest materials known to date (Table 6.3). The hollow, closed morphology of nanotubes results in rather intriguing deformation modes in response to a mechanical stress (Figure 6.51). In particular, it has been suggested that Stone–Wales defects may become mobile under stress, which results in a change in tube diameter/chirality.[68] This also causes a change in the electronic properties of the CNT, opening up possibilities for sensor applications. However, for MWNTs, studies have shown that only the outer graphitic shell is able to support stress,[69] and SWNT bundles/ropes (Figure 6.52) exhibit smaller Young's moduli, relative to isolated SWNTs, due to weak intertube cohesion.[70] Hence, although an individual SWNT has an elastic modulus of 1.2 TPa, the value for bundles is *ca.* 100 GPa for diameters in the range of 15–20 nm. To improve the load distribution throughout the CNT and improve its mechanical strength, there are efforts to create crosslinks between individual shells of MWNTs, and among SWNTs within ropes.

The desirable mechanical properties of CNTs have been exploited in recent years for the structural reinforcement of polymers.[71] This is a natural extension of traditional composite materials used for applications that require both high-strength

Table 6.3. Specific Tensile Strengths of Various Materials[72]

Material	Tensile strength (MPa)	Density (kg dm^{-3})	Specific strength[a] (kN m kg^{-1})	Breaking length[b] (km)
Concrete	10	2.30	4.34	0.44
Rubber	15	0.92	16.3	1.66
Brass	580	8.55	67.8	6.91
Nylon	75	1.15	97.3	9.92
Polypropylene	80	0.90	88.9	9.06
Aluminum	600	2.70	222	22.7
Steel	2,000	7.86	254	25.9
Titanium	1,300	4.51	288	29.4
Silicon carbide	3,440	3.16	1,088	110
Glass fiber	3,400	2.60	1,307	133
Graphite	4,300	1.75	2,457	250
Kevlar[c]	3,620	1.44	5,246	534
CNTs	6,2000	1.34	46,268	4,716

[a]The strength of a material divided by its density.

[b]The length beyond which a strip of the material, of uniform width, would break under its own weight if suspended from both ends.

[c]A synthetic fiber used in bullet-proof vests, comprised of poly(paraphenylene terephthalamide).

Figure 6.51. Illustration of the deformation modes of SWNTs, resulting in a high elasticity. This behavior is likely an artifact of in-plane flexibility of a graphene sheet, and facile rehybridization of carbon atoms from sp^2 to sp^3 geometries. Reproduced with permission from Ajayan, P. M. *Chem. Rev.* **1999**, 99, 1787. Copyright 1999 American Chemical Society.

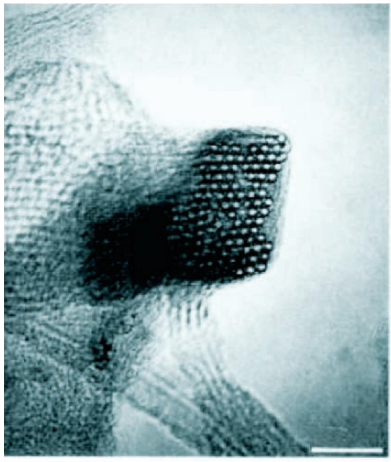


Figure 6.52. High-resolution TEM image of a bundle/rope of single-wall carbon nanotubes, formed spontaneously through the self-assembly of individual SWNTs (scale bar is 10 nm). Reproduced with permission from Thess, A.; Lee, R.; Nikolaev, P.; Dai, H.; Petit, P.; Robert, J.; Xu, C.; Lee, Y. H.; Kim, S. G.; Rinzler, A. G.; Colbert, D. T.; Scuseria, G. E.; Tomanek, D.; Fischer, J. E.; Smalley, R. E. *Science* **1996**, *273*, 483. Copyright 1996 AAAS.

and lightweight materials. There are currently plenty of examples of carbon fiber composites – aircraft and spacecraft parts, racing car bodies, golf club shafts, bicycle frames, fishing rods, automobile springs, sailboat masts, and many others – all incorporating bulk carbon fibers, *ca.* 5–10 µm in diameter. Though this technology

has been around since the 1950s, the much larger aspect ratios of CNTs translates to an even greater impact on the future of materials reinforcement.

In order to broaden the scope of CNTs for textile applications, it is most desirable to convert the as-formed powders into useful fibers and yarns. An intriguing recent precedent consists of shrinking the yarn-spinning process, used by the earliest civilizations, to the nanoregime. In this process, MWNTs from a bamboo-like forest array were drawn into fibers, and weaved into yarns, that were both strong and highly flexible (Figure 6.53). One possible application for these advanced nanotextiles is the design of a "supersuit" for the next generation of soldier. This uniform will feature a number of functionalities that will react appropriately to its surroundings (*e.g.*, deactivation of gaseous warfare agents, "kicking in" artificial muscles, climate control, *etc*. – Figure 6.53, bottom).

In addition to increasing strength and stiffness, the incorporation of CNTs also imparts conductivity to the polymer matrix. The addition of CNTs also enhances the thermal conductivity/stability, solvent resistance, and glass transition temperature of the native polymer. Further, due to the high aspect ratio of CNTs, much lower dopant levels are required to yield the desired properties, relative to standard additives such as carbon black and larger graphitic fibers. There are an endless number of potential applications for CNT composites – imagine future flat-panel displays that are flexible, exhibiting the consistency of common fabrics; or, shirts that are capable of monitoring the external temperature and automatically heating/cooling. The automotive industry currently uses CNT composites for a number of applications, such as increasing the strength of side mirrors. In addition, CNT-doped nylon is being used for fuel lines to reduce the buildup of static electrical charges, and prevent fuel line rupturing during an accident.

It should be noted that reinforcement applications are not limited to organic polymers. Though ceramics are already hard and chemically/thermally resistant, these materials may also be doped with CNTs to improve their inherent brittleness. This has already been proven for the incorporation of 5–10% of CNTs within an alumina matrix. The resultant material exhibits five times the fracture toughness and seven times greater electrical conductivity relative to undoped Al_2O_3, with an added property of unidirectional heat conductivity.[73]

In order to continue the development of intriguing CNT-based composites, three main challenges are being addressed:

(i) Ensuring a homogeneous dispersion of nanotubes throughout the polymer matrix
(ii) Controlling the direction/orientation of the CNTs in the matrix
(iii) Separating individual SWNTs from bundles/ropes (exfoliation)

In order for efficient load delocalization and strength enhancement of a composite, there must be a strong interaction between the dopant fibers and polymer matrix. However, since CNTs are allotropes of carbon, they are inherently insoluble in any solvent. Hence, there are a number of strategies that have been used to modify the surface of CNTs to facilitate their interactions with the surrounding matrix. Not only are these methods essential for the dispersion and self-alignment of nanotubes throughout a polymer host (or on a surface for molecular devices), but may also assist in the

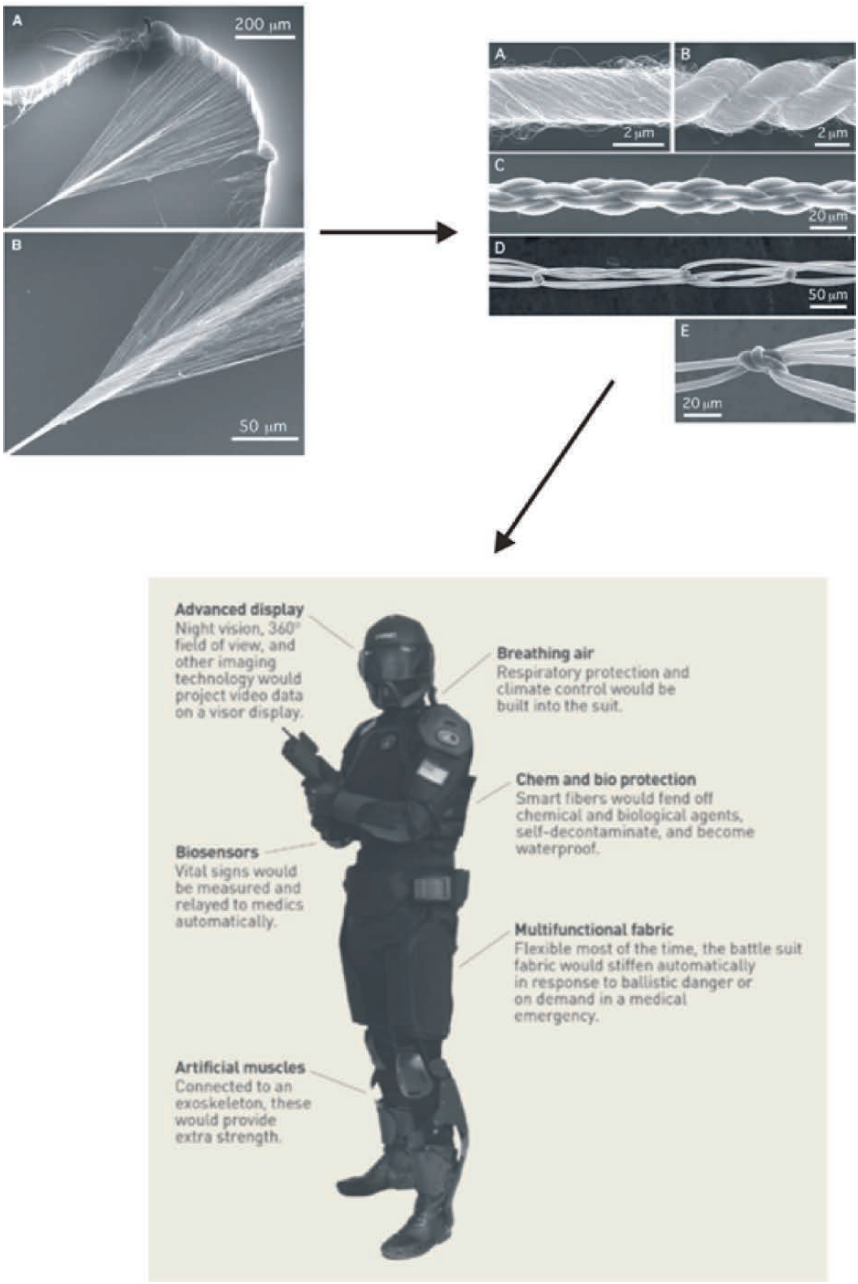


Figure 6.53. SEM images of spinning MWNT arrays into fibers, and subsequent images of MWNT yarns. Shown below is an image of a futuristic uniform to equip the next generation of warfighter. The SEM images were reproduced with permission from Baughman, R. H. *Science* **2004**, *306*, 1358, Copyright 2004 AAAS. The image of the futuristic soldier was furnished by the Army Natick Soldier Center (http://nsc.natick.army.mil) – refer to the Aug. 11, 2003 issue of *Chemical and Engineering News* for more details.

purification of CNTs (*e.g.*, separation of metallic and semiconducting CNTs through selective functionalization, followed by traditional HPLC).

There are two types of surface modifications for CNTs: noncovalent interactions (Figure 6.54) and covalent sidewall (Figure 6.55) or defect-site[74] (Figure 6.56) functionalization.[75] Both methods, as well as physical techniques such as sonication, are successful in separating individual SWNTs from bundles – a process known as *exfoliation*. In general, it is most desirable to incorporate isolated SWNTs in a composite rather than bundles, since the latter features poor intertube interactions resulting in a lower overall strength – especially at low CNT concentrations.

As illustrated in Figure 6.54, noncovalent modifications consist of either wrapping the CNT with polymers or biological macromolecules,[76] or placement of conjugated macromolecules on the surface through π-stacking interactions. As such, these types of interactions have the advantage of not altering the electronic properties of

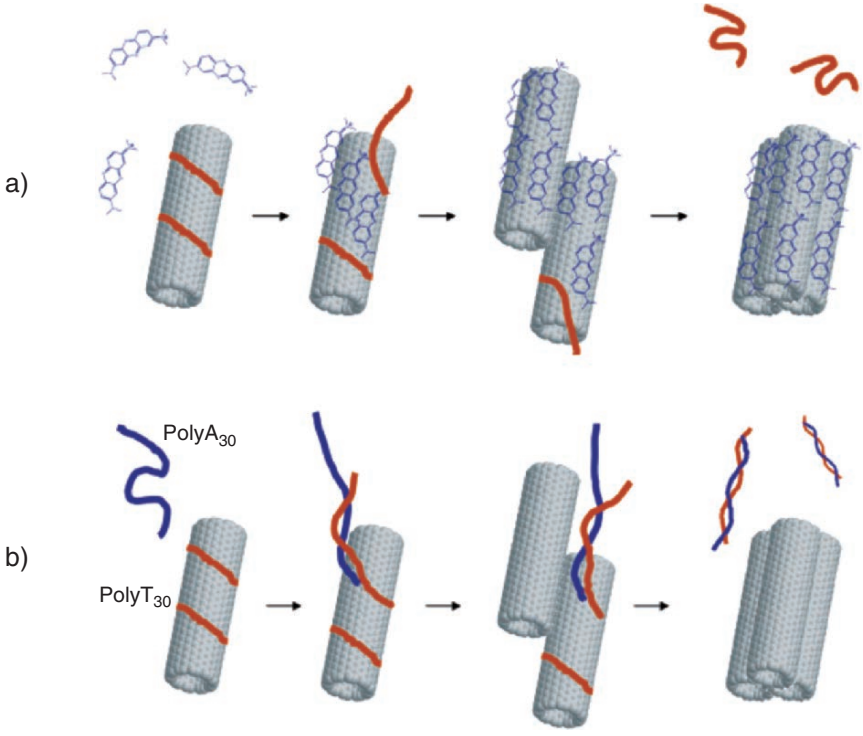


Figure 6.54. Examples of noncovalent interactions with a macromolecule solubilizing agent and CNTs. Shown is (a) the solubilization of SWNTs with PolyT30 single-strand DNA, followed by the displacement of the biomolecule through preferred π–π stacking between the conjugated dye methylene blue and the SWNT surface – causing the controlled precipitation of SWNTs from solution. Also shown is another method for solubilization/precipitation of SWNTs through use of complimentary single-strand DNA. Reproduced with permission from Chen, R. J.; Zhang, Y. *J. Phys. Chem. B* **2006**, *110*, 54. Copyright 2006 American Chemical Society.

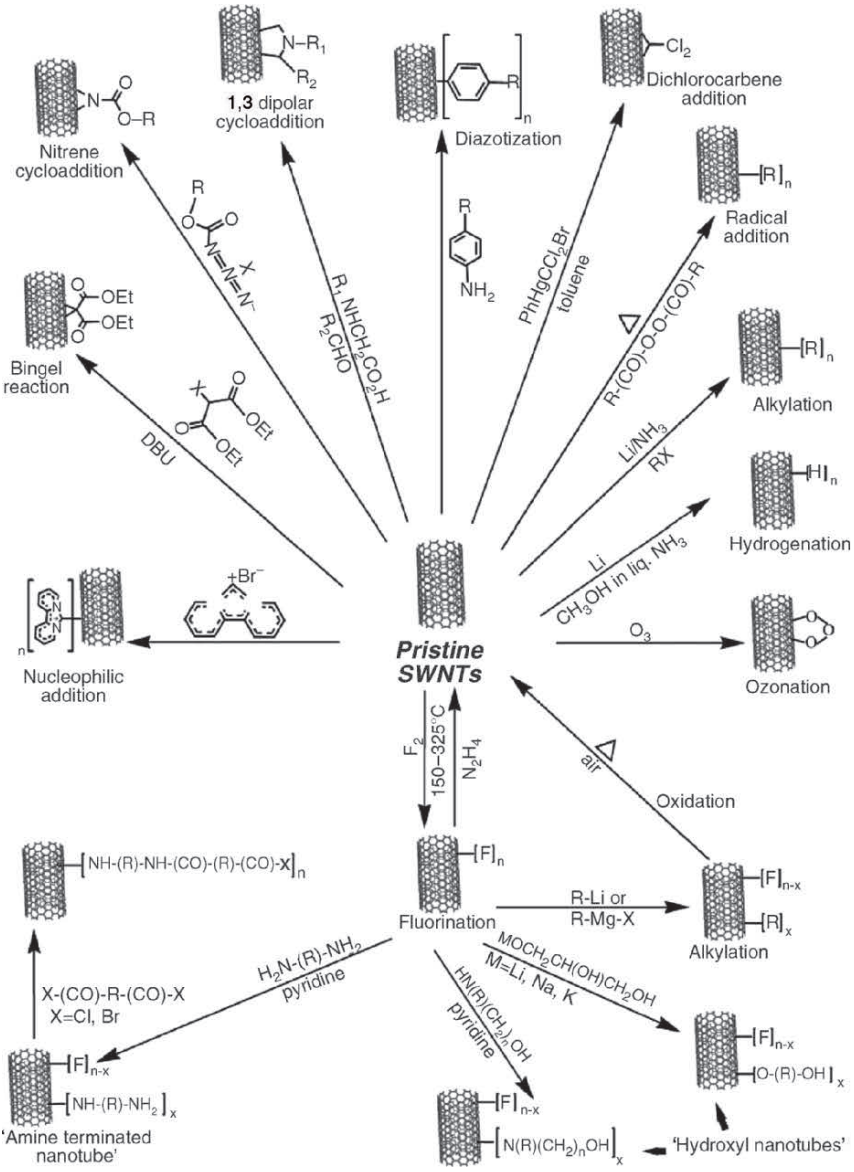


Figure 6.55. Schematic of the various methods to perform sidewall functionalization of SWNTs. Reproduced with permission from Banerjee, S.; Hemraj-Benny, T.; Wong, S. S. *Adv. Mater.* **2005**, 17, 17. Copyright 2005 Wiley-VCH.

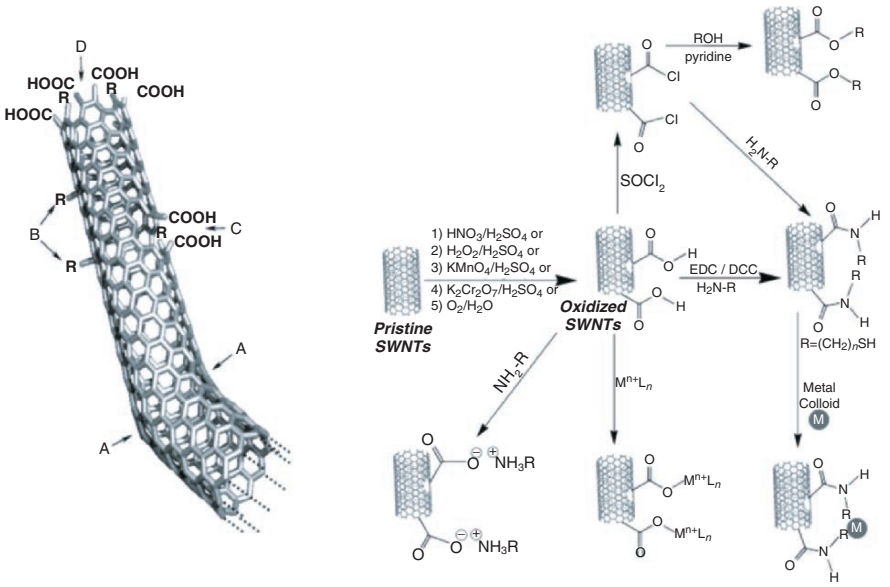


Figure 6.56. Illustration of possible defect sites on the surface of a SWNT, along with methods used for covalent functionalization of SWNTs at defect sites. The types of defects shown are (a) 5- or 7-membered rings in the framework, which results in bending, (b) sp^3-hybridized defects (R = –OH, –H), (c) oxidative surface degradation, with holes capped by –COOH groups, and (d) open end of the SWNT shown with –COOH capping groups (may also be terminated with –OH, = O, $–NO_2$, and –H groups). Reproduced with permission from (a) Hirsch, A. *Angew. Chem. Int. Ed.* **2002**, *41*, 1853 and (b) Banerjee, S.; Hemraj-Benny, T.; Wong, S. S. *Adv. Mater.* **2005**, *17*, 17. Copyright 2002 & 2005 Wiley-VCH.

the CNT (unlike covalent modifications). However, since covalent functionalization places specific chemical groups onto the CNT surface, this route offers a greater potential for the selective tunability of CNT properties as well as molecular control over the organization of CNTs for device fabrication.

The driving force behind the endcap and sidewall functionalization of SWNTs is the strain brought about from the high degree of surface curvature. The deviation of the carbon geometry from planarity is referred to as the *pyramidalization angle*, θ_p, which is defined as 0° for the sp^2 carbon atoms in C_2H_4, and 19.5° (*i.e.*, 109.5°–90°) for the sp^3 carbon in CH_4 (Figure 6.57). Since the θ_p of the carbon atoms in C_{60} is 11.6°, the desired geometry is closer to tetrahedral rather than trigonal planar. As a result, fullerenes readily undergo addition reactions to relieve this high degree of surface strain. Since the endcap of a SWNT may be considered as a half-fullerene structure, with a $\theta_p \geq 9.7°$ regardless of the nanotube diameter,[77] these points are highly reactive toward electrophilic addition.

In contrast, the sidewall pyramidalization angle is much less than the endcap, on the order of 3–6°, depending on the diameter/chirality of the SWNT. For an armchair (5,5) SWNT, two types of C–C bonds are present along the sidewalls – either parallel or perpendicular to the nanotube axis. Each of these bonds exhibits a different

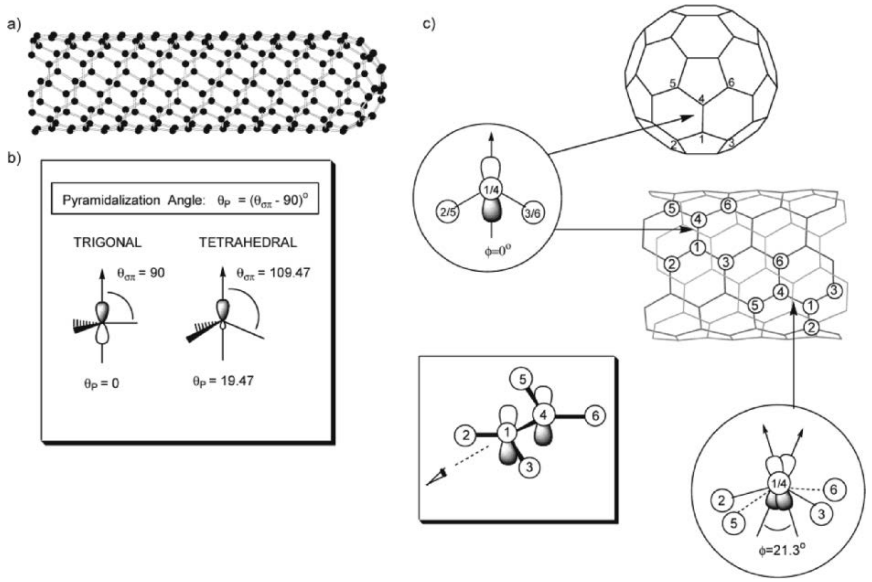


Figure 6.57. Schematic of (a) a metallic (5,5) SWNT, (b) the pyramidalization angles of typical sp^2 and sp^3 carbon atoms, (c) the π orbital misalignment angles along C1–C4 in the (5,5) SWNT and its capping fullerene, C_{60}. Reproduced with permission from Niyogi, S.; Hamon, M. A.; Hu, H.; Zhao, B.; Bhowmik, P.; Sen, R.; Itkis, M. E.; Haddon, R. C. *Acc. Chem. Res.* **2002**, *35*, 1105. Copyright 2002 American Chemical Society.

degree of π orbital overlap with adjacent carbon atoms, as quantified by the *π orbital misalignment angle*, ϕ (Figure 6.57). Whereas $\phi = 0°$ for fullerenes, the misalignment of neighboring π orbitals in SWNTs may be significant – as large as 21.4° for (5,5) species. In general, a SWNT of smaller diameter has a greater sidewall reactivity than larger analogues due to relatively large π orbital misalignment angles (much more influential than θ_p). It should also be noted that the overall curvature of CNTs is less than fullerenes, which make SWNTs significantly less reactive than fullerenes.

Growth of 1D nanostructures

Now that you are familiar with the properties and applications of CNTs, we must now consider the techniques used for their synthesis. In addition to the experimental details of CNT growth, this section will also provide mechanistic details on how these interesting nanostructures form. Fortunately, recent studies have shown that the growth mechanism of CNTs is the same for other 1D nanostructures such as nanowires.

There are three primary methods that are used to grow CNTs: laser evaporation of graphitic targets, arc discharge methods, and chemical vapor deposition (CVD). You might recognize the first two methods as also being used for to generate fullerenes; not surprising, since CNTs are essentially 1D extensions of fullerene clusters. To synthesize SWNTs, a catalytic amount of a metal (*e.g.*, Co, Ni, Fe, Y, Mo, alloys,

etc.) must be present to prevent the leading edge from closing, which would generate spheroidal fullerenes. For MWNT growth, a catalyst is not required; however, the products will contain a large number of other carbonaceous products such as fullerenes and amorphous carbon.[78]

Whereas arc-based methods generally contain a mixture of products, with SWNTs/MWNTs of varying diameters/morphologies, amorphous soot, *etc*., a dual-pulsed laser approach, first used in the mid-1990s, results in SWNT ropes with >70% purity.[79] Though both of these methods may be used to generate small quantities of CNTs, they are not easily scaleable to generate industrial quantities. Further, the CNTs arising from vaporization methods are typically in a tangled array, with other forms of carbon intermixed with remaining catalytic metal (*esp.* affecting subsequent electronic applications).

Consequently, CVD is now the method-of-choice for the synthesis of CNTs. As discussed in Chapter 4, these methods consist of the decomposition (typically thermal) of a hydrocarbon precursor on the surface of catalytic metal nanostructures. Methane and acetylene have been used most extensively as precursors; other alternatives now include CO, C_2H_4, and methanol/ethanol. As with any CVD approach, this method is easily scaleable, and is used to generate kilogram quantities of CNTs for an ever-increasing laundry list of applications.

Most importantly, CVD is most amenable for the facile aligned growth of CNTs from surface-immobilized catalyst nanoclusters. This strategy, pioneered by Dai and coworkers at Stanford, has been used to grow ordered arrays of both MWNTs and SWNTs using a variety of experimental modifications (Figure 6.58).[80] CNT growth emanates from the carefully placed catalyst particle, with resultant tube diameters related to the size of the seed catalyst. The self-assembly of the CNTs takes place through strong intratube/intertube van der Waals interactions (*e.g.*, MWNT arrays, Figure 6.58a), and tube-patterned substrate interactions (*e.g.*, SWNT arrays, Figure 6.58b), or induced by electric fields (*e.g.*, SWNT arrays, Figure 6.58c). These strategies set an important precedent of growing nanostructures along specific growth directions from specific sites – essential for the fabrication of future integrated circuits and other advanced electronic devices.

As previously seen for nanotube-based yarns, the growth of vertically-aligned CNT arrays, also known as "forests," are another example of self-assembly – through intertube van der Waal interactions. Nanotube forests are typically generated by first depositing catalyst species onto a surface, followed by CVD (Figure 6.59). There are also reports of the one-step synthesis of aligned MWNT bundles and Y-junction CNTs using a controlled mixture of metallocene-based precursors ($M(C_5H_5)_2$, M = Fe, Co, Ni) – precluding the use of surface-immobilized catalyst nanoclusters.[81] Since the precursor acts as both the carbon source and catalyst, carbon growth initiates at approximately same time from multiple metal sites. Subsequent growth is self-directed into a parallel array by interactions amongst neighboring CNTs.

For some applications, a main drawback of CVD is that MWNTs are often generated alongside SWNTs. Since the size of the catalyst governs the diameter of resultant tubes, surface-immobilized nanoclusters of iron oxide@PAMAM dendrimer, with well-defined sizes, has been shown to yield only SWNTs with

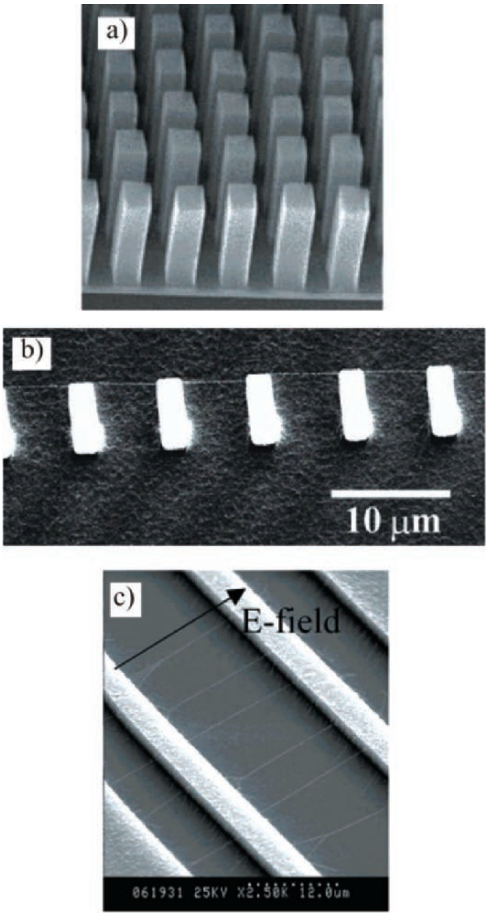


Figure 6.58. Examples of the ordered growth of carbon nanotubes. Shown are (a) MWNT arrays grown from squared regions of iron nanoclusters; (b) Side-view SEM image of a "SWNT power line on Si posts"; (c) aligned SWNT growth through electric-field induction. Reproduced with permission from Dai, H. *Acc. Chem. Res.* **2002**, *35*, 1035. Copyright 2002 American Chemical Society.

an extremely narrow diameter distribution.[82] Two advanced CVD processes, CoMoCAT® (fluidized-bed CVD) and HiPCO® (high-pressure CO CVD), have recently been developed for the commercial production of SWNTs (Figure 6.60). Though the experimental setup of these methods are significantly more complex than standard hot-walled CVD, these techniques are still considered an extension of CVD, as the precursor is decomposed on the surface of the catalyst. Most noteworthy about these and other recent CVD methods is that they offer the ability to fine-tune resultant lengths, diameters, and even chiralities[83] by varying the operating conditions. Think of the possibilities for applications, as it would not be long until you can place an order such as: "10 g of pure (5,5) SWNTs with an average diameter of

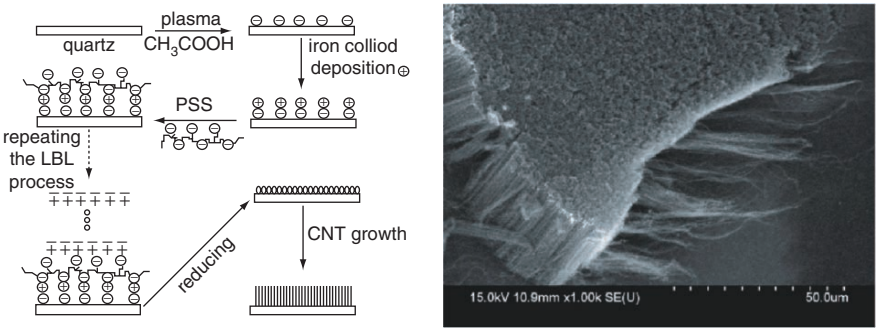


Figure 6.59. Schematic and cross-section SEM image of a layer-by-layer (LbL) approach to deposit iron nanoclusters onto a surface to yield vertically aligned SWNTs. Reproduced with permission from Liu, J.; Li, X.; Schrand, A.; Ohashi, T.; Dai, L. *Chem. Mater.* **2005**, *17*, 6599. Copyright 2005 American Chemical Society.

7 nm and length of 500 nm." However, this degree of control will only be possible once the exact growth mechanism is known, rather than current efforts that invoke "blind" modifications of experimental variables.

Though CNT growth dates back to the early 1990s, details of the growth mechanism are still unclear. The most commonly accepted description is based on the *vapor–liquid–solid* (VLS) model, which was first proposed for the growth of semiconductor whiskers (Figure 6.61).[86] This model assumes that the catalyst liquefies, which then acts as a preferential adsorption site for gaseous precursors. Subsequent growth of the nanotubes/nanowires occurs by supersaturation of the catalyst droplet, and precipitation at the liquid–solid interface. However, the amount of carbon in a supersaturated catalyst nanocluster will never be sufficient to account for the length or number of nanostructures observed in practice. As a result, additional precursor atoms must be supplied continuously to the catalyst nanocluster in order for sustained growth to occur.

By examining the phase diagram for a binary system (precursor and catalyst species), one can easily determine the best temperature range for VLS growth – any value above the eutectic temperature, where the catalyst remains a liquid (Figure 6.62). As we saw earlier for gold nanoparticles, as the diameter decreases, the melting point is significantly decreased relative to the bulk solids. As a result, "size-corrected eutectics" must be determined for the specific nanoparticle diameter in order to determine experimental conditions for nanowire growth.

Though the VLS mechanism is generally sufficient to model the growth of nanowires, there are many recent reports of CNTs being grown at temperatures below the size-corrected eutectic.[87] Though small nanoparticles may exhibit "fluid-like" behavior by altering their surface geometry, empirical data has shown that the catalyst particle may remain crystalline, rather than an amorphous liquid phase, during growth. This suggests that nucleation may be dominated by the catalyst *surface* – not unlike most heterogeneous catalysis processes (*e.g.*, Ziegler–Natta polymerization).[88]

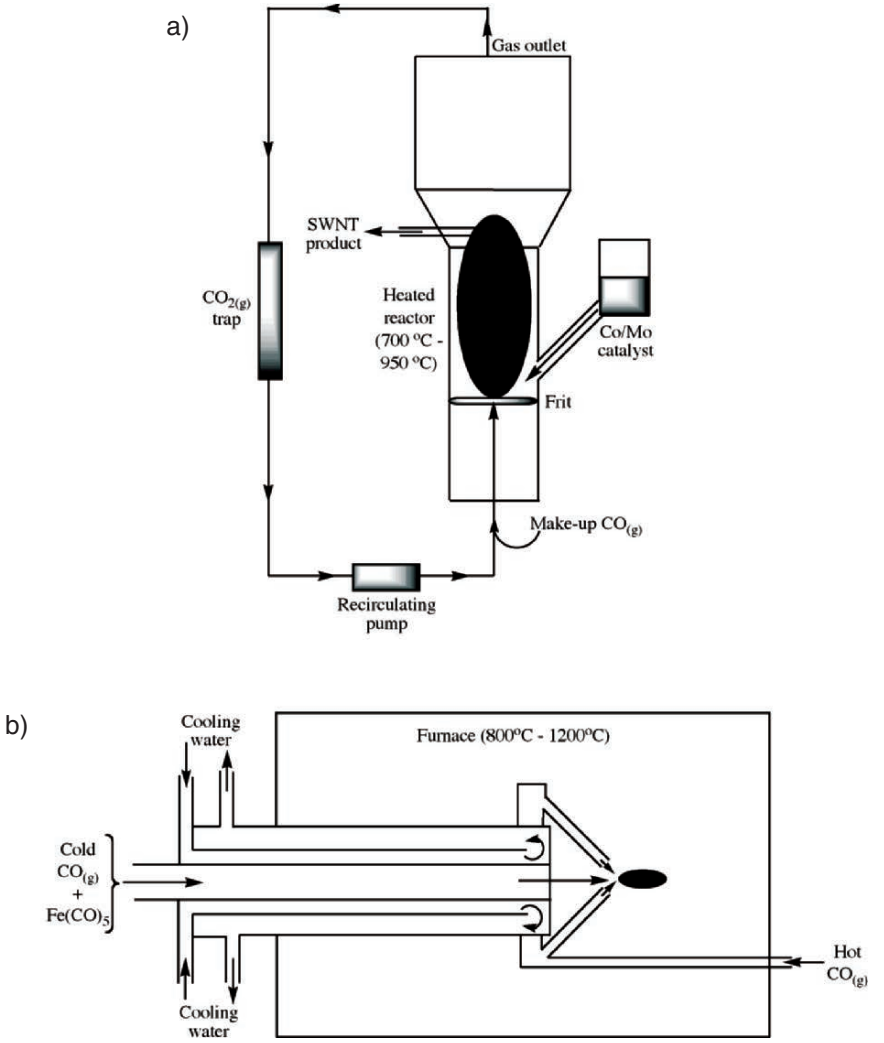


Figure 6.60. Illustration of two methods used for the commercial production of SWNTs. Shown are (a) the CoMoCat fluidized bed method using CO as the precursor and a Co/Mo bimetallic catalyst,[84] and (b) the HiPco "floating catalyst" process using the thermal decomposition of iron pentacarbonyl at pressures of 1–10 atm.[85]

Since the growth of 1D nanostructures is difficult to study *in situ*, a number of computational techniques have been used in an attempt to decipher the growth mechanism(s). For CNT growth, it has been shown that small graphitic islands form on the supersaturated catalyst surface. When the island covers half of the catalyst particle, it lifts off and forms the SWNT – with the same diameter as the catalyst. Based on empirical data and theoretical predictions,[89] the catalytic ability of the metal/alloys

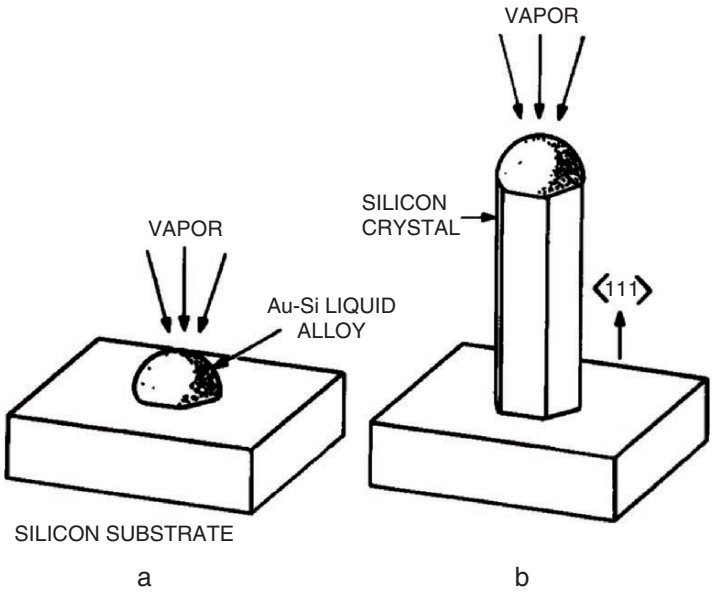


Figure 6.61. The original schematic used to describe vapor–liquid–solid (VLS) growth of semiconductor nanowires. Reproduced with permission from Wagner, R. S.; Ellis, W. C. *Appl. Phys. Lett.* **1964**, 4, 89. Copyright 1964 American Institute of Physics.

for SWNT growth follows the order: Ni/Mo > Ni/Cr > Ni/Co > Ni/Pt > Ni/Rh > Ni/Fe > Ni > Fe/Mo > Fe/Cr > Fe/Co > Fe/Pt > Fe/Rh > Fe > Ni/Mo > Fe/Mo > Co/Mo > Co > Pt > Cu. Depending on whether the mechanism proceeds *via* VLS or surface-addition mechanism, the above order may be related to the ease of carbide formation and carbon diffusion through the nanocluster interior, or the morphology/composition of the catalyst surface and rate of carbon diffusion on the catalyst surface, respectively. Figure 6.63 shows some recent mechanistic proposals, based on experimental and theoretical data. The VLS-based proposals (Figure 6.63, top and middle) illustrate the following basic steps:

(i) At initial growth stages, carbon dissolves in the molten catalytic nanocluster. Calculations indicate that there is a dynamic process of carbon precipitation onto the catalyst surface and redissolution, until a highly supersaturated catalyst is obtained.

(ii) Carbon precipitates on the surface of the highly supersaturated catalyst nanoclusters, forming carbon strings/polygons. This causes a decrease in the dissolved carbon concentration.

(iii) The carbon nuclei form graphitic islands on the surface of the catalyst, which aggregate into larger graphitic clusters.

(iv) At low temperatures, the graphitic islands are not able to lift off the catalyst surface, resulting in graphite-encapsulated metal nanoclusters.[90]

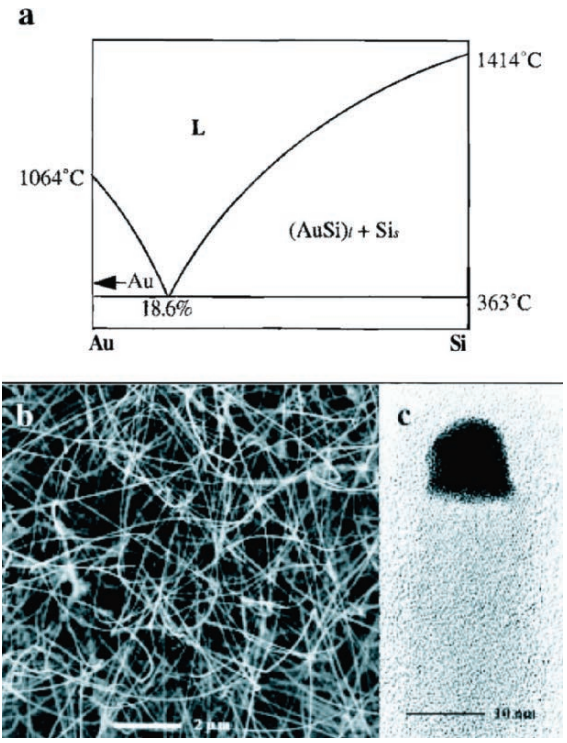


Figure 6.62. Silicon nanowire growth from a gold nanocluster catalyst. Shown is (a) the phase diagram for the Au/Si system, showing the eutectic temperature/composition; (b) SEM image; and (c) high-resolution TEM image of the nanowires grown at a temperature of 450°C. The dark tip of the nanowire is from the gold nanocluster. Reproduced with permission from Hu, J.; Odom, T. W.; Lieber, C. M. *Acc. Chem. Res.* **1999**, *32*, 435. Copyright 1999 American Chemical Society.

(v) At relatively high temperatures (*ca*. 500°C–1,200°C), when the diameter of the island becomes *ca*. 1/2 that of the catalyst, the graphitic nucleus lifts off the catalyst surface to form the SWNT endcap. Subsequent graphitization and growth propagation of the SWNT may occur through two routes (Figure 6.63 (middle)):

(i) "Root Growth" (c–d): carbon atoms precipitate from the molten catalyst and join to the open end of the growing SWNT.

(ii) "Folded Growth" (e–g): carbon atoms precipitate directly onto the catalyst surface, and are added to the graphitic endcap.

In contrast, a surface-governed route is characterized by negligible carbon dissolution in the catalyst bulk. Since the growth is not proposed to occur through supersaturation/precipitation, any species that are chemisorbed onto the catalyst surface will have a dramatic influence on SWNT growth. Figure 6.63 (bottom) shows a proposed route for a surface-based SWNT growth mechanism, which progresses through the continual addition of C_2 units onto the leading edge of the growing nanotube. As

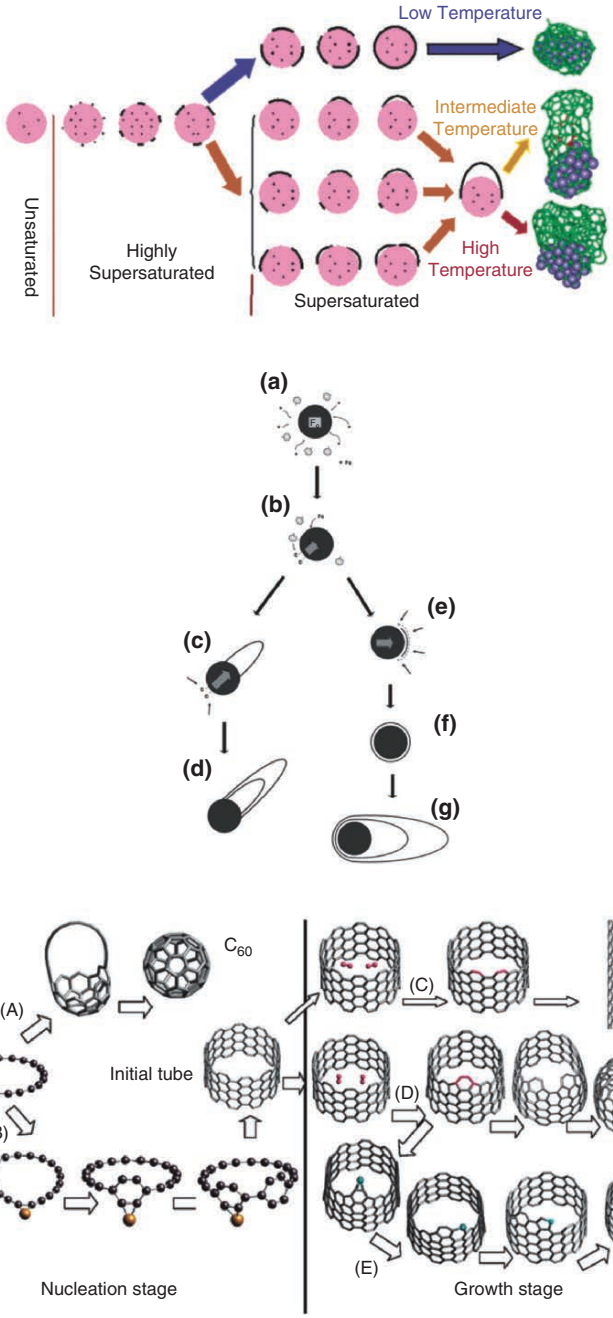


Figure 6.63. Comparison of a VLS-based mechanism (top and middle) and a surface-mediated route[95] (see text for details).

indicated in this scheme, the absence of an active catalyst will result in the formation of fullerenes rather than SWNTs. In this route, nanotube growth is proposed to initiate from surface metal atoms adding to the edge of a polyyne ring, which is subsequently able to add additional carbon atoms.

Whereas path C of Figure 6.63 (bottom) shows the successful addition of C_2 units at the growth edge, path D shows the formation of a defect, which leads to tube closure. It should be noted that this route is 1.3 eV more favorable than route C when there is no metallic catalyst present, thus explaining the preference for closed-cage fullerenes. Most likely, this is a consequence of pentagonal defects that are formed when carbon atoms add to the open end of the growing SWNT. Hence, calculations have shown that catalyst-assisted defect repair is a crucial step in the overall mechanism (Figure 6.63 (bottom, route E). This is proposed to occur through a "scooter mechanism,"[91] whereby the catalyst atoms diffuse along the open-end surface of the SWNT, facilitating the formation of hexagons that prevent premature nanotube closure.

It should be noted that an additional variable may be required to describe the full mechanism for CNT growth. Recent studies suggest that there is likely an important role for metal oxide catalysts, or oxygenated co-reactants, in the overall growth mechanism of SWNTs. For example, the use of water vapor as a co-reactant for CVD results in the growth of a dense forest of vertically aligned SWNTs, with lengths up to 2.5 mm after only 10 min of reaction time (Figure 6.64a).[92] The catalytic enhancement of water vapor is evidenced by the SWNT/catalyst weight ratio of 50,000% – more than 100 times greater than the HiPCO process. The effect of an oxidizing species such as water vapor seems logical from the standpoint of reducing the amount of amorphous carbon deposits, and controlling the degree of carbon precipitation (thus preventing MWNT growth). However, its role may be even more significant – perhaps even providing a new mode of nucleation. In particular, it is proposed that SWNT growth may be initiated from oxygen-etched graphitic shells that encapsulate the catalyst – a route recently termed "nucleation *via* etched carbon shells" (NECS), Figure 6.64b.[93]

The formation of nanotubes is not limited to carbonaceous materials. In fact, a number of inorganic-based nanotubes have been synthesized in recent years – such as metals, oxides, carbides, borides, *etc*.[94] There are many applications for these materials, such as catalysis, sensors, and advanced ceramic nanostructures. Due to the lack of sp^2 hybridization and no growth mode analogous to the rolling of "graphene" sheets, these 1D nanostructures do not self-assemble spontaneously from a catalytic seed. Instead, the most common method to synthesize these structures is by coating a CNT with a thin film of the desired material, followed by the sacrificial removal of the CNT template. As we saw earlier for nanocluster growth, templating routes offer a convenient mode of nanostructural synthesis, as well as surface alignment of the nanoarchitectures on a substrate surface.

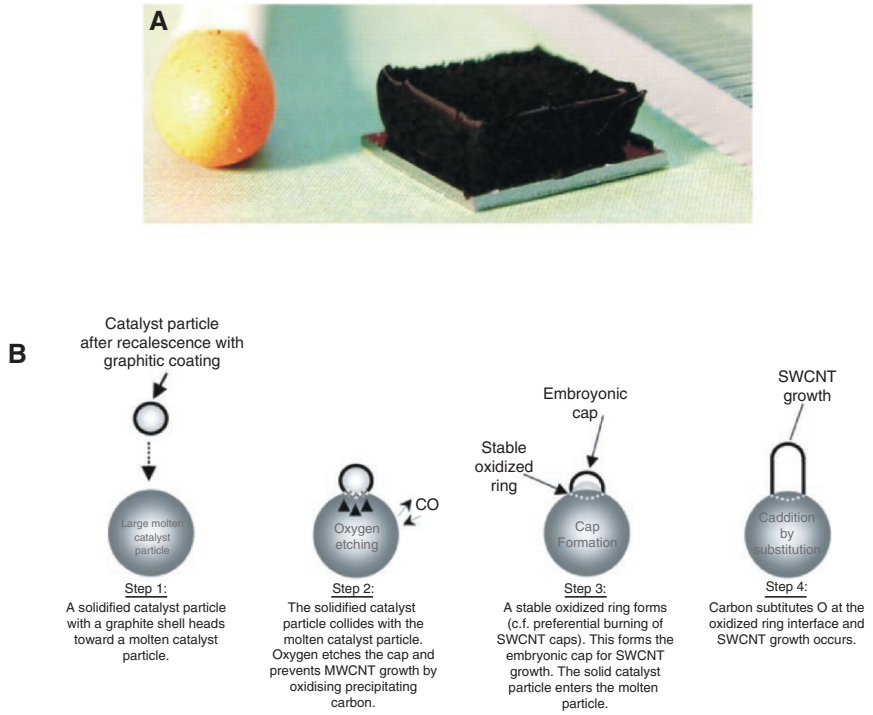


Figure 6.64. (a) Photo of a cm-scale SWNT forest, shown next to a matchstick for comparative purposes. Reproduced with permission from Hata, K.; Futaba, D. N.; Mizuno, K.; Namai, T.; Yumura, M.; Iijima, S. *Science* **2004**, *306*, 1362. Copyright 2004 AAAS. (b) A proposed scheme for the laser ablative growth of SWNTs *via* metal oxide nanocluster catalysts. Reproduced with permission from Rummeli, M. H.; Borowiak-Palen, E.; Gemming, T.; Pichler, T.; Knupfer, M.; Kalbac, M.; Dunsch, L.; Jost, O.; Silva, S. R. P.; Pompe, W.; Buchner, B. *Nano Lett.* **2005**, 5, 1209. Copyright 2005 American Chemical Society.

6.3. TOP-DOWN NANOTECHNOLOGY: "SOFT LITHOGRAPHY"

As discussed in Chapter 4, the design of transistors, with feature sizes currently much less than 100 nm, will need to move beyond traditional photolithography (Figure 6.65a) in order to continue Moore's Law and the miniaturization of electronic devices. In addition to the exorbitant cost of photolithography, this technique is not amenable for the patterning of large and nonplanar substrates – of importance for future 3D devices. Though advanced techniques such as deep UV (DUV), extreme UV (EUV), and focused-ion beam lithographies are able to push the resolution limits of patterning to well below 100 nm, they are much too expensive for the low-cost, high-volume processing that is required for commercial applications.[96]

Within the last decade, comparatively inexpensive and scaleable tchniques known as "soft lithography" have been the focus of much development. Patterning of a substrate is afforded by using a master elastomeric stamp that contains a nanostructured pattern, known as a *relief*, on its surface. Contrary to photolithography, the resolution of the final pattern is not limited by light diffraction, but only depends on the

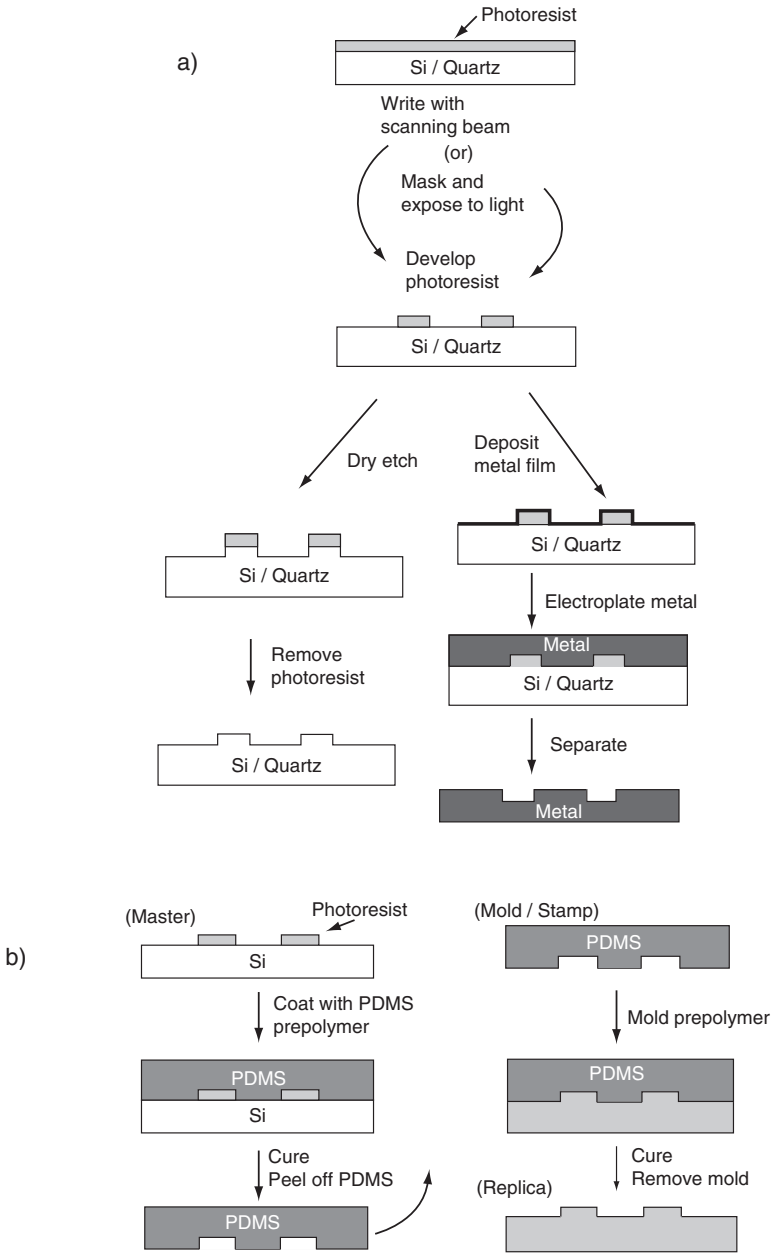


Figure 6.65. Comparison of (a) conventional photolithography/electroplating with (b) soft lithography. Shown in (b) is *replica molding* which consists of the formation of a PDMS stamp, and subsequent replication of a master in a photo- or thermally curable prepolymer. Reproduced with permission from Gates, B. D.; Xu, Q.; Stewart, M.; Ryan, D.; Willson, C. G.; Whitesides, G. M. *Chem. Rev.* **2005**, *105*, 1171. Copyright 2005 American Chemical Society.

dimensions of the relief structures – typically fabricated in the master by electron-beam lithography. Typically, the mold (or stamp) is comprised of poly(dimethylsiloxane) (PDMS), which allows for intimate contact between the mold/substrate surfaces, even if nonplanar substrates are used. More recently, other polymers have been developed for this application such as polyimides, polyurethanes, and a variety of substituted siloxanes – especially fluorinated analogues due to easy release after molding, and lack of swelling by organic solvents.

The technique of replicating a master pattern is aptly termed *replica molding* (Figure 6.65b). In theory, the resolution of the replica will be identical to the master. However, due to the "soft" nature of the mold, the nanoscale features of the relief may become distorted due to polymer shrinkage (*e.g.*, solvent evaporation, *in situ* cross-linking, mechanical deformation), or interfacial phenomena between the mold and master surfaces (*e.g.*, differing thermal expansions, adhesive forces[97]). In contrast, a hard mold of Si or quartz exhibits significantly less distortion due to their solvent/chemical resistance, and thermal stabilities at temperatures sufficient to cause polymer cross-linking. Hard molds, used for step-and-flash imprint lithography (SFIL) and nanoimprint lithography (NIL), are commonly used to pattern materials such as CDs, DVDs, and holographic images on the front of most credit cards.[98]

A common application for elastomeric molds is for micro- or nanocontact printing, where a self-assembled monolayer (SAM) is placed on both planar[99] and curved[100] surfaces *via* contact with the reliefs on the mold (Figure 6.66). SAMs will be an important architecture for the next generation of nanostructured materials.

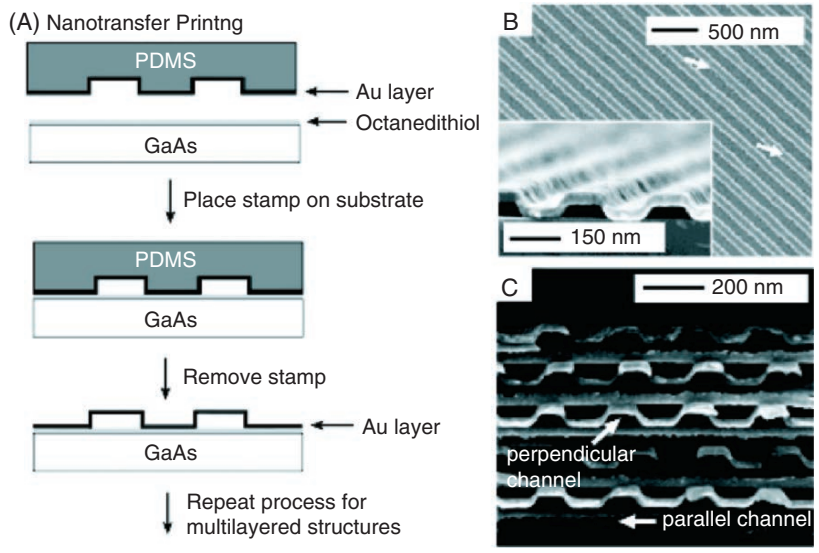


Figure 6.66. Schematic of the general process of nanotransfer printing. SEM image B illustrates a 20-nm grooved gold layer transferred onto a GaAs surface. Image C shows a multilayered stack of 20-nm thick layers of parallel grooves; the channels in adjacent layers are aligned perpendicular to one another.[102] Reproduced with permission from Gates, B. D.; Xu, Q.; Stewart, M.; Ryan, D.; Willson, C. G.; Whitesides, G. M. *Chem. Rev.* **2005**, *105*, 1171. Copyright 2005 American Chemical Society.

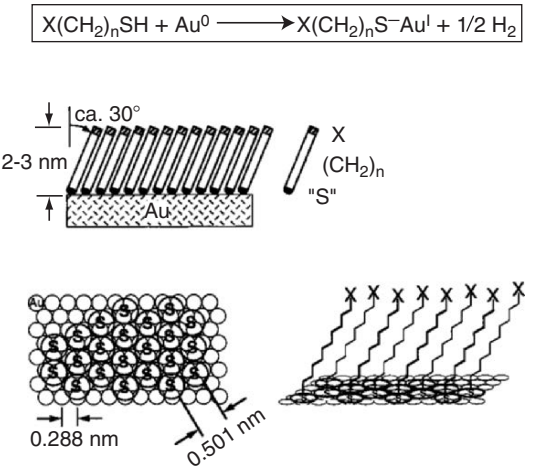


Figure 6.67. Schematic of a self-assembled monolayer (SAM), illustrating the organization of the alkyl chains *via* van der Waal interactions, and the close-packed array of sulfur atoms on the gold surface. Reproduced with permission from Xia, Y.; Whitesides, G. M. *Angew. Chem. Int. Ed.* **1996**, *37*, 550. Copyright 1996 Wiley-VCH.

The archetypical example of a SAM is the chemisorption of alkylthiols on a gold or silver surface, which results in self-assembly/alignment into a 3D forest array (Figure 6.67). Applications for SAMs span a number of fields from sensors to high-density storage; a recent precedent illustrates the selective adsorption and spontaneous alignment of CNTs[101] and directed growth of nanowires[102] from SAMs.

In order to improve the stamping resolution of the elasomeric stamp, there have been recent improvements in both the stamp and "molecular inks" (*e.g.*, alkylthiols, silanes). In particular, traditional PDMS exhibits a relatively high elasticity that limits possible relief linewidths; on the other hand, small molecular weight inks exhibit diffusion during patterning. Hence, the following complementary strategies have been employed:

(i) Using a composite two-layer stamp comprised of a 30 μm hardened PDMS coating on a 2–3 mm thick PDMS support[103]; especially in tandem with sharp, V-shaped grooves[96]
(ii) Using high molecular weight inks such as dendrimers[104] and biological molecules (*e.g.*, proteins[97])

Using a combination of the above modifications has now extended nanocontact printing to well below the 30 nm regime – even as low as 2 nm![105]

One of the more recent techniques for nanocontact printing consists of the direct writing of a molecular ink onto an appropriate substrate *via* the ultra-fine tip of an atomic force microscope[106] (Figure 6.68). The use of such a "nanofountain pen" is known as *dip-pen nanolithography* (DPN), first demonstrated by Mirkin and coworkers in the late 1990s.[107] Though the earliest examples of DPN featured alkylthiols as the ink onto Au surfaces, there are now an increasingly large number of other ink/substrate combinations that have been reported for DPN (Table 6.4).[108]

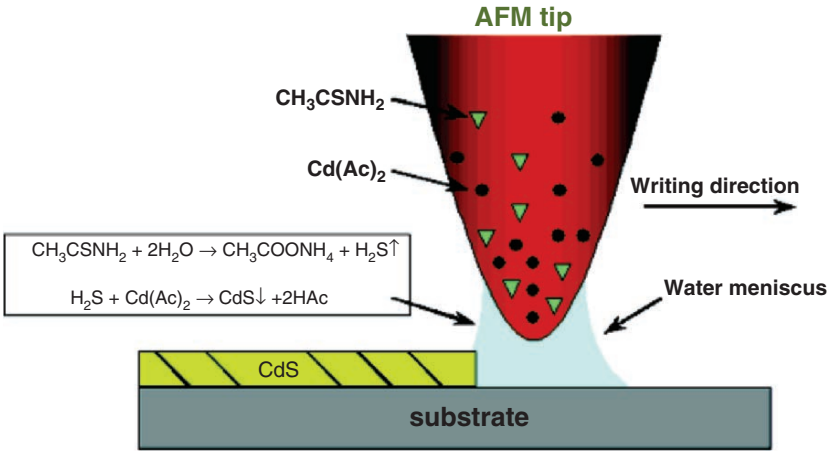


Figure 6.68. Illustration of dip-pen nanolithography, used to write nanofeatures of CdS on mica and SiO_x substrates. Reproduced with permission from Ding, L.; Li, Y.; Chu, H.; Li, X.; Liu, J. *J. Phys. Chem. B* **2005**, *109*, 22337. Copyright 2005 American Chemical Society.

Table 6.4. Summary of the Ink-Substrate Combinations Used to Date for DPN[109]

Molecular ink	Substrate
Alkylthiols (*e.g.*, ODT[a] and MHA[b])	Au
Ferrocenylthiols	Au
Silazanes	SiO_x, GaAs
Proteins	Au, SiO_x
Conjugated polymers	SiO_x
DNA	Au, SiO_x
Fluorescent dyes	SiO_x
Sols	SiO_x
Metal salts	Si, Ge
Colloidal particles	SiO_x
Alkynes	Si
Alkoxysilanes	SiO_x
ROMP materials	SiO_x
Thioacetamide/cadmium acetate[c]	SiO_x, mica

[a] 1-Octadecanethiol.
[b] 16-Mercaptohexadecanoic acid, or thiohexadecanoic acid.
[c] Ding, L.; Li, Y.; Chu, H.; Li, X.; Liu, J. *J. Phys. Chem. B* **2005**, *109*, 22337.

A general benefit of DPN over other soft lithographic techniques is the ability to pattern nanostructures (including biological materials) by a single step without cross-contamination, since the desired chemistry occurs only in a specifically defined location of the substrate.

The mechanism of ink transport from the AFM tip to substrate is currently an item of controversy. Recent models suggest that a meniscus forms between the tip

and substrate, which aids in ink transport. As a result, the transport rate is found to increase concomitantly with the ambient humidity – but only for inks that are soluble in water. In general, the rate of ink transport is found to decrease significantly with increasing contact time, due to the changing surface energy of the substrate. As you might imagine, there are many factors that govern ink transport and the final resolution of the printed nanostructure – tip shape, ink composition/concentration, substrate surface properties, and ambient conditions. It is clear that the current 15-nm resolution of DPN may only be improved once a full picture of ink transport is fully understood.

IMPORTANT MATERIALS APPLICATIONS V: NANOELECTROMECHANICAL SYSTEMS (NEMS)

The past decade or so has brought about tremendous growth in the field of microelectromechanical systems (MEMS) – devices such as sensors, computers, electronics, and machines at the microscale. The design of these devices utilizes a combination of traditional semiconductor processing and mechanical engineering. Typically, the output of an electromechanical device is the movement of the mechanical component; a transducer is used to convert the mechanical energy into electrical/optical signals, or vice versa. A sampling of some current MEMS applications include:

(i) Automotive (*e.g.*, air flow and tire pressure sensors, "smart" suspension, headlight leveling, navigation, vehicle security, automatic seatbelt restraint, *etc*.)
(ii) Micronozzles for inkjet printers
(iii) Microtweezers
(iv) Aerospace navigational gyroscopes
(v) Disposable blood pressure transducers
(vi) Portable skin analysis sensors for cosmetics applications

The extension of MEMS to the nanoregime is referred to as NEMS – representing the ultimate in future devices – with benefits such as lower power dissipation and ultra-sensitive and localized responses. Further, due to the size of NEMS, it will be possible to directly incorporate a number of auxiliary functionalities alongside transistors within a single chip. Indeed, the applications for NEMS will span the fields of sensors, electronics, biotechnology, affecting virtually every aspect of our lives. Though the technology is in place to fabricate NEMS, there are three primary challenges that must be overcome prior to the realization of widespread commercial applications:

(i) How to communicate signals from the nanoscale to the macroscopic world
(ii) Understanding and controlling mechanical responses at the nanoscale
(iii) Developing methods for low-cost, high-volume, and *reproducible* nanofabrication

One recent NEMS device is a mass sensor with a resolution at the zeptogram level ($1\,\text{zg} = 10^{-21}\,\text{g}$).[110] Thus far, the lowest detection limit for this device is 7 zg, or *ca.* 30 Xe atoms. Incredibly, it is suggested that this technology may be combined with nanofluidics for the genetic analysis of the DNA present within a single cell! In

comparison, current methods utilize PCR amplification, whereby small samples of DNA are repeatedly replicated in order to facilitate detection. Such analyses will be of use for the detection of genetic markers for cancer or other diseases such as HIV. The ability to identify proteins will be essential for future diagnostic and forensics applications.

Another area of NEMS that is receiving tremendous attention is the mimicry of biological systems, aptly referred to as *biomimetics*. For instance, in the development of *linear molecular muscles* that undergo contraction and extension movements. Initial work in this field utilized transition metal complexes containing rotaxanes and catenanes, due to the nondestructive redox processes occurring on the metal centers.[111] Though these complexes were actuated by a chemical reaction, the movement was in a noncoherent manner. In order to better mimic skeletal muscle movement, one has to look at the mode of motion within the most efficient molecular machines – in our human bodies.

The cellular unit that is active toward the contraction of skeletal muscles, known as the sarcomere, is comprised of alternatively stacked filaments of the proteins actin and myosin. During muscle contraction, the protein filaments slide past each other as a result of a rowing action of the surface myosin heads (Figure 6.69a).[112] Hence,

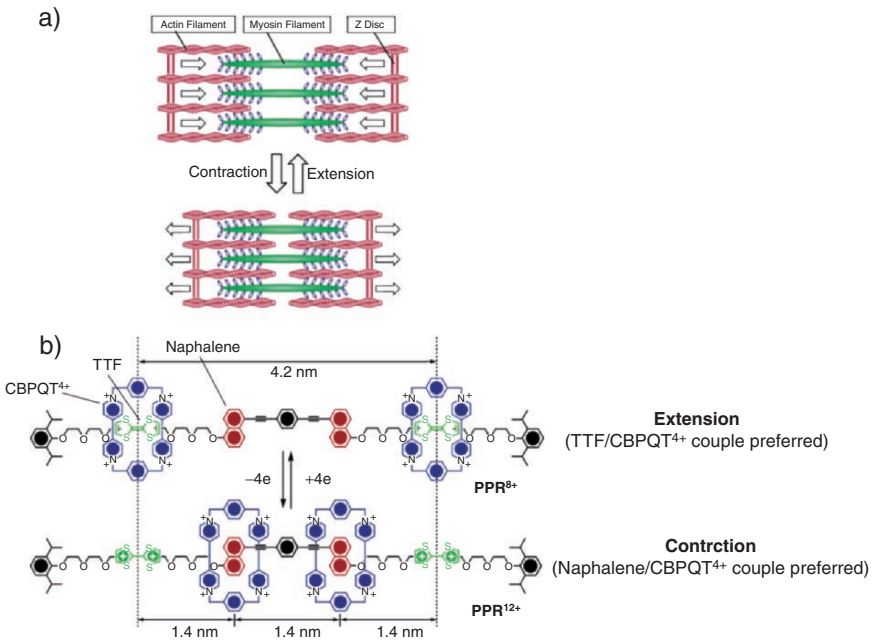


Figure 6.69. A biomimetic approach toward skeletal muscle movement. Shown is (a) the stacked protein filaments of the sarcomere and (b) a redox-controlled molecular analogue. Adapted with permission from Liu, Y.; Flood, A. H.; Bonvallet, P. A.; Vignon, S. A.; Northrop, B. H.; Tseng, H.-R.; Jeppesen, J. O.; Huang, T. J.; Brough, B.; Baller, M.; Magonov, S.; Solares, S. D.; Goddard, W. A.; Ho, C.-M.; Stoddart, J. F. *J. Am. Chem. Soc.* **2005**, *127*, 9745. Copyright 2005 American Chemical Society.

an effective biomimetic approach would entail the design of a linear architecture that features sliding components that will respond to a chemical stimulus. This approach has recently been demonstrated with the design of a rotaxane molecule that exhibits redox-controlled contraction and extension of the molecular architecture, in response to a chemical or electrochemical stimulus (Figure 6.69b).[113]

The controlled movement in this system is controlled by the interaction among redox-active units at the redox-active tetrathiafulvalene (TTF) units. In the neutral, unperturbed state, the tetracationic cyclophane units (cyclobis(paraquat-*para*-phenylene), $CBPQT^{4+}$) are most stable coordinated to the TTF moieties, due to electron donation and π-stacking interactions. However, upon oxidation of the TTF units, the $CBPQT^{4+}$ rings become electrostatically repelled, migrating to the naphalene component (Figure 6.69b). Hence, the "rowing" action exhibited by skeletal muscles is emulated by the Coulombic repulsion and π-donation of the naphalene rings.[114] By attaching disulfide tethers to the $CBPQT^{4+}$ rings, the attachment of the linear muscle may be fashioned to a gold surface en route toward a biomimetic NEMS device.[115]

In the next few years, there will be many exciting developments related to NEMS devices. In particular, replacing the current chemically driven molecular machines with those stimulated by optical or electrical pulses – dramatically extending the range of applications. With so many possible nanobuilding blocks at our disposal, the scope of devices and resultant applications is limited only by our imaginations – an exciting area of discovery awaits!

References and Notes

1 A sampling of some intriguing applications that are already possible using nanomaterials include: self-cleaning fabrics (*via* TiO_2 nanoparticles), automobile clearcoats that prevent scratches (PPG nanoparticle-based coatings), car wash solutions that prevent dirt from adhering to a painted surface, bandages that kill bacteria, drug-release agents and time-release biocidal coatings, and tennis balls that bounce twice as long as conventional balls.

2 Only US-based institutes/centers are listed here; for a more comprehensive list of worldwide nanotechnology efforts, see http://sunsite.nus.sg/MEMEX/nanolink.html, a comprehensive listing of nanorelated websites hosted by the University of Singapore.

3 Now available online at Eric Drexler's "Foresight Institute" website: http://www.foresight.org/EOC

4 For details on the biological effects of CNTs, see: Liu, Z.; Cai, W.; He, L.; Nakayama, N.; Chen, K.; Sun, X.; Chen, X.; Dai, H. *Nat. Nanotechnol.* **2007**, 2, 47, and references therein. The biological effects of dendritic polymers is described in Boas, U.; Heegaard, P. M. H. *Chem. Soc. Rev.* **2004**, *33*, 43, and references therein. Some comprehensive websites on the toxicological effects of nanostructures include: (a) http://orise.orau.gov/ihos/Nanotechnology/nanotech_OSHrisks.html; (b) http://www.cdc.gov/niosh/topics/nanotech; (c) http://www.bnl.gov/cfn; (d) http://www.nanotox.com/nanomaterials-testing.htm; (e) http://membership.acs.org/c/ccs/nano.htm

5 http://www.ethicsweb.ca/nanotechnology

6 http://en.wikipedia.org/wiki/UNIVAC_I. This image is a work of a United States Census Bureau employee, taken or made during the course of an employee's official duties. As a work of the US Federal Government, the image is in the public domain.

7 Taniguchi, N. On the Basic Concept of NanoTechnology. *Proc. ICPE* **1974**.

8 http://www.ipt.arc.nasa.gov/nanotechnology.html

9 For example, see Pishko, V. V.; Gnatchenko, S. L.; Tsapenko, V. V.; Kodama, R. H.; Makhlouf, S. A. *J. Appl. Phys.* **2003**, *93*, 7382.

[10] For an excellent review of transition metal nanocluster formation and nomenclature, as well as the difference between colloids and nanoclusters, see Finke, R. G. *Transition Metal Nanoclusters* in *Metal Nanoparticles: Synthesis, Characterization, and Applications*, Dekker: New York, 2002.

[11] For example, there is a 500% rate difference for the photoreduction of CO_2 using 10 different samples of Pd_n colloids: Wilner, I.; Mendler, D. *J. Am. Chem. Soc.* **1989**, *111*, 1330. Also, see Kohler, J. U.; Bradley, J. S. *Catal. Lett.* **1997**, *45*, 203, wherein they describe a 670% variation in the rate of hydrogenation with PVP-protected Pt_n colloids (due to a widely dispersed composition, with varying numbers of surface Cl^- groups).

[12] Though quantum dots are typically thought of as 0D nanostructures, quantum confinement effects are also exhibited in 1D nanowires and nanorods. Buhro and coworkers have studied the effect on both size and shape on quantum confinement (Yu, H.; Li, J.; Loomis, R. A.; Wang, L.-W.; Buhro, W. E. *Nature Mater.* **2003**, 2, 517). Their work provides empirical data to back up the theoretical order of increasing quantum confinement effects: dots (3D confinement) > rods > wires (2D confinement) > wells (1D confinement). For an example of an interesting nanostructure comprised of both a nanorod and nanodot, see: Mokari, T.; Sztrum, C. G.; Salant, A.; Rabani, E.; Banin, U. *Nature Mater.* **2005**, 4, 855.

[13] (a) Lewis, J. *Chem. Br.* **1988**, *24*, 795. (b) Deeming, A. J. *Adv. Organomet. Chem.* **1986**, *26*, 1.

[14] Shown from left to right are (a) Pd nanoclusters supported on hydroxyapatite: Mori, K.; Hara, T.; Mizugaki, T.; Ebitani, K.; Kaneda, K. *J. Am. Chem. Soc.* **2004**, *126*, 10657. (b) Copper nanoclusters: Williams, G. L.; Vohs, J. K.; Brege, J. J.; Fahlman, B. D. *J. Chem. Ed.* **2005**, *82*, 771.

[15] Huang, J.; Kunitake, T.; Onoue, S.-Y. *Chem. Commun.* **2004**, 1008.

[16] Scher, E. C.; Manna, L.; Alivisatos, A. P. *Philos. Trans. R. Soc. Lond. A.* **2003**, *361*, 241.

[17] Shown is a Ti/O/C nanopowder with individual nanosized grains: Leconte, Y.; Maskrot, H.; Herlin-Boime, N.; Porterat, D.; Reynaud, C.; Gierlotka, S.; Swiderska-Sroda, A.; Vicens, J. *J. Phys. Chem. B* **2006**, *110*, 158.

[18] Shown are submicron particulates (with some nanoparticles also present) of aluminum oxide: Williams, G. L.; Vohs, J. K.; Brege, J. J.; Fahlman, B. D. *J. Chem. Ed.* **2005**, *82*, 771.

[19] For a thorough review of surface plasmon resonance, see Kelly, K. L.; Coronado, E.; Zhao, L. L.; Schatz, G. C. *J. Phys. Chem. B* **2003**, *107*, 668.

[20] Mie, G. *Ann. Phys.* **1908**, *25*, 377. This theory represents the exact solution to Maxwell's equations for a sphere. For details on recent theories to describe scattering from nonspherical nanostructures, see Reference [15], and the references therein.

[21] Haes, A. J.; Stuart, D. A.; Nie, S.; Duyne, R. P. V. *J. Fluoresc.* **2004**, *14*, 355.

[22] For more information/precedents on quantum confinement effects for metallic nanoclusters, see: (a) Rao, C. N. R.; Kulkarni, G. U.; Thomas, P. J.; Edwards, P. P. *Chem. Soc. Rev.* **2000**, *29*, 27. (b) Mohamed, M. B.; Volkov, V.; Link, S.; El-Sayed, M. A. *Chem. Phys. Lett.* **2000**, *317*, 517. (c) Huang, T.; Murray, R. W. *J. Phys. Chem. B* **2001**, *105*, 12498. (d) Link, S.; Beeby, A.; FitzGerald, S.; El-Sayed, M. A.; Schaaff, T. G.; Whetten, R. L. *J. Phys. Chem. B* **2002**, *106*, 3410. (e) Empedocles, S.; Bawendi, M. *Acc. Chem. Res.* **1999**, *32*, 389. (f) El-Sayed, M. A. *Acc. Chem. Res.* **2001**, *34*, 257. (g) Zheng, J.; Petty, J. T.; Dickson, R. M. *J. Am. Chem. Soc.* **2003**, *125*, 7780. (h)Schaaff, T. G.; Shafigullin, M. N.; Khoury, J. T.; Vezmar, I.; Whetten, R. L.; Cullen, W. G.; First, P. N.; Gutierrez-Wing, C.; Ascensio, J.; Jose-Yacaman, M. J. *J. Phys. Chem. B* **1997**, *101*, 7885.

[23] (a) Synthesis of the Iridium complex is reported in: Finke, R. G.; Lyon, D. K.; Nomiya, K.; Sur, S.; Mizuno, N. *Inorg. Chem.* **1990**, *29*, 1784. (b) Aiken, J. D.; Lin, Y.; Finke, R. G. *J. Mol. Catal. A* **1996**, *114*, 29.

[24] For a detailed discussion of the mechanistic steps, see: Besson, C.; Finney, E. E.; Finke, R. G. *J. Am. Chem. Soc.* **2005**, *127*, 8179, and references therein.

[25] For a thorough recent review on nanostructural growth *via* coprecipitation of multiple species (and ways to synthesize/stabilize 0D nanostructures), consult: Cushing, B. L.; Kolesnichenko, V. L.; O'Connor, C. J. *Chem. Rev.* **2004**, *104*, 3893.

[26] A derivation and full explanation of cluster "magic numbers" is given by: Teo, B. K.; Sloane, N. J. A. *Inorg. Chem.* **1985**, *24*, 4545.

[27] Finke, R. G. in *Metal Nanoparticles: Synthesis, Characterization, and Applications*, Feldheim, D. L.; Foss, C. A. eds., Dekker: New York, 2002. Crooks and coworkers determined that a closed-shell metallic nanocluster of Au_{55} has a diameter of 1.2 nm: Kim, Y.-G.; Oh, S.-K.; Crooks, R. M. *Chem. Mater.* **2004**, *16*, 167.

[28] Kroto, H. W.; Heath, J. R.; O'Brien, S. C.; Curl, R. F.; Smalley, R. E. *Nature* **1985**, *318*, 162.

[29] The irradiation of C_{60} with light in the presence of O_2 causes the formation of reactive singlet oxygen ($^1O_2$); for example, see Jensen, A. W.; Daniels, C. *J. Org. Chem.* **2003**, *68*, 207.

[30] Smalley and Curl named this structure after Buckminster Fuller, for his discovery of geodesic domes.

[31] For an interesting book on the history of other serendipitous discoveries in science, see Roberts, R. M. *Serendipity: Accidental Discoveries in Science*, Wiley: New York, 1989.

[32] Kroto, H. *Nanotechnology* **1992**, *3*, 111. A lecture given in the same title is also available as an audio file from http://www.learnoutloud.com/Catalog/Science/Scientists/C60-The-Celestial-Sphere-That-Fell-to-Earth/15201

[33] (a) Kriitschmer. W.; Lamb. L. D.; Fostiropoulos, K.; Huffman. D. R. *Nature* **1990**, *347*, 354. (b) Kratschmer, W.; Fostiropoulos, K.; Huffman, D. R. *Chem. Phys. Lett.* **1990**, *170*, 167. (c) http://www.mercorp.com/mercorp/products1.htm; a company that manages rights to fullerene production technology.

[34] Manolopoulos, D. E. *Chem. Phys. Lett.* **1992**, *192*, 330.

[35] Sitharaman, B.; Bolskar, R. D.; Rusakova, I.; Wilson, L. J. *Nano Lett.* **2004**, 4, 2373.

[36] Tanigaki, K.; Ebbesen, T. W.; Saito, S.; Mizuki, J.; Tsai, J. S.; Kubo, Y.; Kuroshima, S. *Nature* **1991**, *352*, 222.

[37] Zakharian, T. Y.; Seryshev, A.; Sitharaman, B.; Gilbert, B. E.; Knight, V.; Wilson, L. J. *J. Am. Chem. Soc.* **2005**, *127*, 12508.

[38] Ewels, C. P. *Nano Lett.* **2006**, *6*, 890.

[39] Smalley, R. E. *Acc. Chem. Res.* **1992**, *25*, 98.

[40] Heath, J. R. *ACS Symp. Ser.* **1992**, *481*, 1.

[41] (a) V. Z. Mordkovich, V. Z.; Umnov, A. G.; Inoshita, T.; Endo, M. *Carbon* **1999**, *37*, 1855. (b) Mordkovich, V. Z. *Chem. Mater.* **2000**, *12*, 2813.

[42] Sygula, A.; Rabideau, P. W. *J. Am. Chem. Soc.* **1999**, *121*, 7800.

[43] Gluch, K.; Feil, S.; Matt-Laubner, S. M.; Echt, O.; Scheier, P.; Mark, T. D. *J. Phys. Chem. A* **2004**, *108*, 6990.

[44] Wang, C.-R.; Shi, Z.-Q.; Wan, L.-J.; Lu, X.; Dunsch, L.; Shu, C.-Y.; Tang, Y.-L.; Shinohara, H. *J. Am. Chem. Soc.*, **2006**, *128*, 6605.

[45] Gan, L.-H.; Wang, C.-R. *J. Phys. Chem. A* **2005**, *109*, 3980.

[46] (a) Yonezawa, T.; Onoue, S.-Y.; Kimizuka, N. *Langmuir* **2000**, *16*, 5218. (b) Yonezawa, T.; Onoue, S.-Y.; Kimizuka, N. *Chem. Lett.* **2002**, 528.

[47] Crooks, R. M.; Zhao, M.; Sun, L.; Chechik, V.; Yeung, L. K. *Acc. Chem. Res.* **2001**, *34*, 181, and references therein. The first precedent for the use of poly(propylene imine) (PPI) dendrimers is: Floriano, P. N.; Noble, C. O.; Schoonmaker, J. M.; Poliakoff, E. D.; McCarley, R. L. *J. Am. Chem. Soc.* **2001**, *123*, 10545. This also contains many useful references for early precedents for metal@PAMAM nanocomposites.

[48] For an example of trimetallic nanoparticle synthesis (using a nondendritic host), see: Henglein, A. *J. Phys. Chem. B* **2000**, *104*, 6683.

[49] Schaak, R. E.; Sra, A. K.; Leonard, B. M.; Cable, R. E.; Bauer, J. C.; Han, Y.-F.; Means, J.; Teizer, W.; Vasquez, Y.; Funck, E. S. *J. Am. Chem. Soc.* **2005**, *127*, 3506.

[50] (a) Garcia-Martinez, J. C.; Crooks, R. M. *J. Am. Chem. Soc.* **2004**, *126*, 16170–16178. (b) Garcia-Martinez, J. C.; Scott, R. W. J.; Crooks, R. M. *J. Am. Chem. Soc.* **2003**, *125*, 11190–11191. (c) Kim, Y.-G.; Garcia-Martinez, J. C.; Crooks, R. M. *Langmuir* **2005**, *21*, 5485–5491.

[51] (a) Kamata, K.; Lu, Y.; Xia, Y. *J. Am. Chem. Soc.* **2003**, *125*, 2384. (b) Marinakos, S. M.; Shultz, D. A.; Feldheim, D. L. *Adv. Mater.* **1999**, *11*, 34. (c) Chah, S.; Fendler, J. H.; Yi, J. *J. Colloid Interface Sci.* **2002**, *250*, 142. (d) Marinakos, S. M.; Novak, J. P.; Brousseau, L. C., III; House, A. B.; Edeki, E. M.; Feldhaus, J. C.; Feldheim, D. L. *J. Am. Chem. Soc.* **1999**, *121*, 8518.

[52] Chen, M.; Gao, L. *Inorg. Chem.* **2006**, *45*, 5145.

[53] Yin, Y.; Rioux, R. M.; Erdonmez, C. K.; Hughes, S.; Somorjai, G. A.; Alivisatos, A. P. *Science* **2004**, *304*, 711.

[54] Sun, S.; Zeng, H. *J. Am. Chem. Soc.* **2002**, *124*, 8204. Other examples of solution-phase growth of oxide, and other compound 0D nanostructures (including quantum dots) are: (a) Strable, E.; Bulte, J. W. M.; Moskowitz, B.; Vivekanandan, K.; Allen, M.; Douglas, T.*Chem. Mater.* **2001**, *13*, 2201. (b) Frankamp, B. L.; Boal, A. K.; Tuominen, M. T.; Rotello, V. M. *J. Am. Chem. Soc.* **2005**, *127*, 9731. (c) Lemon, B. I.; Crooks, R. M. *J. Am. Chem. Soc.* **2000**, *122*, 12886. (d) Hanus, L. H.; Sooklal, K.; Murphy, C. J.; Ploehn, H. J. *Langmuir* **2000**, *16*, 2621.

[55] Juttukonda, V.; Paddock, R. L.; Raymond, J. E.; Denomme, D.; Richardson, A. E.; Slusher, L. E.; Fahlman, B. D. *J. Am. Chem. Soc.* **2006**, *128*, 420.

[56] It should be noted that in addition to solution-phase methods, quantum dots are frequently synthesized using molecular-beam epitaxy or other vapor-phase technique. For example, see: Wang, X. Y.; Ma, W. Q.; Zhang, J. Y.; Salamo, G. J.; Xiao, M.; Shih, C. K. *Nano Lett.* **2005**, *5*, 1873, and references therein.

[57] (a) Decher, G.; Hong, J. D. *Makromol. Chem. Macromol. Symp.* **1991**, *46*, 321. (b) Decher, G.; Hong, J. D.; Schmitt, J. *Thin Solid Films* **1992**, *210*, 831. For a recent review of electrostatic LbL growth, see: Hammond, P. T. *Adv. Mater.* **2004**, *16*, 1271.

[58] Adamantyl groups were used on the periphery of the dendrimers since they strongly interact with cyclodextrins. For example, see: Rekharsky, M. V.; Inoue, Y. *Chem. Rev.* **1998**, *98*, 1880–1901.

[59] The difference between nanocars and nanotrucks has been described as the former is only able to transport itself, whereas a nanotruck is able to accommodate a load.

[60] (a) Shirai, Y.; Osgood, A. J.; Zhao, Y.; Kelly, K. F.; Tour, J. M. *Nano Lett.* **2005**, *5*, 2330. (b) Shirai, Y.; Osgood, A. J.; Zhao, Y.; Yao, Y.; Saudan, L.; Yang, H.; Chiu, Y.-H.; Alemany, L. B.; Sasaki, T.; Morin, J.-F.; Guerrero, J. M.; Kelly, K. F.; Tour, J. M. *J. Am. Chem. Soc.* **2006**, *126*, 4854.

[61] (a) Iijima, S. *Nature* **1991**, *354*, 56 (first report of MWNTs). (b)Iijima, S. *Nature* **1993**, *363*, 603 (SWNT co-precedent). (c) Bethune, D. S.; Kiang, C. H.; Devries, M. S.; Gorman, G.; Savoy, R.; Vazquez, J.; Beyers, R. *Nature* **1993**, *363*, 605 (SWNT co-precedent).

[62] The term *graphene* designates a single layer of carbon atoms packed into hexagonal units. Though this structure is used to describe properties of many carbonaceous materials (*e.g.*, CNTs, graphite, fullerenes, *etc.*), this planar structure is thermodynamically unstable relative to curved structures such as fullerenes, nanotubes, and other structures found in carbon soot. As such, the isolation of single graphene sheets has only recently been reported through exfoliation from a high purity graphite crystal: Novoselov, K. S.; Geim, A. K.; Morozov, S. V.; Jiang, D.; Zhang, Y.; Dubonos, S. V.; Grigorieva, I. V.; Firsov, A. A. *Science* **2004**, *306*, 666.

[63] (a) The SEM image (low-resolution and high-resolution) of 9,10-antraquinone nanorods is reproduced with permission from (copyright 2004 American Chemical Society): Liu, H.; Li, Y.; Xiao, S.; Li, H.; Jiang, L.; Zhu, D.; Xiang, B.; Chen, Y.; Yu, D. *J. Phys. Chem. B* **2004**, *108*, 7744. (b) The SEM image of GaP–GaAs nanowires is reproduced with permission from (copyright 2006 American Chemical Society): Verheijen, M. A.; Immink, G.; de Smet, T.; Borgstrom, M. T.; Bakkers, E. P. A. M. *J. Am. Chem. Soc.* **2006**, *128*, 1353. (c) The SEM image of carbon nanotubes is reproduced with permission from (copyright 2001 American Chemical Society): Chiang, I. W.; Brinson, B. E.; Smalley, R. E.; Margrave, J. L.; Hauge, R. H. *J. Phys. Chem. B* **2001**, *105*, 1157. (d) The SEM image of TiO_2 nanofibers is reproduced with permission from (copyright 2006 American Chemical Society): Ostermann, R.; Li, D.; Yin, Y.; McCann, J. T.; Xia, Y. *Nano Lett.* **2006**, 6, 1297.

[64] (a) The HRTEM image of V_2O_5 nanorods on TiO_2 nanofibers is reproduced with permission from Reference [54d]. (b) The HRTEM image of GaP–GaAs nanowires is reproduced with permission from reference 54b. (c) The HRTEM image of multiwall carbon nanotubes is reproduced with permission from (copyright 2004 American Chemical Society): Lee, D. C.; Mikulec, F. V.; Korgel, B. A. *J. Am. Chem. Soc.* **2004**, *126*, 4951.

[65] For extensive reviews of molecular electronics see: (a) Tour, J. M. *Molecular Electronics: Commercial Insights, Chemistry, Devices, Architecture and Programming*; World Scientific: River Edge, NJ, 2003. (b) Tour, J. M.; James, D. K. in *Handbook of Nanoscience, Engineering and Technology*; Goddard, W. A., III; Brenner, D. W.; Lyshevski, S. E.; Iafrate, G. J. eds.,;RC: New York, 2003; pp. 4.1–4.28. (c) Tour, J. M. *Acc. Chem. Res.* **2000**, *33*, 791.

[66] Field emission results from the tunneling of electrons from a metal tip into a vacuum, under an applied strong electric field (Chapter 7 will have more details on this phenomenon, and how it is exploited for high-resolution electron microscopy).

[67] (a) Avouris, P. *Acc. Chem. Res.* **2002**, *35*, 1026. (b) Wind, S. J.; Appenzeller, J.; Martel, R.; Derycke, V.; Avouris, P. *Appl Phys. Lett.* **2002**, *80*, 3817. A recent strategy for the bottom-up design of CNT interconnects: Li, J.; Ye, Q.; Cassel, A.; Ng, H. T.; Stevens, R.; Han, J.; Meyyappan, M. *Appl. Phys. Lett.* **2003**, *82*, 2491.

[68] Yakabson, B. I. *Appl. Phys. Lett.* **1998**, *72*, 918.

[69] Micro-Raman spectroscopy has shown that during tension, only the outer layers of MWNTs are loaded, whereas during compression, the load is transferred to all layers.

[70] Salvetat, J.-P.; Briggs, G. A. D.; Bonard, J.-M.; Basca, R. R.; Kulik, A. J.; Stöckli, T.; Burnham, N. A.; Forró, L. *Phys. Rev. Lett.* **1999**, *82*, 944.

[71] For a recent review, see: CNT stabilized polymers.

[72] http://en.wikipedia.org/wiki/Specific_strength. For a very nice summary of specific stiffness/specific strength regions for various materials classes see: http://www-materials.eng.cam.ac.uk/mpsite/interactive_charts/spec-spec/basic.html

[73] Sun, J.; Gao, L.; Li, W. *Chem. Mater.* **2002**, *14*, 5169.

[74] For a nice review regarding defect sites in CNTs, see Charlier, J.-C. *Acc. Chem. Res.* **2002**, *35*, 1063.

[75] For a thorough recent review of the surface chemistry (noncovalent and covalent) of CNTs, see Tasis, D.; Tagmatarchis, N.; Bianco, A.; Prato, M. *Chem. Rev.* **2006**, *106*, 1105.

[76] A interesting recent precedent related to the *reversibly* tunable exfoliation of SWNTs using poly(acrylic acid) at varying pH levels is reported by Grunlan, J. C.; Liu, L.; Kim, Y. S. *Nano Lett.* **2006**, 6, 911.

[77] Knupfer, M.; Reibold, M.; Bauer, H.-D.; Dunsch, L.; Golden, M. S.; Haddon, R. C.; Scuseria, G. E.; Smalley, R. E. *Chem. Phys. Lett.* **1997**, *272*, 38.

[78] Smalley, R. E. *Discovering the Fullerenes*, Nobel Lecture, 1996. May be found online at: http://nobelprize.org/chemistry/laureates/1996/smalley-lecture.pdf (along with the Nobel lectures from Curl and Kroto).

[79] It should be noted that the National Institute of Standards and Technology (NIST) has been recently focused on the development of standard synthesis, purification, and characterization techniques for CNTs. To date, there are a number of competing methods for SWNTs/MWNTs – all citing percent purity values that appear rather arbitrary. Indeed, purchasing a "90% pure SWNT" sample from multiple vendors will result in very different products! In order to continue the rapid progress in CNT synthesis/applications, it is essential that we set up a "gold standard" for CNTs that will immediately tell us what a certain purity level means. That is, if a "60% purity" value is cited, clarifying what the remaining 40% consists of (amorphous carbon, remaining catalytic metal, other nanotube diameters/morphologies, *etc.*)

[80] Dai, H. *Acc. Chem. Res.* **2002**, *35*, 1035.

[81] Rao, C. N. R.; Govindaraj, A. *Acc. Chem. Res.* **2002**, *35*, 998, and references therein. For recent information regarding the role of alumina on the yield/morphology of supported CNT catalysts, see: Jodin, L.; Dupuis, A.-C.; Rouviere, E.; Reiss, P. *J. Phys. Chem. B* **2006**, *110*, 7328. It should be noted that the supported nanoclusters may reside within nanochannels to facilitate 1D growth, examples of these methods, which include both template and "closed space sublimation" (CSS) are (and references therein): (a) Li, J.; Papadopoulos, C.; Xu, J. M.; Moskovits, M. *Appl. Phys. Lett.* **1999**, *75*, 367. (b) Kyotani, T.; Tsai, L. F.; Tomita, A. *Chem. Mater.* **1996**, 8, 2109. (c) Hu, Z. D.; Hu, Y. F.; Chen, Q.; Duan, X. F.; Peng, L.-M. *J. Phys. Chem. B* **2006**, *110*, 8263.

[82] Choi, H. C.; Kim, W.; Wang, D.; Dai, H. *J. Phys. Chem. B* **2002**, *106*, 12361. The first precedent for SWNT growth from gold nanoclusters has been recently reported: Bhaviripudi, S.; Mile, E.; Steiner, S. A.; Zare, A. T.; Dresselhaus, M. S.; Belcher, A. M.; Kong, J. *J. Am. Chem. Soc.* **2007**, *129*, 1516.

[83] The first precedent for (n,m) control of SWNT growth is: Lolli, G.; Zhang, L.; Balzano, L.; Sakulchaicharoen, N.; Tan, Y.; Resasco, D. E. *J. Phys. Chem. B* **2006**, *110*, 2108.

[84] Resasco, D. E.; Alvarez, W. E.; Pompeo, F.; Balzano, L.; Herrera, J. E.; Kitiyanan, B.; Borgna, A. *J. Nanopart. Res.* **2002**, 4, 131.

[85] Nikolaev, P.; Bronikowski, M. J.; Bradley, R. K.; Rohmund, F.; Colbert, D. T.; Smith, K. A.; Smalley, R. E. *Chem. Phys. Lett.* **1999**, *313*, 91. It should be noted that $Fe(CO)_5$ is not the only system in which the precursor acts as the metal catalyst and carbon source. A number of metallocenes

(*e.g.*, ferrocene, cobaltocene, and nickelocene) have also been used; however, they typically result in MWNT growth rather than SWNTs. This is most likely due to the larger number of carbon atoms from cyclopentadienyl groups that must self-assemble, relative to smaller carbon precursors (*e.g.*, CH_4, C_2H_2, *etc.*) used for SWNT growth.

86 Wagner, R. S.; Ellis, W. C. *Appl. Phys. Lett.* **1964**, 4, 89. For a recent review of the solid–liquid–solid (SLS) and supercritical fluid–liquid–solid (SFLS) mechanisms for semiconductor nanowire growth, see: Wang, F.; Dong, A.; Sun, J.; Tang, R.; Yu, H.; Buhro, W. E. *Inorg. Chem.* **2006**, *45*, 7511. A recent precedent for the epitaxial growth of ZnO nanowires at the junction of nanowalls: Ng, H. T.; Li, J.; Smith, M. K.; Nguyen, P.; Cassell, A.; Han, J.; Meyyappan, M. *Science* **2003**, *300*, 1249.

87 The word "generally" is used, since there are also reports of nanowire growth at temperatures below the eutectic. For example, see: Adhikari, H.; Marshall, A. F.; Chidsey, E. D.; McIntyre, P. C. *Nano Lett.* **2006**, 6, 318.

88 Cantoro, M.; Hofmann, S.; Pisana, S.; Scardaci, V.; Parvez, A.; Ducati, C.; Ferrari, A. C.; Blackburn, A. M.; Wang, K.-Y.; Robertson, J. *Nano Lett.* **2006**, 6, 1107.

89 Deng, W.-Q.; Xu, X.; Goddard, W. A. *Nano Lett.* **2004**, 4, 2331.

90 Graphite-encapsulated metal nanostructures are of increasing importance for magnetic applications such as high-density magnetic recording media; for example, see: Flahaut, E.; Agnoli, F.; Sloan, J.; O'Connor, C.; Green, M. L. H. *Chem. Mater.* **2002**, *14*, 2553, and references therein. Encapsulation dominates over CNT growth at low temperatures since the kinetic energy is not sufficient for graphitic islands to lift off the catalyst surface. Hence, encapsulation may easily be limited, which enhances CNT growth, by maintaining elevated temperatures. Experimental results also show that small catalyst nanoclusters (diameters <2 nm) are free of graphite encapsulation since they do not contain a sufficient number of dissolved C atoms. However, for metal nanostructures >3 nm in diameter, calculations suggest that graphite encapsulation is thermodynamically preferred over SWNT growth. This is confirmed by the empirical observation that SWNTs form only on catalyst particles with diameters <2 nm.

91 Lee, Y. H.; Kim, S. G.; Jund, P.; Tomanek, D. *Phys. Rev. Lett.* **1997**, *78*, 2393.

92 Hata, K.; Futaba, D. N.; Mizuno, K.; Namai, T.; Yumura, M.; Iijima, S. *Science* **2004**, *306*, 1362.

93 Rummeli, M. H.; Borowiak-Palen, E.; Gemming, T.; Pichler, T.; Knupfer, M.; Kalbac, M.; Dunsch, L.; Jost, O.; Silva, S. R. P.; Pompe, W.; Buchner, B. *Nano Lett.* **2005**, 5, 1209.

94 For a recent review of inorganic-based nanotubes, see: Goldberger, J.; Fan, R.; Yang, P. *Acc. Chem. Res.* **2006**, *39*, 239, and references therein.

95 The top VLS mechanism was predicted using molecular dynamics calculations. The image was reproduced with permission from Ding, F.; Bolton, K.; Rosen, A. *J. Phys. Chem. B* **2004**, *108*, 17369. The middle VLS mechanism shows both "root growth" (c–d) and "folded growth" (e–g). The image was reproduced with permission from Lee, D. C.; Mikulec, F. V.; Korgel, B. A. *J. Am. Chem. Soc.* **2004**, *126*, 4951. The bottom mechanism, predicted by quantum mechanics/molecular mechanics, is one of the rare examples of an atomic-level picture of CNT growth. The image was reproduced with permission from Deng, W.-Q.; Xu, X.; Goddard, W. A. *Nano Lett.* **2004**, 4, 2331.

96 http://www.asml.com – a recent press release indicates that EUV lithography will likely be implemented for high volume production by 2009, with feature sizes well below 32 nm.

97 In order to reduce the adhesion between a polymeric mold and a silicon/quartz master, the master surface is typically modified with a fluorosilane (*e.g.*, $CF_3(CF_2)_6(CH_2)_2SiCl_{3(g)}$). In addition, the final removal of the mold may also be carried out in the presence of a liquid with a low viscosity such as methanol (solvent-assisted micromolding (SAMIM)).

98 A recent thorough review of nanofabrication using both hard and soft molds, as well as other forms of soft lithography, see: Gates, B. D.; Xu, Q.; Stewart, M.; Ryan, D.; Willson, C. G.; Whitesides, G. M. *Chem. Rev.* **2005**, *105*, 1171.

99 For example, see: Chou, S. Y.; Krauss, P. R.; Renstrom, P. J. *Science* **1996**, *272*, 85.

100 For example, see: Jackman, R. J.; Wilbur, J. L.; Whitesides, G. M. *Science* **1995**, *269*, 664.

101 Im, J.; Kang, J.; Lee, M.; Kim, B.; Hong, S. *J. Phys. Chem. B* **2006**, *110*, 12839.

102 Myung, S.; Lee, M.; Kim, G. T.; Ha, J. S.; Hong, S. *Adv. Mater.* **2005**, *17*, 2361.

103 (a) Odom, T. W.; Thalladi, V. R.; Love, J. C.; Whitesides, G. M. *J. Am. Chem. Soc.* **2002**, *124*, 12112. (b) Odom, T. W.; Love, J. C.; Wolfe, D. B.; Paul, K. E.; Whitesides, G. M. *Langmuir* **2002**, *18*, 5314.

[104] Li, H.-W.; Muir, B. V. O.; Fichet, G.; Huck, W. T. S. *Langmuir* **2003**, *19*, 1963.
[105] Steward, A.; Toca-Herrera, J. L.; Clarke, J. *Protein Sci.* **2002**, *11*, 2179.
[106] Gates, B. D.; Whitesides, G. M. *J. Am. Chem. Soc.* **2003**, *125*, 14986.
[107] We will discuss the operating principle of atomic force microscopy (AFM) and other scanning force microscopies in more detail in Chapter 7. At this point, simply think of this technique as analogous to an antiquated record player, in which the needle gently touches the surface of the record to produce music. Similarly, the AFM tip either gently taps, or hovers immediately above, the surface of a planar substrate.
[108] (a) Piner, R. D.; Zhu, J.; Xu, F.; Hong, S.; Mirkin, C. A. *Science* **1999**, *283*, 661. (b) Hong, S.; Zhu, J.; Mirkin, C. A. *Science* **1999**, *286*, 523. (c) Hong, S.; Mirkin, C. A. *Science* **2000**, *288*, 1808.
[109] A very nice review of the various DPN methodologies is provided by: Ozin, G. A.; Arsenault, A. C. *Nanochemistry: A Chemical Approach to Nanomaterials*, RSC: Cambridge, UK, 2005, pp. 137–157.
[110] Zaumseil, J.; Meitl, M. A.; Hsu, J. W. P.; Acharya, B. R.; Baldwin, K. W.; Loo, Y.-L.; Roger, J. A. *Nano Lett.* **2003**, 3, 1223.
[111] Ginger, D. S.; Zhang, H.; Mirkin, C. A. *Angew. Chem. Int. Ed.* **2004**, *43*, 30. This review contains all of the references for the various experimental conditions.
[112] Yang, Y. T.; Callegari, C.; Feng, X. L.; Ekinci, K. L.; Roukes, M. L. *Nano Lett.* **2006**, 6, 583.
[113] Collin, J.-P.; Dietrich-Buchecker, C.; Gavina, P.; Jimenez-Molero, M.; Sauvage, J.-P. *Acc. Chem. Res.* **2001**, *34*, 477.
[114] Geeves, M. A. *Nature* **2002**, *415*, 129.
[115] Liu, Y.; Flood, A. H.; Bonvallet, P. A.; Vignon, S. A.; Northrop, B. H.; Tseng, H.-R.; Jeppesen, J. O.; Huang, T. J.; Brough, B.; Baller, M.; Magonov, S.; Solares, S. D.; Goddard, W. A.; Ho, C.-M.; Stoddart, J. F. *J. Am. Chem. Soc.* **2005**, *127*, 9745.

Topics for Further Discussion

1. From your knowledge of semiconductor and metallic 0D nanostructures, think about how you would design a coating that would sense the surrounding wall colors of walls and adapt its color to match (*i.e.*, color-adapting furniture!).
2. Mercedes-Benz and other manufacturers feature scratch-resistant clear coats as standard on new vehicles. What are these coatings comprised of, and how does this prevent surface scratching?
3. Carbon nanotubes have been touted as being useful to store large amounts of hydrogen gas for fuel cell applications. From the literature, how is this contained – within the interior or adsorbed along the sidewall?
4. What are some examples of "self-cleaning" coatings? How do these work?
5. What factors govern the tilt angle (between the substrate and alkyl chain) of a SAM on a gold or silver surface?
6. The growth mechanism of carbonaceous nanostructural materials is generally thought to be *via* VLS. Cite some recent examples of carbon nanotube/nanofiber growth at temperatures far below the melting point of the nanoparticulate catalyst species.
7. In the Finke four-step mechanism for nanocluster growth, explain why higher temperatures are most conducive for the growth of monodisperse nanoclusters.
8. Think of a new device that you could fabricate by DPN that would be comprised of both nanoclusters and nanotubes/nanowires. What are some potential applications for this device?
9. How would you synthesize free-standing nanorings using both the top-down and bottom-up approaches?
10. How would you synthesize metal oxide nanotubes, using a sacrificial template? Cite any related precedents from the literature.
11. What are some examples of NEMS devices currently under development?
12. What are the major developmental efforts underway to power nanodevices?
13. How could you deposit a square grid of TiO_2 nanowires with a controllable spacing between adjacent nanowires?
14. Describe how LbL may be used to deposit coatings onto complex, nonplanar substrates. Cite examples for this strategy from the literature.

15. Cite examples of materials designs where *both* "top-down" and "bottom-up" approaches were used.
16. There are recent reports that N-doped CNTs are less toxic than SWNTs or MWNTs (*e.g.*, *Nano Lett.* **2006**, *6*, 1609). Provide some likely rationales for the varying toxicological effects for these doped nanostructures.

Further Reading

1. http://www.nanohub.org (free registration)
2. Ozin, G. A.; Arsenault, A. C. *Nanochemistry: A Chemical Approach to Nanomaterials*, Royal Society of Chemistry Publishing: Cambridge, U.K., 2005.
3. Dresselhaus, M. S.; Dresselhaus, G.; Eklund, P. C. *Science of Fullerenes and Carbon Nanotubes*, Academic: New York, 1996.
4. *Dendrimers and Dendritic Polymers*, Frechet, J. M. J.; Tomalia, D. A. eds., Wiley: New York, 2001.
5. *Advanced Materials* **2004**, *16*, 15. Special issue dedicated to George Whitesides.
6. *MRS Bulletin* **2006**, *31*, 5 – May 2006. Special issue on materials for magnetic data storage.
7. *MRS Bulletin* **2005**, *30*, 12 – December, 2005. Special issue on fabrication of sub-45-nm structures for the next generation of devices.
8. Madou, M. J. *Fundamentals of Microfabrication*, 2nd ed., CRC: Boca Raton, 2002.
9. *Alternative Lithography – Unleashing the Potentials of Nanotechnology*, Sotomayor-Torres, C. M. ed., Kluwer/Plenum: New York, 2003.
10. Wilson, M.; Kannangara, K.; Smith, G.; Simmons, M. *Nanotechnology: Basic Science and Emerging Technologies*, CRC: Boca Raton, 2002.
11. *Carbon Nanotubes*, O'Connell, M. J. ed., CRC: Boca Raton, 2006.
12. *Carbon Nanotubes: Synthesis, Structure, Properties and Applications*, Smalley, R. E.; Dresselhaus, M. S.; Dresselhaus, G.; Avouris, P. eds., Springer: New York, 2001.
13. D. Tomanek. http://www.pa.msu.edu/cmp/csc/nanotube.html. The Nanotube Site (Internet reference).
14. Lyshevski, S. E.; Lyshevski, S. E. *MEMS and NEMS: Systems, Devices, and Structures*, CRC: Boca Raton, 2002.
15. *Nanoparticles: From Theory to Application*, Schmid, G. ed., Wiley: New York, 2004.
16. Rotello, V. *Nanoparticles: Building Blocks for Nanotechnology (Nanostructure Science and Technology)*, Springer: New York, 2004.
17. *The Chemistry of Nanomaterials: Synthesis, Properties, and Applications*, Rao, C. N. R.; Muller, A.; Cheetham, A. K. eds., Wiley-VCH: Berlin, 2004.
18. Wolf, E. L. *Nanophysics and Nanotechnology*, Wiley-VCH: Berlin, 2004.

CHAPTER 7

MATERIALS CHARACTERIZATION

Thus far, we have focused on the relationship between the structure of a material and its properties/applications. However, we have not yet focused on how one is able to determine the structure and composition of materials. That is, when a material is fabricated in the lab, how are we able to assess whether our method was successful? Depending on the nature of the material being investigated, a suite of techniques may be utilized to assess its structure and properties. Whereas some techniques are qualitative, such as providing an image of a surface, others yield quantitative information such as the relative concentrations of atoms that comprise the material. Recent technological advances have allowed materials scientists to accomplish something that was once thought to be impossible: to obtain actual two-dimensional/three-dimensional images of atomic positions in a solid, in real time. It should be noted that the sensitivity of quantitiative techniques also continues to be improved, with techniques now being able to easily measure parts per trillion (ppt) concentrations of impurities in a bulk sample.

This chapter will focus on the most effective and widely used techniques available to characterize solid-state compounds. The primary objective of this chapter is to provide a practical description of the methods used to characterize a broad range of materials. Rather than focusing on the theoretical aspects of each technique, which may be found in many other textbooks (see "Further Reading" section), our treatment will focus on method suitabilities, sample preparation, and anticipated results. In this manner, you will be well informed regarding the best method to use for a particular material. Since techniques such as solution-phase nuclear magnetic resonance (NMR) and infrared spectroscopy (IR) are used throughout undergraduate courses, the background of these methods will not be provided in this textbook. Likewise, it is beyond the scope of this chapter to provide detailed background in optics, electronics, and physical chemistry concepts that underly most of the techniques described herein. For this information, the reader is referred to the "Further Reading" section at the end of this chapter.

7.1. OPTICAL MICROSCOPY

Of the many techniques available for the analysis of solid materials, perhaps the simplest is optical microscopy. Two modes of optical microscopy are typically

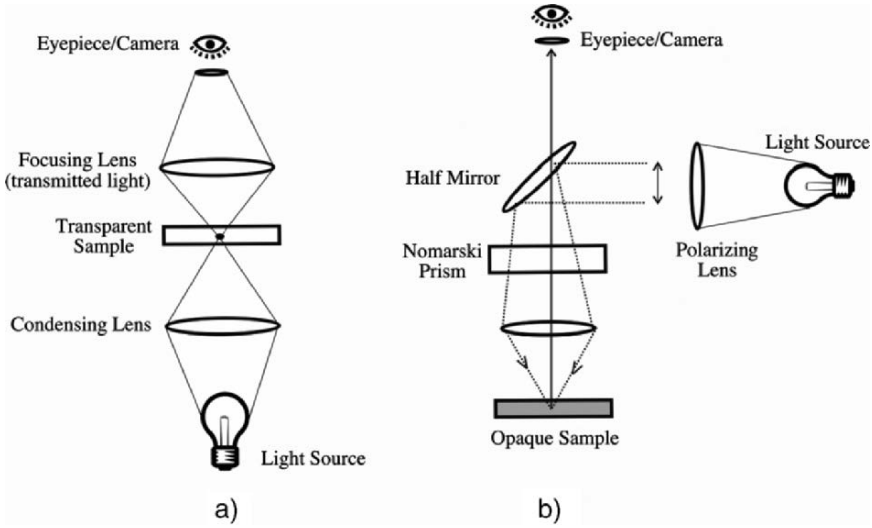


Figure 7.1. Schematic of the components of (a) transmitted light optical microscope and (b) reflected light optical microscope.

employed, based on the measurement of transmitted or reflected light, from a transparent or opaque sample, respectively (Figure 7.1). Often, a microscope is fitted with both modes, allowing one to analyze both types of samples. Most of the solid-state materials discussed thus far are nontransparent in their as-grown/as-deposited states. Further, it is usually difficult to prepare thin cross sections for transmission microscopy. Hence, materials scientists typically employ the reflection mode, also known as episcopic light differential interference contrast (DIC) microscopy.[1] This technique is useful for imaging of a variety of reflective samples including minerals, metals, semiconductors, glasses, polymers, and composites. The semiconductor industry relies heavily on reflective DIC imaging for quality assessment of computer chip components.

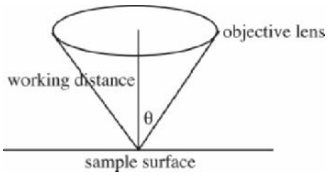
Surface artifacts such as depressions and particulates create optical path differences in the reflected beam. Formation of the final image is the result of interference between two distinct wave fronts that reach the image plane slightly out of phase with each other. Unlike the situation with transmitted light and semitransparent phase specimens, the image created in reflected light DIC can often be interpreted as a true three-dimensional representation of the surface geometry, provided a clear distinction can be realized between raised and lowered regions in the specimen. Oftentimes, the reflected wave fronts emanating from the sample are separated by only fractions of a micrometer, which is much less than the resolution of the objective.

The ability to discern fine details within a magnified image is referred to as the *resolution* of a microscope. Since light is used as the illumination source in optical microscopy, the resolution is expressed in the same units as the wavelength of

light (nm). The theoretical resolution, R, of any optical system may be calculated using Abbe's equation (Eq. 1):

$$(1) \qquad R = \frac{0.61\lambda}{\eta(\sin\theta)},$$

where



The denominator of Eq. 1 represents the numerical aperature (NA) of the objective lens, related to its light gathering ability. Other primary factors that influence the resolution of a lens is the wavelength of light used, the index of refraction (η) of the environment surrounding the lens (*e.g.*, 1.00 for air), and the angle of illumination (θ).

Examining the above equation, one can see that a resolution limit will be reached using visible light (350–700 nm) as the illumination source. That is, using high numerical aperatures (*e.g.*, 1.3–1.4 for oil-immersion lenses where $\eta = 1.5$), the theoretical resolution using polychromatic visible light (*ca.* 500 nm) is on the order of 220 nm. Hence, any sample features that are less than 220 nm apart from one another will appear blurry. It should be noted that the calculated resolution represents the best possible cutoff for clear discernment of small features. In practice, the observed resolution is often worse than the theoretical value, depending on the degree of optical aberrations that are inherent in the lens.

It may be seen by Eq. 1 that higher spatial resolutions (*i.e.*, smaller R values) are possible through use of shorter wavelengths. To illustrate this concept, subsequent sections of this chapter will examine the high resolutions inherent in microscopes that use an electron beam rather than visible light. However, we first must ask ourselves whether it is possible to improve the resolution limits of optical microscopy. If this is possible, the cost of such a modification would be far less than the price of electron microscopes (currently \$600 K–\$1.5 M).

In the late 1920s, Edward H. Synge published a series of articles that conceptualized the idea of an "ultrahigh resolution" optical microscope.[2] His original idea proposed using a screen with an aperature of dimensions much smaller than wavelength of the illuminating source. Upon irradiating the screen with a high-intensity light source, the light is confined to the dimensions of the hole. If this hole is placed in close proximity (nanometer regime) to the sample surface, the light emerging from the aperture could be used to image a specimen before it had time to spread out. By comparison, the distance between the light source and sample in conventional optical microscopy is on the order of millimeters. Not unlike other visionary ideas ahead of

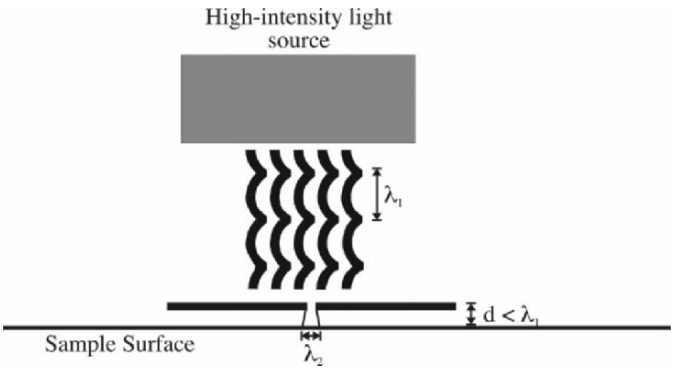


Figure 7.2. Schematic of operating principle of near-field scanning optical microscopy (NSOM). The resulting wavelength of light impinging on the sample (λ_2) has a wavelength much smaller than the illuminating source, resulting in much higher resolution.

their time, the technical difficulties associated with Synge's idea were only recently overcome at IBM in the mid-1980s, which marked the rediscovery of the technique now referred to as near-field scanning optical microscopy (NSOM, Figure 7.2).[3]

A variety of apertures have been used to deliver nanometer-sized spots of light. While early NSOM tips were fabricated out of etched quartz crystals and micropipettes, tapered optical fibers with tip diameters of *ca*. 100 nm are now typically used. A metallic thin film such as aluminum is usually applied around the sides of the tapered region of the NSOM tip to focus the light toward the sample. For optical fibers, the numerical aperature is related to the difference in the indices of refraction of the cladding and core (Eq. 2):

$$\text{NA} \propto \sqrt{(\eta_{\text{core}})^2 - (\eta_{\text{cladding}})^2}. \quad (2)$$

In order to obtain high-resolution optical images with NSOM, the tip must be stabilized within a few nanometers of the sample surface during scanning. Techniques to precisely position the tip above the sample were developed by researchers in 1992, resulting in more applications for both imaging and topographic profiling of surfaces. This latter capability will be discussed in more detail later in this chapter when discussing scanning probe microscopy (SPM).

7.2. ELECTRON MICROSCOPY

Although optical microscopy may be extended into the nanoregime, other techniques must be used to clearly discern components below 100 nm. Indeed, the current "nanotechnology revolution" that we are experiencing would not have been possible if there were not suitable techniques in order to characterize nanomaterials. As we saw in the previous section, in order to improve resolution, we must use source radiation with as small a wavelength as possible.

Louis DeBroglie was the first to assert that matter, like light, could be described as having both particle and wave characteristics (Eq. 3). However, the application of this equation for a macroscopic item such as a golf ball in flight,[4] predicts a wavelength that is too small to be meaningful (Eq. 4). By comparison, subatomic particles have wavelengths comparable to the size of an atom. In particular, electrons that are accelerated in a potential, V, have quantifiable wavelengths[5]; for example, electrons accelerated at 100 kV have a wavelength of 0.037 Å (Eq. 5). This is significantly shorter than the wavelengths of high-energy radiation sources such as X-rays (*e.g.*, 1.54 Å for Cu Kα X-rays), or neutrons (*ca.* 1 nm). Though neutron microscopes have recently been demonstrated,[6] X-ray microscopes are not possible due to the lack of suitable refracting lenses for X-rays. However, we will see later on that X-rays are instrumental for the chemical analysis of a surface:

$$\lambda = \frac{h}{m\upsilon}, \tag{3}$$

$$\frac{6.626 \times 10^{-34}\,\mathrm{J\,s}}{(0.045\,\mathrm{kg})(41\,\mathrm{m\,s^{-1}})} = 3.59 \times 10^{-34}\,\mathrm{m}, \tag{4}$$

$$\lambda = \frac{12.3\,\text{Å}}{\sqrt{V + \frac{V^2}{1\times10^6}}}. \tag{5}$$

The ability to focus the extremely small incident wavelengths of energetic electrons results in an unprecedented spatial resolution compared to optical based microscopy techniques (Figure 7.3). Since the resolution of our eyes is on the order of 0.1 mm, in order for us to observe these features, an appropriate level of magnification must be used. That is, for the increasingly greater resolution capabilities of modern scanning electron microscope (SEM) and transmission electron microscope (TEM) instruments, the sample must be magnified by *ca.* 100–300 K and 1.5–2 M times, respectively (Figure 7.4). However, it should be noted that although the wavelength of electrons are *ca.* 100,000 times smaller than visible light, the resolving power of electron microscopes are only on the order of 1,000 times greater than light microscopes, due to lens aberrations (especially spherical aberration).[7]

Images obtained from electron microscopy are due to the nature/degree of electron scattering from the constituent atoms of the sample. Table 7.1 provides a comparison between electron, X-ray, and neutron sources, pertaining to their utility

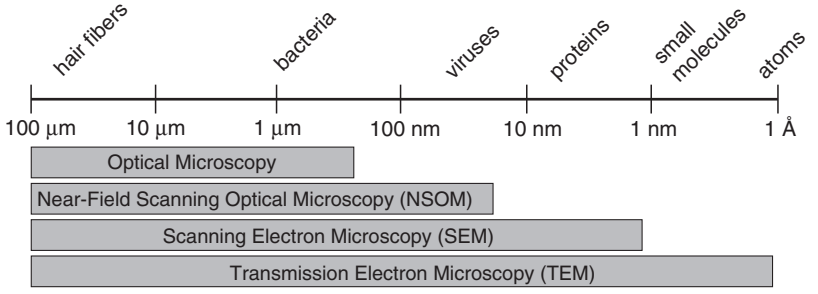


Figure 7.3. Comparison of the characterization size regimes for optical and electron microscopy.

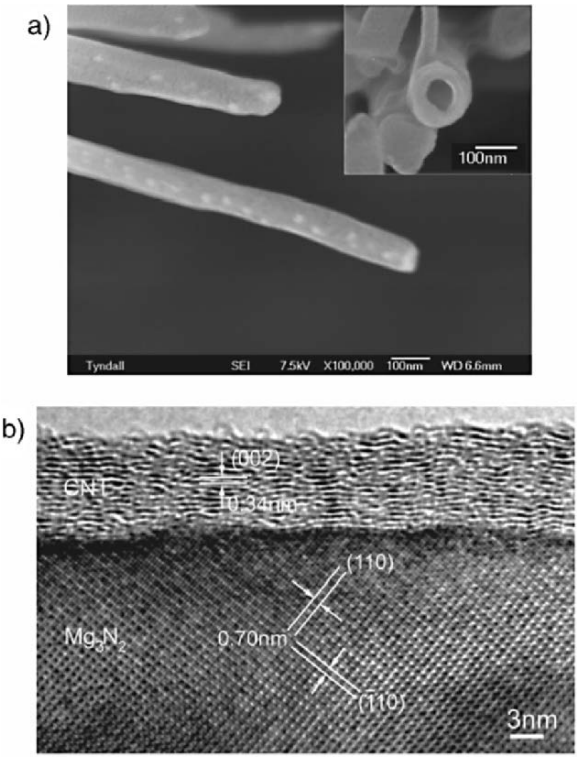


Figure 7.4. High-resolution (HR) electron micrographs. Shown is (a) an SEM image of vanadium oxide nanotubes at a magnification of 100,000×[8] and (b) a TEM image of the interface between a Mg_3N_2 nanowire and a graphitic carbon coating at a magnification of *ca.* 2,000,000×.[9]

Table 7.1. Comparative Characteristics of Energetic Radiation Sources

Source	Brightness[a] (particles cm^{-2} eV^{-1} $steradian^{-1}$)	Elastic Mean Free Path[b] (Å)	Absorption Length[c] (Å)	Minimum Probe Size[d] (Å)
Neutrons	1×10^{24}	1×10^{8}	1×10^{9}	1×10^{7}
X-rays	1×10^{26}	1×10^{4}	1×10^{6}	1×10^{3}
Electrons	1×10^{29}	1×10^{2}	1×10^{3}	1

Data from http://ncem.lbl.gov/team/team_background.htm
[a]Related to the number of particles and the range of incident angles that irradiate a sample.
[b]Related to the distance a particle travels in a solid before elastically colliding with constituent atoms.
[c]Related to the distance a particle travels in a solid before its energy is absorbed through interactions with constituent atoms and electrons.
[d]Related to the minimum diameter of the source that is used to generate the radiation.

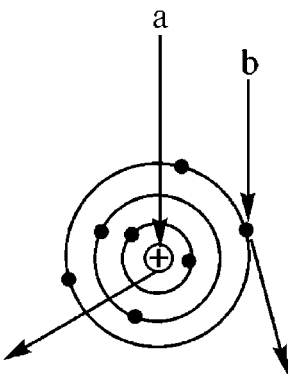


Figure 7.5. Schematic of specific interactions between the electron beam and sample atom. Shown are (a) an elastically scattered electron and (b) an inelastically scattered electron.

for scattering applications. Electrons are much less penetrating than X-rays, and are readily absorbed by air, dictating the use of ultrahigh vacuum (UHV) environments. Since the mass of an electron is 1,000 times smaller than the lightest nucleus, electrons are scattered much more intensely than X-rays or highly penetrating neutrons. This results in high sensitivity from the interaction of electrons with extremely small samples.

There are two types of scattering interactions that are possible when an electron beam impinges on atoms of the sample (Figure 7.5). Interaction between the electron source and atomic nuclei gives rise to *elastic scattering*, which results in a large-angle deflection of the electron beam with little or no energy loss. By comparison, electron–electron interactions between the source and electron clouds of individual atoms cause small-angle deflections with a significant loss of energy, known as *inelastic scattering*. It should be noted that there is a finite probability that the electrons may continue their original trajectory through the atomic structure without scattering. In particular, this is commonly observed with molecules containing light atoms such as H, Li, B, *etc*. As you would expect, a solid consisting of a high density of large atoms such as metallic solids would result in shorter elastic mean free paths, and stronger elastic scattering interactions with incident electrons. By comparison, organic solids such as polyethylene would primarily exhibit inelastic scattering with an electron source.

There are two instruments that are used for electron imaging: TEM and SEM. As the name implies, the transmission mode measures the intensity from an electron source after it has passed through a transparent sample (Figure 7.6a). In contrast, SEM features the scanning of an electron beam over selected regions of an opaque sample (Figure 7.6b).

The basic principles that govern electron microscopy are analogous to optical microscopy. Whereas optical microscopes use light and optical lenses to illuminate and magnify the sample, electron microscopes utilize high-energy electrons and electromagnetic lenses. There are two types of lenses in electron microscopes (see

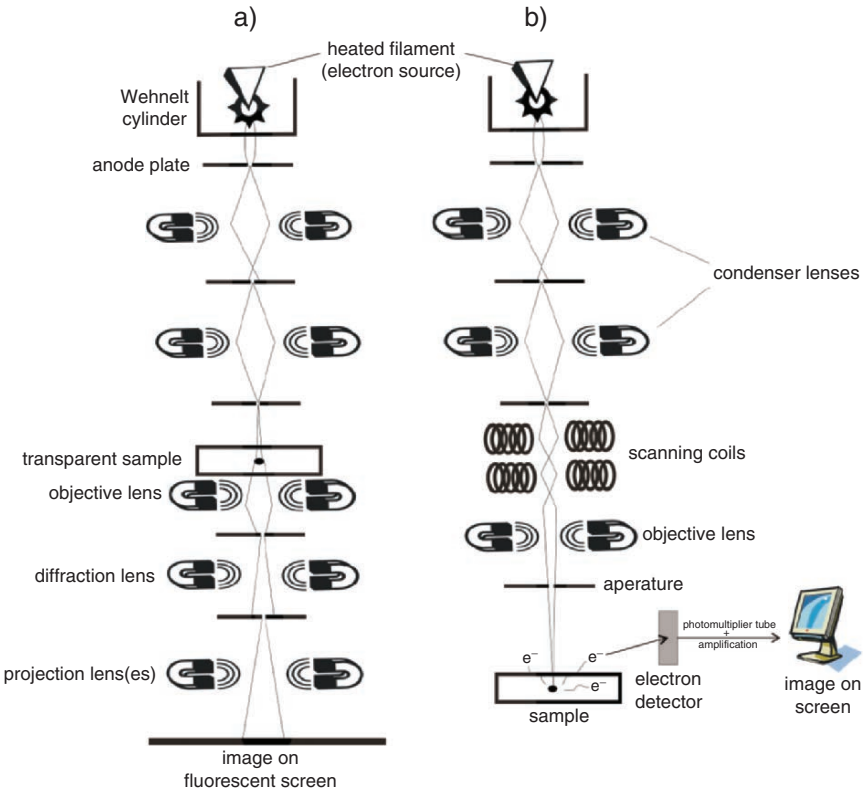


Figure 7.6. Schematic of the major components of (a) transmission electron microscope (TEM) and (b) scanning electron microscope (SEM). Note: stigmator coils are also associated with the lens(es) that reduce beam astigmatism, thereby improving image distortion and resolution. For TEM, the image is shown to be projected onto a viewing screen. This screen is coated with phosphorescent particles such as ZnS and CdS that emit visible light when struck by electrons (hard copies of the image may also be transferred to photographic plates *via* silver halide emulsions). However, modern TEM instruments do not require photographic development, as digital images are observed directly on computer monitors, in real time.

below); TEM has an additional lens, the projector lens, which is used to project an amplified copy of the image onto a screen:

(i) *Condenser* – used to control the illumination of the sample through concentration of the electron beam generated from the source.

(ii) *Objective* – used to magnify the sample (in TEM, the sample is inserted into the objective lens).

(iii) *Projector* – used to project an amplified copy of the image onto a screen (photographic plate, computer screen, fluorescent screen, CCD camera/monitor).

The heart of any electron microscope is the electron source, or gun. In fact, the electron gun within microscopes is identical to that operating within modern televisions. Behind the TV monitor lays a cathode ray tube (CRT), containing a set

of three tungsten filaments. The electrons generated from these filaments are accelerated toward an anode that is coated with phosphor particles, which emit either red, green, or blue light when struck with electrons. This light is focused onto individual pixels that comprise the monitor; if you look close at a television screen, you may notice the millions of tiny dots that comprise individual images. The picture we see on our monitors is a result of the electron guns being scanned sequentially from left to right in a fraction of a second, a process known as *rastering*. This is the exact mode that is operable in SEMs and scanning transmission electron microscopes (STEM), where the focused electron beam is allowed to scan across selected regions of a sample.

Electrons are generated by passing high currents through tungsten or crystalline LaB_6 (or CeB_6) filaments, resulting in temperatures on the order of 2,700–2,900 K or 1,500 K, respectively. The thermal release of electrons at the Fermi level of the material is known as *thermionic emission*, and may only occur once the energy exceeds the work function of the material (4.7 and 2.7 eV for W and LaB_6, respectively). Saturation of the filament occurs when further heating does not increase the number electrons being emitted. Electrons ejected from the filament cathode are focused by a negatively charged Wehnelt cylinder, and drawn toward a positively charged anode plate containing a small aperature (Figure 7.6). The electrons that escape through the pinhole are directed down the column toward the sample. The voltage difference between the filament and the anode plate is referred to as the *accelerating voltage*. This value is directly proportional to the energy of the electron beam that is directed down the column. Whereas the voltage range for SEM is *ca.* 200 V–40 kV, TEMs feature much higher accelerating voltages, ranging from 60 to 400 kV. All electron microscopes must be operated under extremely high vacuum conditions (*ca.* 10^{-6}–10^{-9} Torr), which provides an insulating medium between the cathode and anode, prolongs the lifetime and efficiency of the gun, and allows for sharper beam focusing. In general, the characteristics of an electron gun may be optimized by varying four components: the distance from the filament to the opening of the Wehnelt cylinder, the temperature of the filament, the negative bias of the Wehnelt cylinder with respect to the cathode, and the brightness of the beam.

Lanthanum hexaboride crystal filaments yield a brightness *ca.* 10 times greater than W, with a smaller spot size. This allows for imaging at very low accelerating voltages without a loss of resolution – of importance for beam-sensitive samples. Due to the small spot size, the beam from a LaB_6 source is highly monochromatic, which increases the overall resolution (*i.e.*, less chromatic aberrations). Interestingly, LaB_6 filaments employ the (100) face due to its relatively low work function.[10] The preferential release of electrons from this crystal face is likely due to the dipole existing between the positively charged metal atoms at the crystal surface, and the underlying layer of boride octahedra (Figure 7.7). A high vacuum must be used to prevent oxidation of the crystal surface; adsorption of oxygen raises the work function of LaB_6 to 4.0 eV.[11]

It should be noted that the work function of structural alloys is always less than the work function of the isolated component metals. For instance, the work functions of B and La are 4.5 and 3.3 eV, respectively – much higher than LaB_6. The mechanisms

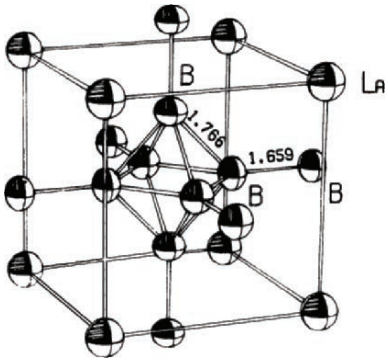


Figure 7.7. Unit cell of LaB_6. Thermal vibration ellipsoids are also shown, illustrating a significant aniostropy in the thermal vibrations of the B atoms. Reproduced with permission from Lundstrom, T. *Pure Appl. Chem.* **1985**, *57*, 1383. Copyright 1985 IUPAC.

Table 7.2. Comparison of Various Electron Sources

	Spot Size (nm)	E Spread (eV)	Brightness ($A\,cm^{-2}\,SR^{-1}$)	Stability (% RMS)	Lifetime (h)
Field Emission					
Schottky					
ZrO_2/W(100)	>0.2	0.35–0.7	5×10^8	<1	>1 year
Cold					
W(310)	>0.2	0.3–0.7	1×10^9	4–6	>1 year
Conventional					
LaB_6	>2	1.5	1×10^7	<1	500
W	>50	3	1×10^6	<1	100

responsible for the reduction of the work function are not completely understood, although it must depend on the three-dimensional arrangement and orbital overlap of electron rich/deficient atoms or ions near the surface of the solid. Once this mechanism is deciphered, many more materials will be isolated as effective candidates for electron emission sources.

In order to increase the intensity and focus of electrons, a *field emission* source may be used. This consists of a single crystal tungsten or $LaB_6(100)$ wire that is sharpened to a tip diameter of *ca.* 100 nm − 1 μm. For crystalline tungsten, the axis is suitably aligned with respect to the optical axis of the microscope. For example, a beam with a diameter < 5 nm is possible from alignment of the filament planes perpendicular to ⟨310⟩ and ⟨111⟩. In addition to W and LaB_6, a number of other materials are proposed for field emission applications, such as silicon, single-walled nanotubes,[12] and ultrananocrystalline diamond (UNCD) or Cu/Li alloy films deposited onto sharpened tips.[13]

Two classifications of field emitters are used in electron microscopes, cold and Schottky sources (Table 7.2). Whereas the former operates through electron

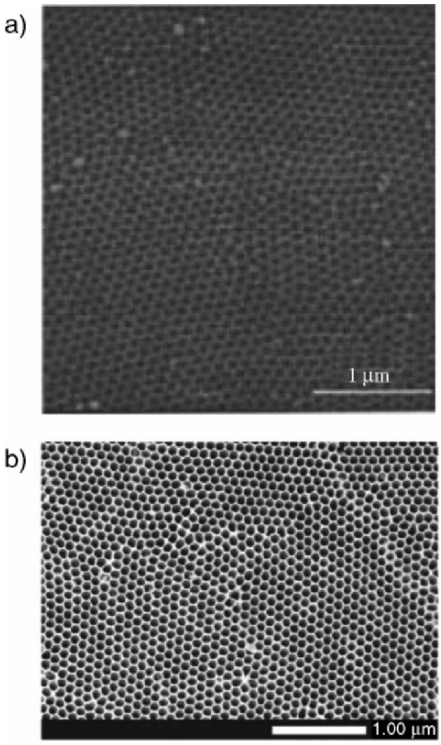


Figure 7.8. Comparison of the resolution for (a) low-resolution SEM,[15] with (b) high-resolution FESEM.[16] Both images are of a nanoporous anodized aluminum oxide template at identical magnification.

tunneling from a cathode wire held at room temperature, the latter features thermionic emission from a ZrO_2-coated sharpened W filament at 1,800 K. In both cases, an electrical field draws electrons from the narrow filament tip into an ultrahigh vacuum (10^{-9} Pa) chamber. When the electrons are accelerated through the magnetic field and focused onto the sample, a dramatically greater resolution is obtained – *ca*. 15 times that of traditional filaments (Figure 7.8). Though cold FE sources feature a higher brightness and smaller spot size than a Schottky FE source, they are not as preferred due to their relatively noisy emission profile. This beam instability arises from the Brownian motion of adsorbed residual gases on the surface of the narrow tip. Though electron microscopes are operated under UHV, these gases are always present in the column – likely originating from the sample itself (residual solvent on grid, e-beam decomposition of the sample, *etc*.). In addition, as electrons strike residual gas molecules in the column, positive ions are generated that may be accelerated back to the electron source. Such ion bombardment will cause a significant deformation of the filament surface, which will lead to beam instabilities. Since Schottky emitters are operated at 1,800 K, these deformations are readily repaired *via* annealing.

However, this is not possible using cold FE sources, which must be periodically "flashed" with a high-temperature pulse in order to repair surface deformations. Though this effectively improves the surface structure of the emitter, the lifetime of the emitter is shorted due to an increase in the tip radius – a consequence of thermal flashing in the absence of an electric field.

7.2.1. Transmission Electron Microscopy

A transmission electron microscope is analogous to a slide projector, with illumination from an electron beam rather than light. When an electron beam is impinged upon a sample, a black and white TEM image is formed from the passage of some electrons through the sample untouched, alongside the combination of interactions between other electrons and sample atoms (*e.g.*, inelastic/elastic scattering, diffraction). If the undiffracted beam is selected to form the image, it is referred to as *bright-field imaging*; in contrast, selection of strongly diffracting regions of the sample, which would appear brighter than the transmitted beam, is known as *dark-field imaging*.[14] It should be noted that electrons may also be absorbed by molecules containing large atoms, or by surface contamination (*e.g.*, dust, grease). The absorption of a high density of electrons in a specific region will cause a buildup of heat, leading to sample destruction and poor image quality.

Analogous to throwing a baseball of varying speeds through a wall, the relative degree of penetration through a particular sample is governed by the energy of the electron source. That is, higher energy electrons (*e.g.*, 200 keV *vs*. 100 keV) will be more penetrating, allowing for the characterization of thicker and/or less transparent samples. In general, increasing the thickness of a sample, or decreasing the energy (*i.e.*, accelerating voltage) of the electron beam, will induce more scattering events through more effective interactions between the electron beam and atoms of the sample. This effect will enhance image contrast, since there is a larger deviation between the path lengths of transmitted and scattered electrons that reach the viewing screen. However, this improvement of image quality is offset by plentiful inelastic collisions that yield a broadened wavelength distribution of the electron beam. Since individual electrons will have differing energies, they will be brought into focus at different points resulting in a blurry image (*i.e.*, decreased resolution).

Sample preparation techniques

Not unlike other materials (or molecular) characterization techniques, the most important and time-consuming aspect of TEM analysis is sample preparation. Specimens for TEM analysis are placed on special micromesh grids of a conductive metal such as Cu, Au, or Ni. The typical dimensions of TEM grids are *ca*. 3 mm in diameter and 10–25 μm thick. The mesh number of a grid indicates the number of grid openings per linear inch. The smaller the grid number, the larger the hole size and the greater the ratio of open area to covered area (Figure 7.9a). For example, a 200 mesh grid has 20 holes along its diameter; a 400 mesh grid has 40 holes. This translates to

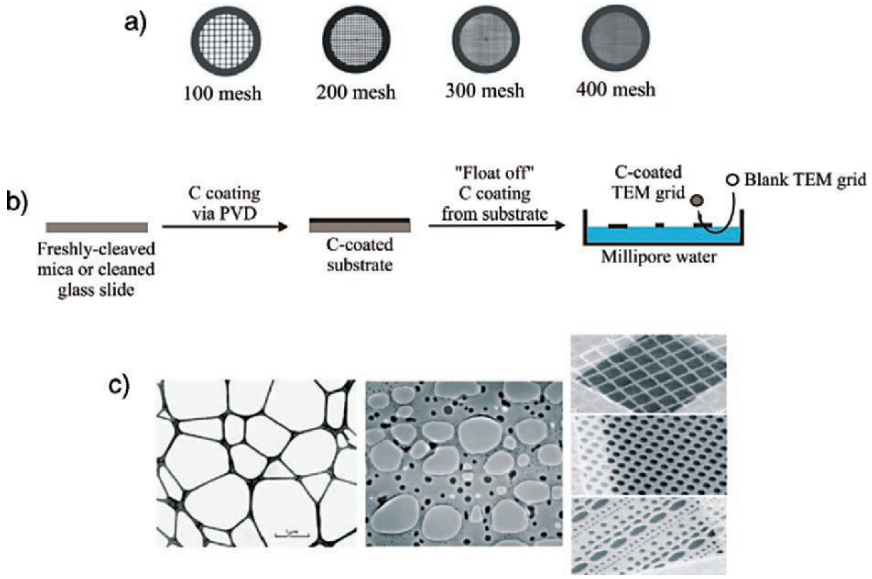


Figure 7.9. (a) A comparison of TEM grid mesh sizes; (b) Illustration of the "float off" method to prepare carbon-coated TEM grids; (c) Comparison of a lacey carbon coating with holey carbon films deposited onto TEM grids.

hole sizes of *ca*. 200 µm for 100 mesh, *ca*. 97 µm for 200 mesh, *ca*. 63 µm for 300 mesh, and *ca*. 42 µm for 400 mesh grids.

Most often, a grid is coated with a support film that holds the sample in place. The film must be as transparent as possible, while providing support for the sample. To prevent interference with electron-sample interactions, films containing light elements (*e.g.*, C, Be) are used. Support films are typically deposited onto the surface of grids through a "floating" technique (Figure 7.9b). Amorphous carbon or plastic (*e.g.*, Formvar – polyvinyl formar, Collodion – nitrocellulose) films are first deposited onto a glass microscope slide or mica surface, and floated off onto the surface of filtered water. For polymer films, the precleaned glass/mica surface is simply immersed in a 0.3% ethylene dichloride solution of the polymer. By contrast, carbon films are deposited onto the glass/mica surface through sputter coating. The coated glass/mica substrate is allowed to dry completely, scored with a razor blade, and slowly immersed into a container of distilled water. When water reaches the score marks, surface tension effects cause the film to be pulled from the glass/mica, resulting in a floating film on the surface of the water. Grids are then dunked into the water and brought up underneath the floating film, lifted out of the water, and allowed to air-dry completely. The coated grids are stored in a dry, dust-free environment.

The typical thickness of amorphous carbon films is 2–5 nm, whereas polymer films have thicknesses on the order of 30 nm. As you might expect, all plastic films are subject to decomposition by the electron beam. Sometimes, this exposure also causes further crosslinking which will cause the film to shrink and become more brittle. The

surface adsorption characteristics of support films vary quite significantly between polymer and carbon varieties. In general, polymer films possess hydrophilic characteristics whereas carbonaceous support films are hydrophobic.

For the imaging of nanomaterials, it is most desirable to have a support film that resembles a piece of lace, featuring many holes on the surface. In this manner, nanostructures such as nanotubes or nanowires are held on the surface of the support film, with unobstructed imaging made possible through holes in the film surface (Figure 7.9c). Such "holey" or "lacey" grids are made from plastic film suspensions to which water or glycerin has been added. Due to the microemulsions created in the solution, when the film is dried, thousands of tiny holes are left in the support film. In this way, the coating itself acts like a mesh grid with the thousands of very tiny open viewing areas.

Due to the high vacuum environment and bombardment of the sample with high-energy electrons, not all samples are appropriate for TEM characterization. In particular, any sample comprising a volatile organic compound, or those containing residual solvents, must be dried using freeze–pump–thaw procedures, or supercritical fluid media. Failure to do so will result in collisions between the electron beam and stray gaseous molecules in the column, resulting in defocusing of the beam and a reduction in image quality. Further, the decomposition of organic matter can severely contaminate the microscope column – especially in finely machined regions such as apertures and pole pieces resulting in blurry images. Some samples may not be directly analyzed even following drying procedures, due to their instability toward the focused electron beam. For these samples (*e.g.*, biological), the specimen is mounted within a plastic support resin, and very thin pieces of the composite are shaved off using an ultramicrotome fitted with a diamond, sapphire, or glass knife (Figure 7.10).

TEM grids of solid samples are typically prepared by pipetting a few drops of a solution containing the suspended solid onto a grid, and allowing it to air-dry. Alternatively, the suspension may be blown onto a grid using an aerosol delivery device (Figure 7.11). If the solid is too thick, or comprises relatively large

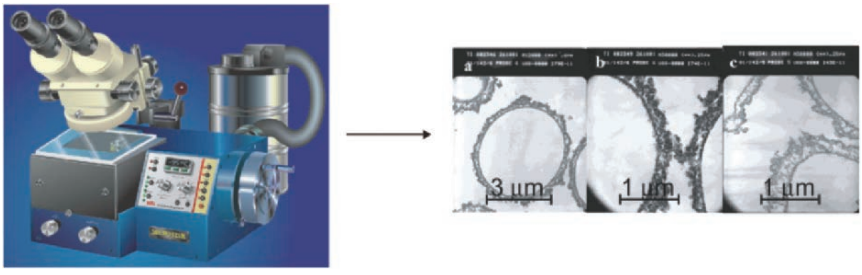


Figure 7.10. Sectioning of a TEM sample using an ultramicrotome instrument. Shown is an ultramicrotome instrument with a built-in nitrogen cryogenic system, used to section samples with a thickness of 25 nm–5 mm.[17] Also shown is a TEM image of polymer capsules that have been sectioned using an ultramicrotome instrument. Image reproduced with permission from Dai, Z.; Mohwald, H. *Langmuir* **2002**, *18*, 9533. Copyright 2002 American Chemical Society.

Figure 7.11. Photo of a glass aerosol delivery system used to spray solvent-suspended particulates onto a TEM grid.

granules, the material may be embedded within a polymeric matrix and cut into thin cross sections. The best embedding medium permits thin sectioning with the least damage during preparation, and gives the least interference during microscopy. While early embedding media included gelatin and paraffin, the <100 nm sections that are required for electron microscopy were not possible to cut using any of these media. Nowadays, polymers such as methacrylates, epoxy and polyester resins are typically employed for sectioning applications, since they are nontoxic, inexpensive, and strong enough to withstand very thin mechanical sectioning. For methacrylates, the degree of final block hardness can be predictably controlled by mixing methyl (hard) and *n*-butyl (soft) functionalities. However, with respect to electron beam stability, epoxy resins are by far superior to both methacrylates and polyester resins.

The "best" section thickness not only depends on the electron opacity of the sample, but also on the operating conditions of the TEM. In general, the higher the accelerating voltage, the thicker the sample cross section may be. For instance, epoxy-based sections of up to 250 nm thick may be used for standard TEM analyses, operating at 100 kV. However, section thicknesses in the range of 50–60 nm are typically used for TEM analysis, in order to improve the resolution through reduced electron scattering.

A recent alternative to (ultra)microtoming is the use of a *focused-ion beam* (FIB) to etch/mill away undesired portions of the sample (Figure 7.12). The most common beam is gallium ions from a liquid gallium source, with an energy of *ca.* 30 keV. Extreme control over the resultant cross section is afforded by varying the beam current/energy, and scanning speed of the ion beam. This technique is especially useful for the study of polymer films that have been deposited onto hard inorganic substrates – using a FIB/lift-out method.[18] Typically, a dual-beam FIB/SEM is used for sample preparation, which affords real-time monitoring of FIB sectioning (Figure 7.12b). A more recent use of a FIB instrument is nanofabrication. In particular, *in situ* CVD of a thin film may be deposited through the introduction of a co-reactant precursor gas during ion bombardment (Figure 7.12c). It should be noted that in contrast to ultramicrotomy, FIB preparation does not induce mechanical stress on the sample surface, which often leads to deformation and interfacial debonding. However,

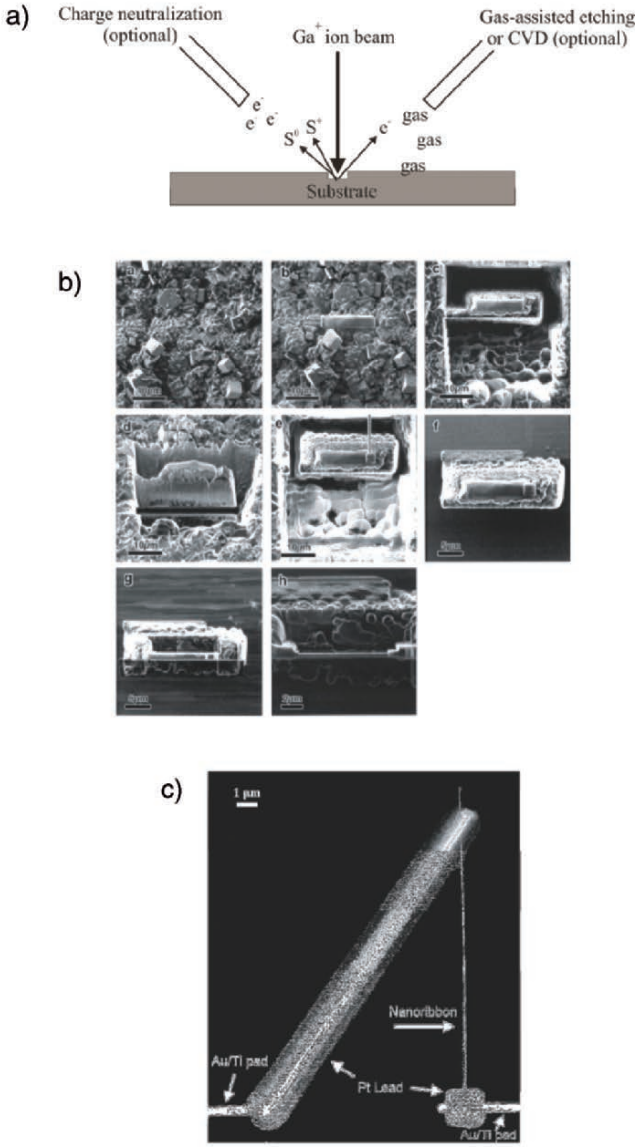


Figure 7.12. Applications for focused-ion beam (FIB). Shown is (a) general illustration of the technique, with a gallium ion source and emission of secondary electrons, as well as charged ions and atoms from the substrate surface (S^+ and S^0, respectively). The optional "charge neutralization" electron beam is used for *in situ* imaging of a surface during FIB (dual-beam FIB/SEM). The images shown in (b) represent the sectioning of a zeolitic membrane (images c and d illustrate FIB milling to a depth of 20 µm, and cutting off the bottom surface; latter images show the manipulation of the milled target and further sectioning by FIB).[19] The SEM image in c illustrates the use of FIB for nanofabrication – the selective deposition of Pt contact leads onto the ends of a Ga_2O_3 nanowire.[20]

the high-energy radiation may result in other surface modifications such as amorphization, shrinkage, and bond cleavage – especially for "soft," polymeric-based samples.[21]

Oftentimes, a lack of image contrast in TEM results from the analysis of very thin sections (*e.g.*, <30 nm), or samples comprising light elements (*e.g.*, hydrocarbons, B, Li, *etc.*). In order to increase the contrast, elements of high atomic weight are introduced to selective regions of the sample – a process referred to as *staining* (Figure 7.13). Due to their high electron density, staining agents are better able to stop or deflect the beam of electrons, whereas elements of low weight allow the beam to pass through relatively unimpeded. Electron staining falls into one of two categories:

(i) *Positive staining* – contrast is imparted to the sample.
(ii) *Negative staining* – contrast is imparted to the area surrounding the sample.

The most common staining agents are aqueous solutions (*ca.* 2 wt% concentration) of OsO_4 (osmium tetroxide), RuO_4 (ruthenium tetroxide), $UO_2(CH_3COO)_2 \cdot 2H_2O$ (uranyl acetate), $Pb(C_6H_2O_7)_2 \cdot 3H_2O$ (lead citrate), or $Na_3[PW_{12}O_{40}] \cdot 30H_2O$ (sodium phosphotungstate). Due to the presence of heavy metals, these agents are considerably toxic and must be handled/disposed with great care. Whereas most of

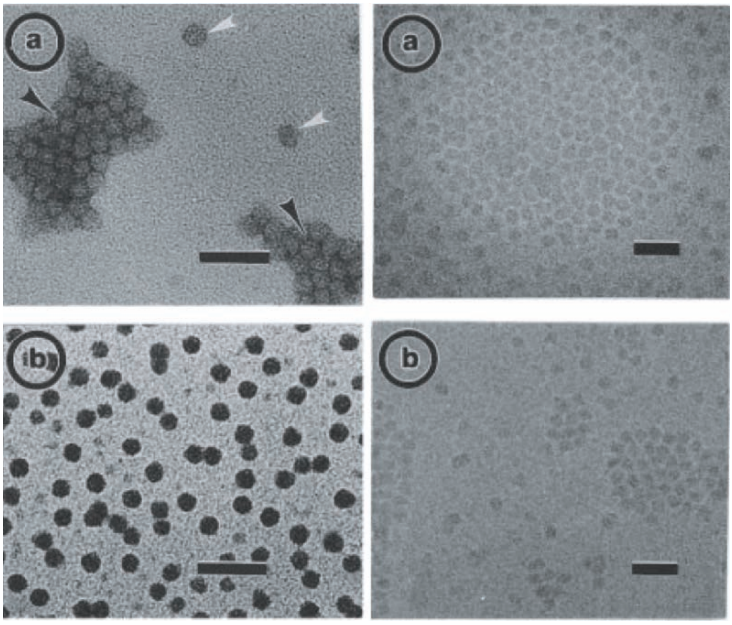


Figure 7.13. (Left, a) TEM image of G10 PAMAM dendrimers stained with 2% methylamine tungstate, showing both positively (white arrows) and negatively (black arrows) stained regions. (Left, b) TEM image of G9 PAMAM dendrimers positively stained with sodium phosphotungstate, showing little agglomeration. (Right, a, b): TEM images showing the lack of contrast from unstained G10 PAMAM dendrimers, imaged in vitrified water using cryo-TEM. The scale bar for all images is 50 nm. Reproduced with permission from Jackson, C. L.; Chanzy, H. D.; Booy, F. P.; Drake, B. J.; Tomalia, D. A.; Bauer, B. J.; Amis, E. J. *Macromolecules* **1998**, *31*, 6259. Copyright 1998 American Chemical Society.

these agents interact most strongly with amino and phosphate groups, lead ions preferentially interact with hydroxyl groups. For biological staining, it has been widely shown that OsO_4 interacts more readily with lipids than proteins, resulting in a structure-specific staining agent. The use of lead citrate requires special considerations due to its high sensitivity to CO_2; hence, NaOH is often used as a CO_2-scavenger for the aqueous staining solution.

Nonimaging applications for TEM

Due to the high spatial resolution and predictive scattering modes, TEMs are often employed to determine the three-dimensional crystal structure of solid-state

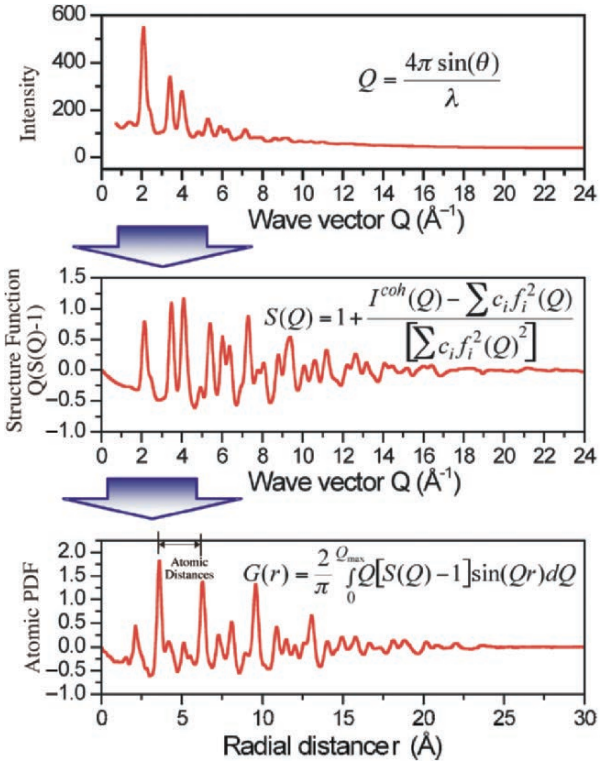
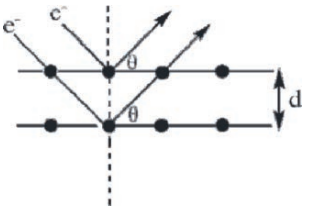


Figure 7.14. The atomic pair distribution function (PDF) used to characterize powders by high-energy X-ray diffraction. Shown (top-bottom) is the conversion of the reflection intensity data into the atomic PDF, which yields information regarding the average inter-atomic spacings of the powder in real space. The three-dimensional structure of the solid is determined by using software, which compares the experimental PDF with structural models that best fit the data. This method is preferred over competing techniques such as Rietveld analysis since the peaks are in real space (not reciprocal space), and both Bragg peaks and diffuse scattering are included in the data. It should be noted that for maximum resolution, Q_{max} in the sine Fourier transform must be as high as possible. Since Q is inversely proportional to the radiation wavelength, a synchrotron source is often utilized ($\lambda = 0.1$ Å, relative to 1.54 Å for a traditional powder X-ray diffractometer). Images provided courtesy of Prof. Valeri Petkov, Department of Physics at Central Michigan University. For more information, see http://www.phy.cmich.edu/people/petkov/

materials. Thus, TEM may be considered as complementary to conventional crystallographic methods such as X-ray diffraction (Figure 7.14). When electrons (or X-rays) interact with a crystalline solid, comprising a regularly spaced three-dimensional array of atoms, the incoming beam is diffracted at specific angles, as predicted by the Bragg equation (Eq. 6). Simply stated, Bragg's Law suggests that for a particular pair of d and λ, diffraction will occur at a specific angle (known as a Bragg reflection):

$$n\lambda = 2d(\sin\theta), \tag{6}$$

where d is the lattice spacings of the crystal planes (*e.g.*, d_{10} would indicate the spacing of the (10) planes in two-dimensions); λ, the wavelength of the incoming beam; and θ is the angle of the incident (and diffracted: $\theta_{in} = \theta_{out}$) beam.



In order to understand the information contained within the diffraction pattern of a crystal lattice, it is necessary to construct a secondary lattice known as a *reciprocal lattice*. This lattice is related to the "real" crystalline array by the following (Figure 7.15):

(i) For a two-dimensional lattice defined by vectors a and b, the reciprocal lattice is defined by vectors a^* and b^*, such that $a^* \perp a$ and $b^* \perp b$.

(ii) The magnitude of the reciprocal lattice vectors are equal to the inverse of the lattice spacings of the associated planes in the real lattice (Eq. 7):

$$\frac{1}{d} = \frac{2}{n\lambda}(\sin\theta). \tag{7}$$

It should be noted that for a three-dimensional reciprocal lattice, a third vector (c^*) is used that is perpendicular to *both* a and b axes of the real lattice.

In order to determine which lattice planes give rise to diffraction, a geometrical construct known as an *Ewald sphere* is used (Figure 7.16). An incident wave that impinges on the crystal is denoted as a reciprocal lattice vector, **k**, which passes through the origin. The Ewald sphere (or circle in two-dimensional) shows which reciprocal lattice points, (each denoting a set of planes) that satisfy Bragg's Law for diffraction of the incident beam. A specific diffraction pattern is recorded for any **k** vector and lattice orientation – usually projected onto a two-dimensional film or CCD camera. In general, very few reciprocal lattice points will be intersected by the Ewald sphere,[22] which results in few sets of planes that give rise to diffracted beams. As a result, a single crystal will usually yield only a few diffraction spots. If a sample is single-crystalline, sharp spots will be observed; a polycrystalline sample

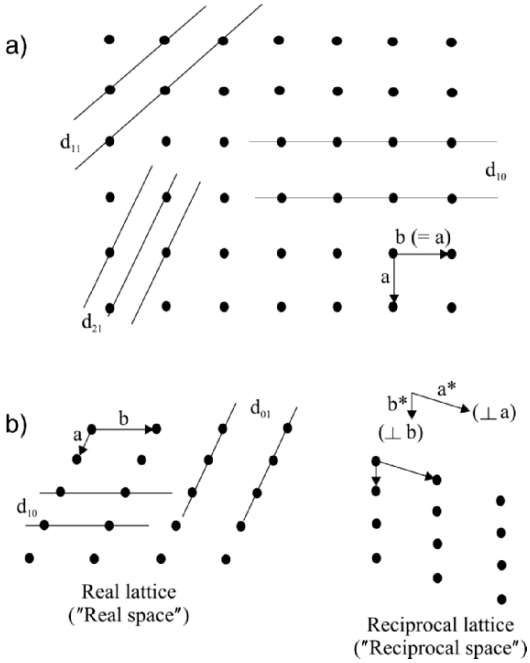


Figure 7.15. (a) Illustration of a two-dimensional square lattice, showing various lattice spacings; (b) a comparison of real space and reciprocal space for a crystalline solid. In the reciprocal lattice, the magnitudes of a^* and b^* are $1/d_{10}$ and $1/d_{01}$, respectively.

will yield many closely spaced diffraction spots. By contrast, an amorphous sample will give rise to diffuse rings.

Since the wavelength of an accelerated electron in a TEM is much smaller than an X-ray beam, the Ewald sphere (radius: $1/\lambda$) is significantly larger for electron diffraction relative to X-ray diffraction studies. As a result, electron diffraction yields much more detailed structural information of the crystal lattice. Information such as lattice parameters and atomic positions in a crystal may be obtained through analyzing the *in situ* electron diffraction pattern from a specimen size of *ca.* >400 nm (>100 nm for a field emission source). This technique is denoted as *selected area electron diffraction* (SAED, Figure 7.17). In addition, the lenses within a TEM allow one to alter the orientation of the incoming electron beam, from a parallel beam to a cone-shaped beam. The latter orientation is known as convergent beam electron diffraction (CBED), and allows one to conduct a diffraction experiment over many incident angles simultaneously to reveal the full three-dimensional reciprocal lattice of the crystal.

In the same manner as X-ray diffraction, by examining the systematic absences of the diffraction spots (Table 7.3), one may easily determine the appropriate Bravais

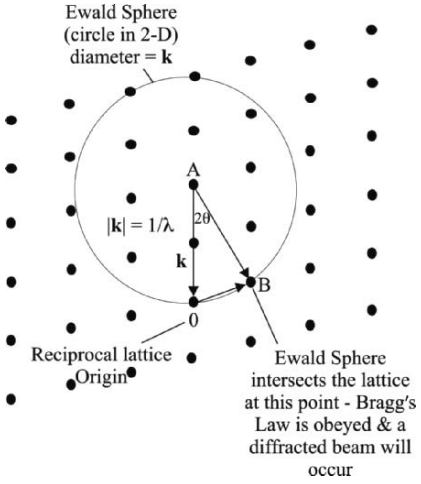


Figure 7.16. Illustration of Ewald sphere construction, and diffraction from reciprocal lattice points. This holds for both electron and X-ray diffraction methods. The vectors AO, AB, and OB are designated as an incident beam, a diffracted beam, and a diffraction vector, respectively.

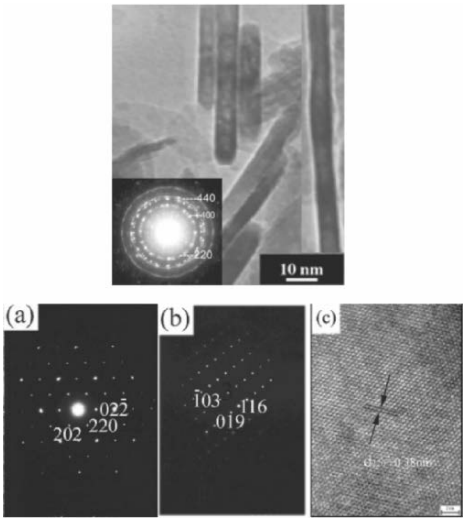


Figure 7.17. (top) SAED pattern from a polycrystalline $MgAl_2O_4$ nanorods.[24] (bottom) SAED patterns (a, b) and high-resolution TEM image (c) of a single-crystalline In_2S_3 nanoparticle. Images (a) and (b) illustrate the diffraction patterns for cubic and tetragonal unit cells, respectively. The TEM image shown in (c) illustrates the lattice fringes and interplanar spacing of 3.81 Å, corresponding to the (220) interplanar spacing of cubic In_2S_3.[25] In high-resolution images, lattice fringes result from the interference of the transmitted and diffracted beams; these are only evident in the TEM image when the lattice spacings are larger than the resolution of the electron microscope.

Table 7.3. Systematic Absences for Electron and X-Ray Diffraction

Cause of absence	Symbol	Absences[a]
Body-centering	I	$h + k + l = 2n + 1$ (odd)
A centering	A	$k + l = 2n + 1$
B centering	B	$h + l = 2n + 1$
C centering	C	$h + k = 2n + 1$
Face-centering	F	*hkl* mixed (not all even or all odd)
Glide plane//(100)	b	$k = 2n + 1$
	c	$l = 2n + 1$
	n	$k + l = 2n + 1$
	d	$k + l = 4n + 1$
Glide plane//(010)	a	$h = 2n + 1$
	c	$l = 2n + 1$
	n	$h + l = 2n + 1$
	d	$h + l = 4n + 1$
Glide plane//(001)	a	$h = 2n + 1$
	c	$k = 2n + 1$
	n	$h + k = 2n + 1$
	d	$h + k = 4n + 1$
Screw axis//*a*	2_1 or 4_2	$h = 2n + 1$
	4_1 or 4_3	$h = 4n + 1$
Screw axis//*b*	2_1 or 4_2	$k = 2n + 1$
	4_1 or 4_3	$k = 4n + 1$
Screw axis//*c*	2_1 or 4_2	$l = 2n + 1$
	3_1, 3_2, 6_2, or 6_4	$l = 3n + 1, 3n + 2$
	4_1 or 4_3	$l = 4n + 1, 4n + 2, 4n + 3$
	6_1 or 6_5	$l = 6n + 1, \ldots, 6n + 5$
Screw axis//(110)	2_1	$h = 2n + 1$

[a]Refers to the Miller indices ((*hkl*) values) that are absent from the diffraction pattern. For instance, a body-centered cubic lattice with no other screw axes and glide planes will have a nonzero intensity for all reflections where the sum of $(h + k + l)$ yields an odd number, such as (100), (111), *etc.*; other reflections from planes in which the sum of their Miller indices are even, such as (110), (200), (211), *etc.* will be present in the diffraction pattern. As these values indicate, there are three types of systematic absences: three-dimensional absences (true for all *hkl*) resulting from pure translations (cell centering), two-dimensional absences from glide planes, and one-dimensional absences from screw axes.[26]

lattice as well as any screw axes and glide planes that are present in the crystal lattice. Since TEM is more user-intensive, involves significant sample preparation, and often results in sample damage from the high-energy electron beam, highly automated X-ray and neutron (useful for light elements such as H) diffraction methods are the preferred methods to yield structural information from a bulk crystalline solid. However, if structural information is desired from an individual nanocrystal, TEM/SAED is the best alternative.[23]

Before leaving the subtopic of electron diffraction, it should be noted that two other techniques known as low-energy electron diffraction (LEED) and reflection high-energy electron diffraction (RHEED) may also be used to gleam structural

information from a crystalline sample. Both techniques are performed independent of a TEM instrument, and utilize an electron gun and fluorescent screen to show the structure and morphology of a crystal surface. The RHEED technique uses a high accelerating voltage and low impact angle in order to focus the electrons to the first few atomic layers of the crystal surface. This technique may be used to monitor crystal growth in real time, or to probe the surface adsorption properties for sensor development. For beam-sensitive crystalline samples (*e.g.*, organic thin films), either LEED or microchannel plate (MCP)–RHEED may be used; the latter features the amplification of the incident electron beam *via* a MCP – resulting in much lower intensities of the incident electron beam.

In addition to structural information from crystalline samples, quantitative analysis may also be carried out in tandem with TEM (and SEM) analyses. The interaction among high-energy electrons and sample atoms results in a variety of emissions (Figure 7.18) that yield important information regarding the surface morphology

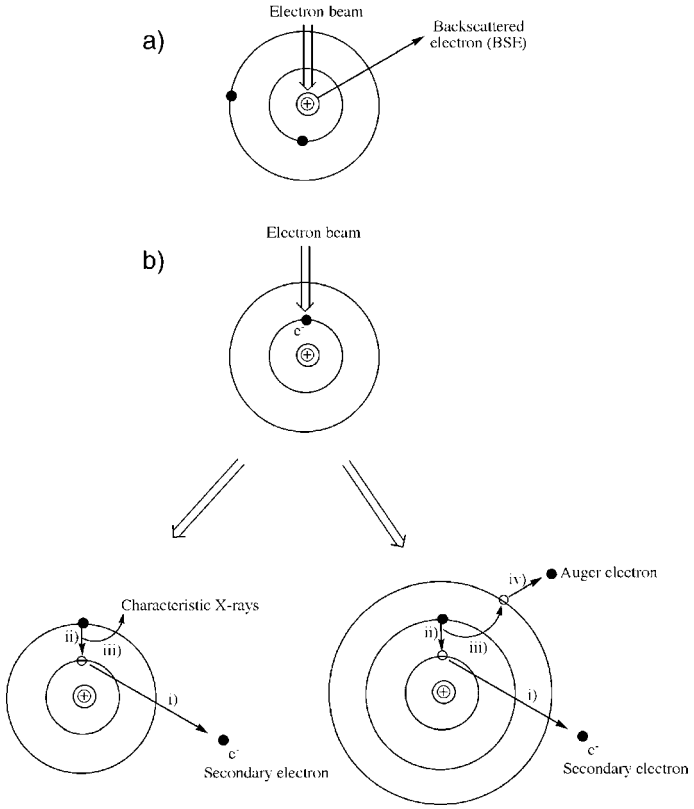


Figure 7.18. Types of possible interactions between the primary electron beam and (a) the nucleus of surface atoms and (b) electrons of surface atoms comprising the sample. For simplicity, only K, L, and M shells (with no L and M subshells) are shown.

and elemental composition of the sample. As previously stated, if the electron beam interacts with the nucleus of surface atoms, the electrons are elastically scattered. In this case, the trajectory of the electron changes with no change in the kinetic energy or velocity, a process known as *backscattering*.

The number of *backscattered electrons* (BSE) that are produced from a given atom is proportional to the atomic number. That is, materials composed of heavy atoms will backscatter more electrons, resulting in brighter gray tones in the image relative to less dense materials. Hence, BSE produce an image that is related to material composition, providing both spatial and chemical information.

The primary electron beam may also be inelastically scattered through interaction with electrons from surface atoms. In this case, the collision displaces core electrons from filled shells (*e.g.*, ns^2 (K) or np^6 (L)); the resulting atom is left as an energetic excited state, with a missing inner shell electron. Since the energies of these *secondary electrons* are sufficiently low, they must be released from atoms near the surface in order to be detected. Electrons ejected from further within the sample are reabsorbed by the material before they reach the surface. As we will see in the next section (re SEM), as the intensity of the electron beam increases, or the density of the sample decreases, information from underlying portions of the sample may be obtained.

In order for the excited-state atom to return to its ground state (within a picosecond or so following secondary electron generation), the vacancy in the K shell is filled with an electron from shells farther from the nucleus (*e.g.*, L shell, Figure 7.19). When such post-ionization atomic relaxation occurs, the excess energy may be released as either characteristic X-rays, or through nonradiative emission of an *Auger electron* (Figure 7.18b; Auger electron spectroscopy (AES) will be discussed later). Especially for elements with $Z > 11$ (*i.e.*, >Na), the electronic shell structure becomes exceedingly complex, which results in electronic transitions that are possible from a number of outer shells (Figure 7.19, bottom).

Typically, a variety of X-rays are produced during atomic relaxation due to a cascading effect. For instance, a vacancy in the K shell may be filled with an electron in an L shell (K_α emission); the resulting vacancy in the L shell may then be filled with an electron from the M shell ($L_{\alpha 1}$ emission), and so forth. By counting the number and energies of X-rays produced from electrons interacting the sample, it is possible to both qualitatively and quantitatively (using suitable standards) determine the chemical composition of the surface being analyzed.

There are two methods used to identify and quantify the X-ray emission: *energy-dispersive X-ray spectroscopy* (EDS), and *wavelength-dispersive X-ray spectroscopy* (WDS).[28] In EDS, all of the characteristic X-ray energies reaching the detector are measured simultaneously. Hence, data acquisition is very rapid across the entire spectrum. By contrast, WDS measures a single wavelength at a time through use of a detecting crystal. As the characteristic X-rays are emitted from the sample, they are diffracted in a regular manner as discussed previously. Not only does this improve the resolution of WDS to an order of magnitude greater than EDS (Figure 7.20),[29] but also improves the count rate and deconvolution of overlapping spectral peaks. Nevertheless, due to its simplicity and speed of analysis, EDS is the standard method

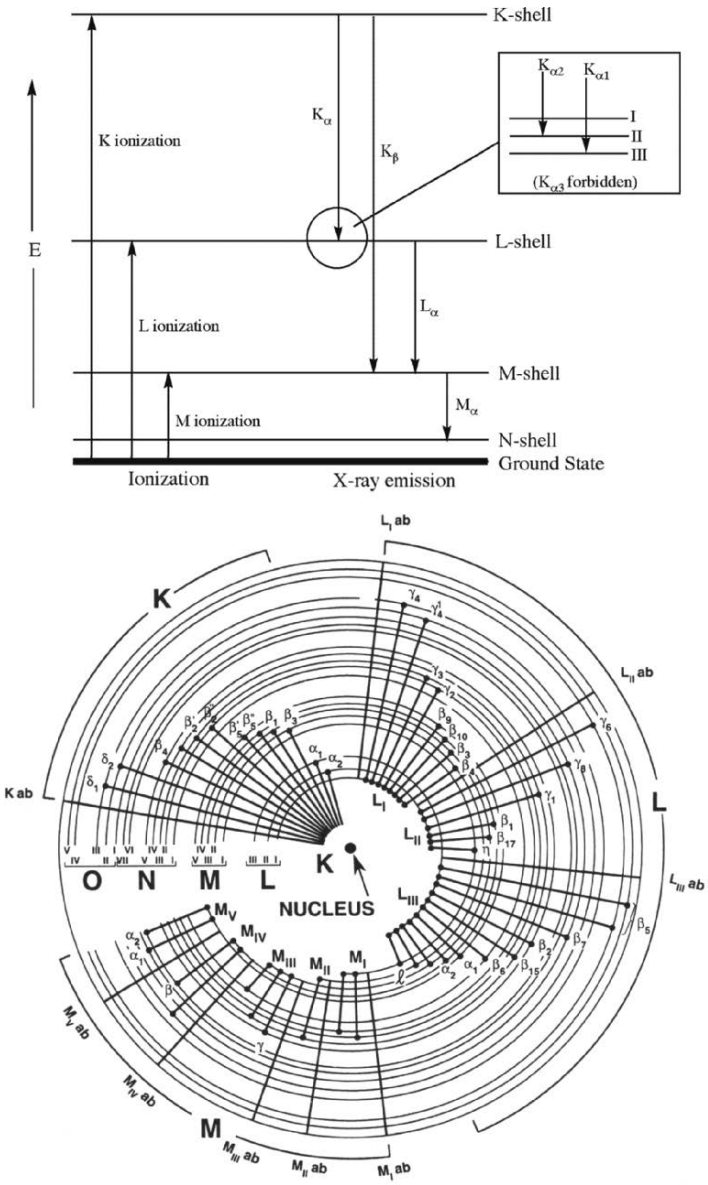


Figure 7.19. Simplified (top) and detailed (bottom) illustration of atomic energy levels, showing the characteristic X-rays emitted as the atom's energy returns to the ground state following ionization. The α and β designations indicate the filling of K, L, or M electron vacancies from adjacent ($n + 1 \rightarrow n$) and ($n + 2 \rightarrow n$) energy levels, respectively.[27] The inset shows transitions between the K shell and L subshells. The illustration of detailed electronic transitions for Cu was reproduced with permission from Goldstein, J.; Newbury, D.; Joy, D.; Lyman, C.; Echlin, P.; Lifshin, E.; Sawyer, L.; Michael, J. *Scanning Electron Microscopy and X-Ray Microanalysis*, 3rd ed., Kluwer: New York. Copyright 2003 Springer Science and Business Media.

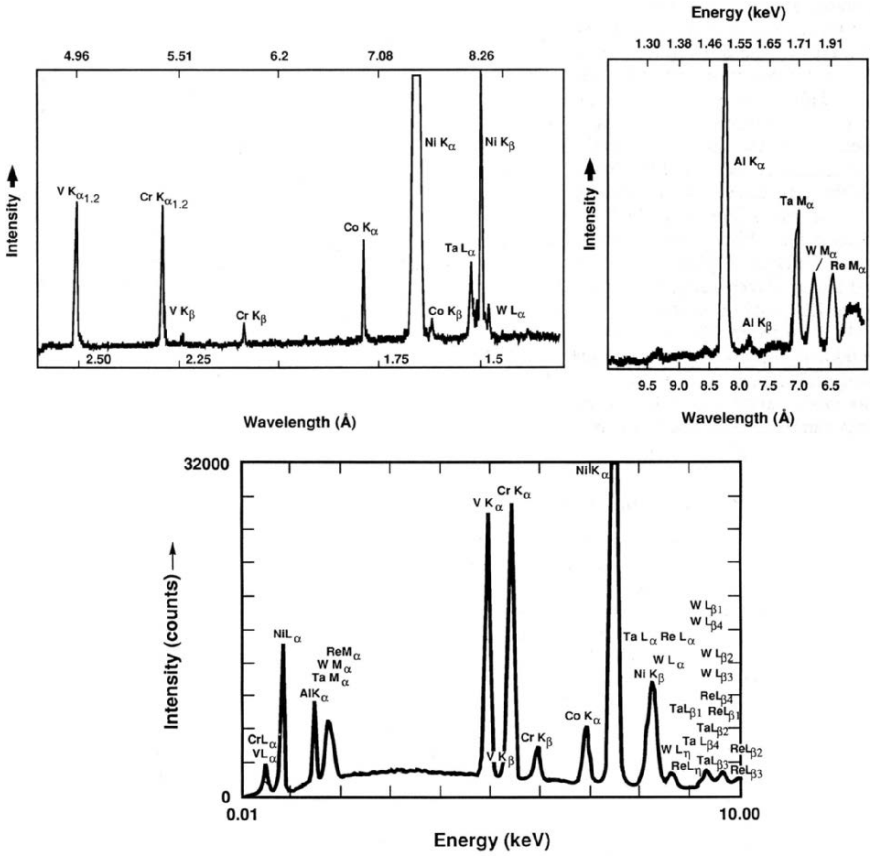


Figure 7.20. Comparison of WDS (top, using LiF and thallium acid phthalate (TAP (1010), respectively) and EDS for the analysis of a superalloy. A number of X-ray spectral lines such as Ta Lα, Ni Kβ, and W Lα are hardly discernible using EDS, but readily visible using WDS analysis. Reproduced with permission from Goldstein, J.; Newbury, D.; Joy, D.; Lyman, C.; Echlin, P.; Lifshin, E.; Sawyer, L.; Michael, J. *Scanning Electron Microscopy and X-Ray Microanalysis*, 3rd ed., Kluwer: New York. Copyright 2003 Springer Science and Business Media.

for chemical analysis within TEM (and SEM) instruments. Typically, if WDS is desired, an instrument known as electron probe microanalyzer (EPMA) is utilized, often in tandem with SEM imaging.

The spatial resolution of X-ray microanalysis may be described by Eq. 8. In general, prerequisites for the best spatial resolution include a high-energy electron beam and extremely thin samples. Though the resolution limits of EDS/WDS will never match the imaging resolution, it is now possible to easily determine the elemental composition of individual nanoclusters in the 1–5 nm size regime – as long as they are suitably dispersed with respect to one another (Figure 7.21):

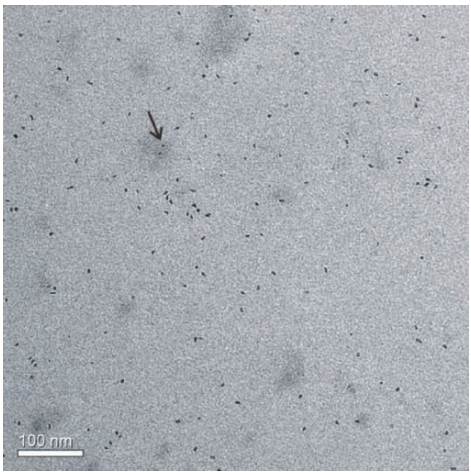


Figure 7.21. Ambiguity in assigning a chemical composition (iron oxide) to individual nanostructures based on TEM/EDS. The small and large nanoparticles are in the same vicinity on the grid; Depending on the area selected for analysis, it may be difficult to definitely state if one (or both) comprise iron oxide (*e.g.*, arrow region). However, by also looking at the relative contrast of the nanoparticles, it is likely that the smaller structures contain iron (higher density) and the larger nanoparticles comprise lighter elements (*e.g.*, a carbonaceous, organic-based nanostructure).[32]

$$R = \frac{d + \sqrt{\left(7.21 \times 10^5 \frac{Z}{E_o}\sqrt{\left(\frac{\rho}{A}\right)}t^{3/2}\right)^2 + d^2}}{2}, \tag{8}$$

where R is the X-ray spatial resolution; d, the beam diameter; E_o, the beam energy (eV); ρ, the specimen density; and t is the specimen thickness (cm).

Scanning transmission electron microscopy

An imaging mode that merges both SEM and TEM is also possible on most modern TEM instruments. This method, referred to as *scanning transmission electron microscopy* (STEM), uses a LaB_6 source that produces a focused electron beam with a high current density and extremely small diameter. Instead of monitoring the transmitted electrons from a static beam as performed in standard TEM imaging, the beam within a STEM is scanned across the sample – analogous to SEM. Due to a higher beam intensity, thicker samples may be analyzed in a STEM; furthermore, staining is generally not necessary for low-Z elements due to a higher sensitivity to sample density/composition. In particular, it is possible to overlay the image with the EDS data – a technique known as *elemental dot-mapping*, widely used for SEM/EDS analysis (Figure 7.22).

The majority of STEM instruments are simply conventional TEMs with the addition of scanning coils. As a result, these "nondedicated STEMs" are capable of TEM/STEM, as well as SEM imaging for thicker samples. The development

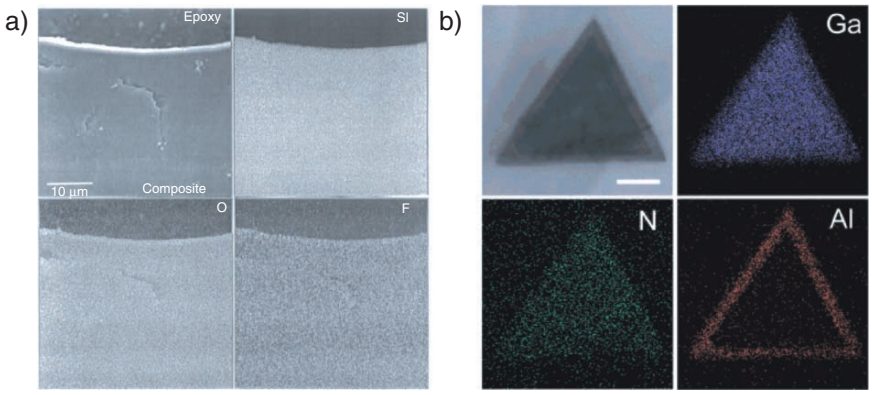


Figure 7.22. Elemental dot-maps. Shown is (a) elemental concentrations of Si, O, and F (as white pixels) overlaid onto the SEM image of a Nafion resin/silica composite[33]; (b) bright-field STEM image of a GaN/AlN/AlGaN nanowire cross section, with elemental mapping of Ga, N, and Al (scale bar is 50 nm).[34]

of HRTEMs and "dedicated" STEMs with lens aberration correction[30] have now pushed the resolution limits to as low as 0.1 nm, *i.e.*, 1 Å, the approximate diameter of the smallest atoms in the Periodic Table![31] In contrast to conventional TEMs, the beam system of dedicated STEMs is reversed, with the placement of the gun on the bottom of the microscope and detectors at the top (Figure 7.23). Since there are no post-sample refocusing lenses, it is not possible to generate a bright-field image as is standard using a conventional (HR)TEM. Instead, the image is based on specific interactions between the electron beam and sample atoms (*i.e.*, elastic/inelastic scattering, unscattered electrons).

In a dedicated STEM, high-angle (elastically) scattered electrons are separated from inelastic/unscattered electrons through use of a high-angle annular dark-field (HAADF) detector (Figure 7.23). Since the incoherently scattered electrons are related to elastic scattering, they are directly related to the structure and chemical composition of the feature being imaged (Figure 7.24).[35] In particular, the incoherent scattering of an electron is described by the Rutherford equation (Eq. 9). This equation predicts that the intensity of the scattered electrons, and resultant image contrast, is strongly dependent on atomic number of the sample atoms. Hence, HAADF–STEM is also referred to as Z-contrast imaging. It should be noted that HAADF–STEM is strongly dependent on variations in sample thickness.[36] That is, thicker regions of a sample will result in higher image intensities, which may be falsely interpreted as the presence of species with relatively high atomic numbers:

$$\frac{\mathrm{d}\sigma(\theta)}{\mathrm{d}\Omega} = \frac{e^4 Z^2}{16(E_\mathrm{o})^2 \sin^4 \dfrac{\theta}{2}}, \tag{9}$$

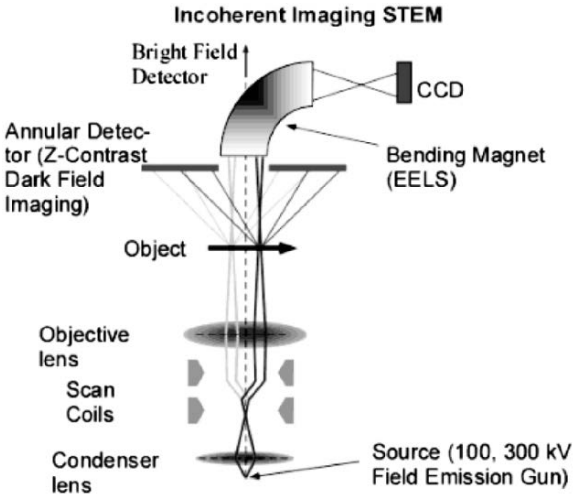


Figure 7.23. Schematic of a dedicated HAADF–STEM. Reproduced with permission from McBride, J. R.; Kippeny, T. C.; Pennycook, S. J.; Rosenthal, S. J. *Nano Lett.* **2004**, 4, 1279. Copyright 2004 American Chemical Society.

where $d\sigma(\theta)/d\Omega$ is the different scattering cross sections as a function of the scattering angle (θ); E_o, the incident beam energy; e, the electron charge (1.602×10^{-19} C); and Z is the atomic number of the scattering nucleus.

A primary limitation of EDS and WDS is the inability to detect light (*i.e.*, low-Z) elements. Since atomic energy levels are closely spaced for low-Z elements, the energies of the emitted X-rays will be relatively low.[37] As a result, they are masked by the broad, continuous background spectrum (known as *bremsstrahlung*[38]) that is most intense at energies below 1 keV. Furthermore, the characteristic X-ray lines are less intense for low-Z elements since they exhibit a low *X-ray fluorescence yield* – favoring nonemissive Auger[39] electron processes rather than X-ray generation (Figure 7.25). Consequently, it becomes increasingly more difficult to observe X-ray lines from elements with $Z < 11$ (Na). Recently, there have been improvements in the design of EDS windows, which separate the detection system from the electron microscope.[40] Since low-energy X-rays are readily absorbed, new detector designs feature ultrathin windows and lightweight compositions that facilitate the detection of elements down as far as beryllium ($Z = 4$). However, the analysis of these elements by EDS/WDS is semiquantitative, at best.

Electron energy-loss spectroscopy

In order to increase the sensitivity toward the detection of light elements, a technique known as *electron energy-loss spectroscopy* (EELS) may be utilized.[41] This method may be carried out within a (S)TEM, and consists of monitoring the loss in energy (due to inelastic scattering) of the beam electrons as they pass through the sample. Since it is more difficult to focus X-rays relative to electrons with appropriate lenses,

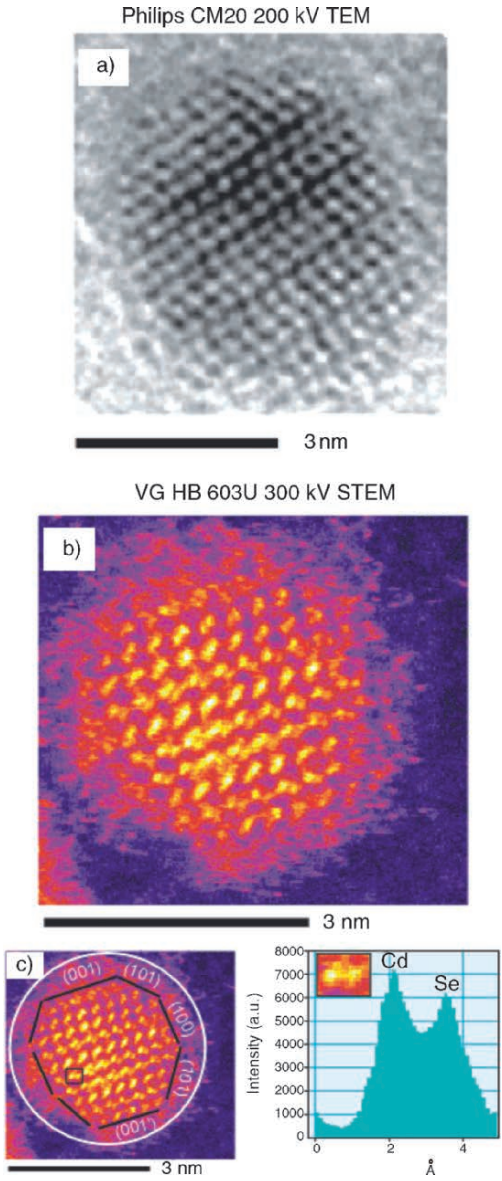


Figure 7.24. Comparison of conventional HRTEM (a), with HAADF–STEM (b). Also shown (c) is the chemical analysis of an individual CdSe "dumbbell." The white circle shows the amorphous oxide region, and the surface of the nanocrystal is outlined in black. Unlike conventional HRTEM, it is also possible to label the individual nanocrystal facets, such as Cd-rich (001) and Se-rich (001′). Reproduced with permission from McBride, J. R.; Kippeny, T. C.; Pennycook, S. J.; Rosenthal, S. J. *Nano Lett.* **2004**, *4*, 1279. Copyright 2004 American Chemical Society.

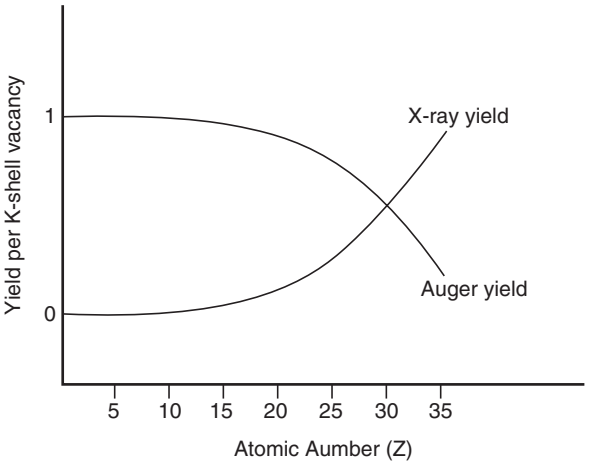


Figure 7.25. Relative probabilities for X-ray and Auger electron emission during the decay of K-electron vacancies.

the collection efficiency for EELS is *ca.* 80–90%, relative to 5–6% for EDS. This leads to a greater sensitivity and spatial resolution for EELS, with elemental mapping of *ca.* $\geq$ 2 Å. The intensity is greater than EDS for light elements since the signal generated by EELS represents the total sum of the number of X-ray photons and Auger electrons emitted from the sample. This technique is useful for elements with $Z > 1$, and like EDS, is amenable for elemental mapping of a sample surface.

As illustrated in Figure 7.23, a dedicated STEM is usually fitted with an EELS detector, which collects the low-angle scattered electrons that pass by the HAADF. Although the beam electrons have energies of several hundred keVs, the electrons being transmitted through the sample only have energies on the order of a few eVs. In order to perform EELS, it is therefore necessary to detect very small differences in the kinetic energies of the electrons. This is accomplished using a magnetic prism that exerts a centripetal force on each electron, causing a circular motion. While in the magnetic field, electrons move along the arc of a circle, whose radius is based on the speed and kinetic energy of the electron. In reality, there is nothing new with this concept; a magnetic prism is actually analogous to the dispersion of white light into a colored spectrum using a glass prism. However, unlike a glass prism, the magnetic field focuses the electrons as they exit the field, generating a spectrum from the grouping of electrons that exhibited identical energy losses.

In addition to detecting/quantifying particular elements in a sample, EELS also provides detailed elemental information such as the electronic structure, bonding, and nearest neighbor distribution of the atoms in the sample.[42] A representative EELS spectrum for a NiO surface is shown in Figure 7.26. The most intense features are peaks corresponding to zero-loss – those electrons that were either unscattered, or elastically scattered, while traversing through the sample. At relatively small energy losses (*ca*. 5–25 eV), a plasmon peak is observed which corresponds to the collective

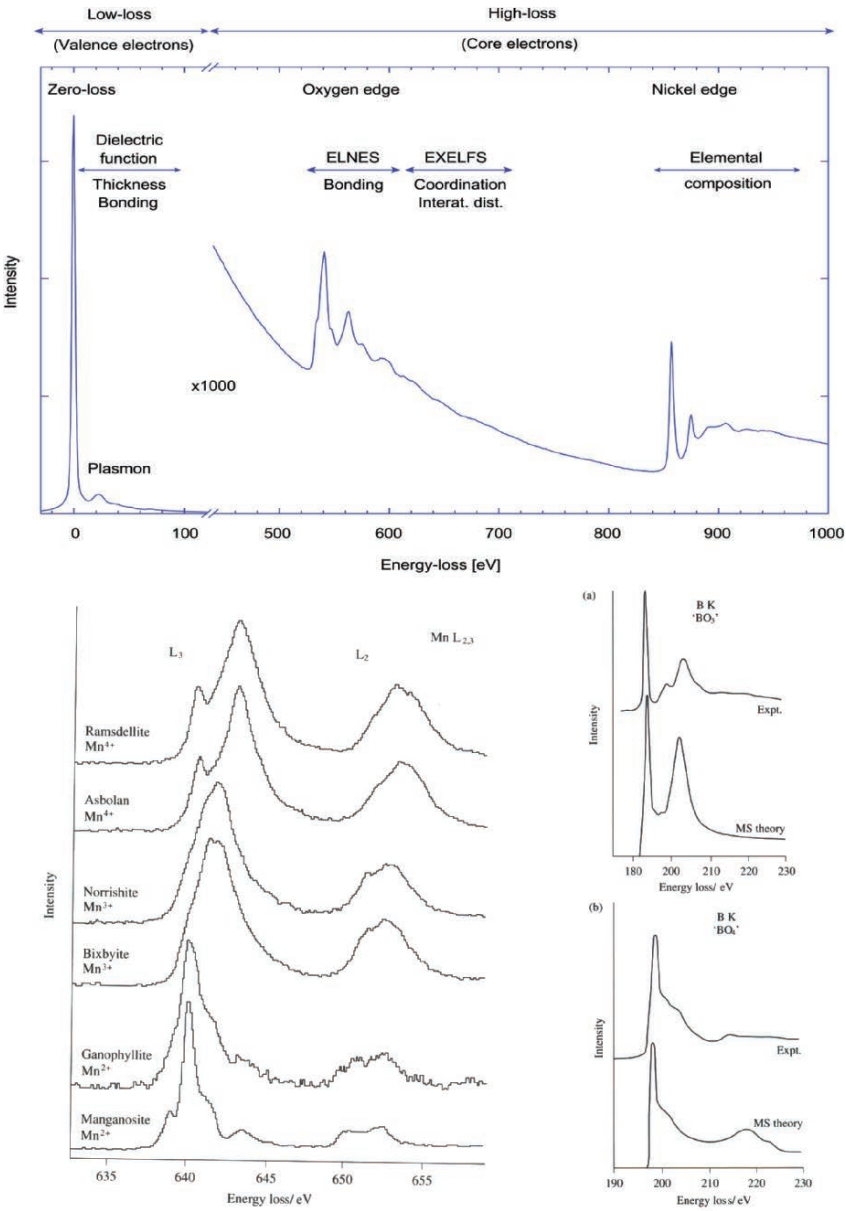


Figure 7.26. Electron energy-loss spectroscopy (EELS) spectra. Shown (top) is a representative EELS spectrum of a nickel oxide sample. A typical EELS spectrum shows a zero-loss peak that represents the unscattered or elastically scattered electrons, the near-edge fine structure (ELNES), and extended energy-loss fine structure (EXELFS). Also shown (bottom) are the "fingerprint" regions of an EELS spectrum, just beyond the core-electron edges, which provide information regarding the detailed bonding and chemical environment of the desired element.[46]

oscillation of many outer-shell (valence or conduction) electrons. The most useful application for this peak is the accurate determination of the sample thickness, of up to several thousand nanometers with a precision of a few percent (Eq. 10).[43] More recently, the plasmon region of the spectrum has been used to delineate variations in the size and geometry of metal nanoparticles.[44] As we saw in Chapter 6, the plasmon resonance frequency is directly related to the effective nuclear charge and size/shape of the charge distribution. Hence, the low-loss region of an EELS spectrum also provides information about bonding interactions and the dielectric function of the sample:

$$T \propto \log\left(\frac{I_p}{I_z}\right), \quad (10)$$

where T is the sample thickness; I_p, the intensity of the plasmon peak; and I_z is the intensity of the zero-loss peak.

At higher energy losses, an EELS spectrum will exhibit a variety of sharp features known as "edges," which are diagnostic for the presence of specific elements. The positions of the edges correspond to the binding energies of the core electrons in the sample. As shown in Figure 7.26 (top), the K-edges for O and Ni are 530 and 860 eV, respectively. Once the background is subtracted, the area under each edge peak(s) is integrated in order to determine the elemental concentrations. The shape of the peak immediately surrounding the edge is aptly referred to as the electron-loss near-edge structure (ELNES). As you might expect, these features are directly dependent on the exact band diagram and density of states (DOS) of the solid being analyzed. As such, this profile may be considered as the electron-scattering counterpart of X-ray absorption near-edge structure (XANES).[45] This region of the spectrum relates to the electronic structure, oxidation state, and bonding hybridization/symmetry of the desired element (Figure 7.26 (bottom)).[47]

Whereas the ELNES region typically extends to *ca.* 20 eV beyond the edge, the extended energy-loss fine structure (EXELFS) provides chemical information from the scattering of electrons by neighboring electrons. Accordingly, this region of the EELS spectrum is the electron-scattering counterpart of extended X-ray absorption fine structure (EXAFS)[42]; both being used to determine nearest-neighbor distances, oxidation states, and coordination numbers of the element being probed (see Figure 7.26 (bottom)).

The spectral fine details from EELS result from dipole-scattering of the incoming electrons. Much like IR spectroscopy, the incoming charged electron is influenced by a vibrating dipole at the sample surface. Hence, the energy loss of the electron is based on the amount of energy that was deposited into the vibrational mode.[48] The compilation of the various regions in an EELS spectrum provides a fingerprint of the surface species being analyzed; hence, this technique is commonly used to probe the localized vibrational modes of chemisorbed molecules on a surface – relevant for the study of any heterogeneous process, including the development of novel catalysts for chemical syntheses, and gas sensors for fuel cell and homeland security applications.[49]

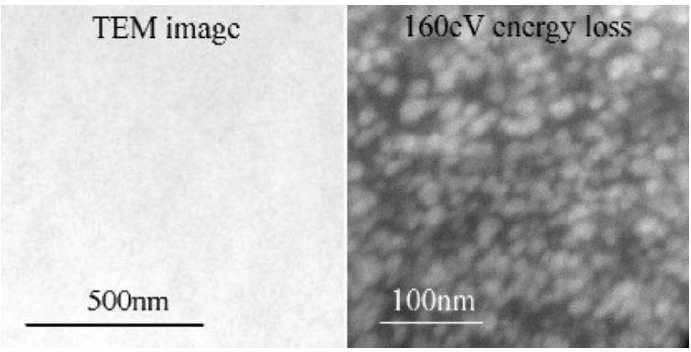


Figure 7.27. Energy-filtered TEM showing silica nanoparticles embedded within an organic coating. The conventional bright-field TEM image (left) shows little/no contrast relative to the EF-TEM image (right).[50]

It should be noted that energy-filtered TEM (EF-TEM) images may be formed with electrons that have lost a specific energy, with respect to a predetermined cut-off energy of the atomic inner shell. Chemical mapping of the surface is possible in this mode, which allows one to determine the exact location of elements in the sample – most useful for surfaces that contain low-Z elements (Figure 7.27). In addition, EF-TEM may be used to illustrate the valence-state distribution of a particular element across a surface, based on slight differences observed in the ELNES region of the EELS spectrum.

7.2.2. Scanning Electron Microscopy (SEM)

In contrast to TEM, with typical sample thicknesses in the range of 10 nm–1 μm, sample depths for SEM often extend into the 10–50 mm range. As such, this technique is most often used to provide a topographic image of the sample surface. However, the electron beam is not confined to the top of the surface, but also interacts with lower depths of the sample. Consequently, SEM provides information regarding the species present at varying depths of the sample (Figure 7.28):

(i) Elastic scattering of electrons by atomic nuclei of the sample results in BSE – useful for generating images based on Z-contrast.
(ii) Inelastic scattering of electrons by sample atoms results in low-energy secondary electrons – useful for providing topographic information regarding the sample surface.
(iii) Inelastic scattering of electrons by sample atoms results in X-ray generation (characteristic and Bremsstrahlung background X-rays) from lower sample depths – useful for chemical analysis of the bulk sample.
(iv) Inelastic scattering results in Auger electrons emitted from sample atoms near the sample surface – useful for surface chemical analysis.

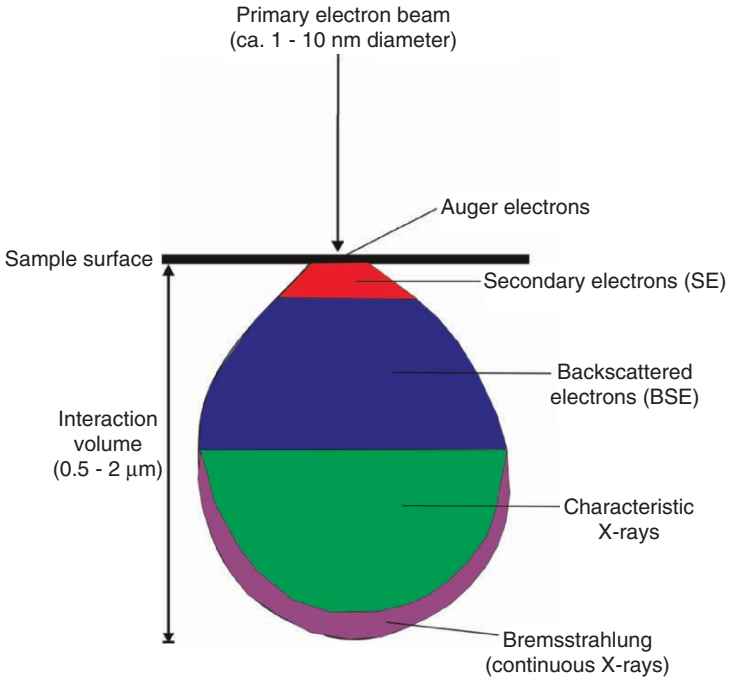


Figure 7.28. Illustration of the sample interaction volume, and the corresponding particles emitted from various sample depths. The exact volume will depend on the accelerating voltage of the electron beam and sample composition.

As you will recall from Figure 7.6b, for imaging and chemical analysis using SEM, the particles generated from the sample must reach the detector situated above the sample surface. Since the relative energies of the generated particles are in the order:

$$\text{Auger electrons} < \text{SE} < \text{BSE} < \text{X-rays}$$

The least energetic emissions will not reach the surface from lower depths of the sample. For instance, Auger electrons that are emitted from deeper regions of the sample lose their energy through collisions with sample atoms before they reach the surface. As a result, AES is a very sensitive technique to probe the chemical composition of only the top 50–100 Å (*i.e.*, 15–30 monolayers). In comparison, the maximum escape depth of secondary electrons has been estimated as 5 nm in metals, and 50 nm in insulators.

Not surprisingly, both the beam current (or accelerating voltage) and sample density will greatly affect the *interaction volume* of the bulk sample with primary electrons (Figures 7.29 and 7.30, respectively). As we saw earlier, the probability for elastic scattering increases with Z^2 (Eq. 9); hence, as the density of the sample increases, the number of BSEs will increase, reducing the number of electrons that may penetrate to deeper regions of the sample. Similarly, as the energy of the incident

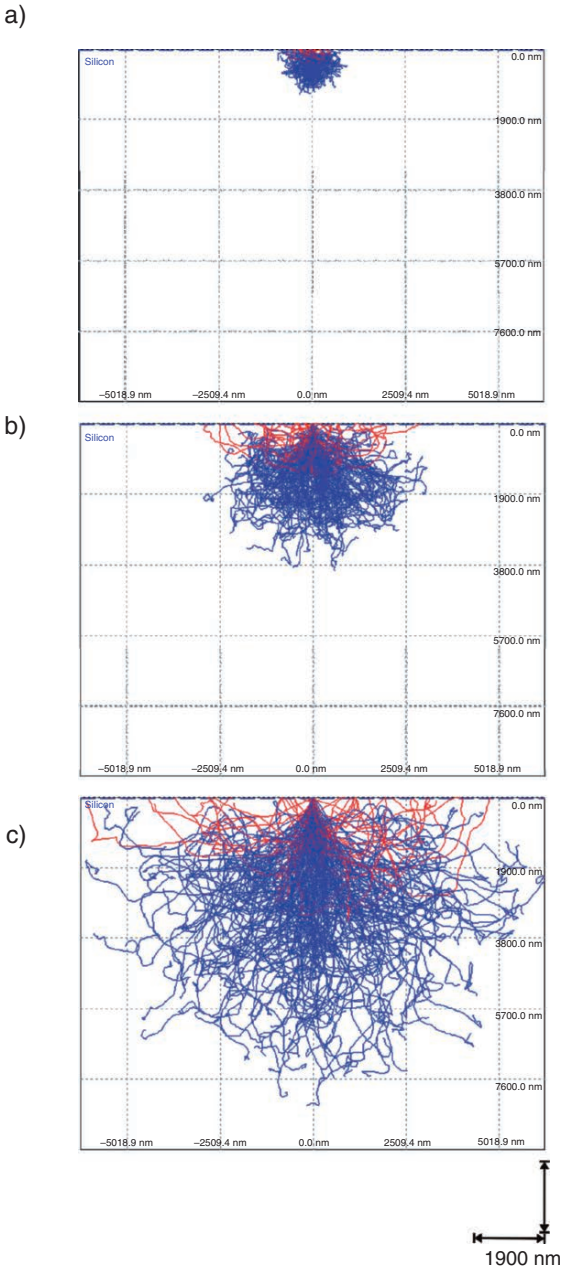


Figure 7.29. Electron-beam penetration volume resulting from varying the beam current. Shown are X-rays (blue) and backscattered electrons (red) generated from a silicon substrate. The beam currents are (a) 10 keV, (b) 20 keV, and (c) 30 keV.[54]

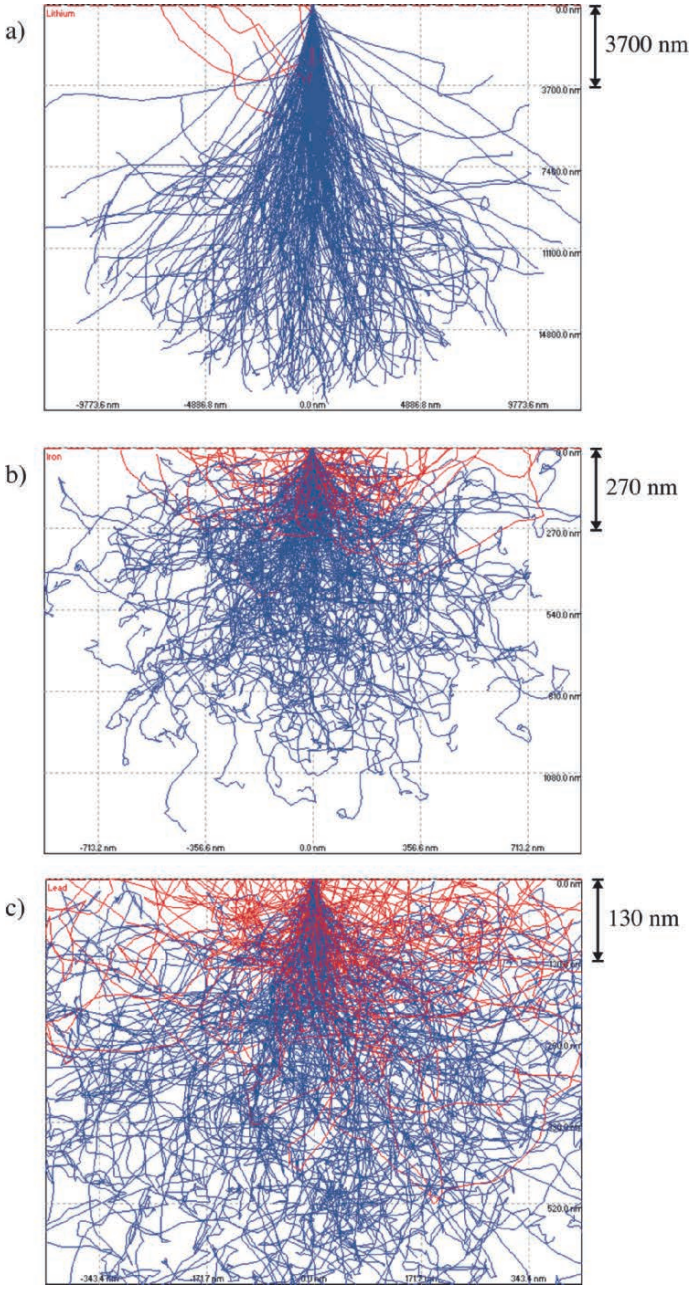


Figure 7.30. Electron-beam penetration volume resulting from varying the sample density. Shown are X-rays (blue) and backscattered electrons (red) resulting from electron beam impingement on (a) Li, (b) Fe, and (c) Pb. The beam current is 20 keV, and beam diameter 10 nm for all metals.[54]

beam decreases, fewer inelastic collisions with sample atoms are needed to bring the electrons to rest, thus decreasing the penetration depth into the sample.

The image formed from SEM is primarily the result of secondary electron emission from sample surface. The Law of Conservation of Energy ensures that any energy lost by the primary beam electrons must be transferred to the secondary electrons that are ejected from the sample atoms.[51] Those with sufficient energy to traverse the sample surface reach an Everhart–Thornley detector, which consists of a scintillator and photomultiplier tube (PMT).[52] The topographical contrast that arises from an uneven surface is due to a differing number of SEs being released from the sample (Figure 7.31).

Structure determination using SEM

In addition to displaying the familiar bright-field images from secondary electron emission, BSE in a SEM may be used to determine the crystallography of (poly)crystalline samples. This technique is referred to as *electron backscattering diffraction* (EBSD) or *backscattering Kikuchi diffraction* (BKD), used to measure individual crystallite orientations, as well as crystallographic parameters of the sample (*e.g.*, interplane spacings/angles, crystal symmetry elements, *etc.*). The EBSD patterns are generated from the interaction of the incident electron beam with a highly tilted (*ca.* 70° from horizontal) planar sample.

When the electrons impinge on the crystalline sample, they interact with individual lattice planes. When these interactions satisfy the Bragg condition, they exhibit backscattering diffraction and (due to the tilted sample) are directed toward a phosphor screen where the fluorescent pattern is detected by a CCD camera. The resulting pattern consists of a large number of intersecting bands, known as Kikuchi lines, which represent the unique crystallographic properties of the crystal

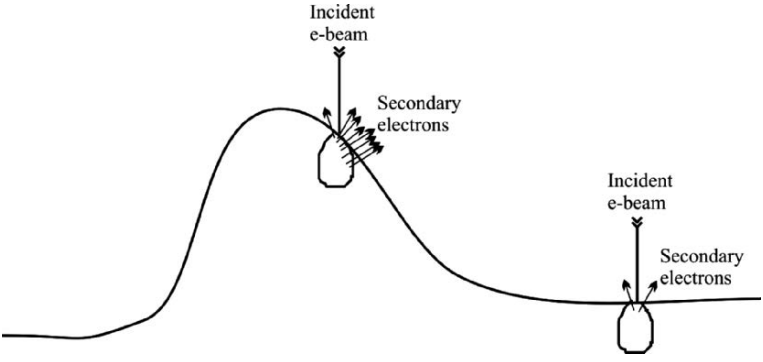


Figure 7.31. Illustration of topographic imaging of a surface using SEM. When the incident beam strikes the side of a feature, many secondary electrons are released from the side and top of the interaction volume ("edge effect"). In comparison, the interaction of the primary beam with a flat substrate releases significantly fewer secondary electrons that originate near the sample surface. The image contrast results from varying numbers of secondary electrons reaching the detector as the beam is rastered across the sample surface.

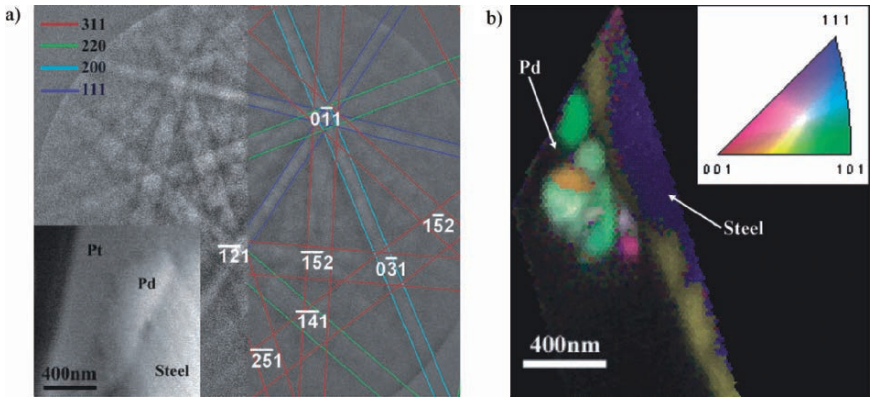


Figure 7.32. (a) SEM micrograph (left bottom inset) of a Pd particulate selected for crystallographic analysis. Also shown are the crisscrossing Kikuchi lines originating from various crystallographic planes. The indices of each plane are indicated in the color legend at the top-left corner. (b) The orientation imaging microscopic (OIM) profile of the Pd particulate, showing a microcrystalline array (color-coated based on the crystallite orientations). Reproduced with permission from Bera, D.; Kuiry, S. C.; Seal, S. *J. Phys. Chem. B* **2004**, *108*, 556. Copyright 2004 American Chemical Society.

(Figure 7.32a). Computer software is used to collect/analyze the resulting patterns to determine the crystallography of the material.[53] In association with compositional data from EDS, the exact phase of the material may be identified from a library of known materials. For a more thorough crystallographic description, *orientation imaging microscopy* (OIM) is often carried out in tandem with EBSD. This technique consists of stepping the beam across the sample, with automatic indexing of the resulting EBSD patterns. The resulting OIM map readily reveals the crystallographic orientations and grain boundaries in three-dimensions (Figure 7.32b) – of use for applications such as sensor/heterogeneous catalyst design and micro/nano defect analyses.

Sample considerations and AES

Sample preparation for SEM analysis is trivial relative to TEM, with the sample simply deposited onto the top of an adhesive fastened to an aluminum stub/holder. Most often, conductive carbon tape is used to sequester the sample; for FESEM, problems with outgasing usually dictates the use of carbon/silver paint, epoxy, or copper/aluminum tape with adequate drying prior to analysis. After allowing the solvent to fully evaporate (if present), a thin conductive film (*ca.* <10 nm) of Au or C is often deposited onto the sample surface.[55] This is especially required if a nonconductive adhesive is used to mount the sample, or if the sample itself is nonconductive. The preparative steps involving conductive mounting materials is important to prevent *charging* – the buildup of electrons on the sample surface, which dramatically affects the imaging ability of the SEM (Figure 7.33). Sometimes, it is even necessary to use an additional amount of carbon paste or tape to create a conductive path between a tall sample and the aluminum holder. It should be noted that metallic

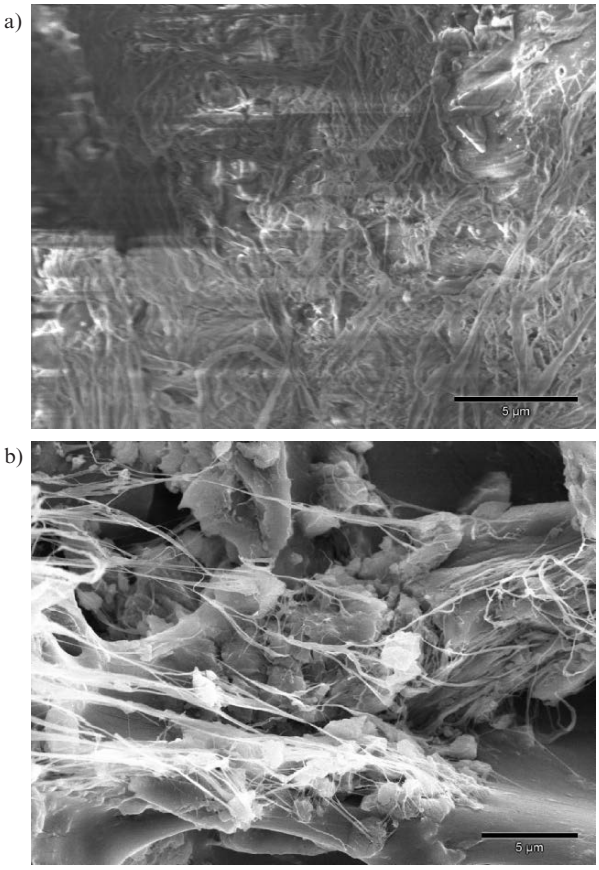


Figure 7.33. SEM images of amorphous carbon nanofibers (nonconductive) grown at room temperature from a dendritic catalyst. Shown is (a) as-formed nanofibers, without a gold coating and (b) after sputtering a thin conductive gold coating on the surface. The uncoated sample exhibited extreme charging (a), which thermally degraded the sample and caused movement of the sample during imaging.[60]

coatings also serve another important use – to increase the SE emission of a sample with a low yield of secondary electrons (*e.g.*, comprising low-Z elements).

In addition to imaging applications, SEM is also widely used for elemental analysis and chemical mapping of surfaces, using EDS/WDS. However, unlike TEM, facile cross-section imaging/chemical mapping may also be performed using specialized sample holders (Figure 7.34). Besides monitoring characteristic X-rays from sample elements, Auger electron emission may be analyzed – a technique known as AES (Figure 7.35a). Analogous to the STEM extension of TEM, scanning auger microscopy (SAM) is also possible, wherein the incident electron beam is scanned across the sample. Perhaps the most intriguing aspect of AES/SAM instruments is the capability of depth profiling. An argon ion beam is used to etch away monolayers of the surface, allowing for compositional studies as a function of sample depth (Figure 7.35b).

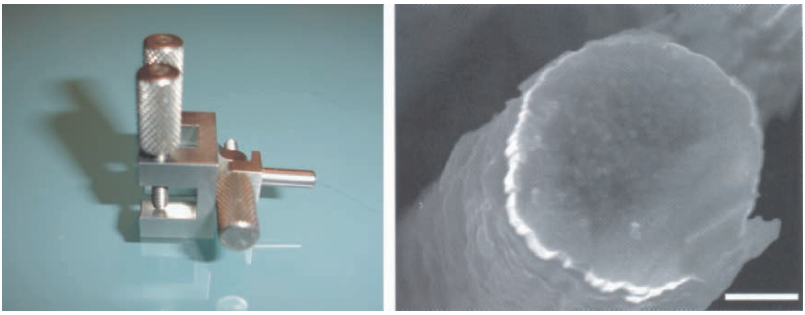


Figure 7.34. Photograph of a variable-angle SEM sample holder, with an example of a cross-section SEM image (scale bar is 2 μm).[61]

A simple way to distinguish between SEM and SAM is the former collects secondary electrons to form an image; the latter collects Auger electrons for chemical analysis. Recent developments have now afforded dual-detector electron microscopes that are equipped with standard FESEM/EDS capabilities in addition to SAM. This powerful combination offers the direct superimposing of both EDS and SAM chemical maps onto the corresponding high-resolution image.[56] In addition to being more sensitive for light elements (±0.5 at% for Li–U), it should be noted that AES is more suited for surface analysis than EDS. Due to the low kinetic energies of Auger electrons (*ca*. 50 eV–3 keV), chemical information is only obtained from sample depths of *ca.* <50 Å (compared to 1–2 μm for EDS). As such, it is not always possible to coat the sample to prevent charging, which explains why SAM is most often applied for compositional studies of conductive samples.[57]

Environmental SEM

Though the analysis of samples in the presence of solvent is normally a *faux pas* for the UHV environment of any electron microscope, there are now instruments known as *environmental SEMs* (ESEMs) that are capable of such studies. These instruments have been in development since the early 1970s,[58] for *in situ* studies of virtually any material (wet/dry, insulating/conducting). Before this major technological development, samples such as paints, inks, and biological specimens had to be dried completely to maintain the integrity of the vacuum system. Typically, an ESEM is not simply a modified SEM, but rather a specially designed instrument that is capable of regular (FE)SEM imaging, as well as "wet mode" operation.[59]

The operation of an ESEM is made possible through use of a differential pumping system (Figure 7.36) that maintains a UHV environment (10^{-7} Torr) required for the electron gun, while allowing the presence of gases in the sample chamber (10–20 Torr). The pressure and temperature of the sample chamber may be strictly controlled, inducing evaporation or condensation events.

The nonvacuum conditions within the sample chamber require a different type of detection system relative to conventional SEMs, referred to as a *gaseous secondary*

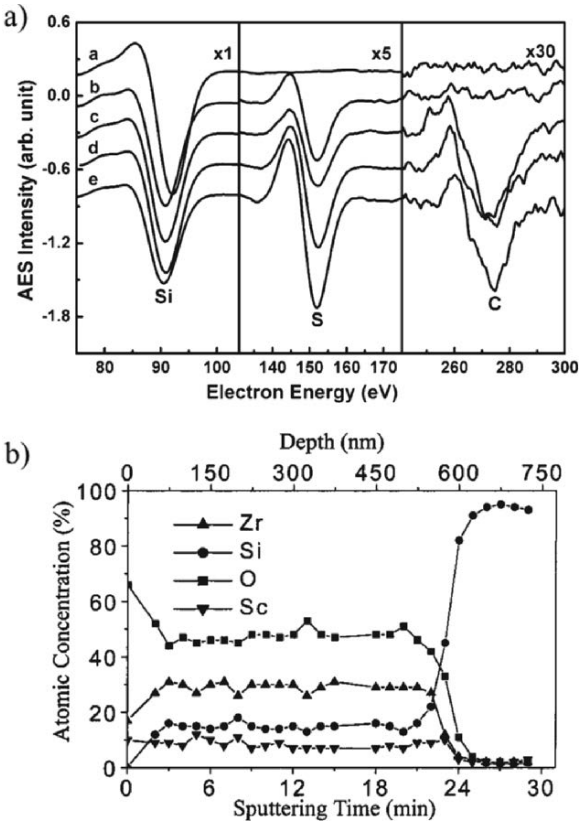


Figure 7.35. (a) AES analysis of organosulfur compounds adsorbed onto Si(100). Shown are the Si(LVV), S(LVV), and C(KLL) transitions for (a) clean Si(100), and after saturation exposure by (b) H_2S, (c) CH_3SCH_3, (d) CH_3SH, and (e) CH_3SSCH_3.[62] (b) AES depth profile for a $(Sc_2O_3)_{0.08}(ZrO_2)_{0.92}$ thin film deposited onto a Si(100) wafer. This profile shows a homogeneous Sc:Zr:O:Si atomic ratio between 75 and 500 nm, with an interdiffusion layer thickness of <100 nm. By examining the rise of the Si peak, the film thickness can be readily determined.[63]

electron detector (GSED). Due to the energetic nature of the primary beam, there is little scatter from its interaction with the gaseous medium about the sample. *En route* toward the positive electrode, the secondary electrons generated from the sample repeatedly collide with gas molecules, generating a cascade of additional electrons and positive ions (Figure 7.36). In addition to amplifying the SE signal, the positive ions migrate back to the sample surface where they dissipate the charge buildup – hence, precluding the need for conductive samples or carbon/gold coating. The aperature diameter through which the primary electron beam is passed determines the overall maximum pressure of the sample chamber. For instance, a 0.5 mm aperature dictates a maximum pressure of 10 Torr about the sample; a 1 mm aperature would lower the maximum pressure to 5 Torr.

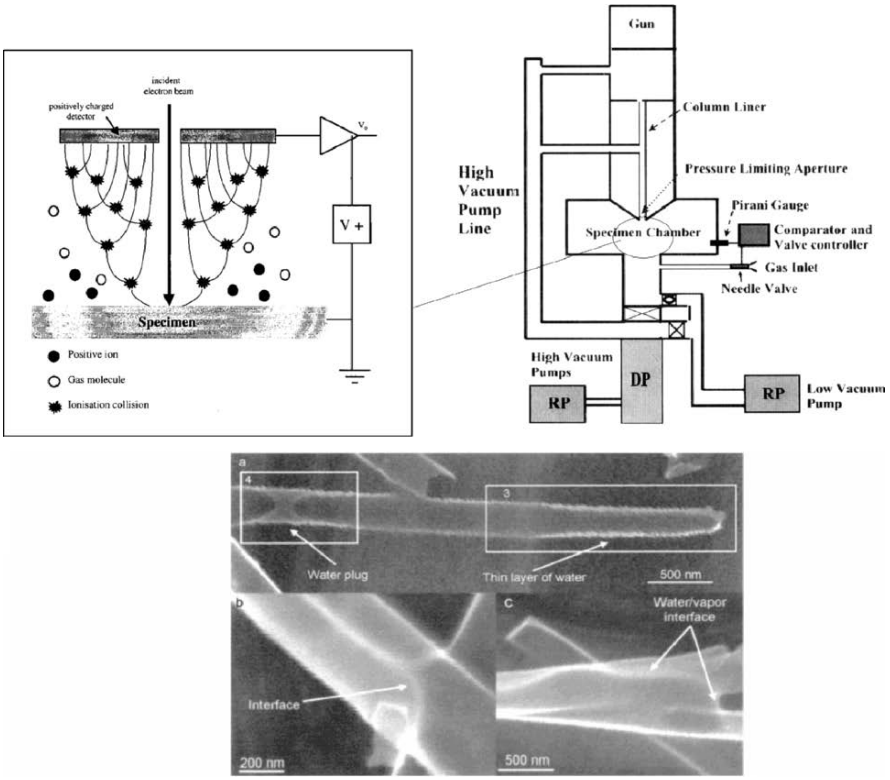


Figure 7.36. Schematic of the differential pumping arrangement of an ESEM (top right)[52] and the cascade amplification process occurring in the gaseous secondary electron detector (GSED) (top left).[64] The abbreviations RP and DP refer to the rough pump and diffusion pump, respectively. Also shown (bottom) is an intriguing application for ESEM – the *in situ* imaging of water condensation inside a carbon nanotube (a), and underneath/around the CNT (b, c).[65]

7.2.3. Photoelectron Spectroscopy

The photoelectric effect, first outlined by Einstein in the early 1900s, refers to the ejection of electrons from a surface due to photon impingement. However, it was not until the 1960s that this phenomenon was exploited for surface analysis – a technique referred to as *X-ray photoelectron spectroscopy* (XPS), or *electron spectroscopy for chemical analysis* (ESCA). This technique consists of the irradiation of a sample with monochromatic X-rays (*e.g.*, Al K_α (1.487 keV), Mg K_α (1.254 keV), Ti K_α (2.040 keV)), which releases photoelectrons from the sample surface (Figure 7.37).[66] Due to the short *free mean path* (FMP) of the photoelectrons in the solid, this technique provides compositional information from only the top 1–5 nm of a sample.

Each atom in the sample has characteristic binding energies of their inner-shell electrons, referred to as *absorption edges*. In order to excite the electrons, the energy

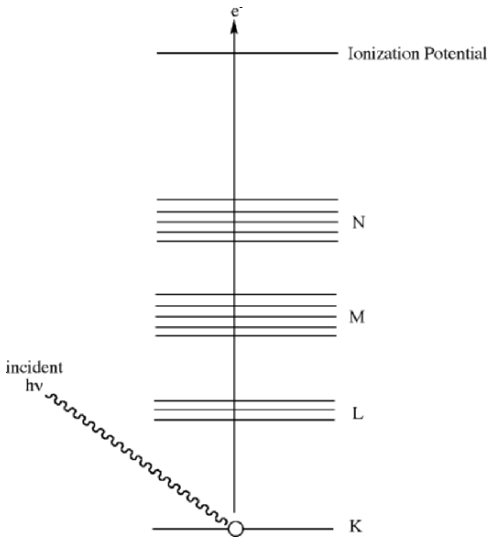


Figure 7.37. The energy-level diagram for X-ray photoelectron spectroscopy (XPS).

of the incident photons must be at least as large as the binding energy of the electrons. When this energy threshold is exceeded, a large absorption of energy takes place, followed by the release of photoelectrons with excess kinetic energy (Eq. 11) in order to relax the atom back to its stable ground state. Since the binding energy of an atom is altered by minute changes in its chemical environment (*e.g.*, oxidation state, hybridization/geometry, *etc*.), XPS provides both elemental quantification and details regarding the chemical environment of the surface atoms (Figure 7.38) [67]:

$$E_k = h\nu - E_b, \quad (11)$$

where E_k is the kinetic energy of the emitted photoelectrons; $h\nu$, the energy of the incident photons; and E_b is the binding energy of the inner-shell electrons.

In addition to using X-rays to irradiate a surface, ultraviolet light may be used as the source for *photoelectron spectroscopy* (PES). This technique, known as *ultraviolet photoelectron spectroscopy* (UPS, Figure 7.38), is usually carried out using two He lines (HeI at 21.2 eV and HeII at 40.8 eV), or a synchrotron source. This technique is often referred to as "soft PES," since the low photon energy is not sufficient to excite the inner-shell electrons, but rather results in photoelectron emission from valence band electrons – useful to characterize surface species based on their bonding motifs. It should be noted that both UPS and XPS are often performed in tandem with an Ar^+ source, allowing for chemical analysis of the sample at depths of $\leq$1 μm below the surface.

Though a typical XPS detector collects all emitted photons, regardless of their ejection angles, it should be noted that *angle-resolved* XPS (ARPES) and UPS (ARUPS) may also be carried out. By detecting photoelectrons emitted from a surface at different emission angles, one obtains the energy of the electrons as a function

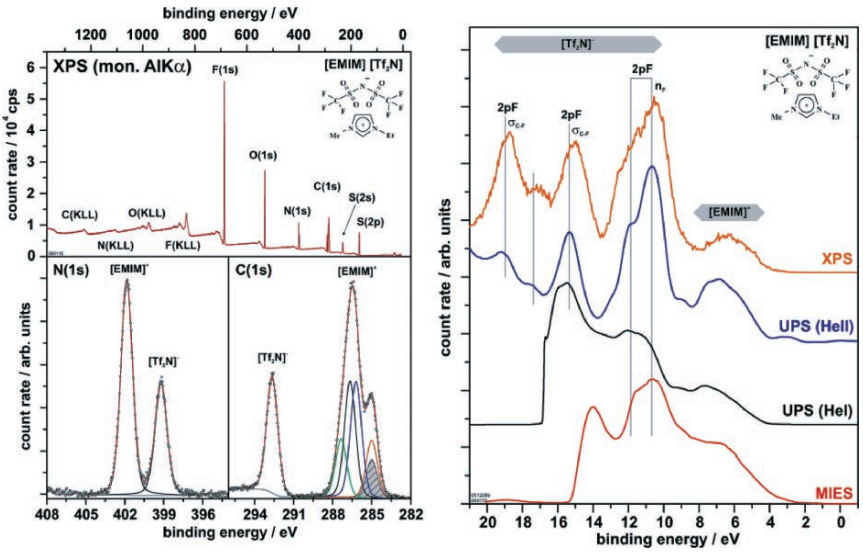


Figure 7.38. XPS spectrum of an ionic liquid, [EMIM][Tf_2N], detailing the C(1s) and N(1s) regions. Since there are no peaks from the Au substrate, the film thickness is likely >10 nm. Also shown (right) is the comparison between XPS, ultraviolet photoelectron spectroscopy (UPS, HeI = 21.2 eV, HeII = 40.8 eV radiation), and metastable impact electron spectroscopy (MIES). Whereas XPS and UPS provide information from the first few monolayers of a sample, MIES is used for zero-depth (surface only) analysis, since the probe atoms are excited He atoms that interact with only the topmost layer of sample. Full interpretations for these spectra may be found in the original work: Hofft, O.; Bahr, S.; Himmerlich, M.; Krischok, S.; Schaefer, J. A.; Kempter, V. *Langmuir* **2006**, *22*, 7120. Copyright 2006 American Chemical Society.

of the momentum vector, **k**. This is referred to as "band mapping," since this analysis probes the electronic structure of crystalline materials.[68]

Structure determination using XAFS

If a tunable X-ray source such as synchrotron radiation[69] is used to generate the incident photons rather than a monochromatic beam, one can select specific absorption edges to gain more detailed information regarding the chemical environment of specific atoms. This technique, known as XAFS, is divided into two spectral regions – XANES and EXAFS – which are analogous to ELNES and ELEFS in EELS studies. Whereas XANES is sensitive to the oxidation state and bonding geometry of the probed element, EXAFS is useful to determine the distances and coordination numbers of neighboring species.[70] The small oscillations that are present in an EXAFS spectrum are indicative of the interference effects that occur as the photoelectrons leave the sample surface. As you might expect, this backscattering effect becomes most pronounced with increasing photon energies, and with smaller interatomic distances between the probed atom and its nearest neighbors. As a further extension of this technique, the X-rays may be reflected from a surface at an angle in order to yield even more structural information. This technique is referred to as

total external reflection EXAFS, or REFLEXAFS,[71] particularly useful to determine chemical information from greater sample depths without the need for destructive Ar^+ etching.

In order to illustrate the complementary information yielded from XPS and XAFS, we will consider the recent work of Gervasini and coworkers, regarding the characterization of CuO catalysts dispersed on silica and silica/alumina supports.[72] The XANES, EXAFS, and XPS spectra for these catalyst species are illustrated in Figure 7.39. The EXAFS spectrum (Cu K-edge) for a $Cu/SiO_2/Al_2O_3$ catalyst, calcined in air, has one primary peak at 1.939 Å corresponding to an (octahedral) environment of six O atoms (Figure 7.39a, thick line).[73] The lack of other O shells suggests that the surface contains isolated Cu ions, likely as a Cu-aluminate phase. In contrast, the same catalyst on a SiO_2 support (Figure 7.39a, thin line) shows the same octahedral environment of O atoms (1.947 Å), and another large peak corresponding to *ca*. 5 Cu atoms at an average Cu-Cu distance of 2.987 Å. Since this is shorter than the Cu-Cu distances in crystalline CuO, the EXAFS spectrum suggests the formation of amorphous CuO aggregates on the SiO_2 support.

After reduction of the catalyst in a H_2 stream at the same temperature as calcination (Figure 7.39b), the Cu/SiO_2 spectrum shows a large peak at 2.547 Å, corresponding to *ca*. 8 Cu atoms. The spectrum closely matches that of the fcc metallic Cu reference, suggesting the presence of small metallic Cu atoms on the surface. On the Si/Al support, the same spectral features are observed; however, line-fitting indicates that the Cu atoms are somewhat electropositive until further reduction takes place at higher temperatures. It was postulated that this resistance toward reduction is likely due to the strong association of surface Cu ions with Lewis acidic aluminum centers of the support.

Not surprisingly, the XANES spectra of Cu/Si and Cu/SiAl (Figure 7.39c) correspond to the absorption edge of Cu^{2+}. However, the spectra are significantly different from the CuO reference. This suggests the presence of a distorted octahedral geometry about the Cu ions, where the Cu may be bound to three different types of O atoms: (i) from the support, (ii) from surface –OH groups, and (iii) from water molecules.

The XPS spectra (Figure 7.39d,e) show a broad peak at 934 eV, corresponding to the Cu L-edge of Cu(2p $\rightarrow$ 3d transition). Since the binding energy for Cu^{2+} and Cu^+ are similar (*i.e.*, 933.6 and 932.5 eV, respectively), the shape of the Cu 2p peaks was examined in detail to determine the valence states of the surface Cu species. Whereas Cu on Si/Al exhibits one contribution, Cu on a SiO_2 support shows the presence of two different Cu species. In particular, this suggests that the Cu^{2+} ions interact with both =O and –OH groups from the silica support. Furthermore, satellite peaks from the spin–orbit components[74] (*i.e.*, $Cu(p_{1/2})$ and $Cu(p_{3/2})$) of the Cu(2p) peak are observed only for Cu on a SiO_2 support; this suggests that the Si/Al support contains Cu atoms in a lower valence state, such as Cu^+ speciation.[75]

Lastly, XPS was also used to gleam information regarding the dispersion of Cu on the two surfaces. The $Cu(2p_{3/2})$, Si(2s), and Al(2p) peaks were integrated to yield the respective elemental concentrations of each supported catalyst. The molar ratios between the surface species (*i.e.*, Cu(2p)/Si(2s) or Cu(2p)/Si(2s) + Al(2p)) and the total metal species (*i.e.*, Cu/Si or Cu/(Si + Al), as obtained from inductively

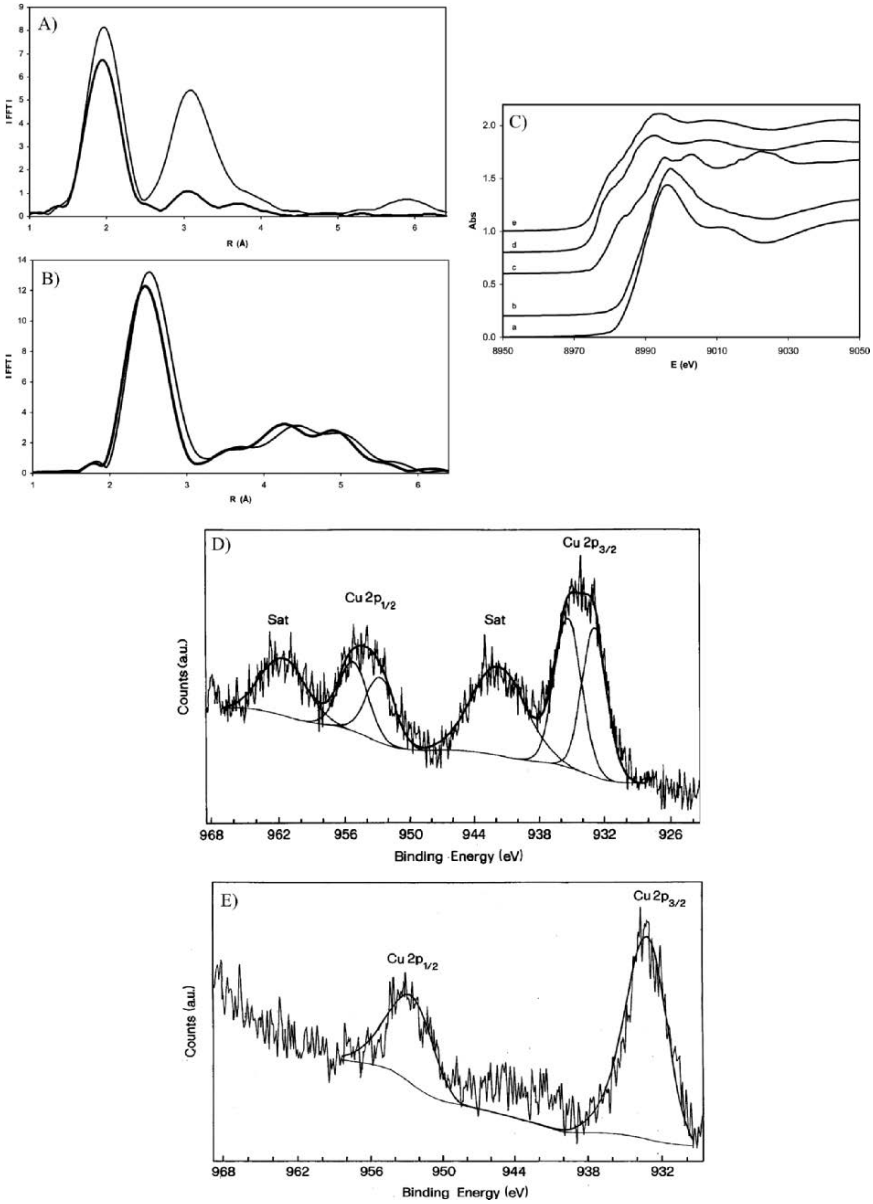


Figure 7.39. XAFS CASE STUDY: surface characterization of dispersed CuO catalysts on silica and alumina/silica supports. Shown are: (A) EXAFS of the Cu K-edge of the catalyst calcined in air at 543 K. The thick and thin lines indicate Si/Al and Si supports, respectively; (B) EXAFS of the Cu K-edge of the catalyst reduced under a H_2 flow at 543 K; (C) Cu K-edge XANES spectra of the calcined catalyst on (a) SiO_2, (b) SiO_2/Al_2O_3 supports, along with references of (c) Cu foil, (d) Cu_2O, and (e) CuO; (D) XPS spectrum of the Cu 2p core level of the calcined catalyst on a Si support; (E) XPS spectrum of the catalyst on a Si/Al support. Reproduced with permission from Gervasini, A.; Manzoli, M.; Martra, G.; Ponti, A.; Ravasio, N.; Sordelli, L.; Zaccheria, F. *J. Phys. Chem. B* **2006**, *110*, 7851.

coupled plasma (ICP) analysis) resulted in values near 1.0 and 0.35 for the SiO_2 and SiO_2/Al_2O_3 supports, respectively. The surface deficiency of Cu on the Si/Al support is proposed to be an artifact of Cu encasement within the pores of aluminum-rich regions, as also suggested from XAFS.

7.3. SURFACE CHARACTERIZATION TECHNIQUES BASED ON ION BOMBARDMENT

Thus far, the majority of surface techniques have employed the "collide and collect" principle, where either electrons or X-rays impinge on the sample, and various emissions are collected/analyzed. However, a number of other ions may also be used for such bombardment-based analyses. For instance, we already saw an example (Figure 7.38) of *metastable impact electron spectroscopy* (MIES), in which metastable He atoms (in an excited $2s^1$ state, with an energy of 19.8 eV and a lifetime of *ca*. 4,000 s) are directed toward a sample. As the atom collides with the sample, the 2s electron tunnels into an empty valence orbital in a sample atom, releasing an Auger electron. A similar surface-relaxation mode may be carried out by incident He ions, which is termed *ion neutralization spectroscopy* (INS). Since the incident atoms/ions are not sufficiently energetic, the beam does not penetrate through the sample surface. Hence, the spectra resulting from these techniques provide information regarding the DOS for the topmost surface monolayer of the sample.[76]

If a high-energy beam (2–4 MeV) of He (α-particles) or H ions[77] collides with a surface, information may be gained regarding the composition and thickness of a surface coating (Figure 7.40). This technique is known as *Rutherford backscattering* (RBS), very similar to the original scattering experiments conducted by Rutherford in the early 1900s – essential in the early development of a structural description of the atom. As the energetic ion penetrates the material, it loses energy through collisions with electrons, and (less frequently) with nuclei. When the positively charged He^+ ion approaches the nucleus of a sample atom, it will be electrostatically repelled. As you would expect, the energy of the backscattered ions will depend on their incident energy, as well as the mass of the surface atom that causes the scattering. For instance, contact with high-Z atoms such as gold, will result in the energy of the backscattered ions being almost equivalent as the incident beam. By measuring the energy spectrum of the backscattered ions, information on the composition of the elements, and their origination depth from the sample may be obtained. Hence, RBS provides facile compositional information regarding a sample, without the need for a reference sample.[78] This technique is most often applied to determine the chemical composition versus depth for heavy elements in a low-Z matrix, amenable for thicknesses of up to 0.5 μm (for He ions) and 10 μm (for H ions).

Based on our previous discussions, it is logical to assume that high-energy ions may dislodge inner-shell electrons, releasing characteristic X-rays (*cf.* EDS). This is the theoretical principle underlying *particle-induced X-ray emission* (PIXE), which uses high-energy H^+ or He^+ ions (at >2 MeV; produced from a van de Graaff accelerator[79]) as the incident beam. This technique offers a nondestructive quantitative analysis of a sample (for elements with $Z \geq 13$), often used by archaeologist

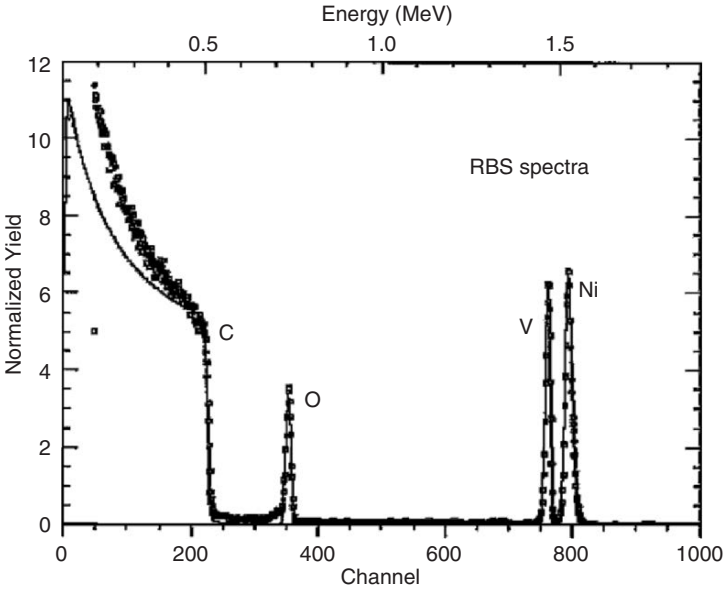


Figure 7.40. Rutherford backscattering (RBS) spectrum of a $LiNiVO_4$ film (thickness of 240(±5) Å) on a carbon substrate. The open squares represent the experimental RBS data, and the continuous line is the simulated data. Reproduced with permission from Reddy, M. V.; Pecquenard, B.; Vinatier, P.; Levasseur, A. *J. Phys. Chem. B* **2006**, *110*, 4301. Copyright 2006 American Chemical Society.

and art curators to assist with dating and authenticity assessments. With low detection limits (*ca*. 1–10 ppm), and ability to analyze a number of elements simultaneously, PIXE has also been widely used for the rapid analysis of multilayer thin films for microelectronics and optoelectronics applications. A recent extension, known as micro-PIXE, allows one to focus the ion beam to a diameter of *ca*. 1 μm, to allow for a more localized analysis (Figure 7.41). The primary benefit of this technique is the wide variety of samples that may be analyzed, from liquids to powders. Though no sample preparation is necessary for PIXE, it should be noted that the analysis is limited to the top 10–50 μm of the sample, depending on its density.

Among the techniques discussed thus far, the best Z-cutoff for elemental composition techniques has been Li (for XPS and AES). One ion-bombardment method, *elastic recoil detection analysis* (ERDA),[80] is able to nondestructively quantify light elements such as hydrogen or deuterium.[81] This technique uses a high-energy ion beam (*e.g.*, He, Au, Cl, C, or O ions), which causes light elements to be elastically ejected from the surface (*i.e.*, "recoiled") due to interactions between the incident beam and their atomic nuclei (Figure 7.42). By choice of the correct angle between the sample surface and detector, the incident beam is blocked by "stopper foil" to increase the signal/noise ratio for the measurement.

The most common mode of chemical analysis presented herein has been the monitoring of elastically/inelastically scattered or recoiled incident beam species, or

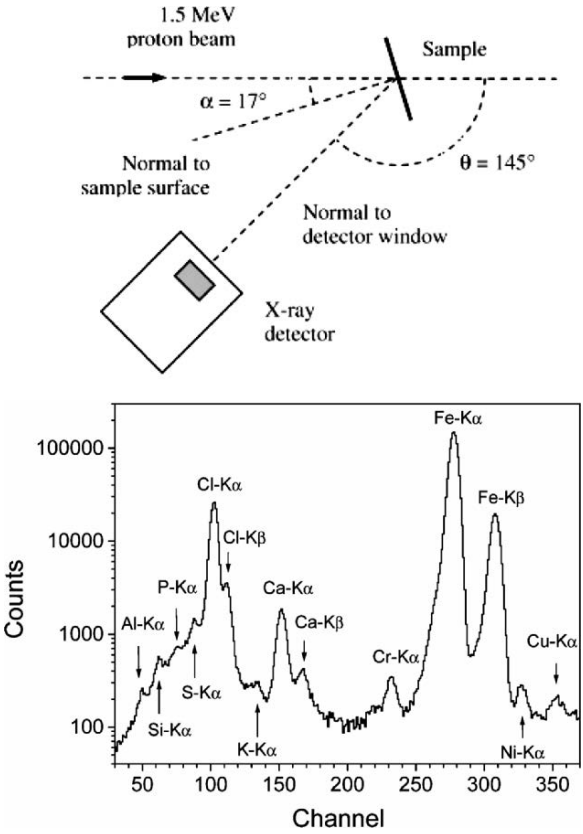


Figure 7.41. Experimental setup and X-ray spectrum resulting from PIXE analysis of single-walled carbon nanotubes. Reproduced with permission from Naab, F. U.; Holland, O. W.; Duggan, J. L.; McDaniel, F. D. *J. Phys. Chem. B* **2005**, *109*, 1415. Copyright 2005 American Chemical Society.

the analysis of a secondary emission pattern. In addition to the release of characteristic X-rays, Auger electrons, and photoelectrons, an incident beam may cause ionization of the sample. This technique is known as *secondary-ion mass spectrometry* (SIMS),[82] which represents the most sensitive surface characterization technique developed to date, with detection limits of 10^{10}–10^{15} atoms cm^{-3} (*i.e.*, 0.1 ppb − 0.1 ppm) and ability to detect/quantify any element in the Periodic Table. In contrast to common bulk MS methods such as matrix-assisted laser desorption/ionization (MALDI)[83] and electrospray ionization (ESI),[84] SIMS analyzes samples in their native state without the need for a suitable matrix solution. As a result, SIMS is the best choice to characterize organic-based thin films and polymer surfaces.

The operating principle of SIMS is not unlike other techniques in this section; that is, a high-energy (1–30 keV) ion source is directed onto a sample surface. However, the absorption of this energy by the top *ca.* 50 Å of the sample results in the sputtering

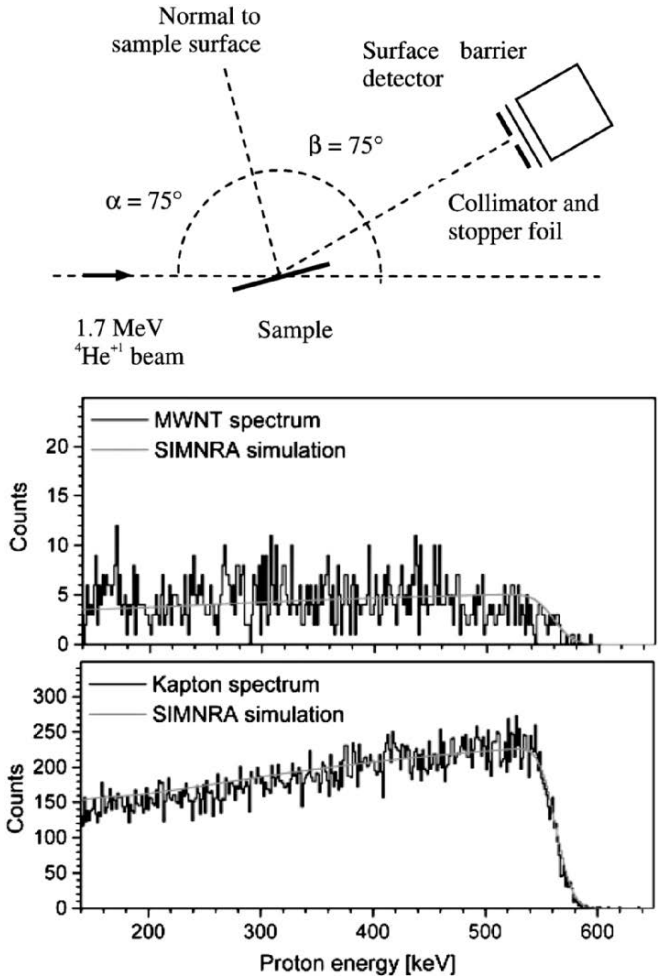


Figure 7.42. Experimental setup and eastic recoil detection (ERD) spectra for multiwalled nanotubes, and a Kapton polyimide film. A computer program, SIMRA, is useful to simulate ERD as well as other non-RBS data. Reproduced with permission from Naab, F. U.; Holland, O. W.; Duggan, J. L.; McDaniel, F. D. *J. Phys. Chem. B* **2005**, *109*, 1415. Copyright 2005 American Chemical Society.

of neutral and charged $(+/-)$ species from the surface.[85] These ejected species primarily include neutral atoms and atomic clusters; however, charged molecular fragments are also released from the surface. Analogous to a traditional mass spectrometer, SIMS analyzes these secondary ions based on their relative mass/charge (m/z) ratios (Figure 7.43).

Depending on the desired species to be analyzed, a variety of primary ion beams may be used such as Cs^+, O_2^+, O, Ar^+, and Ga^+. Whereas a cationic primary beam such as Cs^+ is used to ionize electronegative elements (*e.g.*, O, C, N, chalcogens,

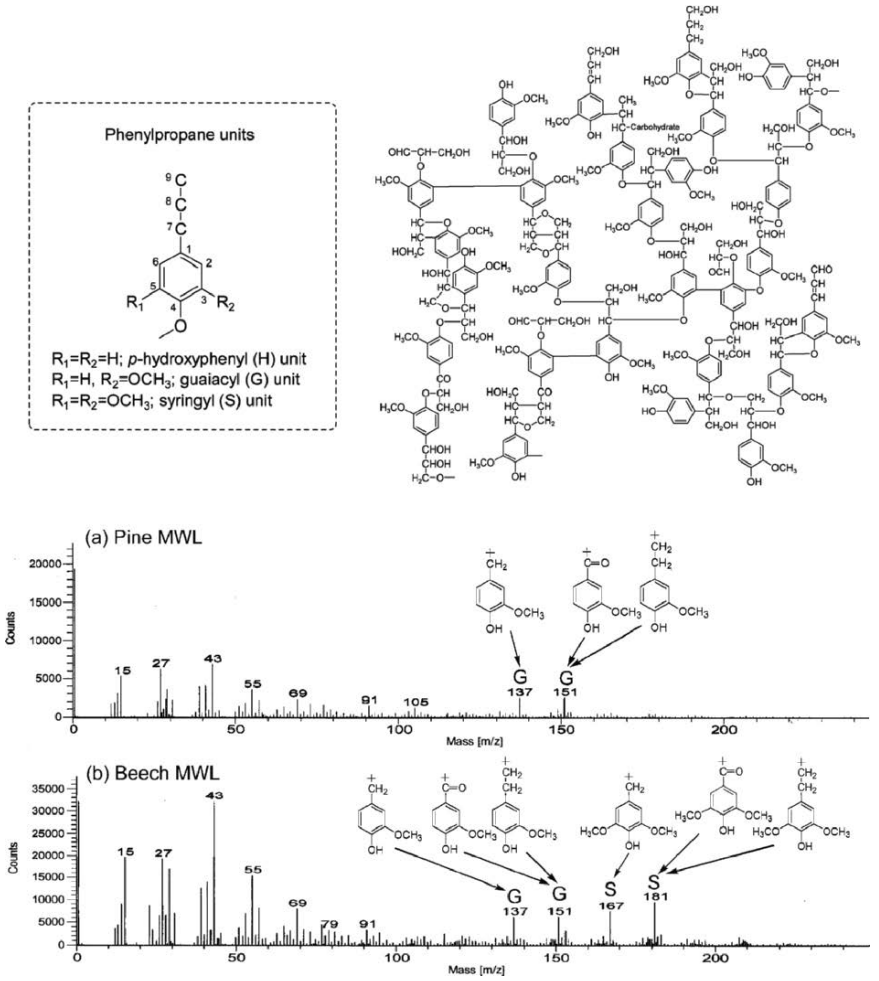


Figure 7.43. SIMS analysis of lignin, the second most abundant biopolymer in nature, following cellulose. Shown is the phenylpropane subunits and a structural model of softwood lignin (top). The secondary-ion mass spectra of pine (softwood) and beech (hardwood) milled wood lignin (MWL, a solvent-extracted form of lignin from beech wood) are also shown (bottom). Reproduced with permission from Saito, K.; Kato, T.; Tsuji, Y.; Fukushima, K. *Biomacromolecules* **2005**, *6*, 678. Copyright 2005 American Chemical Society.

halogens, *etc*.), oxygen atoms are used to ionize electropositive elements (*e.g.*, Li, Mg, Na, B, Al, *etc*.). Liquid metal ion beams (*e.g.*, Ga, Au/Ge) are used for high-resolution studies, since the beam may be focused to a smaller diameter (<50 nm) for micron and nanoscale analyses.

There are two varieties of SIMS – static and dynamic. *Static SIMS* (often referred to as time-of-flight SIMS, TOF-SIMS) is often the method-of-choice, used for

elemental analysis and imaging of the top 2–3 monolayers of a sample; in comparison, *dynamic SIMS* is used to determine elemental concentrations of the sample, as a function of depth. As such, dynamic SIMS is a destructive technique primarily used for depth profiling, whereas TOF-SIMS does not appreciably deteriorate the surface being analyzed. For instance, due to a slow, controllable sputtering rate, the entire analysis may be performed without removing less than 1/10 of an atomic monolayer.

It should be noted that any SIMS analysis will at least result in surface roughness/cratering and elemental mixing within the outermost monolayers. When the primary beam interacts with surface atoms, the incident energy is transferred to target atoms through a series of binary collisions. Some atoms/fragments receive sufficient energy to be recoiled through the sample surface and are removed from the material. However, other less noticeable effects will result such as the implanting of primary ions, and mixing with sample atoms at depths of $\leq$10 nm below the surface. The magnitude of these effects is governed by the incident beam,[86] as well as the crystallinity and surface defects of the sample.

Dynamic SIMS typically uses a quadrupole mass analyzer; however, a TOF mass analyzer offers much higher sensitivities (with limitations, *vide infra*) and mass ranges. In a TOF-SIMS, an ion of known electrical charge, but unknown mass, is accelerated by an electrical field. As a result, all ions of the same charge will have identical kinetic energies. However, the velocity of the ion will depend on the mass/charge ratio, which affects the time it takes for the charged species to reach the detector. That is, light ions will reach the detector before heavier species, allowing for effective mass separation and analysis. It should be noted that each pulse of the primary ion beam produces a full spectrum of secondary ions. Hence, the highest mass resolution is afforded by using the shortest pulse of the primary ion beam, since the time-of-flight for ions may be determined more precisely.

The sputtering event will yield secondary ions with a range of kinetic energies. Whereas atoms will contain their kinetic energy only in translational modes, charged molecular fragments will have the kinetic energy spread over translational, vibrational and rotational modes. By varying the accelerating voltage of the electrostatic ion energy analyzer, one is able to select secondary ions (*e.g.*, monoatomic *vs.* multiatomic (molecular fragments)) based solely on their translational energies. For instance, ions with a lower energy are deflected at larger angles, which prevents them from passing through to the mass analyzer.

Due to more energetic collisions, an incident beams consisting of either heavier ions (*e.g.*, Au^+ *vs.* Ga^+), or polyatomic species (*e.g.*, $C_{60}{}^+$,[87] $Au_3{}^+$, $SF_5{}^+$, $Bi_3{}^+$), will result in much greater secondary ion yields for high molecular components – especially in the range of 500–5,000 Da. For instance, the use of an atomic cluster such as $C_{60}{}^+$ results in a significant number of surface atoms being sputtered from the surface, relative to a smaller beam such as Ga^+ (Figure 7.44). For a 15 keV incident beam, each carbon atom would contain 15,000/60 eV (*i.e.*, 24, 000 kJ mol^{-1}–50 times greater than the C-C bond strength in C_{60}!). Hence, the atomic cluster likely behaves as 60 individual 250 eV atoms with a much broader impact on the surface. However, since the kinetic energy of each C_{60} cluster is much

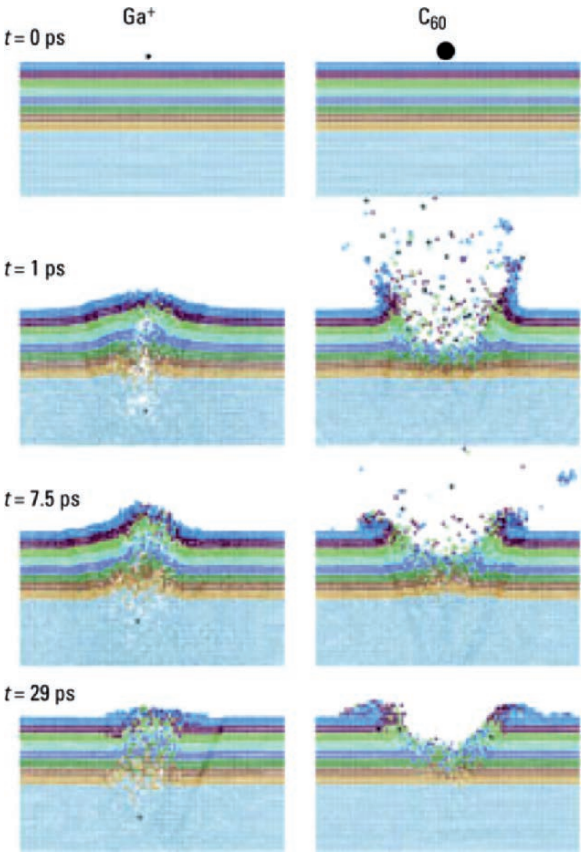


Figure 7.44. Molecular dynamics simulation of the ion bombardment of a Ag crystal surface with a 15 keV Ga^+ beam, and a 15 keV C_{60} atomic cluster beam. The C_{60} beam results in a larger crater and more material removed from the surface. In contrast, the Ga^+ beam results in a destructive effect at greater sample depths, without successful sputtering of the sample. Reproduced with permission from Winograd, N. *Anal. Chem.* **2005**, *77*, 142A. Copyright 2005 American Chemical Society.

less than a Ga^+ ion (15 keV), the sputtering event for $C_{60}{}^+$ is limited to the first few monolayers of the sample.[88]

Another strategy that is used to increase the TOF-SIMS signal for polymeric samples is the application of a metallic coating[89] or nanoparticles[90] onto the sample surface (Figure 7.45). The signal enhancement is likely a result of facilitated electron loss (Eq. 12), as well as the formation of conjugate ions that contain the neutral polymer fragment and metal atoms (Eqs. 13 and 14). The reaction between the polymer fragment and a proton, either in the gas-phase or at the metallic surface, may also result in signal enhancement (Eq. 15).[90]

(12) $\quad M \rightarrow [M]^+ + e^-$

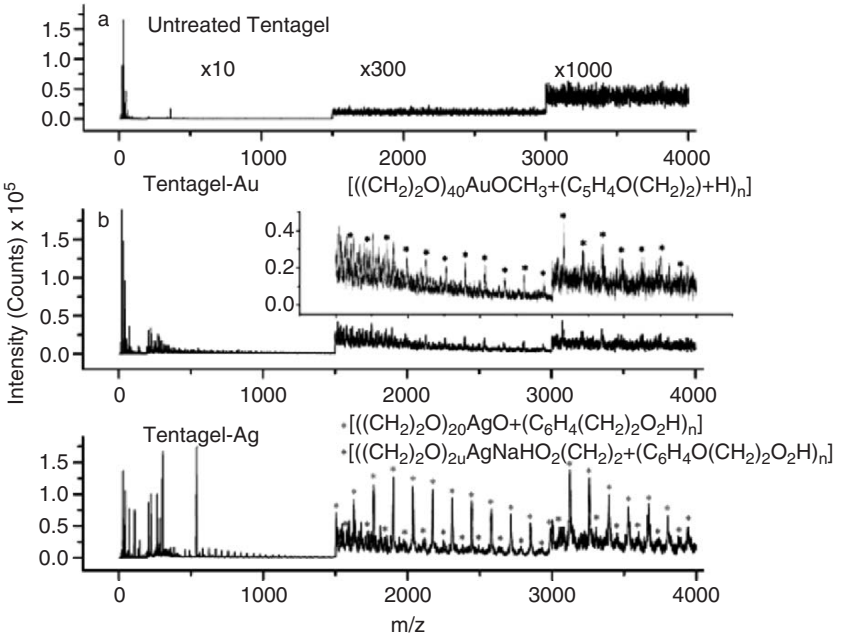


Figure 7.45. SIMS signal enhancement from metal nanoparticle deposition. Shown are comparative spectra in the high molecular weight range for an untreated polymer (a), along with nanoparticle-treated samples (b, c). Reproduced with permission from Marcus, A.; Winograd, N. *Anal. Chem.* **2006**, *78*, 141. Copyright 2006 American Chemical Society.

$$(13)\quad [M + Me]^+ \rightarrow [M]^+ + Me^0$$

$$(14)\quad M + [Me]^+ \rightarrow [M + Me]^+$$

$$(15)\quad M + H^+ \rightarrow [M + H]^+$$

where M is the neutral polymer fragment; Me, neutral metal atom (present as a film or nanoparticles).

In addition to secondary ion generation, SIMS also results in the release of secondary electrons near the sample surface. As previously discussed for SEM analysis, this may result in the buildup of a net electric current for nonconductive surfaces. Analogous to microscopy applications, such surface charging will diffuse the primary beam, making it difficult to perform microanalysis using SIMS. Furthermore, charging will deleteriously affect the detection of secondary ions, by altering the energy distribution of the recoiled ions. These effects are magnified when a sample contains a dielectric surface and a conductive substrate. Due to the strong induced electric field, mobile metal ions in the substrate may migrate toward the surface. Consequently, the data obtained from depth profiling will no longer reflect the original composition of the sample layers. Since information is gleamed from the uppermost portion of the sample, the application of a simple conductive coating is not

feasible for SIMS. Rather, the effects of charging may be minimized by the following techniques:

(i) *Electron bombardment* – typically used in quadrupole (dynamic) SIMS instruments. A low electron beam is used to compensate for a positive charge buildup.
(ii) *Surface conducting grid* – a metallic grid is placed over the sample. When the primary beam impinges this surface, the grid releases secondary electrons that dissipate the buildup of positive charge. Alternatively, a gold or carbon coating may be sputtered onto the sample surface, and then removed from the area to be analyzed – leaving a conductive film that encapsulates the area of interest.
(iii) *Use of negative primary ion beams* – for example, oxygen.

Another widely used application for SIMS is *ion imaging*, which shows secondary ion intensities as a function of spatial location on the sample surface (Figure 7.46). Further, if imaging is performed in tandem with depth profiling, a three-dimensional compositional map of a sample may also be generated (Figure 7.46). Two modes of imaging *via* SIMS are possible:

(i) *Ion microscopy* – the sample is uniformly irradiated and the image is formed directly on a viewing screen *via* optics similar to SEM (max. lateral resolution of *ca*. 0.5 μm).
(ii) *Ion microprobe* – the ion beam is rastered across the surface (max. lateral resolution of *ca*. 20 nm, using liquid metal ion guns).

7.4. SCANNING PROBE MICROSCOPY

The characterization techniques discussed thus far have involved the interaction of an incident beam of radiation (electrons, X-rays, ions, or light) with a sample. Subsequent scattering, transmission, and/or secondary emission is then used to provide an image of the sample, or delineate its chemical composition. This section will discuss a technique that does not involve irradiation, but rather features the scanning of a probe across a surface (hence, termed SPM). A tangible analogy of this technique is actually the precursor to modern CD players – the phonograph/record player (check your local antique shop if you have not seen one before!). The heart of an antiquated phonograph is a cantilever-supported stylus that followed the grooves on the record. The mechanical movement of the stylus within the grooves resulted in an electrical response from a piezoelectric crystal, which was then converted into sound *via* the speaker system. Though this technology was abandoned for stereo applications in the 1980s, SPM represents a next-generation application for a stylus with a much sharper tip (radius of curvature of <50 nm), to follow the surface topography of microscale and nanoscale materials. The resolution of SPM is among the best of all methods to date, with the possibility of atomic scale manipulation/imaging (revisit Figure 6.3 in Chapter 6 – the manipulation and imaging of individual Xe atoms – performed almost 20 years ago!).

Analogous to record players, a SPM tip is supported by a flexible cantilever. During analysis, the tip is slowly rastered across the surface of a material – either

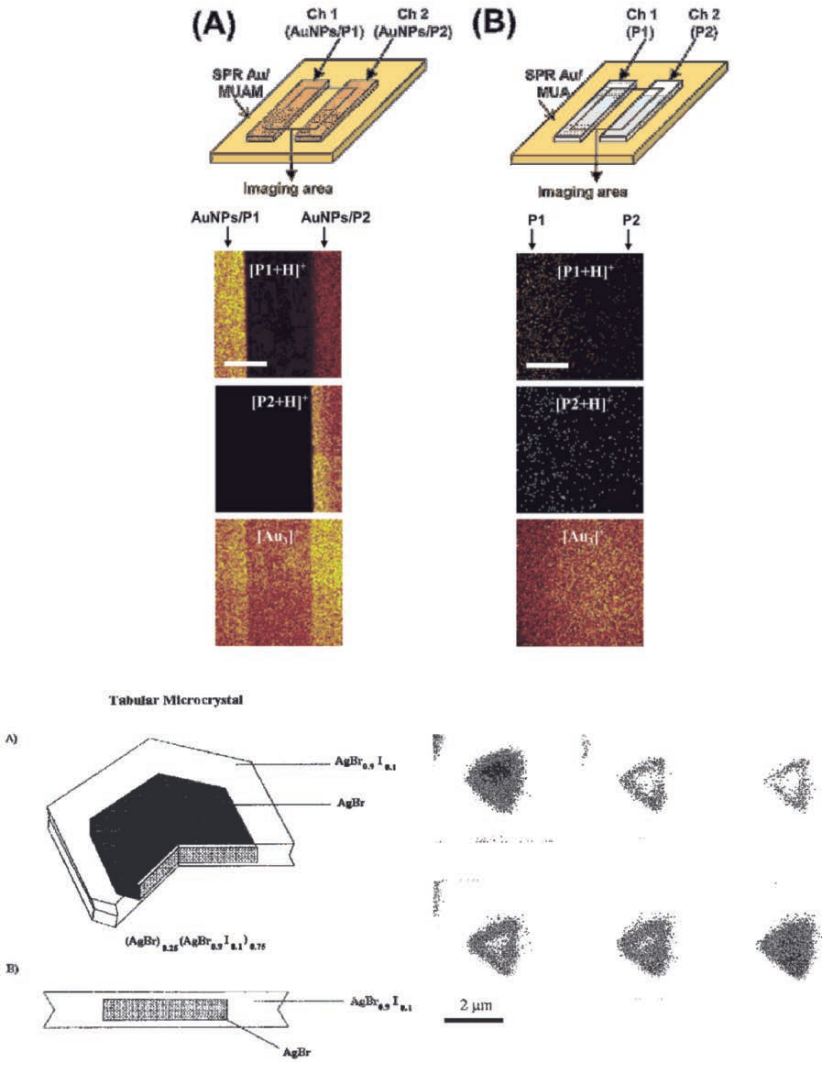


Figure 7.46. Ion imaging using SIMS. (Top) Positive TOF-SIMS images of fragments from peptides (P1, Ac-PRNYVTP-NH_2; P2, Ac-PRNYpVTP-NH_2) patterned on SAMs, pretreated (A) with gold nanoparticles and (B) without gold nanoparticles. The lateral region at both sides (both arrows) represents a microfluidic pattern of different peptides (P1 and P2). TOF-SIMS utilized a Au^+ primary ion gun, with an ion dose $<10^{13}$ ions cm^{-2} at a sample area of $300 \times 300\,\mu m^2$. The scale bar is 100 µm.[91] (Bottom) Schematic of a silver halide microcrystal with a core-shell structure. The diameter is *ca*. 0.5–2 µm; the thickness is *ca*. 150–400 nm. To the right of the illustration are successive I^- ion images (using a Cs^+ primary beam) for the crystal, shown from the top left corner to the bottom right corner, as a function of depth profiling. Whereas the first and last images show a homogeneous iodine-rich array, the middle images show the I^- distribution localized in the shell of the microcrystal.[92]

a few Angstroms away from the surface (noncontact mode), or in contact with the sample (contact mode). There are two primary forms of SPM[93]:

(i) *Scanning tunneling microscopy* (STM) – set the precedent for SPM in the early 1980s. The tip is held a few Angstroms from the sample surface during scans.
(ii) *Atomic force microscopy* (AFM) – the tip may interact with the sample surface through either contact or noncontact modes.

Both types of SPM are most often used to provide an image of the surface topography. Though both techniques are frequently carried out under ambient temperatures/pressures, or even in liquids,[94] they may also be performed in an UHV chamber (*ca.* 10^{-11} Torr)[95] in order to prevent sample contamination, prevent oxidation of the surface, or monitor the deposition of sputtered thin films in real time. As a STM tip is scanned across the sample, a quantum mechanical tunneling current is established between the sample surface and a sharpened W or Pt/Ir tip (Figure 7.47, top). The magnitude of the current is dependent on the distance between the tip and the surface (Δz), as well as the local DOS of the surface. Accordingly, STM is typically performed on conductive and semiconductive surfaces. During topographic imaging, a feedback loop is established to maintain a constant current through varying the distance between the tip and surface ("constant current mode"). In this respect, STM is able to provide real-time, three-dimensional images of the surface with atomic resolution. The use of STM for nanofabrication also represents an important application that is of increasing interest.[96]

In addition to imaging applications, *scanning tunneling spectroscopy* (STS) may be performed, which delineates the local electronic structure of a surface. In this mode, the feedback loop is interrupted, which fixes the distance between tip and sample. A current *vs*. voltage (I–V) curve may then be acquired at a specific (x,y) position on the surface by ramping the bias voltage, and recording the tunneling current. If the I–V curves are collected at every point, a three-dimensional map of the electronic structure may be generated.[97]

Rather than monitoring electronic tunneling phenomena, AFM measures the forces between the tip and surface, which depends on the nature of the sample, the probe tip, and the distance between them (Figure 7.47, bottom).[98] The deflection of the tip in response to surface–tip interfacial forces is recorded by using a laser focused on top of the Si or SiN cantilever, and reflected onto photodetectors. The signal emanating from the photodetector is used to generate a surface topographic map, or the direct measurement of surface intermolecular forces. As with STM, a feedback loop is present in the system, which controls the distance between the tip and sample *via* an electrical current sent to piezoelectric transducers. Such "constant force" scanning is used most frequently, since "constant-height" scanning could result in collisional damage between the tip and surface.

The most common operating modes of AFM are contact, noncontact, and tapping, which are self-explanatory in their manner of interrogation of the surface. In contact-mode AFM, there is a repulsive force between the sample and tip (*ca.* 10^{-9} N); the piezoelectric response of the cantilever either raises or lowers the tip to maintain a constant force. Similarly as STM, the best resolution will be obtained under UHV conditions. That is, in an ambient environment, adsorbed

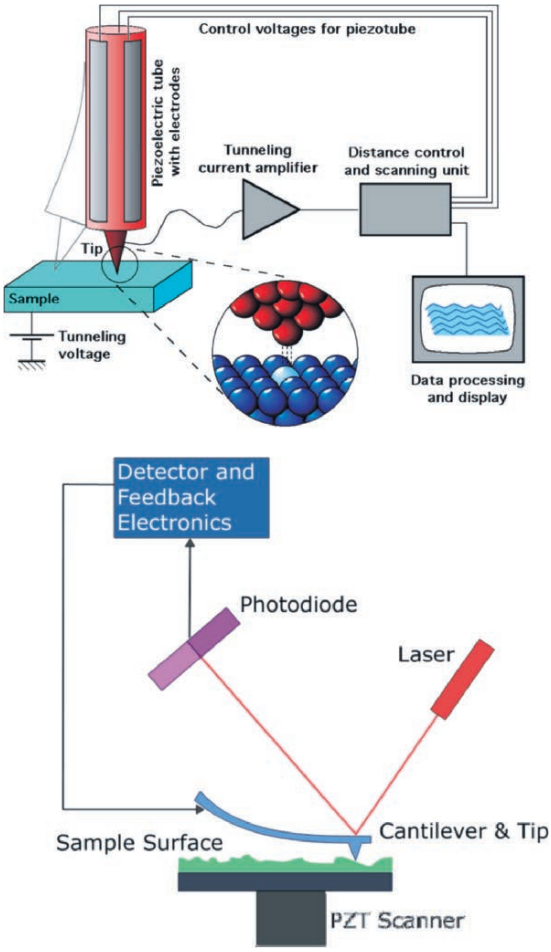


Figure 7.47. Illustrations of STM (top), and AFM (bottom). Reproduced with permission through the Wikipedia Commons agreement. The illustration of STM is courtesy of Michael Schmid, TU Wien, and is available online at: http://www.iap.tuwien.ac.at/www/surface/STM_Gallery/stm_schematic.html. The illustration of AFM is available online at: http://en.wikipedia.org/wiki/Image:Atomic_force_microscope_block_diagram.png

vapors (*e.g.*, N_2, H_2O) form a layer on the surface with a thickness of *ca.* 10–30 monolayers. Consequently, a meniscus will form between the tip and surface, which results in the attraction of the tip toward the surface due to surface tension forces. This force may be neutralized by operating the AFM in a liquid cell, in which the tip is completely immersed in a solvent (Figure 7.48). It should be noted that frictional forces are not always detrimental; lateral force AFM, or the purposeful dragging of the tip along the surface, is useful to determine spatial variations in the composition or phase of a surface.

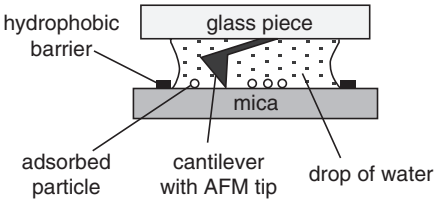


Figure 7.48. Illustration of an AFM system used for *in situ* studies within a liquid. Reproduced with permission from Gliemann, H.; Mei, Y.; Ballauff, M.; Schimmel, T. *Langmuir* **2006**, *22*, 7254. Copyright 2006 American Chemical Society.[100]

Noncontact AFM overcomes the frictional and adhesive forces between the tip and sample by hovering the tip a few Angstroms above the surface. In this mode, the attractive Van der Waal forces between the tip and surface are monitored. As you might expect, these attractive forces are much weaker than those generated in contact mode, often resulting in lower resolution. In order to improve the sensitivity of the tip, the cantilever is oscillated to better detect surface features through small variations in the oscillation wave characteristics (*i.e.*, phase, amplitude, *etc*.). It should be noted that noncontact AFM is often preferred over STM (also noncontact) to study "true" molecular assemblies on surfaces. That is, it has been shown that STM imaging may induce irreversible changes in molecular arrangements on surfaces, especially for high-resolution, constant-height imaging.[99]

During tapping AFM, the cantilever oscillation amplitude remains constant when not in contact with the surface. The tip is then carefully moved downward until it gently taps the surface. As the tip passes over an elevated surface feature, the cantilever has less room to oscillate, and the amplitude of oscillation decreases (*vice versa* for a surface depression). The oscillation frequency is typically 50–500 kHz, with an amplitude of *ca.* 30 nm, which is sufficient to overcome the adhesive forces that are evident in contact (and noncontact) modes (Figure 7.49). Consequently, the tapping mode is most appropriate for soft samples such as organics, biomaterials, *etc*.

Without question, AFM exhibits a much greater versatility for surface analysis than STM. In particular, the following variations are possible, through altering the nature of the tip:

(i) *Chemical force microscopy* (CFM) – uses a chemically modified tip to examine interfacial behavior between the sample and functional groups on the tip surface.[101]

(ii) *Magnetic force microscopy* (MFM) – uses a noncontact magnetic-susceptible tip to map the magnetic properties of a surface, with spatial resolutions of <20 nm.[102]

(iii) *Scanning thermal microscopy* (SThM) – uses a resistive Wollaston wire instead of a conventional AFM probe, which acts as a localized heating source and microthermocouple, used to map the thermal conductivity of a surface.[103]

(iv) *Scanning electrochemical microscopy* (SECM) – based on the electrochemical interaction between a redox-active species produced at the tip, and the substrate

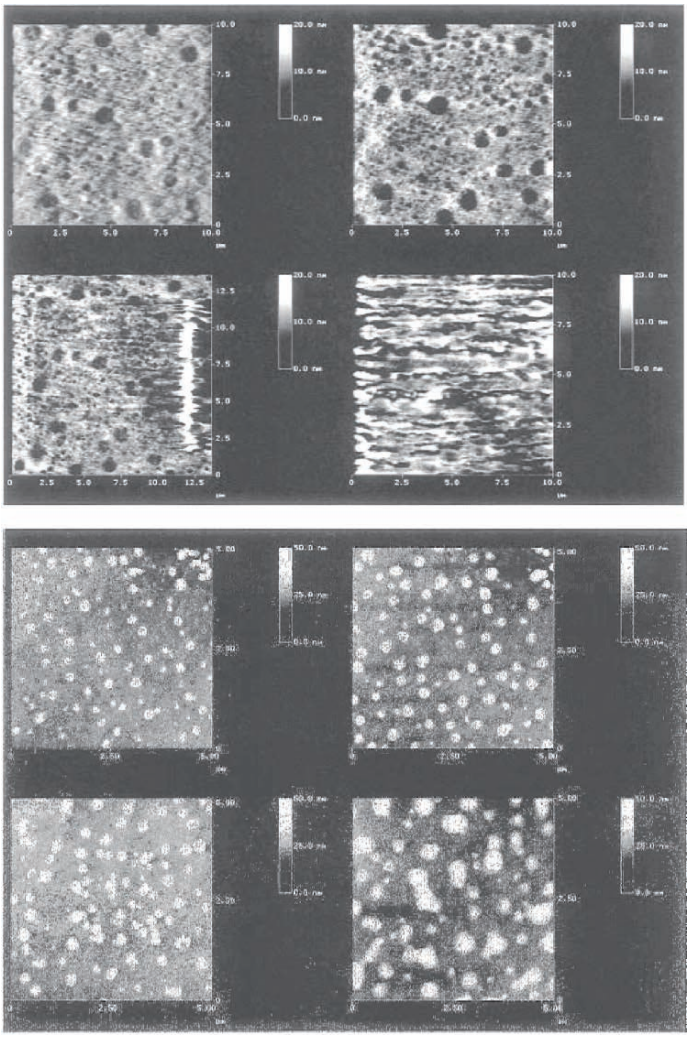


Figure 7.49. Comparative images of contact-mode (top) and tapping-mode (bottom) AFM under an ambient atmosphere. The images represent the surface of freshly cleaved potash–lime–silica glass after the elapsement of time intervals, from upper left to bottom right (top: 150, 264, 264 min (larger scan area), 8 h; bottom: 2, 5, 5 h (larger scan area), 43 h). Humidity-induced pit formation is only observable using contact mode (CM) – note the movement of surface corrosion products by the AFM tip (top, images c and d). In contrast, tapping-mode (TM) revealed round features that became visible after 2 h of exposure time. These surface features (observable only by TM) are likely swelled glass material due to the breakdown of the silicate network through water adsorption and subsequent surface ion-exchange reactions. Contact-mode AFM is not able to image these surface features, since the tip scrapes away the silicate residue during analysis. Reproduced with permission from Schmitz, I.; Schreiner, M.; Friedbacher, G.; Grasserbauer, M. *Anal. Chem.* **1997**, *69*, 1012. Copyright 1997 American Chemical Society.

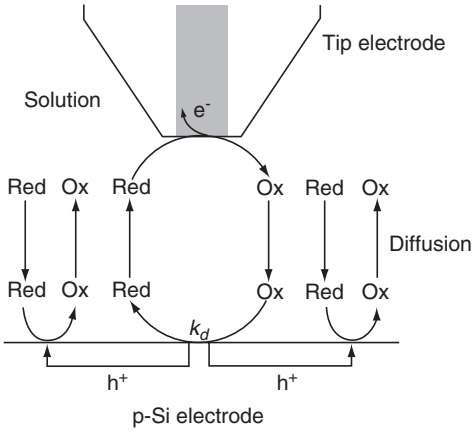


Figure 7.50. Illustration of scanning electrochemical microscopy (SECM). Reproduced with permission from Ghilane, J.; Hauquier, F.; Fabre, B.; Hapiot, P. *Anal. Chem.* **2006**, *78*, 6019. Copyright 2006 American Chemical Society.

being studied (Figure 7.50). Hence, this method offers high chemical selectivity, and is able to correlate localized surface features with their chemical reactivities. [104]

In order to carry out the above applications (as well as standard imaging/force measurements), AFM tips may be selected from among a wide variety of sizes and shapes.[105] A recent discovery that will greatly assist in the characterization of nanomaterials is the fabrication of tips that are terminated with individual gold nanoparticles or single-walled carbon nanotubes (Figure 7.51).[106] These advanced designs will offer significant improvements in resolution (and artifact generation) over conventional tips.[107] Furthermore, analogous tip designs will allow one to probe the physical, thermal, magnetic, and optical properties of individual nanoarchitectures *via* SPM.

7.5. BULK CHARACTERIZATION TECHNIQUES

The majority of characterization techniques discussed thus far have been surface-related, with some capable of analyzing sub-surface depths through *in situ* ion etching. This final section will focus briefly on a selection of common bulk techniques that may be used to characterize as-synthesized materials such as polymers, ceramics, *etc*. More details on these and other techniques not discussed herein may be found in the "Further Reading" section at the end of this chapter. In particular, these additional resources, as well as countless others online, will highlight solid-state characterization techniques such as:

(i) *Solid-state NMR* – chemical environment of NMR-active nuclei; used to obtain physical, chemical, electronic, and structural information about constituent molecules.

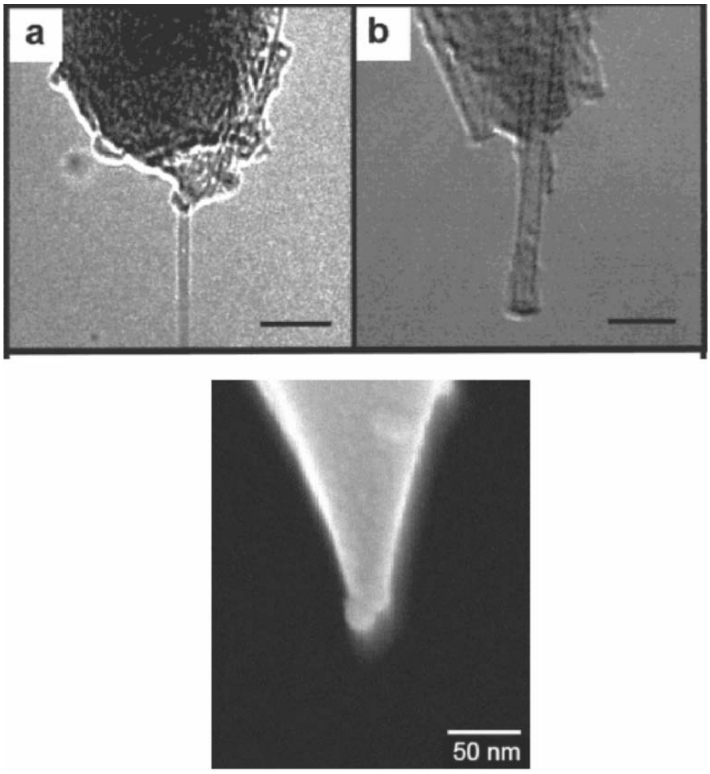
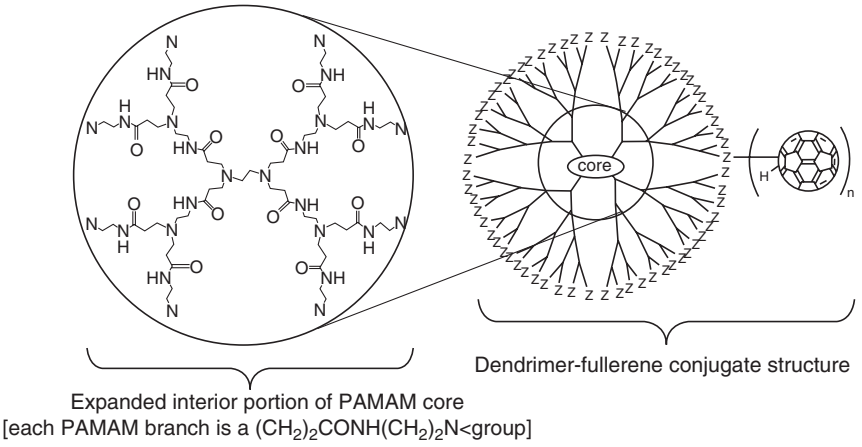


Figure 7.51. Advanced AFM tip designs. Shown (top) are SWNT-terminated AFM tips (scale bars are 10 nm)[106]b and (bottom) an AFM tip terminated with an individual gold nanoparticle (diameter of 14 nm).[106]a

(ii) *Raman spectroscopy* – vibration, stretching, and bending of sample molecules (for the bulk sample, as well as adsorbed surface species); assessing defects in carbon nanotubes.
(iii) *IR spectroscopy* (including surface-characterization modes such as attenuated total reflectance (ATR), diffuse reflectance infrared Fourier transform spectroscopy (DRIFT), and reflection absorption infrared spectroscopy (RAIRS)) – complementary to Raman spectroscopy.
(iv) *UV–Vis spectroscopy* – functional group information; sizes of nanoparticles.
(v) *Mass spectrometry* (MS) – information regarding isotopes, mass, and structure.
(vi) *BET*[108] *surface area analysis* – pore size and surface area of powders.
(vii) *Dynamic light scattering* (DLS) and Coulter counting – particle size.
(viii) *Mössbauer spectroscopy* – chemical environment of ^{57}Fe, ^{129}I, ^{119}Sn, or ^{121}Sb atoms in a sample.
(ix) *Single-crystal and powder X-ray diffraction* – three-dimensional arrangement of atoms/ions/molecules.
(x) Physical testing techniques (tensile strength, flame retardancy, *etc*.).

A primary method that is used to characterize the thermal properties of a bulk material is *thermogravimetric analysis* (TGA). This method provides detailed information regarding the thermal stability and decomposition pathway of a material (*e.g.*, stepwise loss of ligands for an organometallic compound), as well as structural information for complex composites (Figure 7.52). The operating principle of TGA is



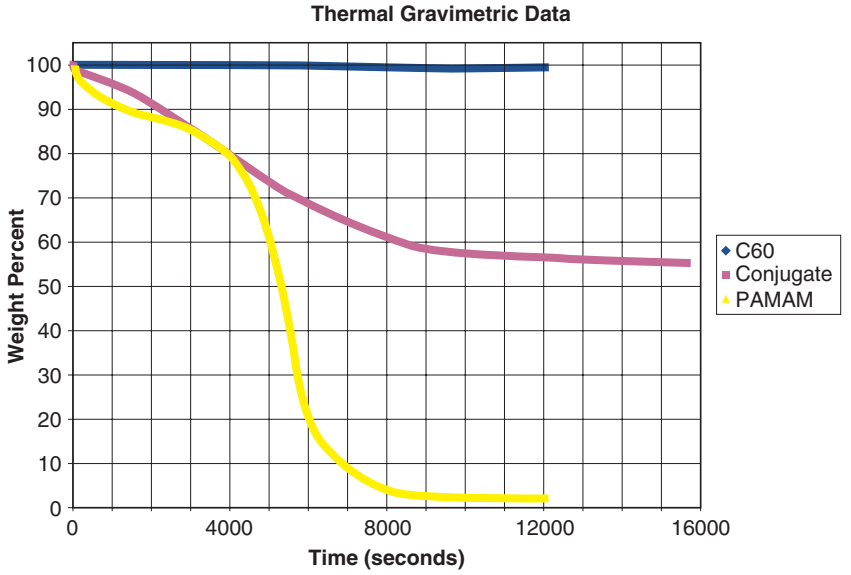


Figure 7.52. Schematic of a PAMAM dendrimer–fullerene conjugate structure (top), with TGA analysis (bottom). Based on the mass loss of the conjugate, it is suggested that each PAMAM dendrimer is surrounded by 30 fullerene units. Reproduced with permission from Jensen, A. W.; Maru, B. S.; Zhang, X.; Mohanty, D. K.; Fahlman, B. D.; Swanson, D. R.; Tomalia, D. A. *Nano Lett.* **2005**, 5, 1171. Copyright 2005 American Chemical Society.

very simple – the solid is placed in a tiny microbalance pan and heated according to preset ramping conditions. The controlled thermolysis of the compound may be carried out *in vacuo*, or in the presence of a carrier gas such as O_2, N_2, or Ar.

In addition to performing simple mass-loss investigations, TGA may also be applied to accurately determine the enthalpy of sublimation (ΔH_{sub}), as well as the sublimation temperature (T_{sub}) of a solid.[109] In these investigations, the mass loss through sublimation (m_{sub}) will be constant at a given temperature, as long as the phase change occurs without appreciable decomposition. Hence, by measuring the mass loss over a variety of isothermal regions, a plot of $\log(m_{sub}T^{1/2})$ *vs.* $1/T$ may be generated, which readily yields values of ΔH_{sub} and T_{sub} from the slope and *y*-intercept, respectively (Eq. 16):

$$\log\left(m_{sub}\sqrt{T}\right) = \frac{-0.0522(\Delta H_{sub})}{T} + \left[\frac{0.0522(\Delta H_{sub})}{T_{sub}} - \frac{1}{2}\log\left(\frac{1,306}{M_w}\right)\right]. \quad (16)$$

A technique that is often used in tandem with TGA[110] is *differential scanning calorimetry* (DSC).[111] This technique monitors the amount of heat that is required to increase the temperature of a sample, relative to a reference. For example, when a sample undergoes an endothermic phase transition (*e.g.*, melting), heat will be absorbed; conversely, an exothermic event (*e.g.*, crystallization) will require less heat to raise the temperature. Accordingly, DSC is used to determine distinct thermodynamic events, as well as subtle changes such as glass transitions that occur during polymer curing (Figure 7.53).

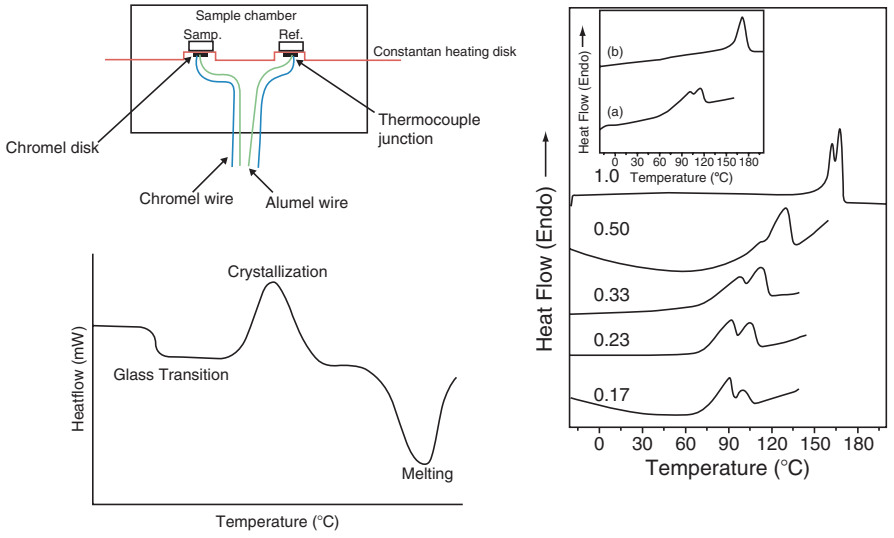


Figure 7.53. Differential scanning calorimetry (DSC). Shown are (a) schematic of the heat-flux sample chamber; (b) an example of a DSC thermogram, showing endothermic events[113]; (c) DSC thermogram of a poly(vinylidene fluoride)–ethyl acetoacetate polymer–solvent system, showing two melting events for the polymer due to its intermolecular interactions with solvent molecules. The inset shows a comparison between the pure polymer (b) and the polymer-solvent (a). Reproduced with permission from Dasgupta, D.; Malik, S.; Thierry, A.; Guenet, J. M.; Nandi, A. K. *Macromolecules* **2006**, *39*, 6110.

The viscoelastic properties of a polymer (high modulus, "glass-like," or low modulus, "rubber-like") may be determined through *dynamic mechanical (thermal) analysis* (DMA or DMTA). This technique involves monitoring the resultant displacement of a polymer following its interaction with an oscillating external force, as the temperature is altered. In addition to readily observing the glass transition, other properties such as stiffness and damping properties are generated from this technique.

Though TGA/DSC and DMA are extremely useful for polymer characterization, these techniques provide no (direct) structural information. As an alternative, structural information such as conformational changes of noncrystalline macromolecules may be determined by *small-angle scattering* (SAS) techniques.[112] This method examines the patterns arising from the elastic scattering of X-rays (SAXS) or neutrons (SANS) from sample components. Whereas SAXS provides information regarding the electron density distribution of the sample, SANS is sensitive toward the sample nuclei. For instance, ^{1}H and ^{2}H (deuterium) exhibit different scattering lengths; hence, many studies perform isotopic labeling to gain additional structural information about the sample.[114] The larger the scattering angle, the smaller the length scale that may be probed; hence, wide-angle X-ray scattering (WAXS) is used to determine structural information on the atomic length scale, and SAXS/SANS are used in the size regime of *ca.* 1–300 nm (Figure 7.54). A wide variety of samples may be analyzed by SAS, most often consisting of powders or solvent suspensions of macromolecules or nanoparticles/colloids.[115]

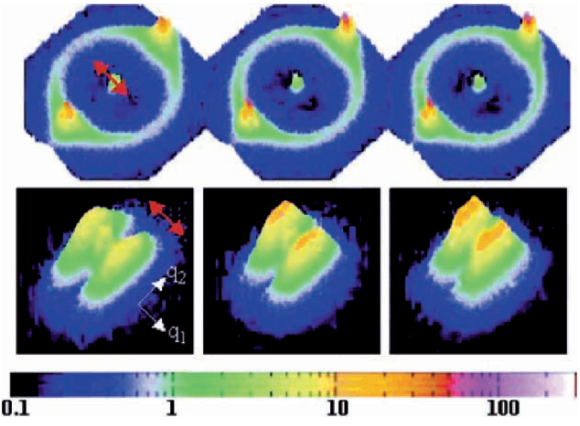


Figure 7.54. Wide-angle (top) and small-angle (bottom) X-ray scattering patterns of drawn polyethylene fibers, annealed for 264 h at 23, 80, and 100°C (from left to right). The red arrows indicate the drawing direction. The peaks in the WAXS pattern remain unchanged during annealing, whereas the pattern changes significantly for SAXS. This is proposed to indicate the preferential orientation of the polymeric chains along the drawing direction, whereas the lamella in the samples exhibit slight shear during annealing. Reproduced with permission from Men, Y.; Rieger, J.; Lindner, P.; Enderle, H. -F.; Lilge, D.; Kristen, M. O.; Mihan, S.; Jiang, S. *J. Phys. Chem. B* **2005**, *109*, 16650. Copyright 2005 American Chemical Society.

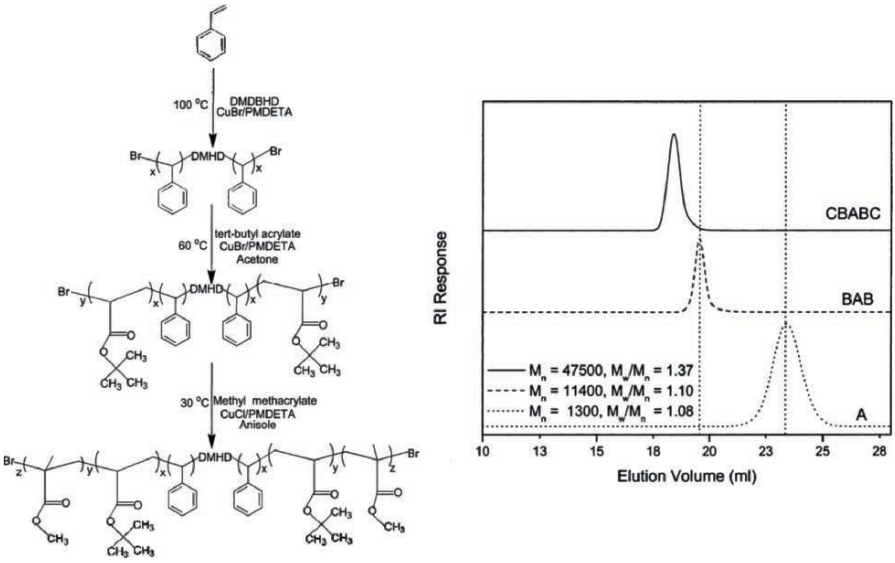
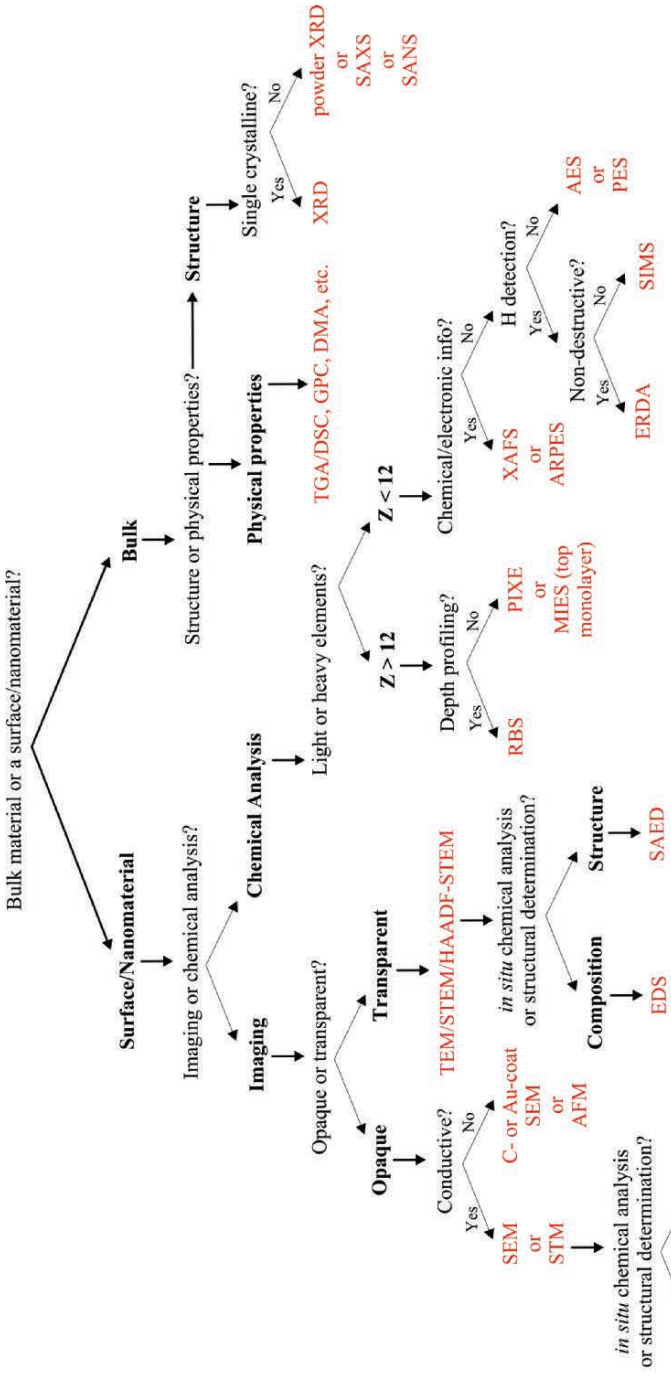


Figure 7.55. Reaction scheme for the room-temperature synthesis of a CBABC pentablock copolymer, and GPC traces of *a*,*z*-dibromo-terminated polystyrene (A, dotted line), poly(*t*BA-*b*-styrene-*b*-*t*BA) (BAB triblock copolymer macroinitiator, dashed line), and poly(MMA-*b*-*t*BA-*b*-styrene-*b*-*t*BA-*b*-MMA) (CBABC, solid line). The GPC data shows evidence of a successful polymerization route – low polydispersity index (PDI), and good control over the number-average molecular weight (M_n). Reproduced with permission from Ramakrishnan, A.; Dhamodharan, R. *Macromolecules* **2003**, 36, 1039. Copyright 2003 American Chemical Society.

An additional method that is used to characterize polymers is *gel-permeation chromatography* (GPC). This technique is a form of *size-exclusion chromatography* (SEC), where components are separated from one another based solely on their sizes (or hydrodynamic volume). GPC is carried out in the same configuration as other HPLC (high-performance liquid chromatography) methods, with a gel stationary phase and a pressurized liquid mobile phase that elutes the dissolved components from the column. Separation occurs as the solution is passed through the gel, which comprises polystyrene crosslinked with divinylbenzene. While the larger molecules pass through the column without significant retention, the smaller molecules are retained by encapsulation within the pores of the gel (*ca.* pore diameters of 100–10^6 nm). The information obtained from gel-permeation chromatographs is mostly used to determine the molecular weight distribution of a polymer, known as the polydispersity index (PDI, Figure 7.55).

IMPORTANT MATERIALS APPLICATIONS VI: SO WHICH ACRONYM SHALL I CHOOSE?!



References and Notes

1 More information on the components and operation of reflection microscopes may be obtained from http://www.microscopyu.com/articles/dic/reflecteddic.html

2 Synge, E. H. *Philos. Mag.* **1928**, 6, 356.

3 For a thorough review of NSOM, see: Dunn, R. C. *Chem. Rev.* **1999**, *99*, 2891.

4 Mass of a Golf Ball: http://hypertextbook.com/facts/1999/ImranArif.shtml

5 It is important to note the general trend of decreasing wavelength (and greater resolution) as the velocity of electrons is increased (*i.e.*, higher accelerating voltages).

6 Cremer, J. T.; Piestrup, M. A.; Gary, C. K.; Pantell, R. H.; Glinka, C. J. *Appl. Phys. Lett.* **2004**, *85*, 494.

7 For example, see: http://accelconf.web.cern.ch/AccelConf/p05/PAPERS/WOAC001.PDF

8 O'Dwyer, C.; Navas, D.; Lavayen, V.; Benavente, E.; Santa Ana, M. A.; Gonzalez, G.; Newcomb, S. B.; Sotomayor Torres, C. M. *Chem. Mater.* **2006**, *18*, 3016.

9 Hu, J.; Bando, Y.; Zhan, J.; Zhi, C.; Golberg, D. *Nano Lett.* **2006**, 6, 1136.

10 For an image of the lattice structure and properties of LaB_6, see: http://www.kimphys.com/cathode/catalog_PDF/LaB6_cathode_ES423.pdf

11 Perkins, C. L.; Trenary, M.; Tanaka, T.; Otani, S. *Surface Sci.* **1999**, *423*, L222. Accessed online at: http://www.chem.uic.edu/mtrenary/LaB6.pdf

12 http://www.physics.uc.edu/~jarrell/COURSES/ELECTRODYNAMICS/Student_Projects/tharanga/review.pdf

13 (a) http://anl.gov/techtransfer/Available_Technologies/Chemistry/uncd_flc.html (b) Krauss, A. R.; Auciello, O.; Ding, M. Q.; Gruen, D. M.; Huang, Y.; Zhimov, V. V.; Givargizov, E. I.; Breskin, A.; Chechen, R.; Shefer, E.; Konov, V.; Pimenov, S.; Karabutov, A.; Rakhimov, A.; Suetin, N. *J. Appl. Phys.* **2001**, *89*, 2958, and references therein.

14 Unless otherwise stated, TEM micrographs are collected as bright-field images.

15 The low-resolution SEM image was reproduced with permission from Li, F.; Zhang, L.; Metzger, R. M. *Chem. Mater.* **1998**, 10, 2470. Copyright 1998 American Chemical Society.

16 The high-resolution SEM image was reproduced with permission from Moon, J. -M.; Wei, A. *J. Phys. Chem. B* **2005**, *109*, 23336. Copyright 2005 American Chemical Society.

17 http://www.microstartech.com/index/cryo.htm

18 The FIB/lift-out technique yields cross sections of both the film and substrate, which maintains the integrity of the interface; in comparison, ultramicrotomy requires the removal of the polymer film and subsequent embedding/sectioning. For a detailed example, see: White, H.; Pu, Y.; Rafailovich, M.; Sokolov, J.; King, A. H.; Giannuzzi, L. A.; Urbanik-Shannon, C.; Kempshall, B. W.; Eisenberg, A.; Schwarz, S. A.; Strzhemechny, Y. M. *Polymer* **2001**, *42*, 1613.

19 Reproduced with permission from Fu, L.; Liu, Y.; Hu, P.; Xiao, K.; Yu, G.; Zhu, D. *Chem. Mater.* **2003**, *15*, 4287. Copyright 2003 American Chemical Society.

20 Reproduced with permission from Liu, Z.; Ohsuna, T.; Sato, K.; Mizuno, T.; Kyotani, T.; Nakane, T.; Terasaki, O. *Chem. Mater.* **2006**, *18*, 922. Copyright 2006 American Chemical Society.

21 Virgilio, N.; Favis, B. D.; Pepin, M. -F.; Desjardins, P. *Macromolecules* **2005**, *38*, 2368.

22 Another way to state this is that for a single crystal, only a few lattice planes will be oriented at their Bragg angle at any one time.

23 An excellent book that contains many high-resolution TEM images and the electron diffraction patterns for a variety of archetypical unit cells see: Sindo, D.; Hiraga, K. *High-Resolution Electron Microscopy for Materials Science*, Springer: New York, 1998.

24 Reproduced with permission from Sun, G.; Sun, L.; Wen, H.; Jia, Z.; Huang, K.; Hu, C. *J. Phys. Chem. B* **2006**, *110*, 13375. Copyright 2006 American Chemical Society.

25 Reproduced with permission from Liu, Y.; Xu, H.; Qian, Y. *Cryst. Growth Design* **2006**, 6, 1304. Copyright 2006 American Chemical Society.

26 An excellent summary of crystallography and systematic absences is given by: http://169.232.130.50/khan/KhanWebClass.ASP?WCI=Nutshell#LAUE%20SYMMETRY%20ANDk%20SYSTEMATIC%20ABSENCES.

[27] The nomenclature for X-ray emission consists of the name of the shell in which the vacancy was created (K, L, M, N), and on the electronic shell that filled the vacancy. For instance, ejection of a K shell electron, filled with a L shell electron is denoted as K_α; if filled with an M shell electron, then K_β is used, and so on. Due to electronic subshells, nomenclature becomes significantly complex, as shown in Figure 7.19.

[28] It should be noted that EDS and WDS are often referred to as EDX and WDX, or XEDS and XWDS, respectively.

[29] The detection limits for EDS are typically 0.1 at% (for elements with $Z > 10$), whereas WDS is able to detect elements present in concentrations of a few ppm.

[30] Electromagnetic lenses suffer from three types of defects: astigmatism, spherical aberrations, and chromic aberrations. Astigmatism may be easily minimized by placing electromagnets, known as stigmator coils, around the column. The current in these coils may be varied, which will reform the distorted electron beam back into a round shape. Chromic aberrations may be reduced by using a field emission source that produces an electron beam with a sharper distribution of electron energies. However, the third type of defect, spherical aberrations (C_S), are not easily circumvented. In fact, this is the primary factor that limits the resolution of electron microscopes to values far below their theoretical values. Recently, techniques have been developed to correct for these defects, often involving the use of multipole lenses (*e.g.*, quadrupole, octapole, *etc*.) with careful control of their fabrication and operation. For a description of techniques used to design a Cs-corrected HRTEM, see: http://www.jeol.com/jeolnews/jeol_news_july2002/htm/02/index.htm

[31] The only atoms with diameters smaller than 1 Å are H and He!

[32] Fahlman, B. D., unpublished results.

[33] Reproduced with permission from Harmer, M. A.; Farneth, W. E.; Sun, Q. *J. Am. Chem. Soc.* **1996**, *118*, 7708. Copyright 1996 American Chemical Society.

[34] Reproduced with permission from Li, Y.; Xiang, J.; Qian, F.; Gradecak, S.; Wu, Y.; Yan, H.; Blom, D. A.; Lieber, C. M. *Nano Lett.* **2006**, 6, 1468. Copyright 2006 American Chemical Society.

[35] Some important precedents that show the utility of Z-contrast imaging: (a) Arslan, I.; Yates, T. J. V.; Browning, N. D.; Midgley, P. A. *Science* **2005**, *309*, 2195. (b) Chisholm, M. F.; Kumar, S.; Hazzledine, P. *Science* **2005**, *307*, 701. (c) Nellist, P. D.; Pennycook, S. J. *Science* **1996**, *274*, 413.

[36] For a recent study that shows the influence of focus, sample thickness, and inner detector angle on the image, see: http://www.jeol.com/jeol_news_july2002/pdf_light/22.pdf

[37] The photon energies for the light elements are: Be(0.109 keV), B(0.185 keV), C(0.282 keV), N(0.392 keV), O(0.523 keV), and F(0.677 keV). Due to these low energies, the emitted X-rays are easily absorbed by the sample or the components of the detector.

[38] Bremsstrahlung is from the German word *bremsen* ("to brake"), and *Strahlung* ("radiation") – thus, "braking radiation." This is characterized by a broad peak that results from deceleration of beam electrons due to scattering from atomic nuclei.

[39] Pronounced: $OH\text{-}zh\bar{a}$.

[40] The energy-dispersing device at the heart of EDS is a semiconducting diode. As an incoming X-ray photon impinges the diode, electron-hole pairs are generated, which yields a measurable electrical current. In order to reduce background noise from photons not originating from the sample, the detector is operated at temperatures of *ca.* 140 K. A protective window comprising Be, or more recently, ultrathin windows of BN, diamond, or a supported polymer (paralene, norvar), is used to prevent the condensation of vapors (*e.g.*, water, organics) onto the cooled diode. A "windowless" detector may also be used for EDS; with careful operation, this modification allows the detection of elements down as far as Be – with a maximum efficiency of only 2%. It should be noted that although UHV conditions are used in a TEM/SEM instrument, there will always be a low concentration of vapors – often originating from the rotary pump fluid, or the sample itself due to beam-induced volatilization/decomposition, residual solvent evaporation, *etc*. The buildup of such a coating on the detector window will reduce the energy of the incoming X-rays, which will drastically reduce the detection sensitivity – especially for low-Z elements.

[41] It should be noted that reflection techniques are also possible through modifying the angle of approach of the incident electron beam of the TEM. Such analysis is referred to as reflection electron microscopy (REM), and accompanying characterizations such as *reflection high-energy electron diffraction* (RHEED) and *reflection electron energy-loss spectroscopy* (REELS)

are also possible, grouped under the umbrella of reflection high-resolution analytical electron microscopy (RHRAEM). (a) For an application of this multifaceted approach to study GaAs(110) surfaces, see: Wang, Z. L. *J. Electron. Microsc. Tech.* **1988**, *10*, 35 (citation found online at: http://www.osti.gov/energycitations/product.biblio.jsp?osti_id=6614173). (b) Additional information may be found in Wang, Z. L. *Reflection Electron Microscopy and Spectroscopy for Surface Analysis*, Cambridge University Press: Cambridge, UK, 1996. (A book synopsis is found online at: http://www.nanoscience.gatech.edu/zlwang/book/book2_intro.pdf)

42 For a thorough summary/examples of the details generated from EELS spectra, see: (a) Thomas, J. M.; Williams, B. G.; Sparrow, T. G. *Acc. Chem. Res.* **1985**, *18*, 324. (b) http://quasi.nims.go.jp/hrtem/recent/eels-e.html.

43 (a) Williams, B. G.; Sparrow, T. G.; Egerton, R. F. *Proc. R. Soc. Lond. Ser. A* **1984**, *393*, 409. (b) Sparrow, T. G.; Williams, B. G.; Thomas, J. M.; Jones, W.; Herley, P. J.; Jefferson, D. A. *J. Chem. Soc., Chem. Commun.* **1983**, 1432. (c) Ferrell, R. A. *Phys. Rev.* **1956**, *101*, 554.

44 For example, see: http://ipn2.epfl.ch/CHBU/papers/ourpapers/Stockli_ZPD97.pdf

45 A very nice tutorial on XANES and XAFS is found online at: http://cars9.uchicago.edu/xafs/xas_fun/xas_fundamentals.pdf

46 (top) Reproduced with permission from Gerald Kothleitner (http://www.felmi-zfe.at), *Electron Energy-Loss Spectroscopy (EELS) for the Hitachi HD-2000,* found online at: www.cs.duke.edu/courses/spring04/cps296.4/papers/EELS.method.pdf (bottom) Reproduced with permission from Brydson, R. *Electron Energy Loss Spectroscopy*, BIOS Scientific Publishers: Oxford, UK. Copyright 2001 Taylor & Francis Group.

47 For a recent application of EELS in determining the nature of C bonding in amorphous carbon nanotubes see: Hu, Z. D.; Hu, Y. F.; Chen, Q.; Duan, X. F.; Peng, L. -M. *J. Phys. Chem. B.* **2006**, *110*, 8263.

48 For information on the analysis of surfaces by IR radiation instead of electrons, a complimentary technique known as reflection absorption infrared spectroscopy (RAIRS), see: (a) http://www.uksaf.org/tech/rairs.html (b) http://www.cem.msu.edu/∼cem924sg/Topic11.pdf

49 For example, the development of fibers/fabrics that will actively adsorb and surface deactivate chemical and biological warfare agents – of increasing importance as new modes of terrorist activity continue to emerge. For more information, see: (a) http://web.mit.edu/isn/(Institute of Soldier Nanotechnologies at M.I.T.). (b) Richards, V. N.; Vohs, J. K.; Williams, G. L.; Fahlman, B. D. *J. Am. Ceram. Soc.* **2005**, *88*, 1973.

50 Simon, C.; Walmsley, J.; Redford, K. *Transmission Electron Microscopy Analysis of Hybrid Coatings*, Proceedings of the 6th International Congress on Advanced Coating Technology, Nuremberg, Germany, April 3–4, 2001. Found online at: http://www.sintef.no/content/page1___5524.aspx

51 Some of the (potential) energy of the incident electrons is required to release an outer electron from its valence or conduction band, with the remaining being transferred into the kinetic energy of the ejected secondary electron.

52 For more details, refer to: Goldstein, J.; Newbury, D.; Joy, D.; Lyman, C.; Echlin, P.; Lifshin, E.; Sawyer, L.; Michael, J. *Scanning Electron Microscopy and X-Ray Microanalysis*, 3rd ed., Kluwer: New York, 2003.

53 For example, see: http://www.edax.com/products/TSL/OIM/OIM-Intro.html

54 Monte Carlo simulations performed by the author using the program CASINO ("monte CArlo SImulation of electroN trajectory in sOlids"), available free-of-charge on the Internet: http://www.gel.usherbrooke.ca/casino/What.html

55 Both carbon and gold coating are performed using a sputter-coating PVD method. The film of choice is most often C since it is less costly and is transparent to X-rays (for EDS). Gold is used to coat very uneven surfaces, and is only useful when EDS is not being performed (strong Au signal would mask other elements present in the sample).

56 An example of a tandem SEM/SAM instrument is the Thermo Microlab 350: http://www.thermo.com/com/cda/product/detail/1,1055,15886,00.html

57 The effect of charging can be controlled by either altering the position of the sample with regard to the incident electron beam, or using an argon ion (Ar^+) gun to neutralize the charge.

58 For a historical background of ESEM development, see: http://www.danilatos.com/

[59] For instance, see: (a) http://www.feicompany.com/Portals/_default/PDFs/content/2006_06_SetSTEM_td.pdf (b) http://www.feicompany.com/Products/ProductTypes/SEM/tabid/67/Default.aspx

[60] The SEM image of charging was taken from unpublished work by the author. The gold-coated SEM image is also from the author's research: Vohs, J. K.; Raymond, J. E.; Brege, J. J.; Williams, G. L.; LeCaptain, D. L.; Roseveld, S.; Fahlman, B. D. *Polym. News* **2005**, *30*(10), 330.

[61] Richards, V. N.; Vohs, J. K.; Williams, G. L.; Fahlman, B. D. *J. Am. Ceram. Soc.* **2005**, *88*(7), 1973.

[62] Auger spectra reproduced with permission from Zhu, Z.; Srivastava, A.; Osgood, R. M. *J. Phys. Chem. B* **2003**, *107*, 13939. Copyright 2003 American Chemical Society.

[63] Auger depth profile reproduced with permission from Zhang, Y. W.; Yang, Y.; Jin, S.; Tian, S. J.; Li, G. B.; Jia, J. T.; Liao, C. S.; Yan, C. H. *Chem. Mater.* **2001**, *13*, 372. Copyright 2001 American Chemical Society.

[64] Schematic reproduced with permission from Miller, A. F.; Cooper, S. J. *Langmuir* **2002**, *18*, 1310. Copyright 2002 American Chemical Society.

[65] Reproduced with permission from Rossi, M. P.; Ye, H.; Gogotsi, Y.; Babu, S.; Ndungu, P.; Bradley, J. -C. *Nano Lett.* **2004**, 4, 989. Copyright 2004 American Chemical Society.

[66] In contrast to SEM, EDS, and AES that use incident X-rays of *ca*. 10 keV, XPS is less damaging to beam-sensitive samples due to the use of "soft" X-rays, of much less energy (1–1.5 keV).

[67] A very detailed example of XPS to distinguish among Li salts for Li-ion battery applications is: Dedryvere, R.; Leroy, S.; Martinez, H.; Blanchard, F.; Lemordant, D.; Gonbeau, D. *J. Phys. Chem. B* **2006**, *110*, 12986.

[68] (a) As you may recall from Chapter 4, the energy bands of crystalline solids (*e.g.*, semiconductors) are denoted as parabolas in an E–k diagram, where momentum (p) is equal to ηk. Interactive E–k diagrams for SiGe and AlGaAs are found online (respectively) at: (i) http://jas.eng.buffalo.edu/education/semicon/SiGe/index.html (ii) http://jas.eng.buffalo.edu/education/semicon/AlGaAs/ternary.html (b) In addition to angle-resolve PES, *photoluminescence spectroscopy* is typically used as a nondestructive means to delineate the electronic properties of materials. For more information, see: (i) http://www.nrel.gov/measurements/photo.html (ii) Glinka, Y. D.; Lin, S. -H.; Hwang, L. -P.; Chen, Y. -T. *J. Phys. Chem. B* **2000**, 104, 8652. (iii) Wu, J.; Han, W. -Q.; Walukiewicz, W.; Ager, J. W.; Shan, W.; Haller, E. E.; Zettl, A. *Nano Lett.* **2004**, 4, 647.

[69] Synchrotron radiation is generated by the acceleration of ultrarelavistic (*i.e.*, moving near the speed of light) electrons through magnetic fields. This is accomplished by forcing the electrons to repeatedly travel in a closed loop by strong magnetic fields. The resulting radiation is orders of magnitude more intense than X-rays generated from X-ray tubes, and is widely tunable in energy (from <1 eV to MeVs). A popular source is the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory, Upton, NY. A Si(111) crystal monochromator is typically used to vary the photon energy incident to the sample.

[70] For an application example of EXAFS, see: Borgna, A.; Stagg, S. M.; Resasco, D. E. *J. Phys. Chem. B* **1998**, *102*, 5077. An example of XPS and XAFS (XANES and EXAFS), see: Chakroune, N.; Viau, G.; Ammar, S.; Poul, L.; Veautier, D.; Chehimi, M. M.; Mangeney, C.; Villain, F.; Fievet, F. *Langmuir* **2005**, *21*, 6788.

[71] For an application example of REFLEXAFS, see: d'Acapito, F.; Emelianov, I.; Relini, A.; Cavotorta, P.; Gliozzi, A.; Minicozzi, V.; Morante, S.; Solari, P. L.; Rolandi, R. *Langmuir* **2002**, *18*, 5277.

[72] Gervasini, A.; Manzoli, M.; Martra, G.; Ponti, A.; Ravasio, N.; Sordelli, L.; Zaccheria, F. *J. Phys. Chem. B* **2006**, *110*, 7851.

[73] Data analysis represents the most essential and time-consuming aspects of these techniques (as well as EELS). Typically, sample spectra are compared to reference samples that contain the probed element with similar valences and bonding motifs. In this example, Cu metal foil was used for the Cu-Cu interactions, and CuO/Cu_2O were used for the Cu-O contributions. A variety of software programs are used for detailed curve-fitting, in order to obtain information regarding the chemical environment of the sample. For example, see: (a) http://cars9.uchicago.edu/~ravel/software/Welcome.html (b) http://www.dragon.lv/eda/ (c) http://www.aecom.yu.edu/home/csb/exafs.htm (d) http://www.xsi.nl/software.html (e) http://www.xpsdata.com/

74 For information regarding the quantum mechanical description of spin–orbit splitting, see: (a) http://hyperphysics.phy-astr.gsu.edu/hbase/quantum/sodzee.html#c1 (b) http://www.technology.niagarac.on.ca/people/mcsele/lasers/Quantum.htm (c) http://gardenofdestiny.com/Physics%20of%20Destiny.htm

75 The presence of these satellites are indicative of Cu^{2+}. For instance, see: (a) Espinos, J. P.; Morales, J.; Barranco A.; Caballero, A.; Holgado, J. P.; Gonzalez-Elipe, A. R. *J. Phys. Chem. B* **2002**, *106*, 6921. (b) Morales, J.; Caballero, A.; Holgado, J. P.; Espinos, J. P.; Gonzalez-Elipe, A. R. *J. Phys. Chem. B* **2002**, *106*, 10185. (c) Fuggle, J. C.; Alvarado, S. F. *Phys. Rev. A* **1980**, *22*, 1615 (describes the cause of peak broadening in XPS spectra, related to core-level lifetimes).

76 Many references exist for MIES studies of surfaces, most often carried out in tandem with UPS (to gain information for both the surface and immediate subsurface of the sample). For example, see: (a) Johnson, M. A.; Stefanovich, E. V.; Truong, T. N.; Gunster, J.; Goodman, D. W. *J. Phys. Chem. B* **1999**, *103*, 3391. (b) Kim, Y. D.; Wei, T.; Stulz, J.; Goodman, D. W. *Langmuir* **2003**, *19*, 1140 (very nice work that describes the shortfall of UPS alone, and the utility of a tandem UPS/MIES approach).

77 Though conventional RBS is carried out with He^{+} ions (which will backscatter from any atom with a greater Z), heavier ions such as C, O, Si, or Cl may be used in order to prevent background backscattering interactions with the matrix. For example, use of incident O ions to eliminate backscattering from lattice O atoms for the RBS analysis of ceramic oxides.

78 Simulations for ion scattering techniques such as RBS are typically compared with actual spectra in order to characterize the surface features. There are many such algorithms; for example: (a) Kido, Y.; Koshikawa, T. *J. Appl. Phys.* **1990**, *67*, 187. (b) Doolittle, L. R. *Nucl. Instrum. Methods* **1986**, *B*15, 227 (RUMP program). (c) http://www.ee.surrey.ac.uk/SCRIBA/ndf/(Ion Beam DataFurnace). (d) http://www.ee.surrey.ac.uk/SCRIBA/ndf/publist.html (publications re RBS simulations). (e) http://www-iba.bo.imm.cnr.it/(a nice compilation of software for ion-beam analyses)

79 http://en.wikipedia.org/wiki/Van_de_Graaff_generator. Manysuch systems exist; some examples include: (a) Western Michigan University (http://tesla.physics.wmich.edu/AcceleratorFacility.php?PG=1). (b) Brookhaven National Laboratory (http://tvdg10.phy.bnl.gov/index.html). (c) National Institute of Standards and Technology (NIST, http://physics.nist.gov/Divisions/Div846/Gp2/graaff.html). (d) Yale University (http://wnsl.physics.yale.edu/)

80 Also known as forward recoil scattering (FRS) or hydrogen forward scattering (HFS).

81 Naab, F. U.; Holland, O. W.; Duggan, J. L.; McDaniel, F. D. *J. Phys. Chem. B* **2005**, *109*, 1415, and references therein.

82 For a very thorough web presentation regarding SIMS see: (a) http://www.eaglabs.com/en-US/presentations/TOFSIMS/Presentation_Files/index.html (b) http://www.eaglabs.com/en-US/research/research.html (other links to SIMS theory, applications, presentations)

83 For background information and recipes to study a variety of polymers using MALDI, see: http://polymers.msel.nist.gov/maldirecipes/maldi.html

84 For a thesis that has a nice background on ESI, see: http://www.diva-portal.org/diva/getDocument?urn_nbn_se_uu_diva-2605-1__fulltext.pdf

85 Typically, the majority of secondary ions are ejected from the top two or three monolayers (10–20 Å) of the sample.

86 For instance, the ion concentration of the impinging ion beam must be $<1\%$ of the number of surface molecules. If this "static limit" is breached, a residue from molecular fragmentation will build up on the surface, which depletes the signal.

87 The use of fullerene ion sources represents an area of increasing interest. For example, see: Cheng, J.; Winograd, N. *Anal. Chem.* **2005**, *77*, 3651, and references therein.

88 Winograd, N. *Anal. Chem.* **2005**, *77*, 142A.

89 For example, see: (a) Delcorte, A.; Medard, N.; Bertrand, P. *Anal. Chem.* **2002**, *74*, 4955. (b) Delcorte, A.; Bour, J.; Aubriet, F.; Muller, J. -F.; Bertrand, P. *Anal. Chem.* **2003**, *75*, 6875.

90 Marcus, A.; Winograd, N. *Anal. Chem.* **2006**, *78*, 141.

91 Images reproduced with permission from Kim, Y. -P.; Oh, E.; Hong, M. -Y.; Lee, D.; Han, M. -K.; Shon, H. K.; Moon, D. W.; Kim, H. -S.; Lee, T. G. *Anal. Chem.* **2006**, *78*, 1913. Copyright 2006 American Chemical Society.

[92] Images reproduced with permission from Verlinden, G.; Janssens, G.; Gijbels, R.; Van Espen, P. *Anal. Chem.* **1997**, *69*, 3772. Copyright 1997 American Chemical Society.

[93] It should be noted that near-field scanning optical microscopy (NSOM) (discussed at the beginning of this chapter) is often grouped alongside other SPM techniques. However, for our discussion, we will focus on AFM and STM since these use *physical* probes to interrogate a surface, rather than focused light.

[94] For example, see: Zhang, J.; Chi, Q.; Ulstrup, J. *Langmuir* **2006**, *22*, 6203.

[95] For example, see: (a) France, C. B.; Frame, F. A.; Parkinson, B. A. *Langmuir* **2006**, *22*, 7507. (b) Li, W. -S.; Kim, K. S.; Jiang, D. -L.; Tanaka, H.; Kawai, T.; Kwon, J. H.; Kim, D.; Aida, T. *J. Am. Chem. Soc.* **2006**, *128*, 10527. (c) Namai, Y.; Matsuoka, O. *J. Phys. Chem. B* **2006**, *110*, 6451.

[96] For a recent precedent, see: Park, J. B.; Jaeckel, B.; Parkinson, B. A. *Langmuir* **2006**, *22*, 5334, and references therein. For a thorough recent review, see: Wan, L. -J. *Acc. Chem. Res.* **2006**, *39*, 334.

[97] For example, see: Alam, M. S.; Dremov, V.; Muller, P.; Postnikov, A. V.; Mal, S. S.; Hussain, F.; Kortz, U. *Inorg. Chem.* **2006**, *45*, 2866.

[98] Examples of some common forces that may exist between a surface and an AFM tip are Van der Waal, electrostatic, covalent bonding, capillary, and magnetic. In addition to providing information regarding the topography of the surface (constant force mode), forces may be applied to understand the morphology of a surface – for example, to determine the frictional force between the tip and surface, or the elasticity/hardness of a surface feature. For instance, see: Tranchida, D.; Piccarolo, S.; Soliman, M. *Macromolecules* **2006**, *39*, 4547, and references therein.

[99] For example, see: O'Dwyer, C.; Gay, G.; Viaris de Lesegno, B.; Weiner, J. *Langmuir* **2004**, *20*, 8172, and references therein.

[100] For a more sophisticated commercial liquid cell AFM system, see: http://www.veeco.com/escope

[101] For example, see: (a) Cho, Y.; Ivanisevic, A. *Langmuir* **2006**, *22*, 1768. (b) Poggi, M. A.; Lillehei, P. T.; Bottomley, L. A. *Chem. Mater.* **2005**, *17*, 4289. (c) Gourianova, S.; Willenbacher, N.; Kutschera, M. *Langmuir* **2005**, *21*, 5429.

[102] For example, see: (a) Takamura, Y.; Chopdekar, R. V.; Scholl, A.; Doran, A.; Liddle, J. A.; Harteneck, B.; Suzuji, Y. *Nano Lett.* **2006**, 6, 1287. (b) Li, Y.; Tevaarwerk, E.; Chang, R. P. H. *Chem. Mater.* **2006**, *18*, 2552.

[103] For example, see: Zhang, J.; Roberts, C. J.; Shakesheff, K. M.; Davies, M. C.; Tendler, S. J. B. *Macromolecules* **2003**, *36*, 1215, and references therein.

[104] For a thorough description of SECM, see: Gardner, C. E.; Macpherson, J. V. *Anal. Chem.* **2002**, *74*, 576A.

[105] For instance, see: https://www.veecoprobes.com/probes.asp

[106] (a) Nanoparticle-terminated tips: Reproduced with permission from Vakarelski, I. U.; Higashitani, K. *Langmuir* **2006**, *22*, 2931. Copyright 2006 American Chemical Society. (b) Nanotube-terminated tips: Reproduced with permission from Hafner, J. H.; Cheung, C. -L.; Oosterkamp, T. H.; Lieber, C. M. *J. Phys. Chem. B* **2001**, *105*, 743. Copyright 2001 American Chemical Society. (c) Nanotube-terminated tips: Wilson, N. R.; Macpherson, J. V. *Nano Lett.* **2003**, *3*, 1365.

[107] That is, an AFM probe responds to the *average* force between the sample surface and a group of tip atoms that are in close proximity to the surface. In order to image individual atoms by SPM, the surface–tip interactions must be limited to the nearest atom(s) on the tip periphery. Hence, an AFM image will not show individual atoms, but rather an average surface, with its ultimate resolution dependent on the sharpness of the tip structure. In contrast, STM is capable of atomic resolution since the tunneling current passes only through the tip atom that is nearest the sample surface.

[108] Named after Brunauer, Emmett, and Teller.

[109] For example, see: Gillan, E. G.; Bott, S. G.; Barron, A. R. *Chem. Mater.* **1997**, *9*, 796.

[110] Tandem TGA/DSC instruments are commercially available, for example: http://www.tainst.com/product.aspx?n=1&id=22

[111] An analogous (older) technique is known as *differential thermal analysis* (DTA), which yields the same information as DSC.

[112] For more information on small-angle scattering, see: (a) http://www.isis.rl.ac.uk/largescale/loq/documents/sans.htm (b) http://www.eng.uc.edu/~gbeaucag/Classes/XRD/SAXS%20Chapter/SAXSforXRD.htm

[113] Schematics (a) and (b) were obtained from the public domain from the website: http://en.wikipedia.org/wiki/Differential_scanning_calorimetry

[114] For a nice example that utilizes SAXS, WAXS, and SANS to determine the structural changes of polyethylene chains following annealing, see: Men, Y.; Rieger, J.; Lindner, P.; Enderle, H. -F.; Lilge, D.; Kristen, M. O.; Mihan, S.; Jiang, S. *J. Phys. Chem. B* **2005**, *109*, 16650.

[115] For an example of a quantitative SAXS study of a block copolymer-solvent system see: Soni, S. S.; Brotons, G.; Bellour, M.; Narayanan, T.; Gibaud, A. *J. Phys. Chem. B* **2006**, *110*, 15157. An example of the use of SAXS to determine the particle size distribution of nanoparticles, see: Rieker, T.; Hanprasopwattana, A.; Datye, A.; Hubbard, P. *Langmuir* **1999**, *15*, 638.

Topics for Further Discussion

1. In our discussion of sample staining for TEM analysis, we mentioned that lead citrate was sensitive to CO_2. What is the balanced reaction for generation of the side-product?
2. What is the difference between XPS and AES?
3. What characterization techniques would be best suited for the following:
   (a) Analysis of a thin film (20 nm thickness) for Na content (very precise value is desired).
   (b) Determination of the Li content of a thin film – comparison of the surface concentration with the Li content at a depth of 100 nm below the surface (diffusion study).
   (c) To image and study the desorption mechanism of SAMs from the surface of Au(111) upon exposure to organic solvents.
   (d) Characterize the film resulting from an attempt to covalently graft poly (ethylene glycol) (PEG) chains on poly(ethylene-*co*-acrylic acid) (EAA) surfaces.
   (e) Characterize the intermediate structures present on a thin film during CVD, at various deposition times.
   (f) Assess whether a new synthetic procedure to grow TiO_2 nanoparticles was successful or not.
   (g) Determine the lattice parameters of metallic ReO_3 nanoparticles.
   (h) Determine the percentage yield and purity for a batch of single-walled carbon nanotubes.
4. What are the differences between STM and AFM, providing examples of their applications?
5. How would you eliminate the presence of "nanobubbles" on Au surface during AFM imaging? Why is the presence of such adsorbates problematic for SPM studies? (hint: see *Langmuir* **2003**, *19*, 10510).
6. What are the benefits of using a liquid cell for AFM studies? Are there any limitations of this technique?
7. What is SECM, and what applications would this technique be used for?
8. Describe what is meant by "charging" during SEM analyses, along with methods that are used to circumvent these effects.
9. Why is the yield of Auger electrons most prevalent for elements with low-Z, relative to heavier elements (where X-ray emission is preferred)?
10. Find literature references for tandem UPS/XPS studies of surfaces. What information is gained from each technique?
11. What techniques are used to prepare thin sample sections for TEM analysis?
12. Find literature precedents for tandem RBS/PIXE surface characterization. What information does each technique yield?
13. What are the differences between EDS and WDS?
14. Why does one need a synchrotron source in order to perform XAFS studies?
15. The Getty museum in Los Angeles has contracted your services to determine if a Monet they recently received is authentic. What techniques, including sample preparation, *etc*. would you use to determine the authenticity of this artwork (handle it carefully!)?
16. You have just discovered a new method to grow carbon nanotubes at incredibly low temperatures. The FESEM and TEM images suggest that the as-grown nanostructures are graphitic, and exhibit very desirable aspect ratios. What other characterization techniques should you perform to fully characterize your product? How would you study the growth mechanism (*in situ* and/or by stopping

the growth at varying stages of growth), assuming that you used a CVD-type reaction with an iron-based nanoparticulate catalyst, at a temperature of 200°C?

17. You have been contracted to design a new polymer that will change its structure in response to temperature changes. What methods would you use to characterize your polymer following its synthesis?
18. You have grown an oxidizable metallic thin film, but must handle it briefly in air before sealing it in a storage container. How would you evaluate the thickness of the native surface oxide, and how deep the oxide diffuses into the thin film, as the sample is annealed?

Further Reading

1. Flegler, S. L.; Heckman, J. W.; Klomparens, K. L. *Scanning and Transmission Electron Microscopy: An Introduction*, W. H. Freeman: New York, 1993.
2. Williams, D. B.; Carter, C. B. *Transmission Electron Microscopy: A Textbook for Materials Science*, Plenum Press: New York, 1996.
3. Egerton, R. F. *Physical Principles of Electron Microscopy: An Introduction to TEM, SEM, and AEM*, Springer: New York, 1986.
4. Goldstein, J.; Newbury, D.; Joy, D.; Lyman, C.; Echlin, P.; Lifshin, E.; Sawyer, L.; Michael, J. *Scanning Electron Microscopy and X-Ray Microanalysis*, 3rd ed., Kluwer: New York, 2003.
5. *Encyclopedia of Materials Characterization – Surfaces, Interfaces, Thin Films*, Brundle, C. R.; Evans, C. A.; Wilson, S. eds., Elsevier: New York, 1992.
6. Campbell, D.; Pethrick, P. A.; White, J. R. *Polymer Characterization*, 2nd ed., Stanley Thornes: Cheltenham, UK, 2000.
7. Criddle, W. J.; Ellis, G. P. *Spectral and Chemical Characterization of Organic Compounds: A Laboratory Handbook*, 3rd ed., Wiley: New York, 1990.
8. Dinardo, N. J. *Nanoscale Characterization of Surfaces and Interfaces*, 2nd ed., Wiley: New York, 2004.
9. *Surface Characterization: A User's Sourcebook*, Brune, D.; Hellborg, R.; Hunderi, O. eds., Wiley: New York, 1997.
10. *Beam Effects, Surface Topography, and Depth Profiling in Surface Analysis (Methods of Surface Characterization)*, Czanderna, A. W.; Madey, T. E.; Powell, C. J. eds., Plenum Press: New York, 1998.
11. *Ion Spectroscopies for Surface Analysis (Methods of Surface Characterization)*, Czanderna, A. W.; Hercules, D. M. eds., Springer: New York, 1991.
12. Brandon, D. D.; Kaplan, W. D. *Microstructural Characterization of Materials*, Wiley: New York, 1999.
13. Pecharsky, V.; Zavalij, P. *Fundamentals of Powder Diffraction and Structural Characterization of Materials*, Springer: New York, 2005.
14. *Concise Encyclopedia of Materials Characterization*, 2nd ed., Cahn, R. ed., Elsevier: San Diego, CA, 2005.
15. *Characterization of Polymers (Materials Characterization)*, Tong, H. -M.; Kowalczyk, S. P.; Saraf, R.; Chou, N. J. eds., Butterworth-Heinemann: New York, 1993.

APPENDIX A

TIMELINE OF MATERIALS AND TECHNOLOGICAL DISCOVERIES

>50,000 B.C.	Brushes are developed to apply pigment to cave walls
30,000 B.C.	Clothing materials are fabricated from animal skins
24,000 B.C.	Ceramic materials are made from animal fat and bone, mixed with bone ash and clay
20,000 B.C.	Ivory and bone are used to make sewing needles
20,000 B.C.	A nonwoven fabric, later termed felt, is made from compressed wool/hair
10,000 B.C.	Gourds, seapods, bones, and clay are used to make ocarinas or vessel flutes
4,000 B.C.	Stones are first used to construct roads in Ur (modern-day Iraq)
3,500 B.C.	Copper metallurgy is invented and used to fabricate a variety of materials
3,500 B.C.	The first reported use of glass in Egypt and Mesopotamia
3,400 B.C.	Linen cloth synthesized from flax is used to wrap mummies in Egypt
3,200 B.C.	Bronze is used for weapons and armor
3,000 B.C.	Egyptians wear clothing comprised of cotton fibers
3,000 B.C.	The Egyptians construct the first stringed musical instrument
3,000 B.C.	Soap is first synthesized in Egypt using wood ash and animal fat
2,600 B.C.	Silk fibers are used for clothing in China
2,000 B.C.	Pewter beginning to be used in China and Egypt
1,600 B.C.	The Hittites develop iron metallurgy
1,600 B.C.	Conceptual designs are invented for bathing suits, fabricated/named bikini in 1946
1,300 B.C.	Invention of steel when iron and charcoal are combined properly
1,000 B.C.	The abacus is created by the Babylonians
1,000 B.C.	Glass production begins in Greece and Syria
900s B.C.	Assyrians develop pontoon rafts for armies to cross rivers
800s B.C.	Spoked wheels are fabricated and used throughout Europe
700 B.C.	Italians invent false teeth
105 B.C.	Paper is first fabricated from bamboo fiber in ancient China
50 B.C.	Glassblowing techniques are developed in Syria
Birth of Christ	
590	Chinese scientists discover explosive mixtures consisting of sulfur, charcoal, and saltpeter (potassium nitrate)
618	Paper money is first put into use during the Tang dynasty of China (618–906)
700s	Porcelain is invented in China

1156	First reported synthesis of perfume by Henchum Seiken
1182	The magnetic compass is developed and widely used in China[1]
1249	Gunpowder is designed/synthesized by Rodger Bacon
1280	The cannon is invented in China
1286	Eyeglasses are first used in Venice
1400	The first use of grenades in France, designed by an unknown inventor
1450	Crystallo, a clear soda-based glass, is invented by Angelo Barovier
1570	The pinhole camera is invented
1590	Glass lenses are developed in Netherlands and used for the first time in microscopes and telescopes
1593	Galileo invents a water thermometer
1608	The Dutch scientist Hans Lippershey invents the telescope
1612	The Flintlock firearm is developed in France
1621	John Napier invents the slide rule
1643	Torricelli makes the first barometer using mercury in a sealed glass tube
1651	The Dutch scientist Anton van Leeuwenhoek develops a microscope
1668	Isaac Newton invents a reflecting telescope
1709	Gabriel Fahrenheit invents an alcohol thermometer (mercury thermometer developed in 1714)
1710	Bathroom bidet is invented in France
1712	The steam engine is first invented in England
1714	The first patent for a typewriter is awarded in England to Henry Mill
1718	The machine gun is developed in England
1738	William Champion patents a process for the production of metallic zinc by distillation from calamine and charcoal
1770	First reported use of porcelain false teeth in France
1774	The electric telegraph is developed by Georges Louis Lesage
1779	Bry Higgins issued a patent for hydraulic cement (stucco) for use as an exterior plaster
1789	Chlorine bleach is developed by Claude Louis Berthollet in France
1800	Alessandro Volta makes a Copper/Zinc acid battery
1815	Humphry Davy invents a safety lamp that is used in coal mines without triggering an explosion
1820	Thomas Hancock develops the first elastic fabrics
1821	Thomas Johann Seebeck invents the thermocouple
1823	Charles Macintosh patents a method for making waterproof garments
1824	Patent issued to Joseph Aspdin for the invention of cement
1825	Hans Christian Orsted produces metallic aluminum
1825	William Sturgeon invents the electromagnet
1837	Wheatstone and Cooke invent the telegraph
1838	Regnault polymerizes vinylidene chloride via sunlight
1839	Goodyear (US), MacIntosh, and Hancock (England) vulcanize natural rubber
1839	Sir William Robert Grove experimented with the first fuel cell, using hydrogen and oxygen gases in the presence of an electrolyte
1842	The facsimile machine is invented by Alexander Bain
1849	Ferroconcrete, concrete reinforced with steel, is invented by Monier
1855	Bessemer process for mass production of steel patented
1856	Invention of the first synthetic dye, mauveine, by William Henry Perkin
1857	Toilet paper is designed and marketed for the first time
1860	Fredrick Walton invents linoleum, comprised of linseed oil, pigments, pine rosin, and pine flour
1861	James Clerk Maxwell demonstrates color photography
1864	Development of flash photography by Henry Roscoe in England

1872	Asphalt is first developed by Edward de Smedt at Columbia University
1872	Polyvinyl chloride (PVC) is first created by Eugen Baumann
1873	Levi Strauss & Co. begin producing blue-jeans out of durable canvas
1876	Nicolaus Otto invents a gas motor engine
1877	Thomas Edison completes the first phonograph
1881	Alexander Graham Bell builds the first metal detector
1883	Charles Fritts makes the first solar cells using selenium wafers
1885	Sunglasses are invented
1885	Karl Benz designs and builds the first gasoline-fueled automobile
1885	The first gasoline pump is manufactured by Sylvanus Bowser
1887	Contact lenses are invented by Eugen Frick in Switzerland
1888	George Eastman introduces a Kodak camera
1890	The zipper is invented by Whitcomb Judson in Chicago, IL
1891	The first commercially produced artificial fiber, Rayon, is invented
1892	Calcium carbide is synthesized, as well as acetylene gas that is generated from the carbide
1893	Edward Goodrich Acheson patents a method for making carborundum (SiC), an abrasive compound
1896	Henry Ford constructs the first horseless carriage
1901	The first mercury arc lamp is developed by Peter Hewitt
1902	August Verneuil develops a process for making synthetic rubies
1902	The neon light is invented in France
1903	Ductile tungsten wire is synthesized by Coolidge
1907	Leo Hendrik Baekeland invents Bakelite (phenol- formaldehyde resins), used in electronic insulation
1908	Cellophane is invented by Brandenberger, a Swiss textile engineer
1909	Leo Baekeland presents the Bakelite hard thermosetting plastic
1916	Jan Czochralski invents a method for growing single crystals of metals
1916	Kotaro Honda discovers a strongly magnetic Co/W alloy
1920	Herman Staudinger (Germany) advances the macromolecular hypothesis – the birth of polymer science
1923	Mercedes introduces the first supercharged automobile, the 6/25/40 hp
1924	Corning scientists invent Pyrex, a glass with a very low thermal expansion coefficient
1924	Celanese Corporation commercially produces acetate fibers
1924	The first mobile, two-way voice-based telephone is invented at Bell Labs
1926	Waldo Semon at B.F. Goodrich invents plasticized PVC known as *vinyl*
1929	Polysulfide (Thiokol) rubber is synthesized
1929	Carothers (du Pont) synthesizes the first aliphatic polyesters, establishes the principles of step-growth polymerization, and develops nylon 6,6
1931	Julius Nieuwland develops the synthetic rubber called *neoprene*
1931	Poly(methylmethacrylate) (PMMA) is synthesized
1932	Hans von Ohain and Sir Frank Whittle file patents for the jet engine
1932	Cathode ray tubes (CRTs) are invented by Allen B. Du Mont
1933	Ernest Ruska discovers the electron microscope; magnification of 12,000×
1933	Fawcett and Gibson develop polyethylene (LDPE)
1936	The first programmable computer, the Z1, is developed by Konrad Zuse
1937	Polystyrene is developed
1937	Chester Carlson invents a dry printing process commonly called *Xerox*
1938	Roy Plunkett discovers the process for making poly-tetrafluoroethylene, better known as *Teflon*™
1940	Thomas and Sparka synthesize isobutylene–isoprene rubber
1940	Butyl rubber is synthesized in the US

1941	Canadian John Hopps invents the first cardiac pacemaker
1942	The synthetic fabric, polyester, is invented
1943	The first kidney dialysis machine is developed
1943	Polyurethanes are synthesized by Otto Baeyer
1944	The first plastic artificial eye is developed in the US
1945	Percy Spencer creates the first microwave oven
1946	Mauchly and Eckert develop the first electronic computer ENIAC (Electronic Numerical Integrator and Computer)
1947	The first transistor is invented by Bardeen, Brattain, and Shockley at Bell Labs
1947	The first commercial application of a piezoelectric ceramic (barium titanate) used as a phonograph needle
1947	Invention of magnetic tape for recording applications
1947	Schlack develops epoxy polymeric systems
1951	Individual atoms seen for the first time using the field ion microscope
1950	The first commercial production of acrylic fibers by du Pont
1951	The computer UNIVAC 1 is developed
1951	Polypropylene is developed by Paul Hogan & Robert Banks of Phillips
1952	The first application of antiperspirant deodorant with a roll-on applicator
1953	Karl Ziegler discovers metallic catalysts which greatly improve the strength of polyethylene polymers
1954	Six percent efficiency silicon solar cells made at Bell Labs
1954	Charles Townes and Arthur Schawlow invent the MASER (microwave amplification by stimulated emission or radiation)
1955	Optical fibers are produced
1956	Liquid Paper™ is formulated by Bette Nesmith Graham
1957	Keller first characterizes a single crystal of polyethylene
1958	Bifocal contact lenses are produced
1959	Pilkington Brothers patent the float glass process
1959	The first commercial production of Spandex fibers by du Pont
1960s	Polymers are first characterized by GPC, NMR, and DSC
1960	The first working laser (pulsed ruby) is developed by Maimam of Hughes Aircraft Corporation. Javan, Bennet, and Herriot make the first He:Ne gas laser
1960	Spandex fibers are synthesized
1962	The first SQUID superconducting quantum interference device is invented
1962	Polyimide resins are synthesized
1963	The first balloon embolectomy catheter is invented by Thomas Fogarty
1963	Ziegler and Natta are awarded the Nobel Prize for 1950's polymerization studies
1964	Bill Lear (of "Lear Jet" fame!) designs the first eight-track player
1965	A bulletproof nylon fabric, Kevlar, is invented at DuPont
1965	James Russell invents the compact disk
1965	Styrene–butadiene block copolymers are synthesized
1966	Fuel-injection systems for automobiles are developed in the UK
1966	Faria and Wright of Monsanto synthesize and test Astroturf
1967	Keyboards are first used for data entry, replacing punch cards
1968	Liquid crystal display is developed by RCA
1968	Allen Breed invents the first automotive air bag system
1969	The scanning electron microscope (SEM) is first used in laboratories to view cells in 3D
1969	George Smith and Willard Boyle invent charge-coupled devices (CCD) at Bell Labs
1970	The floppy disk (8 in.) is invented by Alan Shugart at IBM

1970	The first microfiber (polyester) is invented by Toray Industries in Japan. The first fabric comprised of microfibers, Ultrasuede, is also introduced
1971	The liquid crystal display (LCD) is invented by James Fergason
1971	The first single chip microprocessor, Intel 4004, is introduced
1971	The video cassette recorder (VCR) is invented by Charles Ginsburg
1971	Hydrogels are synthesized
1972	Motorola demonstrates the use of the first portable cellular phone[2]
1973	The disposable lighter is invented by Bic
1973	Magnetic resonance imaging (MRI) is invented by Lauterbur and Damadian[3]
1974	Post-it® notes featuring a low-residue adhesive is invented by 3M
1975	The laser printer is invented
1975	Robert S. Ledley is issued the patent for "diagnostic X-ray systems" (CAT scans)
1976	The inkjet printer is developed by IBM
1977	The Cray-1® supercomputer is introduced by Seymour Cray
1977	Electrically conducting organic polymers are synthesized by Heeger, MacDiarmid, and Shirakawa (Nobel Prize awarded in 2000)
1978	An artificial heart, Jarvik-7, is invented by Robert Jarvik
1978	The first analog video optical disk player is introduced by MCA Discovision
1979	The first cassette Walkman TPS-L2 is invented by Masaru Ibuka of Sony
1980	Compact disk players are introduced by Philips
1981	The world's largest solar-power generating station goes into operation (10 MW capacity)
1981	The scanning tunneling microscope (STM) is invented
1982	The first "personal computer" (PC) is introduced by IBM[4]
1982	Robert Denkwalter *et al.* from Allied Corporation are granted the first patent for dendrimers
1983	US phone companies begin to offer cellular phone service
1983	Steve Jobs of Apple introduces a new computer featuring the first graphical user interface (GUI), named *The Lisa*
1984	The CD-ROM is invented for computers
1984	The first clumping kitty litter is invented by biochemist Thomas Nelson
1985	Donald Tomalia and coworkers at Dow Chemical report the discovery of hyperbranched polymers, named *dendrimers*
1986	Synthetic skin is invented by Gregory Gallico, III
1987	Bednorz and Muller develop a material that is superconducting at $-183°C$
1987	Conducting polymers are developed by BASF
1988	A patent is issued for the Indiglo™ nightlight, consisting of electroluminescent phosphor particles
1989	High-definition television is invented
1989	NEC releases the first "notebook" computer, the NEC Ultralite
1989	A breathable, water- or wind-proof fabric, GORE-TEX®, is introduced
1989	The Intel 486 microprocessor is developed, featuring 1,000,000 transistors
1990	Biotextiles are invented in the US
1991	Iijima of NEC Corporation discovers carbon nanotubes
1992	MiniDiscs (MDs) are introduced by Sony Electronics, Inc.
1992	Prof. Jerome Schentag invents a computer-controlled "smart pill," for drug-delivery applications
1993	The Pentium processor is invented by Intel
1994	The first search engine for the World Wide Web is created by Filo and Yang[5]

1994	Lyocell is introduced by Courtaulds Fibers, consisting of a material derived from wood pulp
1995	Digital Versatile Disk or Digital Video Disk (DVD) is invented
1996	The Nobel Prize in Chemistry is awarded to Richard Smalley, Robert Curl, and Harry Kroto for their 1985 discovery of the third form of carbon, known as *buckminsterfullerene* ("bucky balls")[6]
1996	WebTV is invented by Phillips
1996	The Palm Pilot is debuted by 3Com
1997	The gas-powered fuel cell is invented
1997	A fire-resistant building material, Geobond, is patented
1997	Nokia introduces the Nokia 9000i Communicator. This combines a digital cell phone, hand-held PC, and fax
1998	Motorola introduces Iridium service, the first global satellite-based wireless telephone service[7]
1998	Adam Cohen (19 years old!) develops an "electrochemical paint brush" circuit that uses an STM probe to manipulate copper atoms on a silicon surface
1998	Apple computer introduces the iMac
1998	Geoffrey Ozin at the University of Toronto develops synthetic seashells from SiO_2
1998	Toyota Motor Corporation releases the Prius – the first mass-produced hybrid low-emission vehicle (LEV)
1998	Television stations in the US began to transition from analog to digital signals
1999	Danish physicist Hau is able to control the speed of light, useful for potential applications in communications systems and optical computers
1999	Safeco Field in Seattle opens, featuring a retractable roof, and extensive drainage lines and heating coils to maintain ideal turf conditions[8]
1999	The chemical ingredient used by mussels to anchor themselves to rocks is discovered, and used to synthesize a waterproof adhesive
1999	Molecular-based logic gates are demonstrated to work better than silicon-based gates – an important precedent in the development of a molecular computer
2000	Intel releases the Pentium IV microprocessor, consisting of 42 million transistors
2000	Motorola releases the i1000 Plus – the first cell phone capable of connecting to the internet
2000	Robotic pets (*e.g.*, Poo-Chi, Tekno) are first introduced
2000	The first generation of "digital jukeboxes," the AudioReQuest ARQ1, retails for $800 and is the first device capable of storing thousands of MP3 songs
2001	The AbioCor self-contained artificial heart is implanted into Robert Tools
2001	SmartShirt sensors, to record and report body diagnostics, are designed by SensaTex, Inc. and Georgia Tech Research Corporation
2001	The bioartificial liver is invented by Kenneth Matsumura
2001	A fuel-cell bicycle is developed by Aprilia
2001	Digital satellite radio is developed by XM and Sirius
2001	SunClean self-cleaning glass is introduced by PPG Industries
2001	A wrinkle-free shirt is developed by Corpo Nove (Italy), consisting of Ti-alloy fibers interwoven with nylon
2002	Clothing comprised of nanowhiskers is invented by Nano-tex, LLC to aid in stain resistance
2002	The lightest substance on Earth, known as *Aerogels*, is developed by NASA

2002	Scientists at SUNY, Buffalo, develop a new type of semiconducting material, GaSb/Mn, that will be used for future spintronics-based devices[10]
2003	Scientists discover a method used to commercially produce spider-web silk[9]
2003	Nanoparticles are used for the first time for clearcoat paint finishes (PPG – Ceramiclear^{TM})
2003	Nanofilters are used to purity groundwater in Manitoba, Canada
2003	Digital videodisk recorders (DVRs) are introduced
2003	IBM develops the smallest light-emitting transmitter, comprised of carbon nanotubes (CNTs)
2003	Apple computer releases laptops featuring 17- and 12-in. LCD screens
2003	Bandages are made from fibrinogen, a soluble protein found in blood
2004	The Blue Gene/L produced by IBM is able to perform 70.7 trillion calculations per second, making it the fastest computer in the world, to date
2004	Apple releases the iPod mini – the size of a business card, but able to hold 1,000 songs
2004	Scientists are able to control polymorphism through crystallization within nanopores
2004	A compound in the shape of a Borromean knot is discovered, based on earlier theoretical simulations
2004	Nintendo releases the hand-held gaming system, Nintendo DS
2005	Carbon nanotubes are synthesized in bulk, and spun into a yarn
2005	iPod Nano and a video-capable iPod are introduced by Apple
2005	Motorola releases the ROKR E1 phone, capable of music downloading
2006	High-definition DVD players become commercially available
2006	Apple computer introduces MacBook Pro, MacBook, and iMac product lines that contain Intel dual-core chips – the first to contain over one billion transistors
2006	Flat-panel display technologies employing carbon nanotubes are demonstrated
2006	LG designs cellular phone that has a built-in breathalyzer for sobriety testing; this application is also tested as standard equipment for future automobiles
2007	Apple releases the iPhone, which combines cellular phone, internet, and iPod functionalities
2007	LG releases the first dual HD-DVD/Blu-ray high-definition player

References and Notes

1 The magnetic compass may have been first used during the Qin dynasty in China (*ca.* 200 B.C.).

2 The first call that was made with a portable cellular phone was made by Dr. Martin Cooper, who called his rival, Joel Engel, at Bell Labs!.

3 The Nobel Prize for Medicine was awarded to Paul C. Lauterbur and Sir Peter Mansfield "for their discoveries concerning magnetic resonance imaging." This announcement failed to acknowledge Raymond V. Damadian who was the first person to propose MRI for medical diagnostics. In an effort to become properly acknowledged, Damadian placed full-page ads in major newspapers such as the *Washington Post*, the *New York Times*, and the *Los Angeles Times*. His fight for proper recognition is not unwarranted; the US National Medal of Technology was awarded to Damadian and Lauterbur in 1988, which recognized both scientists as co-inventors. Unfortunately, in an attempt to "make them [the Stockholm Nobel Prize committee] accountable to world opinion," Damadian may only be remembered for his unprecedented fight for personal recognition, rather than his co-discovery.

[4] The first PC (IBM 5150) retailed for $2,880, and was powered by the Intel 8088 microprocessor comprised of a 3-μm circuit, containing 29,000 transistors. This system was capable of performing 4.8 million cycles per second (4.8 MHz). The first microprocessor, the Intel 4004 developed in 1971, consisted of 2,300 transistors in a 3.5-mm circuit. Amazingly, modern computers costing $2,000 are now capable of performing 2+ billion instructions per second (2 GHz), and feature 100s of millions of transistors contained within a circuit size of 130 nm! Computational devices have decreased steadily at about five linear dimensions per decade, but are rapidly approaching a barrier. Chapter 6 on nanotechnology will discuss the future of electronic devices.

[5] This was simply a list of other sites called "Jerry's Guide to the World Wide Web." Within 8 months, more than 100,000 people were using this sit as an index to the Web, and it was eventually renamed Yahoo!.

[6] This award is considered to have generated a new field known as "nanotechnology," as worldwide exposure was instantly aware of these nanoscale molecules, and other developments in this size regime were found shortly thereafter. It should be noted that the discovery of dendrimers by Denkwalter and coworkers from Allied Corporation was disclosed in 1981, 4 years before bucky balls were discovered. It may be expected that these nanopolymeric materials will be extremely influential toward the nanotechnology revolution (see Chapter 5 for more information on dendritic materials).

[7] This service quickly filed for bankruptcy; satellite-based wireless service is not yet widely available.

[8] The price tag was $500M, making Safeco Field the most expensive field built to date in the US.

[9] Spider-web silk is *ca.* five times stronger than steel by weight, and almost as elastic as nylon. Fibers comprised of the synthetic silk were demonstrated to be stronger than Kevlar, and may be useful for biomedical applications such as artificial tendons and ligaments and surgery sutures, as well as lightweight body armor for military applications.

[10] Rather than using electronics to turn switches on/off, spintronic devices use electronic spins to represent information. This will allow these devices to process billions of pieces of information simultaneously, greatly increasing the speed and power of electronic devices. For more information on the future of spintronics, see: http://www.spintronics-info.com/.

APPENDIX B

"THERE'S PLENTY OF ROOM AT THE BOTTOM"

This speech was given by Richard Feynman on 29 December 1959 at the Annual American Physical Society meeting. This text is provided with permission from the *Engineering and Science* magazine, published quarterly by the California Institute of Technology (http://www.pr.caltech.edu/periodicals/EandS/).

"Imagine experimental physicists must often look with envy at men like Kamerlingh Onnes, who discovered a field like low temperature, which seems to be bottomless and in which one can go down and down. Such a man is then a leader and has some temporary monopoly in a scientific adventure. Percy Bridgman, in designing a way to obtain higher pressures, opened up another new field and was able to move into it and to lead us all along. The development of ever higher vacuum was a continuing development of the same kind.

I would like to describe a field, in which little has been done, but in which an enormous amount can be done in principle. This field is not quite the same as the others in that it will not tell us much of fundamental physics (in the sense of, 'What are the strange particles?') but it is more like solid-state physics in the sense that it might tell us much of great interest about the strange phenomena that occur in complex situations. Furthermore, a point that is most important is that it would have an enormous number of technical applications. What I want to talk about is the problem of manipulating and controlling things on a small scale.

As soon as I mention this, people tell me about miniaturization, and how far it has progressed today. They tell me about electric motors that are the size of the nail on your small finger. And there is a device on the market, they tell me, by which you can write the Lord's Prayer on the head of a pin. But that's nothing; that's the most primitive, halting step in the direction I intend to discuss. It is a staggeringly small world that is below. In the year 2000, when they look back at this age, they will wonder why it was not until the year 1960 that anybody began seriously to move in this direction.

Why cannot we write the entire 24 volumes of the Encyclopedia Brittanica on the head of a pin?

Let's see what would be involved. The head of a pin is a sixteenth of an inch across. If you magnify it by 25,000 diameters, the area of the head of the pin is then equal to

the area of all the pages of the Encyclopaedia Brittanica. Therefore, all it is necessary to do is to reduce in size all the writing in the Encyclopaedia by 25,000 times. Is that possible? The resolving power of the eye is about 1/120 of an inch – that is roughly the diameter of one of the little dots on the fine half-tone reproductions in the Encyclopaedia. This, when you demagnify it by 25,000 times, is still 80 angstroms in diameter – 32 atoms across, in an ordinary metal. In other words, one of those dots still would contain in its area 1,000 atoms. So, each dot can easily be adjusted in size as required by the photoengraving, and there is no question that there is enough room on the head of a pin to put all of the Encyclopaedia Brittanica.

Furthermore, it can be read if it is so written. Let's imagine that it is written in raised letters of metal; that is, where the black is in the Encyclopedia, we have raised letters of metal that are actually 1/25,000 of their ordinary size. How would we read it? If we had something written in such a way, we could read it using techniques in common use today. (They will undoubtedly find a better way when we do actually have it written, but to make my point conservatively I shall just take techniques we know today.) We would press the metal into a plastic material and make a mold of it, then peel the plastic off very carefully, evaporate silica into the plastic to get a very thin film, then shadow it by evaporating gold at an angle against the silica so that all the little letters will appear clearly, dissolve the plastic away from the silica film, and then look through it with an electron microscope!

There is no question that if the thing were reduced by 25,000 times in the form of raised letters on the pin, it would be easy for us to read it today. Furthermore; there is no question that we would find it easy to make copies of the master; we would just need to press the same metal plate again into plastic and we would have another copy.

How do we write small?

The next question is: How do we *write* it? We have no standard technique to do this now. But let me argue that it is not as difficult as it first appears to be. We can reverse the lenses of the electron microscope in order to demagnify as well as magnify. A source of ions, sent through the microscope lenses in reverse, could be focused to a very small spot. We could write with that spot like we write in a TV cathode ray oscilloscope, by going across in lines, and having an adjustment which determines the amount of material which is going to be deposited as we scan in lines.

This method might be very slow because of space charge limitations. There will be more rapid methods. We could first make, perhaps by some photo process, a screen which has holes in it in the form of the letters. Then we would strike an arc behind the holes and draw metallic ions through the holes; then we could again use our system of lenses and make a small image in the form of ions, which would deposit the metal on the pin.

A simpler way might be this (though I am not sure it would work): we take light and, through an optical microscope running backwards, we focus it onto a very small photoelectric screen. Then electrons come away from the screen where the light is shining. These electrons are focused down in size by the electron microscope lenses

to impinge directly upon the surface of the metal. Will such a beam etch away the metal if it is run long enough? I don't know. If it doesn't work for a metal surface, it must be possible to find some surface with which to coat the original pin so that, where the electrons bombard, a change is made which we could recognize later.

There is no intensity problem in these devices – not what you are used to in magnification, where you have to take a few electrons and spread them over a bigger and bigger screen; it is just the opposite. The light which we get from a page is concentrated onto a very small area so it is very intense. The few electrons which come from the photoelectric screen are demagnified down to a very tiny area so that, again, they are very intense. I don't know why this hasn't been done yet!

That's the Encyclopaedia Brittanica on the head of a pin, but let's consider all the books in the world. The Library of Congress has approximately 9 million volumes; the British Museum Library has 5 million volumes; there are also 5 million volumes in the National Library in France. Undoubtedly there are duplications, so let us say that there are some 24 million volumes of interest in the world.

What would happen if I print all this down at the scale we have been discussing? How much space would it take? It would take, of course, the area of about a million pinheads because, instead of there being just the 24 volumes of the Encyclopaedia, there are 24 million volumes. The million pinheads can be put in a square of a thousand pins on a side, or an area of about 3 square yards. That is to say, the silica replica with the paper-thin backing of plastic, with which we have made the copies, with all this information, is on an area of approximately the size of 35 pages of the Encyclopaedia. That is about half as many pages as there are in this magazine. All of the information which all of mankind has every recorded in books can be carried around in a pamphlet in your hand – and not written in code, but a simple reproduction of the original pictures, engravings, and everything else on a small scale without loss of resolution.

What would our librarian at Caltech say, as she runs all over from one building to another, if I tell her that, ten years from now, all of the information that she is struggling to keep track of – 120,000 volumes, stacked from the floor to the ceiling, drawers full of cards, storage rooms full of the older books – can be kept on just one library card! When the University of Brazil, for example, finds that their library is burned, we can send them a copy of every book in our library by striking off a copy from the master plate in a few hours and mailing it in an envelope no bigger or heavier than any other ordinary air mail letter.

Now, the name of this talk is 'There is *Plenty* of Room at the Bottom' – not just 'There is Room at the Bottom.' What I have demonstrated is that there *is* room – that you can decrease the size of things in a practical way. I now want to show that there is *plenty* of room. I will not now discuss how we are going to do it, but only what is possible in principle – in other words, what is possible according to the laws of physics. I am not inventing anti-gravity, which is possible someday only if the laws are not what we think. I am telling you what could be done if the laws *are* what we think; we are not doing it simply because we haven't yet gotten around to it.

Information on a small scale

Suppose that, instead of trying to reproduce the pictures and all the information directly in its present form, we write only the information content in a code of dots and dashes, or something like that, to represent the various letters. Each letter represents six or seven 'bits' of information; that is, you need only about six or seven dots or dashes for each letter. Now, instead of writing everything, as I did before, on the *surface* of the head of a pin, I am going to use the interior of the material as well.

Let us represent a dot by a small spot of one metal, the next dash by an adjacent spot of another metal, and so on. Suppose, to be conservative, that a bit of information is going to require a little cube of atoms 5 times 5 times 5 – that is 125 atoms. Perhaps we need a hundred and some odd atoms to make sure that the information is not lost through diffusion, or through some other process.

I have estimated how many letters there are in the Encyclopaedia, and I have assumed that each of my 24 million books is as big as an Encyclopaedia volume, and have calculated, then, how many bits of information there are (10^{15}). For each bit I allow 100 atoms. And it turns out that all of the information that man has carefully accumulated in all the books in the world can be written in this form in a cube of material one two-hundredth of an inch wide – which is the barest piece of dust that can be made out by the human eye. So there is *plenty* of room at the bottom! Don't tell me about microfilm!

This fact – that enormous amounts of information can be carried in an exceedingly small space – is, of course, well known to the biologists, and resolves the mystery which existed before we understood all this clearly, of how it could be that, in the tiniest cell, all of the information for the organization of a complex creature such as ourselves can be stored. All this information – whether we have brown eyes, or whether we think at all, or that in the embryo the jawbone should first develop with a little hole in the side so that later a nerve can grow through it – all this information is contained in a very tiny fraction of the cell in the form of long-chain DNA molecules in which approximately 50 atoms are used for one bit of information about the cell.

Better electron microscopes

If I have written in a code, with 5 times 5 times 5 atoms to a bit, the question is: How could I read it today? The electron microscope is not quite good enough, with the greatest care and effort, it can only resolve about 10 angstroms. I would like to try and impress upon you while I am talking about all of these things on a small scale, the importance of improving the electron microscope by a hundred times. It is not impossible; it is not against the laws of diffraction of the electron. The wave length of the electron in such a microscope is only 1/20 of an angstrom. So it should be possible to see the individual atoms. What good would it be to see individual atoms distinctly?

We have friends in other fields – in biology, for instance. We physicists often look at them and say, 'You know the reason you fellows are making so little progress?' (Actually I don't know any field where they are making more rapid progress than

they are in biology today.) 'You should use more mathematics, like we do.' They could answer us – but they're polite, so I'll answer for them: 'What *you* should do in order for *us* to make more rapid progress is to make the electron microscope 100 times better.'

What are the most central and fundamental problems of biology today? They are questions like: What is the sequence of bases in the DNA? What happens when you have a mutation? How is the base order in the DNA connected to the order of amino acids in the protein? What is the structure of the RNA; is it single-chain or double-chain, and how is it related in its order of bases to the DNA? What is the organization of the microsomes? How are proteins synthesized? Where does the RNA go? How does it sit? Where do the proteins sit? Where do the amino acids go in? In photosynthesis, where is the chlorophyll; how is it arranged; where are the carotenoids involved in this thing? What is the system of the conversion of light into chemical energy?

It is very easy to answer many of these fundamental biological questions; you just *look at the thing!* You will see the order of bases in the chain; you will see the structure of the microsome. Unfortunately, the present microscope sees at a scale which is just a bit too crude. Make the microscope one hundred times more powerful, and many problems of biology would be made very much easier. I exaggerate, of course, but the biologists would surely be very thankful to you – and they would prefer that to the criticism that they should use more mathematics.

The theory of chemical processes today is based on theoretical physics. In this sense, physics supplies the foundation of chemistry. But chemistry also has analysis. If you have a strange substance and you want to know what it is, you go through a long and complicated process of chemical analysis. You can analyze almost anything today, so I am a little late with my idea. But if the physicists wanted to, they could also dig under the chemists in the problem of chemical analysis. It would be very easy to make an analysis of any complicated chemical substance; all one would have to do would be to look at it and see where the atoms are. The only trouble is that the electron microscope is one hundred times too poor. (Later, I would like to ask the question: Can the physicists do something about the third problem of chemistry – namely, synthesis? Is there a *physical* way to synthesize any chemical substance?

The reason the electron microscope is so poor is that the f-value of the lenses is only 1 part to 1,000; you don't have a big enough numerical aperture. And I know that there are theorems which prove that it is impossible, with axially symmetrical stationary field lenses, to produce an f-value any bigger than so and so; and therefore the resolving power at the present time is at its theoretical maximum. But in every theorem there are assumptions. Why must the field be symmetrical? I put this out as a challenge: Is there no way to make the electron microscope more powerful?

The marvelous biological system

The biological example of writing information on a small scale has inspired me to think of something that should be possible. Biology is not simply writing information; it is *doing something* about it. A biological system can be exceedingly small.

Many of the cells are very tiny, but they are very active; they manufacture various substances; they walk around; they wiggle; and they do all kinds of marvelous things – all on a very small scale. Also, they store information. Consider the possibility that we too can make a thing very small which does what we want – that we can manufacture an object that maneuvers at that level!

There may even be an economic point to this business of making things very small. Let me remind you of some of the problems of computing machines. In computers we have to store an enormous amount of information. The kind of writing that I was mentioning before, in which I had everything down as a distribution of metal, is permanent. Much more interesting to a computer is a way of writing, erasing, and writing something else. (This is usually because we don't want to waste the material on which we have just written. Yet if we could write it in a very small space, it wouldn't make any difference; it could just be thrown away after it was read. It doesn't cost very much for the material).

Miniaturizing the computer

I don't know how to do this on a small scale in a practical way, but I do know that computing machines are very large; they fill rooms. Why can't we make them very small, make them of little wires, little elements – and by little, I mean *little*. For instance, the wires should be 10 or 100 atoms in diameter, and the circuits should be a few thousand angstroms across. Everybody who has analyzed the logical theory of computers has come to the conclusion that the possibilities of computers are very interesting – if they could be made to be more complicated by several orders of magnitude. If they had millions of times as many elements, they could make judgments. They would have time to calculate what is the best way to make the calculation that they are about to make. They could select the method of analysis which, from their experience, is better than the one that we would give to them. And in many other ways, they would have new qualitative features.

If I look at your face I immediately recognize that I have seen it before. (Actually, my friends will say I have chosen an unfortunate example here for the subject of this illustration. At least I recognize that it is a *man* and not an *apple*.) Yet there is no machine which, with that speed, can take a picture of a face and say even that it is a man; and much less that it is the same man that you showed it before – unless it is exactly the same picture. If the face is changed; if I am closer to the face; if I am further from the face; if the light changes – I recognize it anyway. Now, this little computer I carry in my head is easily able to do that. The computers that we build are not able to do that. The number of elements in this bone box of mine are enormously greater than the number of elements in our 'wonderful' computers. But our mechanical computers are too big; the elements in this box are microscopic. I want to make some that are *sub*microscopic.

If we wanted to make a computer that had all these marvelous extra qualitative abilities, we would have to make it, perhaps, the size of the Pentagon. This has several disadvantages. First, it requires too much material; there may not be enough germanium in the world for all the transistors which would have to be put into this

enormous thing. There is also the problem of heat generation and power consumption; TVA would be needed to run the computer. But an even more practical difficulty is that the computer would be limited to a certain speed. Because of its large size, there is finite time required to get the information from one place to another. The information cannot go any faster than the speed of light – so, ultimately, when our computers get faster and faster and more and more elaborate, we will have to make them smaller and smaller.

But there is plenty of room to make them smaller. There is nothing that I can see in the physical laws that says the computer elements cannot be made enormously smaller than they are now. In fact, there may be certain advantages.

Miniaturization by evaporation

How can we make such a device? What kind of manufacturing processes would we use? One possibility we might consider, since we have talked about writing by putting atoms down in a certain arrangement, would be to evaporate the material, then evaporate the insulator next to it. Then, for the next layer, evaporate another position of a wire, another insulator, and so on. So, you simply evaporate until you have a block of stuff which has the elements – coils and condensers, transistors and so on – of exceedingly fine dimensions.

But I would like to discuss, just for amusement, that there are other possibilities. Why can't we manufacture these small computers somewhat like we manufacture the big ones? Why can't we drill holes, cut things, solder things, stamp things out, mold different shapes all at an infinitesimal level? What are the limitations as to how small a thing has to be before you can no longer mold it? How many times when you are working on something frustratingly tiny like your wife's wrist watch, have you said to yourself, 'If I could only train an ant to do this!' What I would like to suggest is the possibility of training an ant to train a mite to do this. What are the possibilities of small but movable machines? They may or may not be useful, but they surely would be fun to make.

Consider any machine – for example, an automobile – and ask about the problems of making an infinitesimal machine like it. Suppose, in the particular design of the automobile, we need a certain precision of the parts; we need an accuracy, let's suppose, of 4/10,000 of an inch. If things are more inaccurate than that in the shape of the cylinder and so on, it isn't going to work very well. If I make the thing too small, I have to worry about the size of the atoms; I can't make a circle of 'balls' so to speak, if the circle is too small. So, if I make the error, corresponding to 4/10,000 of an inch, correspond to an error of 10 atoms, it turns out that I can reduce the dimensions of an automobile 4,000 times, approximately – so that it is 1 mm. across. Obviously, if you redesign the car so that it would work with a much larger tolerance, which is not at all impossible, then you could make a much smaller device.

It is interesting to consider what the problems are in such small machines. Firstly, with parts stressed to the same degree, the forces go as the area you are reducing, so that things like weight and inertia are of relatively no importance. The strength of material, in other words, is very much greater in proportion. The stresses and

expansion of the flywheel from centrifugal force, for example, would be the same proportion only if the rotational speed is increased in the same proportion as we decrease the size. On the other hand, the metals that we use have a grain structure, and this would be very annoying at small scale because the material is not homogeneous. Plastics and glass and things of this amorphous nature are very much more homogeneous, and so we would have to make our machines out of such materials.

There are problems associated with the electrical part of the system – with the copper wires and the magnetic parts. The magnetic properties on a very small scale are not the same as on a large scale; there is the 'domain' problem involved. A big magnet made of millions of domains can only be made on a small scale with one domain. The electrical equipment won't simply be scaled down; it has to be redesigned. But I can see no reason why it can't be redesigned to work again.

Problems of lubrication

Lubrication involves some interesting points. The effective viscosity of oil would be higher and higher in proportion as we went down (and if we increase the speed as much as we can). If we don't increase the speed so much, and change from oil to kerosene or some other fluid, the problem is not so bad. But actually we may not have to lubricate at all! We have a lot of extra force. Let the bearings run dry; they won't run hot because the heat escapes away from such a small device very, very rapidly. This rapid heat loss would prevent the gasoline from exploding, so an internal combustion engine is impossible. Other chemical reactions, liberating energy when cold, can be used. Probably an external supply of electrical power would be most convenient for such small machines.

What would be the utility of such machines? Who knows? Of course, a small automobile would only be useful for the mites to drive around in, and I suppose our Christian interests don't go that far. However, we did note the possibility of the manufacture of small elements for computers in completely automatic factories, containing lathes and other machine tools at the very small level. The small lathe would not have to be exactly like our big lathe. I leave to your imagination the improvement of the design to take full advantage of the properties of things on a small scale, and in such a way that the fully automatic aspect would be easiest to manage.

A friend of mine (Albert R. Hibbs) suggests a very interesting possibility for relatively small machines. He says that, although it is a very wild idea, it would be interesting in surgery if you could swallow the surgeon. You put the mechanical surgeon inside the blood vessel and it goes into the heart and 'looks' around. (Of course the information has to be fed out.) It finds out which valve is the faulty one and takes a little knife and slices it out. Other small machines might be permanently incorporated in the body to assist some inadequately-functioning organ.

Now comes the interesting question: How do we make such a tiny mechanism? I leave that to you. However, let me suggest one weird possibility. You know, in the atomic energy plants they have materials and machines that they can't handle directly because they have become radioactive. To unscrew nuts and put on bolts and so on, they have a set of master and slave hands, so that by operating a set of levers here,

you control the 'hands' there, and can turn them this way and that so you can handle things quite nicely.

Most of these devices are actually made rather simply, in that there is a particular cable, like a marionette string, that goes directly from the controls to the 'hands.' But, of course, things also have been made using servo motors, so that the connection between the one thing and the other is electrical rather than mechanical. When you turn the levers, they turn a servo motor, and it changes the electrical currents in the wires, which repositions a motor at the other end.

Now, I want to build much the same device – a master-slave system which operates electrically. But I want the slaves to be made especially carefully by modern large-scale machinists so that they are one-fourth the scale of the 'hands' that you ordinarily maneuver. So you have a scheme by which you can do things at one-quarter scale anyway – the little servo motors with little hands play with little nuts and bolts; they drill little holes; they are four times smaller. Aha! So I manufacture a quarter-size lathe; I manufacture quarter-size tools; and I make, at the one-quarter scale, still another set of hands again relatively one-quarter size! This is one-sixteenth size, from my point of view. And after I finish doing this I wire directly from my large-scale system, through transformers perhaps, to the one-sixteenth-size servo motors. Thus I can now manipulate the one-sixteenth size hands.

Well, you get the principle from there on. It is rather a difficult program, but it is a possibility. You might say that one can go much farther in one step than from one to four. Of course, this has all to be designed very carefully and it is not necessary simply to make it like hands. If you thought of it very carefully, you could probably arrive at a much better system for doing such things.

If you work through a pantograph, even today, you can get much more than a factor of four in even one step. But you can't work directly through a pantograph which makes a smaller pantograph which then makes a smaller pantograph – because of the looseness of the holes and the irregularities of construction. The end of the pantograph wiggles with a relatively greater irregularity than the irregularity with which you move your hands. In going down this scale, I would find the end of the pantograph on the end of the pantograph on the end of the pantograph shaking so badly that it wasn't doing anything sensible at all.

At each stage, it is necessary to improve the precision of the apparatus. If, for instance, having made a small lathe with a pantograph, we find its lead screw irregular – more irregular than the large-scale one – we could lap the lead screw against breakable nuts that you can reverse in the usual way back and forth until this lead screw is, at its scale, as accurate as our original lead screws, at our scale.

We can make flats by rubbing unflat surfaces in triplicates together – in three pairs – and the flats then become flatter than the thing you started with. Thus, it is not impossible to improve precision on a small scale by the correct operations. So, when we build this stuff, it is necessary at each step to improve the accuracy of the equipment by working for awhile down there, making accurate lead screws, Johansen blocks, and all the other materials which we use in accurate machine work at the higher level. We have to stop at each level and manufacture all the stuff to go

to the next level – a very long and very difficult program. Perhaps you can figure a better way than that to get down to small scale more rapidly.

Yet, after all this, you have just got one little baby lathe four thousand times smaller than usual. But we were thinking of making an enormous computer, which we were going to build by drilling holes on this lathe to make little washers for the computer. How many washers can you manufacture on this one lathe?

A hundred tiny hands

When I make my first set of slave 'hands' at one-fourth scale, I am going to make ten sets. I make ten sets of 'hands,' and I wire them to my original levers so they each do exactly the same thing at the same time in parallel. Now, when I am making my new devices one-quarter again as small, I let each one manufacture ten copies, so that I would have a hundred 'hands' at the 1/16th size.

Where am I going to put the million lathes that I am going to have? Why, there is nothing to it; the volume is much less than that of even one full-scale lathe. For instance, if I made a billion little lathes, each 1/4,000 of the scale of a regular lathe, there are plenty of materials and space available because in the billion little ones there is less than 2 percent of the materials in one big lathe. It doesn't cost anything for materials, you see. So I want to build a billion tiny factories, models of each other, which are manufacturing simultaneously, drilling holes, stamping parts, and so on.

As we go down in size, there are a number of interesting problems that arise. All things do not simply scale down in proportion. There is the problem that materials stick together by the molecular (Van der Waals) attractions. It would be like this: after you have made a part and you unscrew the nut from a bolt, it isn't going to fall down because the gravity isn't appreciable; it would even be hard to get it off the bolt. It would be like those old movies of a man with his hands full of molasses, trying to get rid of a glass of water. There will be several problems of this nature that we will have to be ready to design for.

Rearranging the atoms

But I am not afraid to consider the final question as to whether, ultimately – in the great future – we can arrange the atoms the way we want; the very *atoms*, all the way down! What would happen if we could arrange the atoms one by one the way we want them (within reason, of course; you can't put them so that they are chemically unstable, for example).

Up to now, we have been content to dig in the ground to find minerals. We heat them and we do things on a large scale with them, and we hope to get a pure substance with just so much impurity, and so on. But we must always accept some atomic arrangement that nature gives us. We haven't got anything, say, with a 'checkerboard' arrangement, with the impurity atoms exactly arranged 1,000 angstroms apart, or in some other particular pattern.

What could we do with layered structures with just the right layers? What would the properties of materials be if we could really arrange the atoms the way we want

them? They would be very interesting to investigate theoretically. I can't see exactly what would happen, but I can hardly doubt that when we have some *control* of the arrangement of things on a small scale we will get an enormously greater range of possible properties that substances can have, and of different things that we can do.

Consider, for example, a piece of material in which we make little coils and condensers (or their solid state analogs) 1,000 or 10,000 angstroms in a circuit, one right next to the other, over a large area, with little antennas sticking out at the other end – a whole series of circuits. Is it possible, for example, to emit light from a whole set of antennas, like we emit radio waves from an organized set of antennas to beam the radio programs to Europe? The same thing would be to *beam* the light out in a definite direction with very high intensity. (Perhaps such a beam is not very useful technically or economically.)

I have thought about some of the problems of building electric circuits on a small scale, and the problem of resistance is serious. If you build a corresponding circuit on a small scale, its natural frequency goes up, since the wave length goes down as the scale; but the skin depth only decreases with the square root of the scale ratio, and so resistive problems are of increasing difficulty. Possibly we can beat resistance through the use of superconductivity if the frequency is not too high, or by other tricks.

Atoms in a small world

When we get to the very, very small world – say circuits of seven atoms – we have a lot of new things that would happen that represent completely new opportunities for design. Atoms on a small scale behave like *nothing* on a large scale, for they satisfy the laws of quantum mechanics. So, as we go down and fiddle around with the atoms down there, we are working with different laws, and we can expect to do different things. We can manufacture in different ways. We can use, not just circuits, but some system involving the quantized energy levels, or the interactions of quantized spins, etc.

Another thing we will notice is that, if we go down far enough, all of our devices can be mass produced so that they are absolutely perfect copies of one another. We cannot build two large machines so that the dimensions are exactly the same. But if your machine is only 100 atoms high, you only have to get it correct to one-half of one percent to make sure the other machine is exactly the same size – namely, 100 atoms high!

At the atomic level, we have new kinds of forces and new kinds of possibilities, new kinds of effects. The problems of manufacture and reproduction of materials will be quite different. I am, as I said, inspired by the biological phenomena in which chemical forces are used in repetitious fashion to produce all kinds of weird effects (one of which is the author).

The principles of physics, as far as I can see, do not speak against the possibility of maneuvering things atom by atom. It is not an attempt to violate any laws; it is something, in principle, that can be done; but in practice, it has not been done because we are too big.

Ultimately, we can do chemical synthesis. A chemist comes to us and says, 'Look, I want a molecule that has the atoms arranged thus and so; make me that molecule.' The chemist does a mysterious thing when he wants to make a molecule. He sees that it has got that ring, so he mixes this and that, and he shakes it, and he fiddles around. And, at the end of a difficult process, he usually does succeed in synthesizing what he wants. By the time I get my devices working, so that we can do it by physics, he will have figured out how to synthesize absolutely anything, so that this will really be useless.

But it is interesting that it would be, in principle, possible (I think) for a physicist to synthesize any chemical substance that the chemist writes down. Give the orders and the physicist synthesizes it. How? Put the atoms down where the chemist says, and so you make the substance. The problems of chemistry and biology can be greatly helped if our ability to see what we are doing, and to do things on an atomic level, is ultimately developed – a development which I think cannot be avoided.

Now, you might say, 'Who should do this and why should they do it?' Well, I pointed out a few of the economic applications, but I know that the reason that you would do it might be just for fun. But have some fun! Let's have a competition between laboratories. Let one laboratory make a tiny motor which it sends to another lab which sends it back with a thing that fits inside the shaft of the first motor.

High school competition

Just for the fun of it, and in order to get kids interested in this field, I would propose that someone who has some contact with the high schools think of making some kind of high school competition. After all, we haven't even started in this field, and even the kids can write smaller than has ever been written before. They could have competition in high schools. The Los Angeles high school could send a pin to the Venice high school on which it says, 'How's this?' They get the pin back, and in the dot of the 'i' it says, 'Not so hot.'

Perhaps this doesn't excite you to do it, and only economics will do so. Then I want to do something; but I can't do it at the present moment, because I haven't prepared the ground. It is my intention to offer a prize of $1,000 to the first guy who can take the information on the page of a book and put it on an area 1/25,000 smaller in linear scale in such manner that it can be read by an electron microscope.

And I want to offer another prize – if I can figure out how to phrase it so that I don't get into a mess of arguments about definitions – of another $1,000 to the first guy who makes an operating electric motor – a rotating electric motor which can be controlled from the outside and, not counting the lead-in wires, is only 1/64 inch cube.

I do not expect that such prizes will have to wait very long for claimants.

APPENDIX C

MATERIALS-RELATED LABORATORY EXPERIMENTS

This section describes the experimental details for six modules that are appropriate for undergraduate curricula. These experiments were reproduced with permission from the Journal of Chemical Education, published by the Division of Chemical Education of the American Chemical Society (http://jchemed.chem.wisc.edu/index.html). Herein, we provide a few representative modules; the breadth of experiments will be expanded in future editions of this textbook, to include other materials classes such as ceramics, metals, polymers, semiconductors, superconductors, and magnetic materials.

C.1. CHEMICAL VAPOR DEPOSITION OF CARBON NANOTUBES

(Fahlman, B. D. *J. Chem. Ed.*, **2002**, *79*, 203)

C.1.1. Background Information

There are two primary methods used to produce carbon nanotubes on the laboratory scale: the carbon-arc process and transition metal-catalyzed decomposition of organic precursors. The latter method offers a relatively facile setup and scale-up for large-scale nanotube synthesis. Hence, this method has now become the method of choice for nanotube growth. In 1998, only three reported methods for nanotube growth involved CVD; however, since this time, the number of reported instances using this technology has increased by at least an order of magnitude. This experiment is designed to introduce students to both CVD technology and nanotube characterization techniques by focusing on the variation of nanotubes that may be deposited as a function of the catalyst composition.

Similar catalysts prepared from alumina nanoparticles and ferric nitrate were reported to show the speciation of iron as Fe, FeO, and α-Fe_2O_3 prior to CVD and as α-Fe_2O_3 and FeO after nanotube growth. However, in this reported synthesis, hydrogen gas was used; our method did not use a reducing atmosphere and the iron in these catalysts has been described elsewhere as being present solely as Fe_2O_3 after heating to 1,000°C for 10 min in an argon atmosphere.

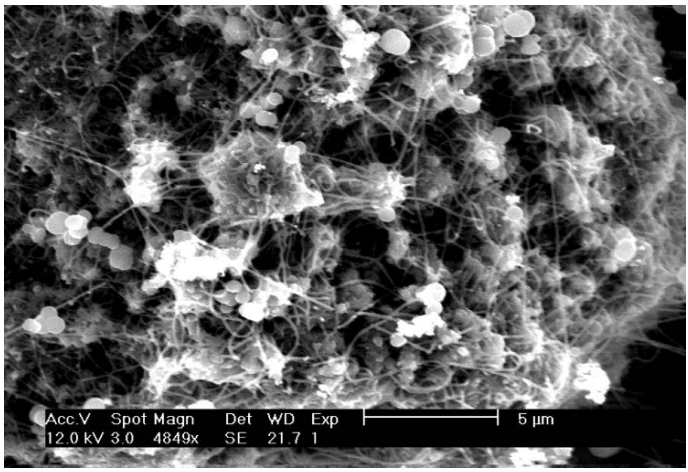


Figure C.1. FESEM image of the raw nanotube felt obtained from catalyst particles containing 7.5 at% Fe.

The CVD method used in this experiment to grow nanotubes was accomplished using the pyrolysis of methane, catalyzed by iron-doped alumina catalyst (Eq. 1).

$$(1) \qquad CH_{4(g)} \xrightarrow{Fe_2O_3/Al_2O_3,1{,}000^\circ C} C_{(s)} + H_{2(g)}$$

The hydrogen gas produced in the decomposition is thought to eliminate the surface-adsorbed carbon-containing species, $CH_{x(ads)}(x = 0\text{–}3)$ via Eq. 2. This would explain the relative lack of amorphous carbon and graphitic deposits found in our samples (Figure C.1), a problem that has plagued alternate CVD experiments.

$$(2) \qquad \frac{(4-x)}{2} H_{2(g)} + CH_{x(ads)} \leftrightarrow CH_{4(g)}$$

The deposited nanotubes are found within bundles on the surface of the catalyst particles. Methane is the most difficult hydrocarbon to thermally decompose; hence, precursor decomposition likely takes place through a traditional CVD mechanism of adsorption and surface decomposition/migration rather than preliminary gas-phase pyrolysis. A wealth of information has been reported on the growth of graphitic carbon over metal substrates. Active metals such as iron, nickel, and cobalt are recognized to best catalyze the decomposition of volatile carbon compounds by forming metastable carbide intermediate species. Most importantly, for nanotube growth, these metals allow facile and rapid diffusion of carbon through and over their surfaces. This suggests that if decomposition of the gas-phase organic molecules would have occurred away from the metal surface, only amorphous carbon deposits would have been formed.

In the first institution of this laboratory module, an assessment of the overall deposition environment of the nanotubes in association with catalyst particles was

to be made. Therefore, no purification methods were used to separate nanotubes from catalyst particles. Although methods such as sonication in mineral acids for 15 min have been used to completely remove metal and support particles for MgO-supported nanotubes, this method is not so easily performed for alumina or silica substrates. For these latter supports, more vigorous purification techniques such as prolonged digestion in strong acids or gas-phase methods would need to be utilized to completely separate the nanotubes from the raw felt.

The variance of nanotube diameters with catalyst concentrations is in accord with models for carbon nanotube growth using CVD. It is now well established that the diameter of individual catalyst particles will dictate the diameter of the grown nanotubes. Particles with diameters in the range 1–2 nm catalyze SWNT growth whereas larger particles on the order of 10–50 nm produce multiwall nanotubes (MWNTs). The increase in nanotube diameters with increasing metal concentrations in the catalyst has been attributed to a greater opportunity for the reduction of M^{n+} ions, occurring within a reducing atmosphere, to form relatively large metal particles. However, in our system, such a reducing atmosphere was not used and the increase of nanotube diameters was most likely due to larger supported Fe_2O_3 particles being formed through agglomeration from more concentrated solutions used in the catalyst synthesis. Nevertheless, the physical characteristics of the catalyst particles do not solely govern the nanotube diameter; the partial pressure of the hydrocarbon precursor(s) also plays a role. This is verified from reports by Kong *et al.* who were able to obtain predominantly SWNTs using Fe_2O_3/alumina catalysts prepared similarly as our 2.4 at% Fe catalysts but using a much greater methane flow (6,150 L min^{-1}).

C.1.2. Procedure

CAS Registry Numbers for Chemicals:

Iron(III) nitrate nonahydrate: 7782-61-8
Aluminum oxide nanoparticles: 1344-28-1
Methanol: 67-56-1
Methane: 74-82-8

Synthesis of the supported catalyst

Materials:

Safety glasses
Electronic balance and weighing paper
Fumed alumina nanoparticles (Degussa)
$Fe(NO_3)_3 \cdot 9H_2O$
100-mL round-bottom flasks
Cork flask stands
Methanol

Rotary evaporator/water bath with appropriate connections

1. Transfer 1.0 g of fumed alumina nanoparticles to a 100-mL round-bottom flask.
2. Calculate the amount of $Fe(NO_3)_3 \cdot 9H_2O$ necessary to prepare 0.020, 0.050, 0.075, and 0.10 M solutions in methanol (use 30.0 mL as delivered from a volumetric pipette). Accurately weigh out one of these amounts into a clean, dry beaker, add the methanol, and ensure the solid is completely dissolved. Add this solution to the alumina weighed out in step 1. *Note*: Each person in the group will prepare a catalyst from solutions containing a different iron concentration (*i.e.*, unless there is a group of more than four students).
3. Place a magnetic stir bar in the bottom of the round-bottom flask and allow the solution to stir for 90 min at room temperature.
4. Remove the methanol from the solution using a rotary evaporator and a water temperature of *ca.* 50°C. This should only take on the order of 2–3 min.
5. Place the product in an oven at 130°C overnight. The solid should now be a brownish-orange color with little/no white particles present in the mixture. Wearing safety goggles and gloves, grind the product into a fine powder using a mortar and pestle within a fume hood.
6. Place the finely ground product in a 20-mL scintillation vial for the CVD growth experiment you will carry out during your next lab section.

CVD growth

Materials:

Ceramic boats
Tongs
Heat-proof gloves
Tube furnace
Quartz tube
Methane and argon cylinders with regulators and tubing
Rotometer flow controller

Oil bubbler with appropriate connections (assembled in a fume hood away from ignition sources)

1. Transfer the synthesized catalyst powder to a ceramic boat and place inside the quartz tube within the tube furnace. Since the final products will be black powders, regardless of the iron content of the catalyst, ensure you record the position of your catalyst in order to differentiate it from your colleagues' samples. Securely fasten the valve joint to the quartz tube with rubber bands to prevent gas leakage during the reaction.
2. Turn on the argon valve and maintain a constant flow (at 50 mL min^{-1}) by monitoring the rotometer beads. Set the furnace to 1,000°C in all three zones. After 1,000°C is reached, allow argon to flow through the chamber for 10 additional minutes. Turn on the nitrogen valve and ensure that a stream of nitrogen gas is flowing onto the point where the oil bubbler tubing is fastened to the reaction

tube. Failure to do this will cause the tubing to melt due to the high temperature of the evolved gases.

3. Open the methane valve to the same flow rate as argon, close the argon valve, and allow methane to flow for 10 min. During this time, the tubing leading to the bubbler, as well as the mineral oil inside the bubbler, will become black from carbon deposits. After 10 min, reopen the Ar valve, close the methane valve, and allow argon to flow through the chamber. Immediately program the oven to return to room temperature (no cooling ramp is necessary).

Characterization by field-emission SEM

Many techniques are used to characterize carbon nanotubes. Of these, scanning electron microscopy (SEM) is used most frequently to quickly assess the quantity and quality of the nanotubes present in the sample. The extremely fine electron source of the field-emission system to be used in this section will enable the attainment of much higher resolution images than a conventional SEM. Useful magnifications in excess of 200,000 times are often obtainable, which translates to a resolution of 3–5 nm at an accelerating voltage of 30 kV. The high brightness of this source also allows high-resolution imaging and characterization of beam-sensitive materials (*e.g.*, fragile plastics and integrated circuits) at high magnifications in excess of 100,000 times at very low accelerating voltages (*i.e.*, 0.2–5 kV). Insulating samples may also be examined without a conventional conductive coating of gold or carbon.

You will prepare samples for SEM analysis using two methods:

Method 1. Place a small amount (*ca.* 2 mg) of the synthesized material in a glass vial and add 10 mL of methanol. Place the vial in a sonicator for half an hour. Place a drop of the suspension on carbon tape stuck to a mounting pin. Allow the methanol to evaporate completely and insert into the sample chamber. If FESEM images indicate an agglomeration of nanotubes, try preparing your sample from a chloroform suspension.

Method 2. Simply place a small amount (spatula tip) of the black powder onto carbon tape fastened to an aluminum-mounting pin. Tap down the powder with the flat end of the spatula and shake off the excess prior to mounting in the SEM vacuum chamber.

C.2. SUPERCRITICAL FLUID FACILITATED GROWTH OF COPPER AND ALUMINUM OXIDE NANOPARTICLES

(Williams, G. L.; Vohs, J. K.; Brege, J. J.; Fahlman, B. D. *J. Chem. Ed.*, **2005**, *82*, 771)

The burgeoning field of *nanotechnology* research is focused on objects with architectural dimensions of less than 100 nm. Nanoparticles of metals or metal oxides have a variety of applications for next-generation electronic devices, advanced sensors, inks, antitumor targeting agents, and chemical/biological warfare sensing devices. This experiment is designed to introduce students to many exciting areas of current

industrial/research interest including SCF technology, nanoparticle growth, and nanoscale characterization techniques.

C.2.1. Procedure

CAS Registry Numbers for Chemicals:

Sodium aluminate: 1302-42-7
Copper (II) chloride dihydrate: 10125-13-0
1,4-phenylenediamine: 106-50-3
Ethanol, 200-proof: 64-17-5
Carbon dioxide (bone-dry grade): 124-38-9

Synthesis of aluminum oxide nanoparticles

1. Clean a silicon wafer by rinsing with distilled water, and dry with KimWipes. Place the clean silicon substrate in a tall beaker, and place in the supercritical fluid reactor using long tweezers. With assistance from the instructor, fasten the reactor to the supercritical fluid system.
2. Using solid sodium aluminate powder, prepare a 4% (w/v) aqueous aluminate solution in an 8-mL vial, and shake vigorously until dissolved. Filter this solution into a clean 8-mL vial through a disposable pipette fitted with a cotton plug.
3. In a separate 8-mL vial, place 0.2 g of the surfactant (*ask instructor which one to use*) in 6 mL of heptane (if AOT: sodium bis(2-ethylhexyl)sulfosuccinate) or water (if fluorinated surfactant: ammonium carboxylate perfluoropolyether – $[CF_3O(CF_2CF(CF_3)O)_3CF_2COO]^-[NH_4]^+$). Shake vigorously until the surfactant is completely dissolved. Add two drops of the filtered aluminate solution to the surfactant solution, and shake vigorously until one phase is obtained (*i.e.*, the final solution should not be cloudy).
4. With consultation with the instructor, prime the cosolvent pump and set the appropriate reactor temperature and pressure (see Figure C.2).
5. Once the reactor pressure and temperature has stabilized, inject the microemulsion solution using the cosolvent pump.
6. After the microemulsion solution has been added to the reactor, allow the pressure to remain at its initial setting. After this time, flush the reactor for 5 min with a dynamic flow of carbon dioxide.
7. Vent the system and disconnect the reactor. Remove the silicon wafer using tweezers and place in a vial until characterization studies are performed. Dip half of the coated wafer in heptane (if AOT was used) or water (if fluorinated surfactant was used) and allow to air dry. Store the coated wafer until analysis by placing it in a plastic weighing tray (polished side up), and place another weighing tray on top. Completely tape the two-weighing trays together to prevent dust from forming on the coated wafer.

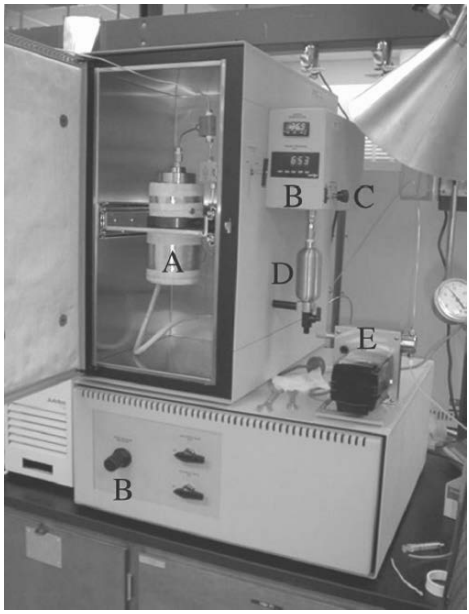


Figure C.2. Photograph of the supercritical fluid system used for nanoparticle synthesis. Shown is the 300-mL high-pressure reactor (A), with pressure/temperature controllers (B). The system is rated for safe operation at temperatures and pressures below 200°C and 10,000 psi, respectively. The vessel may be slowly vented, or exposed to a dynamic CO_2 flow, using a multiturn restrictor valve (C), which provides a sensitive control over system depressurization, allowing for the collection of CO_2-solvated species in the stainless steel collector (D). For deposition using the rapid expansion of the supercritical solution (RESS), nanoparticles were blown onto a TEM grid that was placed under the stopcock below D. Also shown is the cosolvent addition pump (E) used for the synthesis of aluminum oxide nanoparticles, capable of delivering liquids into the chamber against a back-pressure of $\leq$5,000 psi.

Synthesis of copper nanoparticles

1. Prepare a 0.05 M copper solution using $CuCl_2 \cdot 2H_2O$ in absolute ethanol (200 proof), and transfer this to a clean 8-mL vial. Also prepare a 0.04 M aqueous solution of 1,4-phenylenediamine, and transfer to a second clean 8-mL vial. Position the vials within the supercritical fluid reactor, avoiding any contact between the solutions. With consultation with your instructor, allow the system to reach the desired pressure and temperature. Maintain these conditions for 45 min.
2. Vent the system and remove the vials from the reactor. Note color changes, turbidity, and phases (*i.e.*, is there one liquid present, or does there appear to be a layer of two liquids?) of both solutions. Remove both vials and keep tightly sealed until characterization studies are performed. Comment on what you observe in the Results and Discussion section of your formal report.
3. Repeat step 1 using ethanol for both copper and phenylenediamine solutions. This time, when you vent the system after 45 min, place a carbon TEM grid at the base of the collector vessel (the instructor will point this out). This technique, known

as *RESS* (rapid expansion of the supercritical solution), is a unique benefit of performing reactions in supercritical fluids. When the system is vented, the gas/liquid carbon dioxide expands, being rapidly converted from its original high-pressure environment to a low-pressure gas. Place the grids in a vial for characterization studies.

C.3. SYNTHESIS AND CHARACTERIZATION OF LIQUID CRYSTALS

(Van Hecke, G. R.; Karukstis, K. K.; Li, H.; Hendargo, H. C.; Cosand, A. J.; Fox, M. M. *J. Chem. Ed.*, **2005**, *82*, 1349)

This experiment will introduce you to some techniques of chemical synthesis based on acid-base reactions, purification by recrystallization, and brief characterization by melting point. Also on the characterization side, you will investigate the absorption of polarized light by matter. A liquid crystal is a state of matter neither liquid nor crystal but a state in-between. Liquid crystals are often called *mesophases* after the Greek mesos for middle. Normally when a crystal melts it forms an ordinary liquid phase. However, a substance which exhibits liquid crystalline behavior melts at least twice, first into the liquid crystalline or mesophase, and second, into the ordinary liquid.

Molecules in a mesophase exhibit orientation with respect to each other, and, in certain types of liquid crystals, exhibit some extent of ordered position with respect to each other. Thus, mesophases are described by degrees of orientational and positional order. Using ordering as a basis, liquid crystals fall into two types: *nematic* and *smectic*. Nematic mesophases possess only orientation order while smectic mesophases possess orientational and some positional order (Figure C.3). Also shown in Figure C.3 is a special type of nematic phase, the *chiral nematic* or *cholesteric* phase (cholesteric is the older and still more common name but chiral nematic is now the correct name). In this phase, not only do the molecules comprising the phase point more or less in the same direction on a local scale, but also on a larger scale that direction changes following a helix. The quantity called *pitch* is defined as the physical distance required for one complete revolution about the optical axis. Cholesteric phases were the first liquid crystals ever observed, and for many years were thought to be a separate type of liquid crystal. Now it is known that what made cholesteric liquid crystalline materials different is the fact that they are composed of optically active or chiral molecules. The cholesteric phase derives its name from the fact that the first substances observed to exhibit what we today call liquid crystallinity were derivatives of the naturally occurring substance cholesterol. While many commercial liquid crystal devices employ the chlolesteric phase to function, cholesterol and its derivatives are important objects of study in medicine since cholesterol derivatives are components of the deposits that form on the wall of arteries and lead to hardening of the arteries.

A nematic mesophase forms when rod-like molecules orient themselves on average in a given direction. The director of the phase is a vector that points in a direction determined by looking at the average of the directions in which all

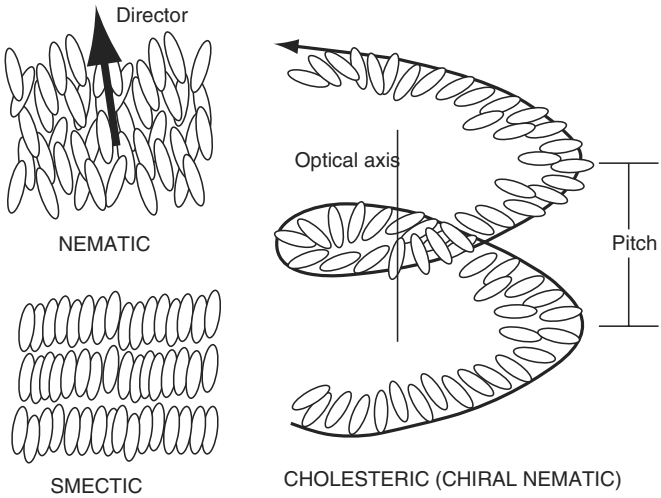


Figure C.3. Schematic of three mesophases: the nematic and cholesteric (which have only orientation order) and the smectic phase (which has orientational and positional order).

the molecules point or are oriented. The director's orientation is a function of temperature, pressure, and applied electric and magnetic fields. The many practical applications of liquid crystals depend critically on modifying the director by external influences. This experiment will be partly concerned with what factors affect the director of the cholesteric phase, which is the direction in which any local collection of molecules points. What happens to the cholesteric director as a function of temperature will be studied by determining the pitch of the phase as a function of temperature.

Substances in their liquid crystalline phase behave like liquids in that they flow and fill any shape container in which they are placed. However, even though they are fluids, they retain certain optical properties characteristic of a solid, especially those optical properties which are influenced by polarized light. For this experiment how the plane of polarization of plane polarized light changes on passage of such light through a mesomorphic medium is particularly crucial. In the cholesteric mesophase, the director, which remember reflects the local organization of a collection of molecules, follows a helical path. The sense of this helix, left or right, determines the angle of rotation of plane polarized light incident on a cholesteric sample. The pitch of the helix also depends on external influences. In fact, the exact pitch is dependent not only on temperature and pressure, but also in the case of mixtures of liquid crystals, also on composition. You may choose to explore the effect of composition during the second week of this experiment.

During the first week of this experiment, you will synthesize cholesteryl nonanoate, which is an ester derivative of cholesterol, and perform some simple characterizations to test your successful synthesis. Cholesteryl nonanoate (ChNon) is known to exhibit a chiral nematic, *i.e.*, a cholesteric, liquid crystalline phase.

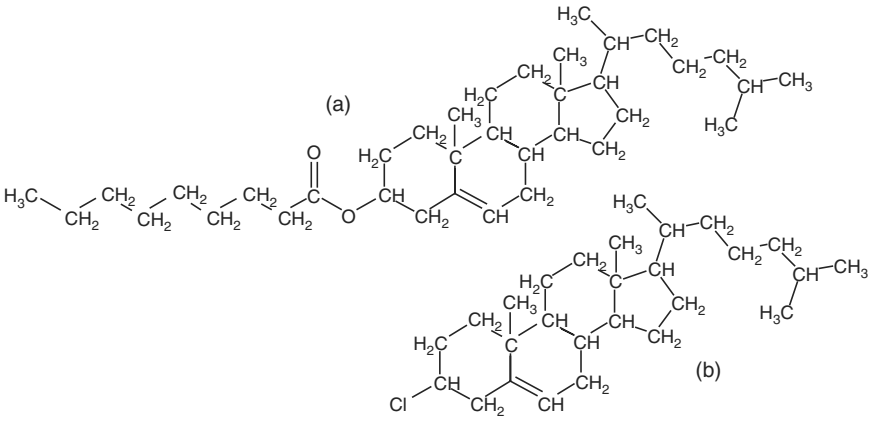


Figure C.4. Molecular structures of (a) cholesterol nonanoate (ChNon) and (b) cholesteryl chloride (ChCl). Pure cholesterol has the structure of cholesteryl chloride with the –Cl replaced by a –OH group.

While you will synthesize and characterize cholesterol nonanoate, you will actually study mixtures of two liquid crystalline materials each of which is a derivative of cholesterol. The other derivative is cholesteryl chloride (ChCl) whose structure along with that of cholesterol nonanoate is shown in Figure C.4. The reason to study mixtures here is for convenience of handling. Though both pure materials exhibit liquid crystalline phases (ChNon: at 74°C, solid:smectic; 80°C smectic:cholesteric; 93°C cholesteric:isotropic, and ChCl: about 70°C, solid:cholesteric), appropriate mixtures are liquid crystalline at much lower and more manageable temperatures, for example, room temperature.

The driving force for the chemical reactions used in the synthetic scheme outlined below is most readily understood in terms of Lewis acid-base chemistry. To begin, cholesterol is dissolved in pyridine. According to the Lewis definition of acids and bases, in which acids and bases are described in terms of their propensity to accept or donate electron pairs, cholesterol is a weak acid (an electron pair acceptor) and pyridine is a strong base (an electron pair donor). These two chemicals, then, undergo an acid-base reaction, transferring a proton (an electron acceptor) to yield the pyridinium ion and a negatively charged cholesterol species (which results as the proton accepts the electrons from the pyridine leaving behind a negative charge on the cholesterol).

The next step involves the addition of nonanoyl chloride and is also understood using the Lewis definition. The carbonyl (C═O) group on nonanoyl chloride is very polar, yielding an electron-deficient carbon, which is then a Lewis acid. The negative cholesterol ion is an electron donor, or Lewis base, and reacts with the positive carbon center to form a new bond and release a chloride ion. This chloride ion is then neutralized by the pyridinium ion in another Lewis acid-base reaction to form pyridine hydrochloride. Pyridine hydrochloride is soluble in water, while the newly synthesized cholesteryl nonanoate is not. Addition of the reaction mixture to aqueous sulfuric acid causes only the desired product, cholesteryl nonanoate,

to precipitate. The resulting solid, collected by filtration, contains impurities which you will remove by recrystallization. While the reaction is straightforward, there are certain hazards which must be avoided. Both pyridine and nonanoyl chloride react readily with water in the air, forming the pyridinium ion and nonanoic acid. Even a small amount of decomposition of these essential reactants will cause the intended reaction to fail. For this reason, these reactants should be distilled prior to use, and stored in such a way to minimize their contact with air.

C.3.1. Procedure

CAS Registry Numbers for Chemicals:

Cholesterol: 57-88-5
Dry nonanoyl chloride: 764-85-2
Cholesteryl chloride: 910-31-6
Reagent acetone: 67-64-1
1 M sulfuric acid: 7664-93-9
Dry pyridine: 110-86-1

Filter flask (125 mL)
Buchner funnel
Two dram vial with screw cap
Glass stir rod
Graduated cylinder (50 mL)
Stir hot plate with magnetic stirrer
Rubber septum
Disposable syringes and needles
Microscope slides
Mylar film and tape
Digital voltmeter
Polarizer sheet and $\lambda/4$ wave plate
Melting point apparatus
Beckman DU650 spectrophotometers with temperature controllers and cell holders
Abbe refractometers with temperature controllers
Thermocouple

1. Place about 1.5 g of cholesterol in a 125-mL Erlenmeyer flask. The cholesterol should be recrystallized from acetone prior to acquiring a pure sample. Add 5–8 mL of dry pyridine (obtained by distillation over potassium hydroxide) and a stir bar to flask. Work quickly and carefully when transferring the pyridine to the flask, to minimize water absorption from the ambient atmosphere. Once pyridine is transferred, immediately cap flask with a rubber septum and drying needle.
2. Stir solution in an ice/water bath for about 10 min. The cholesterol should be dissolved. Use a plastic dish for the ice bath container.

3. Ensure that the nonanoyl chloride to be used is dried via distillation *in vacuo*, and stored in a sealed stock bottle for maximum purity. Acid chlorides, such as nonanoyl chloride, readily react with water to form the corresponding carboxylic acid. Withdraw 1.2 mL of nonanoyl chloride from the stock bottle using a fresh, clean/dry plastic syringe fitted with a syringe needle. Pass the syringe needle through the rubber septum on the reaction flask and add the nonanoyl chloride dropwise and slowly. Since the amount of nonanoyl chloride is not the limiting reagent, you do not need to be extremely concerned with the exact amount added provided the amount is adequate. The amount added may be estimated by noting the volume delivered by the syringe and the volume should be about 1.2 mL. A precipitate should form on the addition of the acid chloride.
4. Stir the reaction mixture in an ice bath for 30 min using a magnetic stirrer. Remove solution from ice bath and continue stirring at room temperature for an additional 30 min on the magnetic stirrer.
5. Pour 50 mL of cold 1 M H_2SO_4 into the solution. Stir vigorously by hand using a glass stir rod to break up oil globules. Collect the solid which is your impure product using a typical suction filtration apparatus. Draw air through the sample for several minutes. Visual inspection should suggest when the sample is relatively dry and ready for the next step.
6. To purify your product, recrystallize your collected solid from acetone. To accomplish this, transfer the solid to a 125-mL flask. Add a small amount (20 mL to start) of acetone and a stir bar. Heat to fully dissolve the solid. Add small amounts of acetone, order 1 mL at a time, as needed, to aid in the dissolution. When solid is dissolved, remove from heat. Cool with the aid of an ice bath. Crystals should form. If they do not, try scratching the inside of the flask with a glass rod. Or, boil off some solvent and recool. Remove these crystals by suction filtration just as you did to collect your impure product. Wash these crystals with a cold 5 mL portion of acetone and allow them to dry on the filter paper. Transfer to a preweighed vial. Calculate yield of cholesteryl nonanoate from cholesterol and calculate the conversion of cholesterol to cholesteryl nonanoate.
7. Using appropriate melting point apparatus, characterize the melting behavior of the purified solid. Cholesteryl nonanoate exhibits a chiral nematic, more commonly known as a *cholesteric*, liquid crystalline phase around 85°C and melts to the isotropic liquid around 93°C. The cholesteryl nonanoate first melts to form a smectic phase around 75°C and with further heating transforms to the cholesteric phase around 85°C.
8. In order to calculate the pitch of the liquid crystal, you will need the average refractive index at several temperatures. Use an Abbe refractometer to measure the refractive indices of your liquid crystal

at a number of different temperatures, about every 10° from room temperature to 60–70°C.

9. As mentioned above you will be studying mixtures of cholesteryl nonanoate and cholesteryl chloride rather than the pure materials. Mixtures are used for the convenience of working at temperatures closer to room temperature. To prepare a mixture of a given composition, use the mole fraction composition scale. A mole fraction of a component in a mixture is defined as the number of moles of that component in the mixture divided by the total number of moles of all components in the mixture. To actually prepare a mixture, add the desired amounts of each component to a screw cap vial. Mix the components physically with a spatula to make a reasonably uniform mixture of the solids. Place the vial in a sample oven whose temperature has been set to about 100°C and allow the solids to melt and mix. Mix the sample while hot, again with a spatula. At the oven temperature the sample should appear clear since it should be in the isotropic phase. As the sample cools note what happens.
10. Full details regarding the determination of selective reflection and refractive indices may be found online (http://www.jce.divched.org/Journal) in the Supplementary Documentation for the above article.

C.4. TEMPLATE SYNTHESIS AND MAGNETIC MANIPULATION OF NICKEL NANOWIRES

(Bentley, A. K.; Farhoud, M.; Ellis, A. B.; Lisensky, G. C.; Nickel, A. –M. L.; Crone, W. C. *J. Chem. Ed.*, **2005**, *82*, 765)

C.4.1. Procedure

Whatman Anodisc alumina membranes with polypropylene support rings (25-mm membrane diameter) 0.02-μm pore diameter

Tweezers
Electrical tape
Cu sheet ($12'' \times 12''$ is enough to make approximately 15 substrates)
Nickel wire, 1 mm diameter × 10 m long, 99.5%
AA batteries
Battery holders
Wires with alligator clips
Nickel electroplating solution
Compound optical microscope
Microscope slides and coverslips
Bar magnets
SEM specimen mounts

Note. The alumina membranes have 20-nm diameter pores on the upper face that quickly widen within 5 μm of the surface to 200 nm throughout the membrane. Thus one face has 20-nm pores and the other has 200-nm pores. The membranes are packed with the 20-nm pores facing upward, but the two faces cannot be easily distinguished visually. However, the clear polypropylene support ring is wider on the 20-nm pore side of the membrane, which is the side that will be coated with the conductive layer.

1. The Cu plates and battery holders should be prepared before the lab period and can be reused indefinitely. Cut rectangles of copper sheet approximately $1.25'' \times 2''$ and then cut out corners of the rectangle to form the stem that is used for electrical contact. We used a variety of copper sheets that ranged from $0.03''$ to $0.06''$ in thickness. To assemble the battery holder, first cut a double-ended test lead in half using wire cutters. Then strip the cut ends and solder each lead to the clip on one end of a battery holder.
2. Remove an alumina membrane from the package by holding the polypropylene support ring with tweezers (discard the paper circles separating the membranes from each other). Always use tweezers to hold the membranes by the support ring; the alumina will crack if handled directly. Cracked membranes cannot be used for deposition, as the solution will leak through the membrane and deposit on the copper plate. Dispose of cracked membranes in a glass disposal box. There are two methods of applying a conductive backing to the membranes. If a sputter coater is available (such as those used to coat samples with gold for SEM), then a 250-nm thick layer of silver can be sputtered onto each membrane before the lab period. An alternative method that can be used during the lab period is to have students paint the upward-facing side of the membrane (the side with 20-nm pores; see above) with GaIn eutectic using a cotton-tipped stick. Students will not need to dip the cotton end into the GaIn more than once or twice. The GaIn is applied by gently rubbing the GaIn-coated cotton tip over the membrane surface. Students can check for gaps in the GaIn coating by looking at the noncoated face of the membrane: any areas without GaIn will appear light blue in color. Areas with GaIn will appear white or opaque. We have observed similar success rates with either silver or GaIn as the conductive coating.
3. After the conductive layer is applied, mount the membrane on a copper plate with the conductive, metal-coated side facing the copper by taping the membrane's support ring to the copper using insulating electrical tape, as shown in Figure C.5 (top). Little if any electric tape should come in contact with the alumina membrane, as it is hard to remove. It is essential to completely cover both sides of the copper plate with the electrical tape to prevent any unwanted electrodeposition on the copper. Leave the stem uncovered so electrical contact can be made with the alligator clip, but this section of the copper should not make contact with the electrolyte.
4. Set a AA battery inside the battery holder, making sure to align the positive and negative ends of the battery with the labels on the holder. Clip the negative lead from the battery to the exposed part of the copper and clip the positive lead to

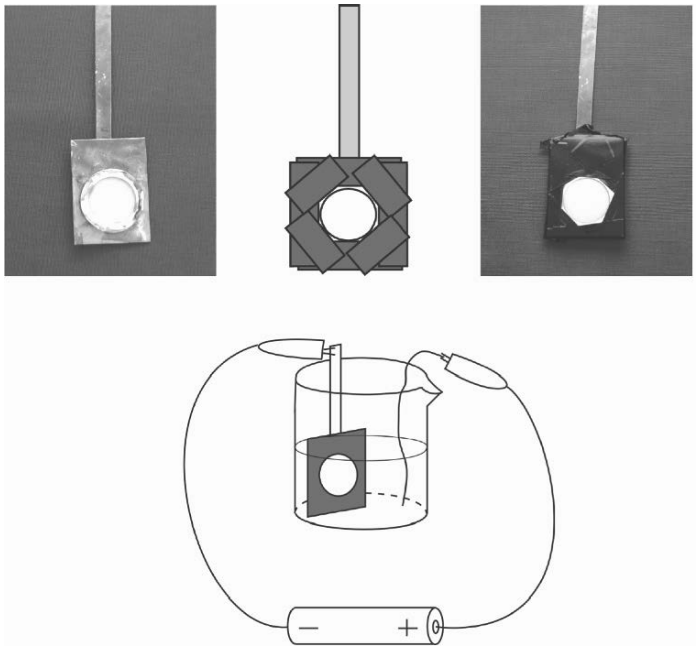


Figure C.5. (top) Placement of the membrane on the copper electrode, fastened with electrical tape. (bottom) The electrochemical cell used to synthesize Ni nanowires.

the nickel wire. Place the two electrodes in a 50-mL beaker with the membrane surface facing the nickel wire (Figure C.5 (bottom)).

5. After the electrodes are arranged in the beaker and connected to the battery, pour in enough Ni-plating solution to just cover the top of the membrane (approximately 40 mL). The electrodeposition of nickel will begin as soon as the solution completes the circuit.
6. As the electrodeposition takes place, the membrane surface changes color from a pale yellow/orange (for silver backing) or yellow/white (for GaIn backing) to black. Allow the cell to sit for 50 min to make nanowires that are $\sim 50\,\mu$m long. To make shorter nanowires, stop the reaction sooner. The minimum time required to produce nanowires that are visible using an optical microscope is 10 min.
7. To halt deposition, remove the battery from the cell and disconnect the wires from the electrodes. Remove the Cu plate from the solution and rinse the membrane surface with deionized water. The nickel-plating solution can be returned to its original container and reused indefinitely. We have used the same 500 mL of plating solution for approximately 50 depositions without any noticeable change in quality.
8. Soak the copper plate and nickel-filled membrane in a beaker of acetone for 5 min to dissolve the electrical tape's adhesive. Carefully detach the membrane

from the copper plate and the electrical tape. If the copper plate is discolored, it can be cleaned with nitric acid before use in another deposition.

9. To free the nanowires from the alumina template, the GaIn or silver cathode must be removed and the alumina membrane dissolved. To remove the cathode layer, mount the membrane with the GaIn or silver side facing outward on a glass microscope slide using one piece of electrical tape on the support ring. Dip a cotton-tipped stick in concentrated nitric acid and rub it over the surface once or twice. The GaIn or silver will dissolve as it is oxidized and the membrane will appear black as the nickel underneath is exposed.
10. Immediately soak the cotton-tipped stick in water to prevent formation of nitrocellulose. Rinse the membrane surface many times with deionized water to remove excess acid. Peel off the electrical tape that is holding the membrane in place and put the membrane in a 50-mL beaker. Cracks in the membrane are not a concern at this point.
11. Add 5 mL of 6 M NaOH and let the membrane soak for 10 min. Swirl the contents of the beaker occasionally. As the base dissolves the alumina, the support ring will float freely in the solution.
12. Remove the support ring from the solution using tweezers, rinse in water, and discard. Black material will collect at the bottom of the beaker. Use tweezers or a Pasteur pipette to break up any clumps of black material until the solution is a cloudy gray color.
13. Set a magnet next to the side of the beaker and wait 1 min while the nickel nanowires collect against the magnet. Use a Pasteur pipette to remove as much of the alkaline solution as possible (without removing any of the black nanowire material) and then add 15-mL deionized water to the beaker. Resuspend the nanowires in the fresh water using a Pasteur pipette and collect the nanowires against a side again using a magnet. Remove the water again and rinse once more.
14. After removing water from the last washing, add 2–3 mL of ethylene glycol. Other solvents can also be used, such as water or ethanol, but we have found that the wires stay suspended longer in ethylene glycol, allowing the students to better see the response of their wires to magnetic fields using the optical microscope. Transfer the suspension to a small vial. Use a Pasteur pipette or shake the vial to disperse the nanowires; the solution will appear cloudy and gray. The nanowire suspension can be stored indefinitely.
15. Transfer a drop of the nanowire suspension to a microscope slide and cover with a coverslip. If clumps remain in the solution, try to avoid transferring them. Even if the solution does not appear to be very cloudy, there are probably still a large number of nanowires suspended, since each membrane has on the order of one billion pores.
16. To examine the nanowires using SEM, place a drop of nanowires suspended in water directly on an aluminum SEM specimen mount and allow the drop to dry. (To obtain a smoother background in the images, attach a piece of silicon to the specimen mount using carbon tape and place the drop of nanowire suspension on the silicon.) No conductive coating needs to be added. We imaged

the nanowires using a LEO FE-SEM with a 3 kV accelerating voltage and 3 mm working distance.

17. To examine the nanowires by X-ray diffraction (XRD) while they are still embedded in the membrane, tape the membrane to a glass microscope slide with the Ni deposition side facing upward. To examine a sample of liberated nanowires, suspend the nanowires in ethanol and place a few drops on a microscope slide. After the solvent evaporates, add a few more drops and repeat this process until there is a visible pile of nanowires on the slide. Our patterns were obtained using a Scintag PADV powder X-ray diffractometer with Cu K-α radiation, a tube voltage of 40 kV and a tube current of 35 mA.

C.5. INTRODUCTION TO PHOTOLITHOGRAPHY

(Berkowski, K. L.; Plunkett, K. N.; Yu, Q.; Moore, J. S. *J. Chem. Ed.*, **2005**)

C.5.1. Procedure

CAS Registry Numbers for Chemicals:

Isobornyl acrylate (IBA): 5888-33-5
Bisphenol-A-glycidyldimethyacrylate (Bis-GMA): 1565-94-2
2,2-dimethoxy-2-phenylacetophenone (DMPA): 24650-42-8
Oil Red O (Solvent Red 27): 1320-06-5
Oil Blue N (Solvent Blue 14): 2646-15-3
Fluorescent Yellow 3G (Solvent Yellow 98): 12671-74-8
Ethanol: 64-17-5

To prevent premature polymerization of the photoresist, store it in an amber vial or a clear vial wrapped with aluminum foil in a chemical refrigerator, if available, and minimize its exposure to light.

Photoresist preparation

1. Use a spatula to mass approximately 2.0 g of Bis-GMA into an amber vial. Bis-GMA is extremely viscous and sticky and you may need an additional spatula to push it to the bottom of the vial.
2. If some of the Bis-GMA sticks to the sides or top of the vial, tightly screw on the vial lid and completely submerge the vial in a beaker full of hot tap water until the Bis-GMA flows to the bottom.
3. Add approximately 3.5 g of IBA to the vial containing Bis-GMA.
4. Add approximately 0.18 g (3% wt) of DMPA to the vial containing IBA and Bis-GMA.
5. Sonicate the mixture for 30 min.
6. Swirl the solution to make sure it is homogeneous. If all components are not dissolved, sonicate the solution for an additional 10 min.
7. Repeat step 6 until the solution is homogeneous.

8. Add the appropriate color dye (1–1.5 mg or a small spatula tip) to the resulting photoresist.
9. Sonicate this solution for 5–10 min.
10. Store the photoresist in an amber vial with a screw-top dropper in a chemical refrigerator, if available. A clear vial wrapped with aluminum foil may also be used.

Photomask preparation

1. Create a 3 × 3 in. black box in Microsoft Word to use as a template.
2. Copy and paste six of these black boxes on a single page to create multiple photomasks.
3. Text-containing photomasks: create a text box, format it for white font and no fill, and center the text box inside the black box to allow at least a 0.5 in. black border surrounding the text.
4. Picture-containing photomasks: manipulate shapes in Microsoft Paint by pasting a white shape on a black background and copy this figure to the center of the black box on your template.
5. Print the photomasks on transparency paper using a black and white printer at a resolution of 600 dpi or higher.

Polymer patterning (Figure C.6)

1. Cut a transparency paper into four equal pieces of 5.5 × 4.25 in. and place one of these flat on the table.
2. Align two thin glass coverslips (22 × 50 mm) on the top and bottom of the transparency.
3. Apply 15–30 drops of the photoresist to the center of the transparency leaving about 2.5 in. between each coverslip and the liquid photoresist.
4. Cover the photoresist with a thick glass slide (75 × 50 × 1 mm) by placing the slide in the center of the transparency and allowing its top and bottom to rest on top of the thin coverslips.
5. Wait until the entire space between the thick glass slide and the transparency is filled with photoresist.
6. Place the photomask on top of the thick glass slide and make sure that photoresist is visible through all parts of the photomask.
7. Place a second thick glass slide on top of the photomask to keep it in place.
8. Polymerize the exposed photoresist by shining UV light (365 nm) from the handheld lamp 0.5 in. above the sample for 20 s (time may be dependent on the UV lamp).
9. Remove the photomask and carefully peel away the bottom transparency from the thick glass slide.
10. Wash away the unpolymerized photoresist into a beaker with a wash bottle containing ethanol.
11. Dry the resulting polymer by gently dabbing it with a paper towel.

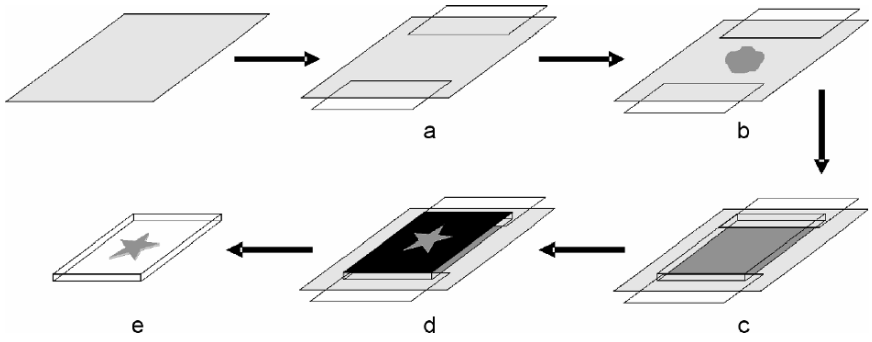


Figure C.6. Overview of the fabrication process. Two coverslips are placed at the top and bottom of a transparency (a) and prepolymer solution is dropped in the center (b). A thick glass slide is positioned on top of the prepolymer solution, allowing it to rest on the coverslips, and causing the prepolymer solution to flow and fill the gap (c). A photomask is aligned on top of the glass slide (d), a second glass slide is placed on top of the photomask to keep it in place, and the exposed photoresist is polymerized using UV light. After rinsing, an insoluble, patterned polymer results (e).

Troubleshooting

1. Photoresist solution is not homogeneous – increase sonication time.
2. Photoresist solution polymerizes in container – store in an amber container or a container wrapped with aluminum foil in a chemical refrigerator, if available.
3. Photoresist does not flow to edges of glass slide or polymer pattern is too thin – increase the initial amount of prepolymer solution on the transparency film.
4. Bubbles form in the polymer product:
   (a) Poke away all bubbles with a pipette tip when photoresist is dropped on the transparency.
   (b) Carefully place the glass slide on top of the photoresist.
5. Incomplete or no polymerization upon exposure to UV light:
   (a) Increase UV exposure time.
   (b) Increase crosslinker concentration (Bis-GMA).
   (c) Increase photoinitiator concentration (DMPA).
   (d) Decrease amount of dye.
6. Brittle or delaminating polymer structures:
   (a) Decrease UV exposure time.
   (b) Decrease crosslinker concentration (Bis-GMA).
   (c) Decrease photoinitiator concentration (DMPA).
7. Polymer forms in areas that were not exposed to UV light:
   (a) Decrease UV exposure time.
   (b) Increase space between structures and letters on photomask.
   (c) Blacken dark areas of photomask with a marker to suppress light transmission.

C.6. SYNTHESIS OF GOLD NANOCLUSTERS

(Adapted from Weare, W. W.; Reed, S. M.; Warner, M. G.; Hutchison, J. E. *J. Am. Chem. Soc.*, **2000**, *122*, 12890)

A number of techniques may be used to synthesize nanoclusters. In this experiment, a gold precursor is transformed into nanoclusters with diameters <2 nm via a two-step reaction. The first step consists of phase-transfer into an organic layer where it reacts with triphenylphosphine. In the second step, the mixture is reduced by sodium borohydride.

C.6.1. Procedure

CAS Registry Numbers for Chemicals:

$HAuCl_4 \cdot 3H_2O$: 16961-25-4
Tetraoctylammonium bromide: 14866-33-2
Triphenylphosphine: 603-35-0
$NaBH_4$: 16940-66-2
Toluene: 108-88-3
Hexanes: 110-54-3
Methanol: 67-56-1
Chloroform: 67-66-3
Pentane: 109-66-0
$NaNO_2$, saturated solution: 7632-00-0

1. Dissolve hydrogen tetrachloroaurate trihydrate (1.00 g, 2.54 mmol) and tetraoctylammonium bromide (1.60 g, 2.93 mmol) in a degassed water/toluene mixture (50 mL/65 mL).
2. When the golden color is transferred into the organic phase, add triphenylphosphine (2.32 g, 8.85 mmol) with vigorous stirring for *ca.* 10 min until the organic phase is white and cloudy.
3. Add aqueous sodium borohydride (1.41 g, 37.3 mmol, dissolved in 10 mL of water immediately prior to use) to the organic phase. The organic phase will immediately turn dark purple. Continue to stir the solution for 3 h under a nitrogen flow.
4. Separate the toluene layer, washing it with water (2×100 mL). Remove the solvent *in vacuo*, or with a stream of nitrogen to yield a black solid.
5. Wash the resulting solid with a series of solvents (hexanes, saturated aqueous sodium nitrite, and a 2:3 methanol:water mixture) to remove the phase-transfer catalyst, byproducts, and unreacted starting materials.
6. Dissolve the solid in chloroform, and slowly add pentane to remove salts such as $AuCl(PPh_3)$. Perform this precipitation procedure three times; after purification, the yield is *ca.* 170 mg of purified nanoparticle from 1 g of $HAuCl_4$.

7. To prepare the sample for TEM analysis, put a drop of the chloroform suspension on a TEM grid and allow the solvent to evaporate. Alternatively, you may spray the solvent suspension onto a grid using an aerosol delivery device.

Nice Papers (One Thorough Review Article) Related to Nanocluster Growth

1. Sun, S.; Zeng, H.; Robinson, D. B.; Raoux, S.; Rice, P. M.; Wang, S. X.; Li, G. *J. Am. Chem. Soc.*, **2004**, 126, 273.
2. Cushing, B. L.; Kolesnichenko, V. L.; O'Connor, C. J. *Chem. Rev.*, **2004**, 104, 3893.

Index

D

Q

R