

**MOLECULAR MACHINERY AND MANUFACTURING
WITH APPLICATIONS TO COMPUTATION**

by

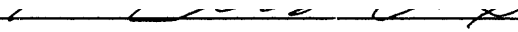
K. ERIC DREXLER

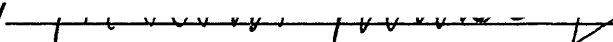
S.B. Interdisciplinary Science, MIT (1977)

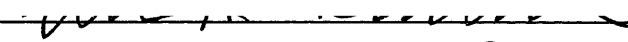
S.M. Engineering, MIT (1979)

Submitted to the Media Arts and Sciences Section,
School of Architecture and Planning, in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in an Interdepartmental Program
in the field of Molecular Nanotechnology

at the
Massachusetts Institute of Technology
September 1991

Signature of Author 
August 9, 1991

Certified by 
Marvin L. Minsky, Thesis Supervisor
Professor of Computer Science and Engineering
Toshiba Professor of Media Arts and Sciences

Accepted by 
Stephen A. Benton
Chairperson - Departmental Committee on Graduate Students

© K. Eric Drexler, 1991. All rights reserved.

The author hereby grants to MIT permission to reproduce
and to distribute copies of this thesis document in whole or in part.

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

OCT 09 1991

LIBRARIES

ARCHIVES

Molecular Machinery and Manufacturing with Applications to Computation

by
K. Eric Drexler

Submitted to the Media Arts and Sciences Section,
School of Architecture and Planning, August 9, 1991,
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in an Interdepartmental Program
in the field of Molecular Nanotechnology

Abstract

Studies were conducted to assemble the analytical tools necessary for the design and modeling of mechanical systems with molecular-precision moving parts of nanometer scale. These analytical tools were then applied to the design of systems capable of computation and of molecular-precision manufacturing.

Part I draws on classical, statistical, and quantum mechanics, together with empirical force-field models developed in chemistry, to select and develop a set of practical models describing the key engineering properties of nanometer-scale mechanical systems. These properties include potential energy functions; positional uncertainties resulting from the combined effects of quantum mechanics and thermal excitation; transition rates and damage rates resulting from these same combined effects and from ultraviolet and ionizing radiation; and energy dissipation resulting from acoustic radiation, phonon scattering, thermoelastic effects, phonon viscosity, and transitions occurring between potential wells in disequilibrium states. Part I concludes with an analysis of the capabilities of mechanosynthesis, that is, chemical synthesis directed by devices capable of moving and positioning reactive moieties with atomic-scale precision.

Part II draws on the tools developed in Part I, analyzing the mechanical properties of representative structural components and of mobile nanomechanical components such as gears and bearings. Using components and devices with these properties, nanomechanical computational systems (comprising logic gates, signal transmission elements, registers, I/O mechanisms, power supply, power distribution, and clocking) are described and analyzed, yielding estimated component densities $> 10^{19}/\text{cm}^3$ and power dissipation levels $< 10^{-9}$ those of current transistor-logic devices. Part II concludes with an analysis of the capabilities of molecular manufacturing systems based on mechanosynthesis performed by nanomechanical systems, concluding that assembly cycle times of $\sim 10^{-6}$ s will commonly be compatible with error rates of $< 10^{-12}$.

Part III summarizes a study of the feasibility of performing positionally-controlled chemical synthesis using a modified atomic force microscope, concluding that assembly cycle times of ~ 1 s and error rates of $\sim 10^{-5}$ can be anticipated from devices based on combinations of existing molecules and mechanisms. Molecular assembly mechanisms in this class can provide one of several paths forward from our present capabilities toward more advanced molecular technologies of the sort assumed in the body of the present work.

Thesis supervisor: Marvin L. Minsky
Professor of Computer Science and Engineering
Toshiba Professor of Media Arts and Sciences

Thesis Committee
of K. Eric Drexler

Stephen A. Benton
Professor of Media Technology

Rick Lane Danheiser
Professor of Chemistry

Steven H. Kim
Assistant Professor of Mechanical Engineering

Marvin Lee Minsky
Professor of Computer Science and Engineering
Toshiba Professor of Media Arts and Sciences

Alexander Rich
Professor of Biophysics
William Thompson Sedgwick Professor of Biology

Gerald Jay Sussman
Professor of Electrical Engineering

Acknowledgements

The author wishes to thank the members of the Interdepartmental Doctoral Program committee for their help and patience in an unusual program, and (in particular) Stephen Benton and the Media Lab for providing an environment in which new ideas flourish. He further wishes to thank Ralph Merkle for productive conversations and computer support; Jeffrey Soreff for productive conversations, two mathematical models in Chapter 7, and careful checking of the rest of that chapter; and, most of all, my spouse and partner Christine Peterson, for support of most imaginable kinds (and a few more besides) over many years. First and last thanks go to Marvin Minsky, for getting this doctoral program both started and finished.

Biographical Note

The author entered MIT as an undergraduate in 1973 and received an SB degree from the Department of Interdisciplinary Science in 1977. In 1977, he entered the MIT Department of Aeronautics and Astronautics, supported by an NSF Graduate Fellowship in the field of space industrialization. After receiving an SM degree in 1979, he pursued research interests in molecular nanotechnology as a Research Affiliate of the MIT Space Systems and Artificial Intelligence laboratories, and then as a Visiting Scholar in the Department of Computer Science at Stanford University, where he taught the course “Nanotechnology and Exploratory Engineering.” In spring 1989, he re-enrolled at MIT to complete a doctoral program in molecular nanotechnology through an interdepartmental program.

Table of contents

Abstract.....	2
Acknowledgements	5
Biographical note	6
1. Introduction	
1.1. What is molecular nanotechnology?	13
1.2. Comparisons	20
1.3. The approach in this volume.....	23
Part I. Physical principles	
2. Classical magnitudes and scaling laws	
2.1. The role of classical continuum models.....	33
2.2. Scaling of classical mechanical systems.....	34
2.3. Scaling of classical electromagnetic systems	40
2.4. Scaling of classical thermal systems.....	45
2.5. Beyond the classical continuum model.....	47
3. Potential energy surfaces	
3.1. The PES concept	49
3.2. Quantum theory and approximations.....	50
3.3. Molecular mechanics	56
3.4. Potentials for chemical reactions	80
3.5. Continuum representations of surfaces.....	83
3.6. Molecular models and the continuum approximation	88
3.7. Further reading.....	90
4. Molecular dynamics	
4.1. Models of dynamics	93
4.2. Non-statistical mechanics	93
4.3. Statistical mechanics.....	96
4.4. PES revisited: accuracy requirements.....	110
4.5. Further reading.....	114

5. Positional uncertainty

5.1. Uncertainty in engineering.....	115
5.2. Thermally excited harmonic oscillators.....	116
5.3. Elastic extension of thermally excited rods	122
5.4. Bending of thermally excited rods.....	133
5.5. Piston displacement in a gas-filled cylinder	141
5.6. Longitudinal variance from transverse rod deformation	144
5.7. Conclusions.....	150

6. Transitions, errors, and damage

6.1. Overview.....	153
6.2. Transitions between potential wells.....	154
6.3. Placement errors.....	166
6.4. Thermomechanical damage	172
6.5. Photochemical damage	192
6.6. Radiation damage.....	197
6.7. Device and system lifetimes	200

7. Energy dissipation

7.1. Overview.....	205
7.2. Radiation from forced oscillations.....	206
7.3. Phonons and phonon scattering	216
7.4. Thermoelastic damping and phonon viscosity.....	229
7.5. Compression of square and harmonic potential wells	232
7.6. Transitions among time-dependent wells	238
7.7. Conclusion	243

8. Mechanosynthesis

8.1. Overview.....	245
8.2. Perspectives on solution-phase organic synthesis.....	248
8.3. Solution-phase synthesis and mechanosynthesis.....	251
8.4. Reactive species	272
8.5. Forcible mechanochemical processes	283
8.6. Mechanosynthesis of diamondoid structures	306
8.7. Conclusions.....	320

Part II. Components and systems

9. Nanoscale structural components

9.1. Overview.....	325
9.2. Nanomechanical components in a structural context.....	326
9.3. Surface effects on stiffness in nanoscale components	327
9.4. Control of shape in nanoscale components.....	333
9.5. Nanoscale components of high rotational symmetry.....	336
9.6. Conclusions.....	339

10. Mobile nanomechanical components

10.1. Overview.....	341
10.2. Spatial fourier transforms of nonbonded potentials.....	342
10.3. Sliding of irregular objects over regular surfaces	346
10.4. Symmetrical sleeve bearings.....	355
10.5. Other sliding-interface bearings (and bearing systems).....	375
10.6. Atomic-axle bearings	378
10.7. Gears, rollers, and belts.....	379
10.8. Barriers in extended systems	387
10.9. Dampers, detents, and clutches	388
10.10. Conclusions.....	389

11. Nanomechanical computational systems

11.1. Overview	391
11.2. Digital signal transmission with mechanical rods	392
11.3. Gates and logic rods.....	393
11.4. Registers.....	409
11.5. Combinational logic systems and finite-state machines	415
11.6. Clocking and power distribution for CPU-scale systems	421
11.7. Power supply systems.....	426
11.8. Cooling.....	432
11.9. Interfacing to conventional microelectronics.....	433
11.10. Conclusion	435

12. Molecular manufacturing systems

12.1. Overview.....	437
12.2. Molecule acquisition and concentration	438

12.3. Molecule sorting	440
12.4. Ensuring that sites are occupied.....	440
12.5. Molecule processing	441
12.6. Reagent application.....	444
12.7. Larger-scale assembly.....	445
12.8. Conclusion	446

Part III. Implementation strategies

13. Positional synthesis exploiting AFM mechanisms

13.0. Abstract	449
13.1. Introduction.....	449
13.2. Tip-array geometry and forces.....	450
13.3. Molecular tips in AFM.....	453
13.4. Imaging with molecular tips	455
13.5. Positional synthesis	456
13.6. Summary	391

Appendix

Comparison with other work

A.1. Overview	459
A.2. How related fields have been divided.....	460
A.3. Mechanical engineering and microtechnology	461
A.4. Chemistry	461
A.5. Molecular biology	462
A.6. Protein engineering	463
A.7. Proximal probe technologies.....	464
A.8. Feynman's 1959 talk.....	465
A.9. Conclusion	466

References	469
-------------------------	------------

Summaries vs. new results

The structure of this thesis reflects its role as an intermediate stage in the preparation of a textbook introducing the new subject of nanomechanical systems engineering; accordingly, it includes extensive surveys of existing knowledge, though from an unusual perspective. What, then, are the new results presented in this work? A non-exhaustive list includes the demonstration of the feasibility of nearly-thermodynamically-reversible mechanochemical processes (including bond cleavage and hydrogen abstraction) in Chapter 8, the analyses of sufficient conditions for smooth sliding motion in irregular bearing interfaces (and the atomically-specified designs for regular bearing structures) in Chapter 10, the designs and analyses presented in the description of mechanical nano-computer systems (including logic rods, registers, power distribution and clocking, power supply, input and output) in Chapter 11.

Chapter 1

Introduction and overview

1.1. What is molecular nanotechnology?

New fields commonly require new terms to describe their characteristic features, and so it may be excusable to begin with a few definitions: *Molecular nanotechnology* comprises the characteristic techniques and products of *molecular manufacturing*, the construction of objects to complex, atomic specifications by sequences of chemical reactions directed by non-biological molecular machinery. *Mechanosynthesis* refers to mechanically-guided chemical synthesis, including operations performed in molecular manufacturing. The most significant characteristic of mechanosynthesis will be the positional control of chemical reactions on an atomic scale by means other than the local steric and electronic properties of the reagents; it is thus distinct from (for example) enzymatic processes and present techniques for organic synthesis.*

At the time of this writing, positional chemical synthesis is at the threshold of realization: precise placement of atoms and molecules has been demonstrated (e.g., Eigler and Schweizer 1990), but flexible, extensible techniques remain in the domain of design and theoretical study, as does the longer-term goal of molecular manufacturing. Accordingly, the implementation of molecular nanotechnologies like those analyzed in Part II of the

* Considering the words in isolation, the terms “molecular nanotechnology” and “molecular manufacturing” could instead be interpreted to include much of chemistry, and “mechanosynthesis” could be interpreted to include substantial portions of enzymology and molecular biology. These established fields, however, are already named; the above terms will serve best if reserved for the fields they have been coined to describe, or for borderline cases that emerge as these fields are developed.

present work awaits the development of a future generation of tools. This volume* is addressed to those concerned with identifying promising directions for current research, and to those concerned with understanding and preparing for future technologies.

The following chapters form three parts: Part I describes the physical principles of importance in molecular mechanical systems and mechanosynthesis; Part II uses these principles in the design and analysis of components and systems; Part III describes an implementation strategy emerging from the current technology base. The analysis presented in Part I indicates that existing models of molecular structure and dynamics, although limited in their scope and accuracy, are adequate to describe a functionally-diverse set of nanomechanical components. Part II examines a set of components that analysis indicates can be combined to form a wide range of systems offering characteristics and capabilities such as:

- Mechanosynthesis at $> 10^6$ operations/device-second
- Mechanochemical processes dissipating $< 10^{-21}$ J per operation
- Mechanochemical power conversion at $> 10^{10}$ W/m³
- Electromechanical power conversion at $> 10^{12}$ W/m³
- Tensile load-bearing capacities $> 5 \times 10^{10}$ Pa
- Logic gate volumes $\sim 10^{-26}$ m³
- Energy dissipation per switching event $< 10^{-21}$ J

Of these capabilities, several are qualitatively novel, and others improve on present engineering practice by one or more orders of magnitude. The stated logic-gate volumes and switching energies, for example, are consistent with the construction of a 100 W desktop computer with a parallel architecture delivering $> 10^9$ times the computational power of a 1990 mainframe computer. It can be anticipated that the techniques and products of molecular manufacturing will eventually displace the techniques and products of semiconductor lithography.

1.1.1. Example: a nanomechanical bearing

As discussed in Section 1.2, molecular nanotechnology of the sort described in the present work is related to, yet distinct from, such fields as mechanical engineering, microtechnology, chemistry, and molecular biology. An example may serve as a better introduction than would a general definition or description.

* Which is being written, in part, for use as a textbook.

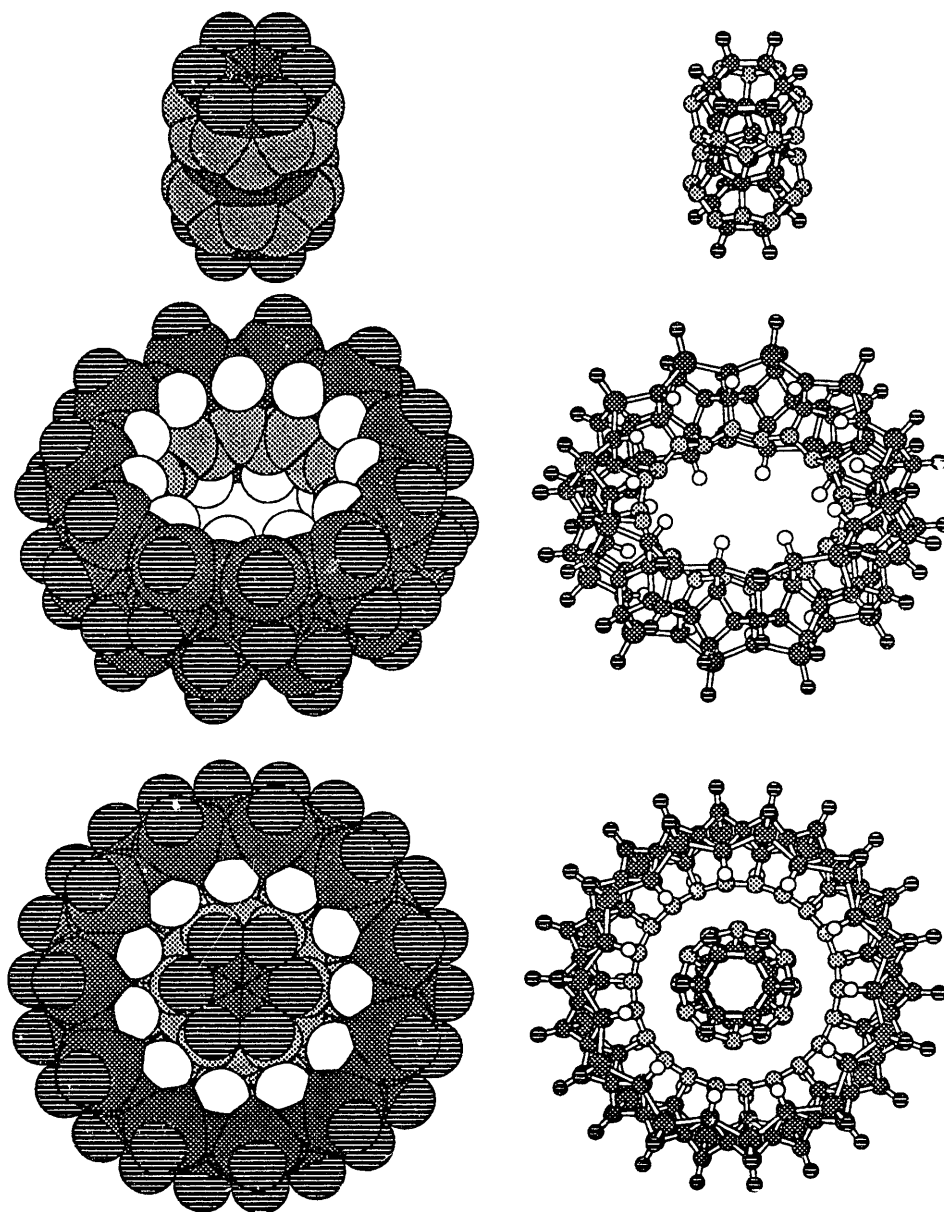


Figure 1.1. End views and exploded views of a sample steric-repulsion bearing design (in both ball-and-stick and space-filling representations, to the same scale), energy minimized with the MM2/C3D+ molecular mechanics model. Note the six-fold symmetry of the shaft structure and the eleven-fold symmetry of the surrounding ring; this relatively-prime combination results in low energy barriers to rotation of the shaft within the ring. This and other bearing structures are discussed further in Chapter 11. (“MM2/C3D+” denotes the Chem 3D Plus implementation of the MM2 molecular mechanics force field; the MM2 model is discussed in detail in Chapter 3.)

Figure 1.1 shows several views of a single design for a nanomechanical bearing discussed in greater depth in Chapter 10. In a functional context, many of the bonds shown as hydrogen terminated would instead link to other moving parts or to a structural matrix. Several characteristics are worthy of note:

- The components are polycyclic, more nearly resembling the fused-ring structures of diamond than the open-chain structures of biomolecules such as proteins.
- Accordingly, each component is relatively stiff, lacking the numerous possibilities for internal rotation about bonds that make conformational analysis difficult in many biomolecules.
- Repulsive, non-bonded interactions strongly resist both displacement of the shaft from axial alignment with the ring, and displacement either along that axis or perpendicular to it.
- Rotation of the shaft about its axis within the ring encounters negligible energy barriers, showing a nearly complete absence of static friction.

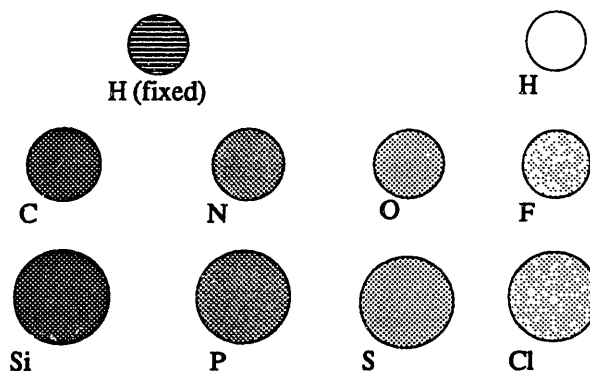


Figure 1.2. Relative sizes and grey-scale values for different atom types, as used in diagrams throughout this volume. Shading indicates valence, and radius differentiates atoms from different rows of the periodic table; hydrogen atoms with horizontal bars (fixed) represent bonds to an extended covalent structure that have been modeled as hydrogen atoms fixed in space. All radii are set equal to the values for 0.1 nN compressive contacts given in Chapter 3.

- The combination of stiffness in five degrees of freedom with facile rotation in the sixth makes the system a good bearing, in the standard mechanical engineering sense of the term.

- The lack of significant static friction in a system that places bumpy surfaces in firm contact with no intervening lubricant cannot be understood in standard mechanical engineering terms.

- Neither of the components of the bearing is a plausible target for synthesis using reagents diffusing in solution; their construction would require mechanosynthetic control.

- This particular design, to achieve compactness, exploits structures (e.g., chains of sp^3 nitrogen atoms) that would be of dubious stability in an ordinary chemical context, but will be sufficiently stable in the given structural context and contemplated chemical environment.

How typical are these characteristics? Stiff, polycyclic structures are ubiquitous in the designs presented in Part II. Many components will be designed to exhibit stiff constraints in some degrees of freedom and nearly free motion in others, thereby fulfilling roles familiar in mechanical engineering; nonetheless, a detailed understanding of how those roles are fulfilled requires analyses based on uniquely molecular phenomena. Finally, the designs in Part II (unlike those described in Part III) will consistently be of a scale and complexity that precludes synthesis using present techniques, and will often exploit structures that would be unstable if free to diffuse in a solution.

The bearing shown in Fig. 1.1 is at least suggestive of other systems that are described in Part II. For example, the combination of a bearing and shaft suggests the concept of extended systems of power-driven machinery. The outer surface of the bearing is reminiscent of a molecular-scale gear. The controlled motion of the shaft within the ring, together with the concept of extended systems of machinery, suggests the possibility of controlled molecular transport and positioning, a requirement for advanced mechanosynthesis.

1.1.2. A chemical perspective

Chemistry today (and organic synthesis in particular) is overwhelmingly focused on the behavior of molecules in solution, moving by diffusion and encountering one another in random collisions. Reaction rates in diffusive chemistry are determined by a variety of influences, including molecular concentrations and local steric and electronic factors.

Although based on the same principles of physics, molecular manufacturing of the sort described in this volume is fundamentally different from diffusive chemistry. Concepts developed to describe either immobile molecules in a solid phase or diffusing molecules in a gas or liquid require modification when describing systems characterized by non-diffusive mobility. The concept of “concentration,” for example, in the familiar sense of “number of molecules of a particular type per unit of macroscopic volume” plays no role. Local steric and electronic effects remain significant, but the decisive influence on reaction rates becomes mechanical positioning. Where differences from solid, liquid, and gas phase systems are to be emphasized, it will sometimes be useful to speak of *machine phase* systems:

- *A machine-phase system is one in which all atoms follow controlled trajectories (within the limits of thermal excitation)*
- *Machine-phase chemistry comprises systems in which all potentially-reactive moieties follow controlled trajectories (again, within the limits of thermal excitation).*

The useful distinction between liquid and gas is blurred by the existence of supercritical fluids; the useful distinction between solid and liquid is blurred by the existence of glasses, liquid crystals, and gels. Where machine-phase chemistry is concerned, definitional ambiguities are chiefly associated with the words “all” and “controlled.” In a conventional chemical reaction or an enzymatic active site, a moderate number of atoms in a small region may be said to follow somewhat-controlled trajectories, but this example falls outside the intended bounds of the definition. In a good example of a machine-phase system, large numbers of atoms will follow paths that seldom deviate from a nominal trajectory by more than an atomic diameter, while executing complex motions in an extended region from which freely-diffusing molecules are rigorously excluded. The latter conditions are fundamentally unlike those that prevail in existing chemical systems.

Table 1.1. Typical characteristics of conventional machining, micromachining, diffusive chemistry, biochemistry, and molecular manufacturing.

	Conventional fabrication	Micro- fabrication	Diffusive chemistry	Bio- chemistry	Molecular manufacturing
Molecular precision?	–	–	yes	yes	yes
Positional control?	yes	yes	–	some	yes
Typical feature scale	1 mm	1 μ	0.3 nm	0.3 nm	0.3 nm
Typical product scale	1 m	10 mm	1 nm	10 nm	1 μ
Typical defect rate	10^{-4}	10^{-7}	10^{-2}	10^{-11}	10^{-12}
Typical cycle times	10 s	100 s	1000 s	10^{-3} s	10^{-6} s
Products described by	materials and shapes	materials and shapes	atoms and bonds	monomer sequences	atoms and bonds

Note: The defect rate in the biochemistry column corresponds to a high-reliability DNA replication process including kinetic proofreading (Watson, Hopkins et al. 1987); most biochemical defect rates are higher.

Mechanosynthesis of the sort discussed in Chapters 8 and 12 is a machine-phase process (Part III discusses mechanosynthesis within a solvent). Machine-phase mechanosynthesis will afford novel chemically capabilities, including positional discrimination of chemically-equivalent sites, effective suppression of side reactions, and new sources of activation energy.

Chemistry in the machine phase shares characteristics of gas-, solution-, and solid-phase chemistry, but also displays unique characteristics; these similarities and differences are discussed further in Chapters 6 and 8. Since experience shows that the habits of thought developed in the study of liquid- and gas-phase systems can yield misleading conclusions if hastily applied to machine-phase systems, frequent recourse to fundamental principles is required.

1.2. Comparisons

Molecular nanotechnology and molecular manufacturing share many features of other fields, yet differ substantially. A discussion of similarities can aid understanding by illustrating the applicability of existing knowledge; a discussion of differences can likewise aid understanding by warning of false analogies and consequent misunderstandings.

As an introduction, Table 1.1 compares several existing production processes—conventional fabrication, microfabrication, diffusive chemistry, and biochemistry—to molecular manufacturing. The following sections examine molecular manufacturing and molecular nanotechnology in comparison to these other processes and their products. An appendix discusses previous theoretical and experimental work related to molecular manufacturing.

1.2.1. *Mechanical engineering*

1.2.1.1. *Similarities: components, systems, controlled motion, manufacturing*

A wide range of mechanical engineering concepts apply directly to nanomechanical systems. As shown in Chapters 3–6, models based on classical mechanics are adequate for much of the required analysis. As shown in Chapters 9–11, beams, shafts, bearings, gears, motors, and the like can all be constructed on a nanometer scale to serve familiar mechanical roles. Accordingly, mechanical engineering and nanomechanical engineering share many design issues and analytical techniques.

Like conventional manufacturing systems, molecular manufacturing systems will use machines to perform essentially deterministic patterns of motion, moving, reshaping, and joining components to build complex three-dimensional structures. Both can manufacture machines, including machines useful in manufacturing.

1.2.1.2. *Differences: scale, molecular phenomena*

Despite these similarities, nanomechanical engineering forms a distinct field. The familiar model of objects as made of homogeneous materials, while still useful, must frequently be replaced by models that treat objects as sets of bonded atoms (Chapters 2 and 3). Thermally-excited vibrations are of major importance, and quantum effects are sometimes significant (Chapters 5, 6, and 7). Further, molecular phenomena permit (and demand) novel bearings (Chapter 10); scaling laws favor electrostatic over electromag-

netic motors (Chapters 2 and 11); novel damage mechanisms become important (Chapter 6); and the basic unit operations in manufacturing are chemical transformations (Chapter 8).

1.2.2. Microtechnology

1.2.2.1. Similarities: small scale, electronic quantum effects

Microtechnology has enabled the fabrication of micron-scale mechanical devices, which share basic scaling properties with nanomechanical devices; as a consequence, electrostatic motors are preferred over electromagnetic motors in both micro- and nanotechnology. Further, microelectronic technology has begun to grapple with quantum transport phenomena, and devices of kinds now being explored using microfabrication technologies may be of interest as targets for molecular manufacturing.

1.2.2.2. Differences: fabrication, scale, molecular phenomena

Microfabrication techniques rely on a technology base almost unrelated to that of molecular manufacturing and unable (for example) to make a structure like the bearing in Fig. 1.1. In a sense, these fields are moving in opposite directions: microfabrication attempts to make bulk-material structures *smaller* without losing the resolution necessary to define device geometries; molecular manufacturing will emerge from efforts to make molecular structures *larger* without losing the atomic precision characteristic of stereospecific chemical synthesis. Making structures containing a few dozen precisely-arranged atoms seems unachievable using the etching and deposition processes characteristic of microfabrication, but it is routine in chemical synthesis. The gears, bearings, and motors described in Chapters 10 and 11 differ in volume from their closest microfabricated counterparts by a factor of $\sim 10^9$, and rely on molecular structures and phenomena for their operation.*

* The term “nanotechnology” was first introduced into widespread use (Drexler 1986) to refer to what is here termed “molecular nanotechnology,” but has increasingly been used to refer to the incremental extension of conventional microfabrication techniques into the submicron size range. Accordingly, some recent discussions of the history, status, and prospects of “nanotechnology” have confused essentially dissimilar concepts, as if “ornithology” were used to describe the study of flying things, thereby stirring birds, balloons, and bombers together into a single conceptual muddle.

1.2.3. Chemistry

1.2.3.1. Similarities: molecular structure, processes, fabrication

Chemistry is the discipline most applicable to the study of the unit operations of molecular manufacturing, in which each basic step will consist of a chemical transformation; it is also applicable to the study of the stability of the products, and is the field that has originated the most useful tools for modeling the mechanical behavior of molecular objects. Organic chemistry is of particular importance, owing to the mechanical advantages of carbon-based structures. Such fundamental chemical concepts as bonding, strain, reaction rates, transition states, orbital symmetry, steric hindrance, and equilibrium constants are all applicable; familiar chemical entities such as alkanes, alkenes, aromatic rings, functional groups, radicals, and carbenes are all of use.* Solution-based organic synthesis, like mechanochemistry, is a technique for making precisely-structured molecular objects. It has even been used to make molecular gears (Mislow 1989), although of a kind having no obvious utility for nanomechanical engineering.

1.2.3.2. Differences: machine-phase systems, mechanosynthesis

The chief differences between the present subject and conventional chemistry stem from the properties of machine-phase systems and mechanosynthesis. These have been summarized in Section 1.1.2, and are discussed at length in Chapter 8.

* Writing for a general audience, by occasionally speaking of controlling and building with “individual atoms” (Drexler 1986) have fostered the notion that proposals for molecular manufacturing involve (or require) somehow prying loose, moving, and putting down highly *unfamiliar* chemical entities, such as individual, *unbonded*, *highly-reactive* carbon atoms. This rightly strikes chemists as implausible. Indeed, free atoms would be difficult to position, and conventional reagents and reactive intermediates have always seemed to be the appropriate models. When describing molecular manufacturing, (Drexler 1986) speaks of using “reactive molecules” as tools “to bond atoms together. . . a few at a time”; the first paper on the subject (Drexler 1981) speaks of positioning “reactants” and “reactive groups.” Controlling the trajectories and reactions of molecules, of course, implies controlling the trajectories and destinations of their individual constituent atoms.

1.2.4. Molecular biology

1.2.4.1. Similarities: molecular machines, molecular systems

Molecular biology, like molecular nanotechnology, embraces the study of molecular machines and systems of molecular machines. Ribosomes, like flexible molecular manufacturing systems, can be described as numerically-controlled machine tools, following a series of instructions to produce a complex product. Molecular biology and biochemistry inspired the train of thought that led to the concept of molecular manufacturing (Drexler 1981), and they offer one of several paths for the development of molecular manufacturing systems.

1.2.4.2. Differences: materials, machine-phase systems, general mechanosynthesis

Molecular biology is a product of evolution rather than design, and is organized in a manner quite unlike the machine-phase systems described here. Unlike molecular manufacturing systems, ribosomes and the other molecular machines found in cells lack the ability to construct generalized covalent objects of broadly diamond-like structure, being restricted to producing small molecules and a stereotyped set of polymers. With few exceptions, the resulting structures must therefore acquire their three-dimensional shapes as the result of weak forces (hydrogen bonds, salt bridges, van der Waals attraction, hydrophobic effects), unlike the strongly-bonded covalent solids discussed in Part II; as a consequence, biological and contemplated nanomechanical components differ by orders of magnitude in such mechanical properties as strength and modulus. Further, biochemical systems make extensive use of diffusion in a liquid phase; although they incorporate molecular machines, they are not machine-phase systems. The bearings, gears, motors, and computers discussed in Part II are likewise without biological parallel, being quite different from the bacterial flagellar motor, the actin-myosin system, systems of neurons, and so forth.

1.3. The approach in this volume

1.3.1. Purpose, level, and approach

As Section 1.2 indicates, the study of molecular nanotechnology crosses several disciplinary boundaries. This circumstance has hampered both evaluation of existing concepts and research aimed at extending and superseding them. A basic purpose of the

present volume is to put a large portion of the requisite core knowledge between two covers in a form that does not assume specialized knowledge in the component disciplines. An effort has been made to make chemical concepts accessible to nonchemists, solid-state physics concepts accessible to nonphysicists, and so forth, assuming only a basic background in both chemistry and physics (and a willingness to skip past the occasional obscure observation aimed at a reader in a different discipline). The contribution of this work does not consist in extending the often-esoteric boundaries of established fields, but in combining their basic principles to delineate a new field of design and enquiry.

To facilitate understanding, a variety of mathematical results in Part I are derived from first principles. Many of these results appear in existing textbooks; others (so far as is known) are novel, being motivated by new questions. The exposition of these mathematical models includes an unusually large number of graphs that illustrate equations in the text. The chief reason is to facilitate design, which is a synthetic as well as an analytic process. In the analysis of a given system, a calculation based on an equation and a single set of parameter values is frequently adequate. In synthesis, however, the designer typically wishes to understand how system properties will vary with the controllable parameters; for this, a graph is often far more useful than a bare equation.

1.3.2. Scope and assumptions

The present volume works within several assumptions and limitations that will not necessarily apply to future engineering practice. Each excludes possibilities that are presently difficult to analyze, but that may prove worth implementing. The following assumptions and limitations are thus conservative, resulting in underestimates of future capabilities:

1.3.2.1. A narrow range of structures

From the broad range of forms of matter (metals, ionic crystals, molecular crystals, etc.), the present work selects the class of covalent solids as its focus. Covalent solids include the strongest and stiffest structures known to be possible at ordinary pressures (Kelly 1973), making them attractive on engineering grounds. Carbon-rich structures are superior in this regard. Further, since they can be viewed as large polycyclic organic molecules, much of the vast base of knowledge developed by organic chemists is immediately applicable. Small components will be subject to large surface effects, but typical

organic molecules are, in effect, *all* surface; accordingly, surface effects are an integral part of molecular models.

1.3.2.2. Mechanical and electromechanical devices

On a macroscale, mechanical systems are clearly distinct from electronic systems: they involve the motion of materials, rather than of electrons and electromagnetic fields. On a nanoscale, mechanical motions are identified with the motions of nuclei, yet electronic activity has nuclear correlates. Still, many systems (e.g., the bearing in Fig. 1.1) can be described by molecular mechanics models that take no explicit account of electronic degrees of freedom, subsuming them into a potential energy function defined in terms of the positions of nuclei. The focus of the present work is on such mechanical systems, which are frequently well-approximated by classical models.

Other systems are strongly electronic in character, relying on changes of electronic state to change other electronic states, with the associated nuclear motions being of small amplitude (e.g., resulting from relaxation or phonon-mediated coupling processes). Molecular nanotechnologies will include nanometer-scale electronic devices (nanoelectronics) exploiting quantum phenomena to achieve (for example) switching and computation. Research relevant to this class of devices is increasing. Nanoelectronic devices are beyond the scope of the present work, although they will likely prove to be important products of molecular manufacturing. Accordingly, the nanocomputers discussed here are based on mechanical devices, although electronic devices will surely give greater speed.

Some attention is given (Chapter 11) to nanoscale electromechanical systems. In these, electron transport through conductors and across tunneling junctions is used to drive motors and actuators. Quantum phenomena are significant in this process, but the gross results (interconversion of electrical and mechanical energy) do not depend on subtle quantum effects. Finally, despite their likely utility, machine-phase electrochemical processes will be touched on in only a cursory fashion.

1.3.2.3. Machine-phase chemistry

Because it permits the elimination of many potential modes of chemical reaction by tightly constraining molecular motion, machine-phase chemistry can be simpler (in certain respects) than is solution-phase chemistry. Mechanosynthesis and other operations could be conducted by systems of molecular machines exposed to a solution environment, and there may sometimes be advantages to doing so. These less-controlled, more-

complex chemical systems fall beyond the scope of the present work, save for a discussion of implementation strategies in Part III.

1.3.2.4. Room temperature processes

Operation at reduced temperatures reduces thermally-excited displacements (Chapter 5), thermal damage rates (Chapter 6), and phonon-mediated drag (Chapter 7). The technological opportunities and problems associated with low-temperature systems are nonetheless not explored in the present work. Operation at elevated temperatures will be desirable in many circumstances, and can facilitate chemical reactions (both desired and undesired). The technological problems and opportunities associated with high-temperature operation are likewise neglected here.

1.3.2.5. No photochemistry

Photochemical damage mechanisms are discussed in Chapter 6. Design of molecular machines for photochemical damage resistance is a challenge worthy of attention, but for simplicity, the following will instead assume operation of devices in a shielded environment. Optical shielding requirements for photochemically-sensitive systems are discussed.

1.3.2.6. The single-point failure assumption

Design of components that can tolerate atomic-scale damage and defects is likewise a challenge worthy of attention, but the following will instead assume that any such damage or defect causes component-level (and typically subsystem-level) failure. Damage tolerance through redundant subsystems is assumed only at higher levels of organization. Diverse damage mechanisms are reviewed and modeled in Chapter 6.

1.3.2.7. Few claims of optimality

In mature fields of technology, competitive pressures encourage a search for designs that are (where possible) nearly optimal. In the exploratory phase, however, the more modest goal of workability is sufficient. The present work describes an approach to molecular manufacturing and molecular nanotechnology that seems workable in light of current knowledge of physical law and molecular systems, and that is attractive by the standards of current technology. Other broad approaches (to say nothing of other specific devices) may prove to be superior, that is, more nearly optimal, thereby further extending

the bounds of known technological possibilities. This development would in no way invalidate the present conclusions regarding what is workable.

1.3.2.8. False-negative errors in analysis

In modeling and analyzing proposed designs, the ideal goal would be to distinguish workable from unworkable designs with perfect accuracy. Since models are seldom exact, though, errors are inevitable. These errors may be of two kinds: *False-positive* evaluations wrongly accept an unworkable design; *false-negative* evaluations wrongly reject a workable design. In exploring a new domain of technology, conclusions regarding the feasibility of systems rest on conclusions regarding the feasibility of subsystems, forming a hierarchical structure of analysis that may have several layers. A substantial rate of false-positive assessments at the subsystem level will make false-positive assessments at the system level quite likely, thus invalidating the conclusions. False-negative assessments, in contrast, are relatively benign. Correcting a false-negative assessment regarding a subsystem does not invalidate the analysis of a system, it merely shows that a previously-rejected alternative is also workable. It is desirable to minimize errors of both kinds, but where uncertainties remain, the analyses in the present work are biased toward safe, false-negative conclusions.

1.3.3. Overview of later chapters

1.3.3.1. Part I

Chapter 2 summarizes classical scaling laws for mechanical, electrical, and thermal systems, describing the magnitudes they imply for various physical parameters, and describing where and how these laws break down, requiring that they be replaced with molecular and quantum mechanical models.

Chapter 3 provides an overview of molecular and inter-surface potential energy functions, describing in some detail the molecular mechanics models that later chapters apply to the description of molecular machines.

Chapter 4 gives an overview of various models of molecular dynamics, and describes the basis for the choice of models made in later chapters.

Chapter 5 examines several classes of mechanical structures and derives relationships describing the positional uncertainties resulting from the combined effects of quantum mechanics and thermal excitation; these results are later used to derive models describing error rates.

Chapter 6 examines a variety of mechanisms that result in transitions between different potential wells in a nanomechanical system, with results including errors and structural damage; these mechanisms include thermal excitation, mechanical stress, photochemistry, and ionizing radiation.

Chapter 7 examines a variety of mechanisms that result the degradation of mechanical energy to heat, resulting in frictional losses; these include acoustic radiation, phonon scattering, shear-reflection drag, phonon viscosity, thermoelastic damping, nonisothermal compression of mobile components, and transitions among time-dependent potential wells.

Chapter 8 compares and contrasts the known capabilities of diffusive chemistry to those to be expected from mechanochemical systems, in terms of speed, efficiency, versatility, and reliability; it also examines an illustrative set of mechanochemical processes in detail.

Overall, Part I expends significant effort in describing physical effects that will be of negligible importance in nanomechanical design. In established fields, such as conventional engineering, training and experience usually focus attention on the important physical effects; trivial effects are automatically ignored. In surveying a new field, however, insignificant effects must often be given a cursory inspection, precisely to establish their insignificance.

1.3.3.2. Part II

Chapter 9 uses molecular mechanics models and extrapolations from bulk properties to describe the mechanical and geometric properties of nanoscale structural components, including generalizations regarding surface corrections to bulk properties.

Chapter 10 uses molecular mechanics models and analytical models developed in Part I to describe the mechanical properties and performance characteristics of active devices such as bearings and gears, including an analysis of the conditions required to guarantee the availability of low-friction motion.

Chapter 11 describes the use of mechanical components to build digital logic systems that can serve as building blocks for computers, surveying particular electromechanical components (motors for power, electrostatic actuators and tunneling junctions for I/O), and considering such system-level issues as speed, computational density, energy dissipation, and cooling.

Chapter 12 builds on the conclusions of Chapters 10 and 11 to show how complex, controlled patterns of motion can be used to guide mechanosynthetic processes in the construction of complex systems, including systems of molecular machinery.

Starting with the assumption that molecular manufacturing systems can build essentially any kinetically-stable structure, Part II describes nanomechanical systems of increasing complexity, ending with a description of molecular manufacturing systems. This supports the non-trivial proposition that molecular manufacturing systems are feasible, given molecular manufacturing systems to build them. Part III describes a potential step toward building a bridge between our existing technology base (ca. 1991) and a technology base capable of equaling or exceeding the capabilities described in Part II.

1.3.3.3. Part III

Chapter 13 describes current capabilities in molecular-scale mechanical positioning and sensing via atomic force microscopy, and discusses means for exploiting this, together with chemical techniques, to develop systems capable of positional synthesis.

1.3.3.4. Appendix

An appendix discusses previous work.

Part I.

Physical Principles

Chapter 2

Classical magnitudes and scaling laws

2.1. The role of classical continuum models

When used with caution, classical continuum models are of substantial value in the design and analysis of nanoscale systems. They represent the simplest level in a hierarchy of approximations of increasing accuracy, complexity, and difficulty. The present chapter uses classical continuum models to introduce basic scaling laws and to indicate the expected magnitudes of physical quantities, noting where more accurate approximations are obligatory.

Experience teaches the value of approximation in design. The design process starts with the generation and preliminary evaluation of options, then proceeds by elaborating and further evaluating a few preferred options, finally settling on a detailed specification and analysis of a chosen design. The first steps entail little commitment to a particular approach. Initially, the ease of exploring and comparing many qualitatively distinct approaches is at a premium, and drastic approximations can suffice to screen out the worst options. Even the final design and analysis does not require an exact calculation of physical behavior; a combination of approximations and safety margins suffices. Accordingly, the design process can make use of different approximations at different stages, moving up the hierarchy of accuracy.

Approximation is also inescapable. Real physical systems are governed by the laws of quantum mechanics, and devices typically contain many electrons and many nuclei. This suffices to render exact solutions unavailable, even if one regards present physical theory as correct. As discussed in the the next chapter, the quantum mechanical models of

molecular systems in common use today rely on approximate solutions to wave equations which are approximations to the Schrödinger equation, which approximates the Dirac equation, which approximates quantum electrodynamics. Even with these approximations, accurate quantum mechanical calculations become impracticably expensive for large molecules, where “large” describes structures far smaller than a protein or a typical nanomechanism.

In macromechanical design, engineers employ approximations based on classical mechanics, neglecting quantum mechanics, the thermal excitation of mechanical motions, and the molecular structure of matter. Since macromechanical engineering blends into nanomechanical engineering with no clear line of demarcation, the approximations of macromechanical engineering can provide a point of departure for exploring the nanomechanical realm. In some circumstances, as will be seen, these approximations provide an adequate basis for design and analysis. In a broader range of circumstances, they provide an adequate basis for exploring design options and for conducting a preliminary analysis. In a yet broader range of circumstances, they provide a crude description to which one can compare more sophisticated approximations.

Chapters 3–7 examine nanomechanical systems in terms of molecular structure, thermal excitation, and (where significant) quantum mechanical effects. The present chapter examines how physical parameters vary with scale in the classical continuum approximation and outlines how this approximation fails as the scale becomes sufficiently small.

2.2. Scaling of classical mechanical systems

Nanomechanical systems will be basic to molecular manufacturing and will find use in many of its products. The widespread use of the molecular mechanics approximation in chemistry (Chapter 3) indicates the utility of describing nanoscale mechanical systems in terms classical mechanics. Chapters 4, 5, and 6 will examine quantum corrections to these laws and the effects of thermal excitation; Chapter 3 will describe forces at the molecular level. The present section describes scaling laws and magnitudes using the conventional engineering approximations of classical mechanics and continuous media.

2.2.1. Basic assumptions

The present section considers only mechanical systems, neglecting fields and currents. Like the following sections, it examines the consequences of varying a length parameter, L , while holding constant all material properties (e.g., strengths, moduli, den-

sities, coefficients of friction) and all parameters describing shape.

A description of scaling laws must begin with choices that determine the scaling of dynamical variables. The most natural choice appears to be that of constant stress. As shown, this ensures scale-independent constancy of the shape of mechanically-loaded parts; by resulting in scale-independent speeds, it also implies constancy of the space-time shapes describing the trajectories of moving parts. Some sample magnitudes are provided, based on material properties consistent with those of diamond: density $\rho = 3.5 \times 10^3 \text{ kg/m}^3$; Young's modulus $E = 10^{12} \text{ N/m}^2$; working stress (~ 0.2 times tensile strength) $\sigma = 10^{10} \text{ N/m}^2$.

2.2.2. Magnitudes and scaling

Given constancy of stress and material strength, the strength of a component or system and the force it exerts both scale with its cross-sectional area

$$\text{area} \propto \text{total strength} \propto \text{force} \propto L^2. \quad (2.1)$$

Nanoscale devices accordingly exert minute forces: a stress of 10^{10} N/m^2 equals 10^{-8} N/nm^2 , or 10 nN/nm^2 . Compressive and shear stiffnesses depend on both area and length

$$\text{compressive stiffness} \propto \text{shear stiffness} \propto \frac{\text{area}}{\text{length}} \propto L \quad (2.2)$$

and vary less rapidly with scale; a cubic nanometer block of $E = 10^{12} \text{ N/m}^2$ has a stretching stiffness of 1000 N/m . Bending stiffness scales in the same fashion

$$\text{bending stiffness} \propto \frac{\text{radius}^4}{\text{length}^3} \propto L. \quad (2.3)$$

Given the above scaling relationships, the magnitude of the deformation under load

$$\text{deformation} \propto \frac{\text{force}}{\text{stiffness}} \propto L \quad (2.4)$$

is proportional to scale, and hence the shape of deformed structures is scale invariant.

The assumption of constant density makes mass scale with volume,

$$\text{mass} \propto \text{volume} \propto L^3, \quad (2.5)$$

and the mass of a cubic nanometer block of $\rho = 3.5 \times 10^3 \text{ kg/m}^3$ equals $3.5 \times 10^{-24} \text{ kg}$.

The above expressions yield the acceleration scaling relationship

$$\text{acceleration} \propto \frac{\text{force}}{\text{mass}} \propto L^{-1}. \quad (2.6)$$

A cubic-nanometer mass subject to a net force equaling the above working stress applied to a square nanometer will experience an acceleration of $\sim 3 \times 10^{15} \text{ m/s}^2$. Accelerations in nanomechanisms will typically be large by macroscopic standards, though aside from special cases (such as transient acceleration during impact and steady acceleration in a small flywheel) they will be less than the value just calculated. (Note that terrestrial gravitational stresses and accelerations usually have negligible effects on nanomechanisms.)

Modulus and density determine the acoustic speed, which is a scale-independent parameter (along a slim rod, the speed is $[E/\rho]^{1/2}$; in bulk material, somewhat higher). The vibrational frequencies of a mechanical system are proportional to the acoustic transit time

$$\text{frequency} \propto \frac{\text{acoustic speed}}{\text{length}} \propto L^{-1}. \quad (2.7)$$

The acoustic speed in diamond is $\sim 1.75 \times 10^4 \text{ m/s}$. Some vibrational modes are more conveniently described in terms of lumped parameters of stiffness and mass,

$$\text{frequency} \propto \sqrt{\frac{\text{stiffness}}{\text{mass}}} \propto L^{-1}, \quad (2.8)$$

but the scaling relationship is the same. The stiffness and mass associated with a cubic nanometer block (as calculated above) yield a vibrational frequency characteristic of a stiff, nanometer-scale object: $[(1000 \text{ N/m})/(3.5 \times 10^{-24} \text{ kg})]^{1/2} \approx 1.7 \times 10^{13} \text{ rad/s}$.

Characteristic times are inversely proportional to characteristic frequencies

$$\text{time} \propto \text{frequency}^{-1} \propto L. \quad (2.9)$$

The speed of mechanical motions is constrained by strength and density. Its scaling can be derived from the above expressions

$$\text{speed} \propto \text{acceleration} \cdot \text{time} = \text{constant}. \quad (2.10)$$

A characteristic speed (which will seldom be exceeded in practical mechanisms) is that at

which the rim of a flywheel in the form of a slim hoop is subject to the chosen working stress as a result of its centripetal acceleration. This occurs when $v = (\sigma/\rho)^{1/2} \approx 1.7 \times 10^3$ m/s.

The frequencies characteristic of mechanical motions scale with transit times

$$\text{frequency} \propto \frac{\text{speed}}{\text{length}} \propto L^{-1}. \quad (2.11)$$

These frequencies scale in the same manner as vibrational frequencies, hence the assumption of constant stress leaves speed and frequency ratios as scale invariants. At the above characteristic speed, transit of a 1 nm distance would take $\sim 6 \times 10^{-13}$ s; the assumption of a large speed makes this shorter than the motion times anticipated in typical nanomechanisms. A modest 1 m/s speed, however, still yields a transit time of only 1 ns, indicating that nanomechanisms can operate at frequencies typical of modern micron-scale electronic devices.

The above expressions yield relationships for the scaling of mechanical power,

$$\text{power} \propto \text{force} \cdot \text{speed} \propto L^2, \quad (2.12)$$

and mechanical power density

$$\text{power density} \propto \frac{\text{power}}{\text{volume}} \propto L^{-1}. \quad (2.13)$$

Combining a 10 nN force and a 1 nm^3 volume yields a power of $17 \text{ } \mu\text{W}$ and a power density of $1.7 \times 10^{22} \text{ W/m}^3$ (at a speed of 1.7×10^3 m/s) or 10 nW and 10^{19} W/m^3 (at a speed of 1 m/s). The combination of strong materials and small devices promises mechanical systems of extraordinarily high power density (e.g., the power transmitted by a gear, divided by its volume), if speeds of motion are in the range described.

Mechanical systems typically employ bearings to permit motion across an interface. Macromechanical systems frequently use liquid lubricants, but (as noted by (Feynman 1961)), this poses problems on a small scale. The above scaling assumptions hold speeds constant; reducing the thickness of the lubricant layer then results in increasing shear rates and hence increasing viscous shear stress,

$$\text{viscous stress (at constant speed)} \propto \text{shear rate} \propto \frac{\text{speed}}{\text{thickness}} \propto L^{-1}. \quad (2.14)$$

In Newtonian fluids, shear stress is proportional to shear rate. Liquids can remain nearly Newtonian up to shear rates in excess of 100 m/s across a 1 nm layer (Ashurst 1975), but depart from bulk viscosity (or even from liquid behavior) when film thicknesses are less than 10 molecular diameters (Israeliachvili, McGuiggan et al. 1988; Schoen, Rhykerd et al. 1989), owing to interface-induced alterations in liquid structure. Feynman suggested the use of low-viscosity lubricants (e.g., kerosene) for micromechanisms (Feynman 1961); from the perspective of a typical nanomechanism, however, kerosene is better thought of as a collection of bulky molecular objects than as a liquid. If one nonetheless applies the classical approximation to a 1 nm film of low viscosity fluid ($\eta = 10^{-3}$ N·s/m²), the viscous shear stress at a speed of 1.7×10^3 m/s is 1.7×10^9 N/m²; the shear stress at a speed of 1 m/s, 10^6 N/m², is still substantial, dissipating energy at a rate of 1 MW/m².

For systems using liquid lubricants, an alternative (and also undesirable) scaling principle would hold viscous stresses constant, resulting in a scale-dependent speed,

$$\text{viscous speed} \propto \frac{\text{force} \cdot \text{thickness}}{\text{area}} \propto L, \quad (2.15)$$

and a scale-independent frequency of mechanical motion

$$\text{viscous frequency} \propto \frac{\text{length}}{\text{viscous speed}} = \text{constant}. \quad (2.16)$$

The problems of liquid lubrication motivate consideration of dry bearings (as suggested by (Feynman 1961)). Assuming a constant coefficient of friction,

$$\text{frictional force} \propto \text{force} \propto L^2, \quad (2.17)$$

and both stresses and speeds are once again scale-independent. In terms of conventional mechanical engineering, the use of dry bearings would seem to present problems. The frictional power,

$$\text{frictional power} \propto \text{force} \cdot \text{speed} \propto L^2, \quad (2.18)$$

is proportional to the total power, implying scale-independent mechanical efficiencies. In the absence of lubrication, however, those efficiencies would be expected to be low. Further, static friction in dry bearings would be expected to cause problems of jamming and vibration.

A yet more serious problem for unlubricated systems would seem to be wear. Assuming constant interfacial stresses and speeds, as implied by the above scaling relationships, the surface erosion rate should be independent of scale. Assuming that wear life is determined by the time required to produce a certain fractional change in shape,

$$\text{wear life} \propto \frac{\text{thickness}}{\text{erosion rate}} \propto L, \quad (2.19)$$

and a centimeter-scale part having a ten-year lifetime would be expected to have a 30 s lifetime if scaled to nanometer dimensions.

Design and analysis have shown that dry bearings with atomically-precise surfaces need not suffer these problems. As shown in Chapters 6, 7, and 11, dynamic friction can be low, and both static friction and wear can approach zero. The scaling laws applicable to such bearings are compatible with the constant-stress, constant-speed expressions derived above.

2.2.3. Major corrections

The above scaling relationships treat matter as a continuum with bulk values of strength, modulus, and so forth. They readily yield results for the behavior of iron bars scaled to a length of 10^{-12} m, although such results are utterly meaningless since a single atom of iron is over 10^{-10} m in diameter. They also neglect the influence of surfaces on mechanical properties, which can be substantial when objects are only a few atomic layers thick. Finally, they give at best crude estimates for objects of conventional molecular scale, in which some dimensions may be no more than a single atomic diameter.

From the perspective of conventional mechanical engineering, the ability to engineer structures at a molecular level will yield a variety of unusual results. These will emerge from more detailed analysis in later chapters.

Aside from the molecular structure of matter, major corrections to the conventional mechanical engineering perspective include statistical mechanical and quantum mechanical uncertainties in position. These are examined in detail in Chapter 5. Thermal excitation superimposes random velocities on those that result from the planned operations of a mechanism. These random velocities depend on scale, such that

$$\text{thermal speed} \propto \sqrt{\frac{\text{thermal energy}}{\text{mass}}} \propto L^{-3/2} \quad (2.20)$$

and, for $\rho = 3.5 \times 10^3 \text{ kg/m}^3$, the mean thermal speed of a cubic nanometer object at 300 K is $\sim 55 \text{ m/s}$. Random thermal velocities (typically occurring in vibrational modes) will often exceed the velocities imposed by planned operations, and cannot be ignored in analyzing nanomechanical systems.

Quantum mechanical uncertainties in position and momentum parallel statistical mechanical uncertainties in their effects on nanomechanical systems. The importance of quantum mechanical effects in vibrating systems depends on the ratio of the characteristic quantum and thermal energies, $\hbar\omega/kT$; this varies directly with the frequency of vibration, that is, with L^{-1} . At 300 K, a cubic nanometer object with a characteristic frequency $1.7 \times 10^{13} \text{ rad/s}$ (see above) has a value of $\hbar\omega/kT \approx 0.4$, and quantum mechanical effects will be smaller than thermal effects, but still significant (Chapter 5).

2.3. Scaling of classical electromagnetic systems

2.3.1. Basic assumptions

In considering the scaling of electromagnetic systems, it is convenient to assume that electrostatic field strengths (and hence electrostatic stresses) are scale-independent. With this assumption, the above constant-stress, constant-speed scaling laws for mechanical systems continue to hold for electromechanical systems, so long as magnetic forces are neglected. If electrostatic fields are limited by field emission from conductors (a reasonable assumption for small-scale systems), fields at negative electrodes can be $\sim 10^9 \text{ V/m}$ (see Chapter 11).

2.3.2. Major corrections

A broad treatment of quantum electronic systems lies beyond the scope of the present work, despite their undoubted importance for nanotechnology (particularly where computation is concerned). Chapter 11 will, however, consider several nanometer scale electromechanical systems.

Corrections to classical continuum models are more important in electromagnetic systems than in mechanical systems: they become dominant at a larger scale, and at small scales they can render classical continuum models useless even as crude approximations. Electromagnetic systems on a nanometer scale are often characterized by extremely high frequencies, resulting in large values of $\hbar\omega/kT$. Electronic transitions in molecules typically absorb and emit light in the visible to ultraviolet range, rather than the infrared

range characteristic of thermal excitation at room temperature. The mass of an electron is less than 10^{-3} that of the lightest atom, hence for comparable binding potentials, electron wave functions are more diffuse and permit longer-range tunneling. This small mass nonetheless has significant inertial effects at high frequencies which are neglected in the usual macroscopic expressions for electrical circuits. Accordingly, many of the following classical continuum scaling relationships fail in nanometer-scale systems.

2.3.3. Magnitudes and scaling: steady-state systems

Given a scale-invariant electrostatic field strength,

$$\text{voltage} \propto \text{electrostatic field} \cdot \text{length} \propto L. \quad (2.21)$$

At a field strength of 10^9 V/m, a one nanometer distance yields a 1 V potential difference. A scale-invariant field strength implies a force proportional to area,

$$\text{electrostatic force} \propto \text{area} \cdot (\text{electrostatic field})^2 \propto L^2, \quad (2.22)$$

and a 10^9 V/m field applied to a square nanometer area yields a force of $\approx 4.4 \times 10^{-12}$ N.

Assuming a constant resistivity,

$$\text{resistance} \propto \frac{\text{length}}{\text{area}} \propto L^{-1}, \quad (2.23)$$

and a cubic nanometer block with the resistivity of copper would have a resistance of 17 Ω . This yields an expression for the scaling of currents,

$$\text{ohmic current} \propto \frac{\text{voltage}}{\text{resistance}} \propto L^2, \quad (2.24)$$

which leaves current density constant. In present microelectronics work, current densities in aluminum interconnections are limited to $< 10^{10}$ A/m² or less by electromigration, which redistributes metal atoms and eventually interrupts circuit continuity (Mead and Conway 1980). This current density equals 10 nA/nm².

For field emission into free space, current density depends on surface properties and the electrostatic field intensity, hence

$$\text{field emission current} \propto \text{area} \propto L^2, \quad (2.25)$$

and field emission currents scale with ohmic currents. Where surfaces are close enough

together for tunneling to occur from conductor to conductor, rather than from conductor to free space, this scaling relationship breaks down.

With constant field strength, electrostatic energy scales with volume:

$$\text{electrostatic energy} \propto \text{volume} \cdot (\text{electrostatic field})^2 \propto L^3. \quad (2.26)$$

A 10^9 V/m field has an energy density of $\sim 4.4 \times 10^{-21}$ J per cubic nanometer ($\sim kT$ at room temperature).

Scaling of capacitance follows from the above,

$$\text{capacitance} \propto \frac{\text{electrostatic energy}}{(\text{voltage})^2} \propto L, \quad (2.27)$$

and is independent of assumptions regarding field strength. The calculated capacitance per square nanometer of a parallel plate vacuum capacitor with a 1 nm separation is $\sim 9 \times 10^{-21}$ F; note, however, that electron tunneling will lead to substantial conduction through so thin an insulating layer.

In electromechanical systems dominated by electrostatic forces,

$$\text{electrostatic power} \propto \text{electrostatic force} \cdot \text{speed} \propto L^2, \quad (2.28)$$

and

$$\text{electrostatic power density} \propto \frac{\text{electrostatic power}}{\text{volume}} \propto L^{-1}. \quad (2.29)$$

These scaling laws are identical to those for mechanical power and power density. Like them, they suggest high power densities for small devices (see Chapter 11).

The power density associated with resistive losses scales differently, given the above current density:

$$\text{resistive power density} \propto (\text{current density})^2 = \text{constant}. \quad (2.30)$$

The current density needed to power an electrostatic motor, however, scales differently from that derived from the above constant-field scaling arguments. In an electrostatic motor, surfaces are charged and discharged with a certain frequency, hence

$$\text{motor current density} \propto \frac{\text{charge}}{\text{area}} \cdot \text{frequency} \propto \text{field} \cdot \text{frequency} \propto L^{-1}, \quad (2.31)$$

and the resistive power losses climb sharply with decreasing scale:

$$\text{motor resistive power density} \propto (\text{motor current density})^2 = L^{-2}. \quad (2.32)$$

Accordingly, the efficiency of electrostatic motors will decrease with decreasing scale. The lack of long conducting paths (as in electromagnets) makes resistive losses smaller to begin with, however, and a detailed examination (see Chapter 11) shows that efficiencies remain high in absolute terms even for motors on a scale of tens of nanometers. The above relationships show that electromechanical systems cannot be scaled in the simple manner suggested for purely mechanical systems, even in the classical continuum approximation.

The scaling of fields in electromagnets is far less attractive for small-scale systems, since

$$\text{magnetic field} \propto \frac{\text{current}}{\text{distance}} \propto L. \quad (2.33)$$

At a distance of 1 nm from a conductor carrying 10 nA, the field strength is 2×10^{-6} T. The associated forces are highly scale-dependent,

$$\text{magnetic force} \propto \text{area} \cdot (\text{magnetic field})^2 \propto L^4, \quad (2.34)$$

and are minute in small scale systems: two parallel, 1 nm long segments of conductor, separated by 1 nm and carrying 10 nA exert a force on one another of only 2×10^{-23} N—a force 14 orders of magnitude too small to break a covalent bond and 11 orders of magnitude smaller than the characteristic electrostatic force calculated above. Magnetic forces between nanoscale current elements will generally be negligible. Magnetic fields associated with magnetic materials, in contrast, are scale-independent; accordingly, the associated forces, energies, and so forth follow the scaling laws described for constant-field electrostatic systems. Nanoscale current elements interacting with fixed magnetic fields can produce more significant (though still small) forces: a 1 nm long segment of conductor carrying a 10 nA current will experience a force of up to 10^{-17} N when immersed in a 1 T field.

The magnetic field energy associated with nanoscale current elements is very small:

$$\text{magnetic energy} \propto \text{volume} \cdot (\text{magnetic field})^2 \propto L^5. \quad (2.35)$$

The scaling of inductance can be derived from the above, but is independent of assumptions regarding the scaling of currents and magnetic field strengths:

$$\text{inductance} \propto \frac{\text{magnetic energy}}{(\text{current})^2} \propto L. \quad (2.36)$$

The inductance per nanometer of length for a fictitious solenoid with a 1 nm^2 cross sectional area and one turn per nm of length would be $\sim 10^{-15} \text{ h}$.

2.3.4. Magnitudes and scaling: time-varying systems

In systems with time-varying currents and fields, skin depth effects increase resistance at high frequencies; these effects complicate scaling relationships and are ignored here. The following simplified relationships are included chiefly to illustrate trends and magnitudes that *preclude* the scaling of classical AC circuits into the nanometer size regime.

For LR circuits,

$$\text{inductive time constant} \propto \frac{\text{inductance}}{\text{resistance}} = L^2. \quad (2.37)$$

Combining the characteristic 17Ω resistance and 10^{-15} h inductance calculated above yields an LR time constant of $\sim 6 \times 10^{-17} \text{ s}$. This time constant is unphysical: it is, for example, short compared to the electron relaxation time in a typical metal at room temperature ($\sim 10^{-14} \text{ s}$). Current will actually decay more slowly because of electron inertia and finite electron relaxation time.

Within the approximation of scale-independent resistivity,

$$\text{capacitative time constant} \propto \text{resistance} \cdot \text{capacitance} = \text{constant}. \quad (2.38)$$

The time required for a capacitor to discharge through a resistor in a pure RC circuit is thus be scale-independent; the scale-dependence of the LR time constant, however, can change a structure with fixed proportions from a nearly pure RC circuit (if built on a small scale) to a nearly pure LR circuit (if built on a large scale). The nanometer-scale RC time constant is $(17 \Omega) \times (9 \times 10^{-21} \text{ F}) \approx 1.5 \times 10^{-19} \text{ s}$, which is again unphysical owing to the neglected effects of electron inertia and relaxation time.

The LC time constant defines an oscillation frequency

$$\text{oscillation frequency} \propto \sqrt{\frac{1}{\text{inductance} \cdot \text{capacitance}}} \propto L^{-1}. \quad (2.39)$$

The characteristic inductance and capacitance calculated above would yield an LC circuit with an angular frequency of $\sim 3 \times 10^{17}$ rad/s. Alternatively, in structures such as waveguides,

$$\text{oscillation frequency} \propto \frac{\text{wave speed}}{\text{length}} \propto L^{-1}. \quad (2.40)$$

To propagate in a hypothetical waveguide 1 nm in diameter, an electromagnetic wave would require a frequency of $\sim 9 \times 10^{17}$ rad/s or more. Even the lower of the two frequencies just mentioned corresponds to quanta with an energy of $\sim 3 \times 10^{-17}$ J, that is, to photons in the x-ray range with energies of ~ 200 eV. These frequencies and energies are inconsistent with physical circuits and waveguides (metals are transparent to x-rays; electrons are stripped from molecules at energies far below 200 eV, etc.), hence quantum effects force the classical scaling laws to fail at sizes well above the nanometer range.

Scale also affects the quality of an oscillator:

$$Q \propto \text{oscillation frequency} \frac{\text{inductance}}{\text{resistance}} \propto L. \quad (2.41)$$

Since Q is a measure of the characteristic oscillation time divided by the characteristic damping time, small AC circuits will be heavily damped unless non-classical effects intervene.

Where they consider electromagnetic systems at all, the following chapters examine systems with nearly steady-state currents and fields, and time-varying systems characterized by RC behavior. High-frequency quantum electronic devices will undoubtedly be of great importance in nanotechnology, but are beyond the scope of the present work.

2.4. Scaling of classical thermal systems

2.4.1. Basic assumptions

For thermal systems, the classical continuum model assumes the scale-invariance of volumetric heat capacities and thermal conductivities. Since heat flows in systems will

typically be consequences of other physical processes, no independent assumptions are made regarding the scaling of thermal gradients or fluxes.

2.4.2. Major corrections

Classical, diffusive models for heat flow in solids can break down in a number of ways. On sufficiently small scales (which can be macroscopic for crystals at low temperatures) thermal energy is transferred ballistically by phonons whose mean free path would, in the absence of interfaces, exceed the dimensions of the structure in question. In the ballistic transport regime, interfacial properties analogous to optical reflectivity and emissivity become significant. Radiative transport of heat will be modified when the separation of surfaces becomes small compared to the characteristic wavelength of blackbody radiation, owing to coupling of non-radiative electromagnetic modes in the surfaces. In gases, separation of surfaces by less than a mean free path again modifies conductivity. The following will assume classical thermal diffusion, which should be a good approximation for liquids and for solids of low thermal conductivity, even on scales approaching the nanometer range.

2.4.3. Magnitudes and scaling

With a scale-independent volumetric heat capacity,

$$\text{heat capacity} \propto \text{volume} \propto L^3. \quad (2.42)$$

A cubic nanometer volume of a material with a not-atypical volumetric heat capacity of $10^6 \text{ J/m}^3 \cdot \text{K}$ will have a heat capacity of 10^{-21} J/K .

Thermal conductance scales like electrical conductance, with

$$\text{thermal conductance} \propto \frac{\text{area}}{\text{length}} \propto L, \quad (2.43)$$

and a cubic nanometer of material with a not-atypical thermal conductivity of $10 \text{ W/m} \cdot \text{K}$ will have a thermal conductance of 10^{-8} W/K .

Characteristic times for thermal equilibration follow from the above relationships, yielding

$$\text{thermal time constant} \propto \frac{\text{heat capacity}}{\text{thermal conductance}} \propto L^2. \quad (2.44)$$

For a cubic nanometer block separated from a heat sink by a thermal path with the above conductance, the calculated thermal time constant is $\sim 10^{-13}$ s, which is comparable to the acoustic transit time. (In an insulator, a calculated thermal time constant approaching the acoustic transit time would signal a breakdown of the diffusion model for transport of thermal energy and the need for a model accounting for ballistic transport; in the ballistic regime, time constants scale in proportion to L .)

The scaling relationships for frictional power dissipation can be used to derive a scaling law for the temperature elevation of a device in thermal contact with its environment,

$$\text{mechanical system temperature increment} \propto \frac{\text{frictional power}}{\text{thermal conductance}} \propto L. \quad (2.45)$$

This indicates that a nanomechanical system will be more nearly isothermal than an analogous system of macroscopic size.

2.5. Beyond the classical continuum model: atoms, statistics, and quantum mechanics

This chapter has examined the scaling laws implied by classical continuum models for mechanical, electromagnetic, and thermal systems, together with the magnitudes they imply for the physical parameters of nanometer scale systems. It has also considered some of the limits to the validity of classical models imposed by statistical mechanics, quantum mechanics, and the molecular structure of matter. Different classical models fail at different length scales, with the most dramatic failures appearing in AC electrical circuits.

The following chapters go beyond the classical, continuum model. The next chapter surveys models of molecular structure, dynamics, and statistical mechanics from a nanomechanical systems perspective. Chapters 5 and 6 examine the combined effects of quantum and statistical mechanics on the nanomechanical systems, first analyzing positional uncertainty in systems subject to a restoring force, and then analyzing the rates of transitions, errors, and damage in systems that can settle in alternative states. Chapter 7 examines mechanisms for energy dissipation. These chapters provide the foundation for discussions of specific nanomechanical systems. Later chapters examine not only nanomechanical systems, but a few specific electrical and fluid systems; where analysis of the latter must go beyond a classical, continuum approximation, the needed principles are discussed in context.

Chapter 3

Potential energy surfaces

3.1. The PES concept

The notion of a molecular potential energy surface (PES) is fundamental to practical models of molecular structure and dynamics. The PES describes the potential energy in terms of the molecular geometry defined by the set of nuclear positions. In the classical approximation, molecular motions result from forces defined by gradients of the PES, and equilibrium molecular structures correspond to minima of the PES. The term “surface” stems from a visualization in which a potential energy function in N dimensions is visualized as a surface in $N + 1$ dimensions, with energy corresponding to height. (When $N > 2$, the “visualization” is necessarily abstract.)

Modern physical theory seems adequate, in principle, to describe all mechanical, chemical, and electronic properties of structures made of ordinary matter at ordinary energies. In practice, this theory is, in its exact form, mathematically and computationally intractable in all but the simplest cases (e.g., the hydrogen atom). Physicists and chemists accordingly use a spectrum of approximations of varying accuracy and applicability; the PES concept itself is one such approximation. The following sections move from extremely accurate but impractical theories to less accurate but more useful approximations.

For nanomechanical engineering, empirical PES approaches like those described in Sections 3.3–3.5 provide the most directly useful approximations. The content of these sections is shaped by the demands of nanomechanical engineering, which sometimes differ from the demands of chemistry and molecular physics. These demands are discussed at greater length in a later section: although the next chapter discusses molecular dynamics, it returns to the topic of potential energy surfaces and ends by describing the differing

requirements of differing applications. Most of the topics introduced in this chapter are discussed at length in the literature; suggestions for further reading are appended.

3.2. Quantum theory and approximations

The design and analysis of nanomechanical systems requires models for the behavior of molecular scale systems. These requirements more closely resemble those of chemistry and materials science than (say) those of high energy physics. For perspective, it may be useful to view the hierarchy of approximations from the heights of modern physical theory (in which there is no molecular PES), then move into the domain of practical calculation.

3.2.1. Overview of quantum mechanics

In the late 1940s, Feynman, Schwinger, and Tomonaga each independently developed formulations of the theory of quantum electrodynamics (QED), which describes electromagnetic fields and electrons in a unified way. Where the mathematics of QED can be manipulated to yield precise predictions, as for the magnetic properties of free electrons and the spectrum of the hydrogen atom, its predictions have been confirmed to the last measurable detail. It correctly predicts that the g -value of the electron is 2.0023 rather than 2 as predicted by previous theory, and it correctly predicts the Lamb shift in the hydrogen spectrum (which shifts an energy level by less than one part in 10^6) to an accuracy of many decimal places. There is every reason to think that the theory would be equally precise in other areas, if it could be applied. In practice, owing to the difficulty of applying it, chemists and material scientists do not use QED. Its great contributions have been in high energy physics and in its use as a model for other physical theories, such as quantum chromodynamics.

Earlier, in 1931, Dirac had developed a fully relativistic quantum mechanics which predicted electron spin (with a g -value of 2) and provided a correct explanation for the splitting of certain spectral lines in hydrogen, reflecting shifts in energy levels by about one part in 10^4 . Relativistic effects are large in the inner electron shells of heavy elements, but these shells are so tightly bound to the nucleus that they are chemically inert; among the chemically-active outer electrons, effects remain on the order of one part in 10^4 . Since Dirac's theory is difficult to apply and the relativistic effects are small, it is not used in standard quantum mechanical calculations in chemistry and materials science. (Relativistic effects in heavy atoms do result in strong spin-orbit coupling, which affects

the rate of electronic transitions of chemical interest, such as the flipping of unpaired electron spins in free radicals. For typical molecules of interest, however, spin-orbit coupling does not significantly affect structure or dynamics before or after the transition.)

Earlier still, in 1926, Schrödinger had developed a formulation of non-relativistic quantum mechanics that remains the basis for all practical quantum chemistry calculations. Schrödinger described matter in terms of a wave equation (here shown in a mass-weighted form in the presence of a scalar potential)

$$\frac{\hbar^2}{2} \sum_i \frac{\partial^2}{\partial r_i^2} \psi(\mathbf{r}, t) + \mathcal{V}(\mathbf{r}, t) \psi(\mathbf{r}, t) = -i\hbar \frac{\partial}{\partial t} \psi(\mathbf{r}, t) \quad (3.1)$$

where a system of N particles with masses $m_0, m_1, m_2, \dots, m_{N-1}$ is described by the coordinate vector \mathbf{r} with components $r_0, r_1, r_2, \dots, r_{3(N-1)}$ equalling $x_0 m_0^{-1/2}, y_0 m_0^{-1/2}, z_0 m_0^{-1/2}, x_1 m_1^{-1/2}, y_1 m_1^{-1/2}, z_1 m_1^{-1/2}, x_2 m_2^{-1/2}, y_2 m_2^{-1/2}, z_2 m_2^{-1/2}, \dots, z_{(N-1)} m_{(N-1)}^{-1/2}$. This is a partial differential equation in a $3N$ -dimensional configuration space, and physically valid solutions, $\psi(\mathbf{r}, t)$, are subject to a set of boundary, continuity, normalization, and particle-exchange symmetry conditions. Molecular structure calculations seek solutions describing bound, time-invariant systems, but this problem cannot be solved exactly even for a molecule as simple as H_2 . Nonetheless, because no simpler theory gives an acceptable description of the quantum mechanical behavior of molecular matter, the Schrödinger equation has been the basis of a host of approximation schemes. These approximations represent steps backward from the point of view of fundamental physical theory, but steps forward from the point of view of understanding real physical systems.

In order to understand the general nature of these approximation schemes, it is necessary to say a bit more about the structure of the Schrödinger theory as applied to molecules. In calculations of the structure of isolated systems, the potential function is time-independent, and is simply the total Coulomb energy for the interaction of each pair of charged particles

$$\mathcal{V}(\mathbf{r}) = \sum_{i < j} \frac{q_i q_j}{4\pi\epsilon_0 d_{ij}} \quad (3.2)$$

where d_{ij} is the distance between particles i and j , and ϵ_0 is the permittivity of free space. The potential energy function is thus based on a picture of particles with precise positions in ordinary space. Since the probability density function

$$f_{r,i}(\mathbf{r}, t) = \psi^*(\mathbf{r}, t) \psi(\mathbf{r}, t) \quad (3.3)$$

is defined over the $3N$ dimensional configuration space, it yields not only the expected particle densities (and hence the charge density) at each point in physical space, but also the probability density associated with each spatial configuration of particles. Owing to electrostatic repulsion (and to particle-exchange symmetry conditions), particle positions are not independent, but correlated. This prevents the $3N$ dimensional problem from being reduced to a set of N coupled problems in three dimensions. (Note that the correlation of particle positions and motions is described by a wave function that is time independent, save for phase.)

Monte Carlo path-integral methods have recently been applied to yield approximate evaluations of exact $3N$ dimensional wave functions (Ceperley and Alder 1986). As is usual with Monte Carlo methods, the quality of the approximation can be increased without limit by increasing the expenditure of computational time, with the expected error scaling as $(\text{time})^{-1/2}$. These methods converge toward exact results (within the limitations of the Schrödinger model). Because they compute the full wave function, they can in principle be used to determine any physical property of the system being modeled. For a given level of accuracy, computational costs scale as N^2 , but feasible values of N (electrons and nuclei) are still small. In chemical calculations, these methods are not presently competitive with established molecular orbital methods, discussed below. Monte Carlo methods are still evolving, in part though the introduction of judicious physical approximations.

3.2.2. The Born–Oppenheimer PES

In chemistry and nanomechanical engineering, the full wave function gives more information than necessary; indeed, the wave function *per se* is of little interest. Approximations and partial solutions can accordingly be of great value. Because each particle adds three dimensions and substantial computational cost, it is useful to partition problems into simpler sub-problems wherever the resulting inaccuracies are not too severe.

Most computational techniques exploit the Born–Oppenheimer approximation, which treats the motion of electrons and nuclei separately. Even the lightest nucleus has 1836 times the inertia of an electron. The characteristic speeds and frequencies of nuclear motion are accordingly much lower than those of electronic motion. The *Born–*

Born–Oppenheimer approximation treats nuclei as motionless, and permits computation of the wave function and energy for a system of electrons in the presence of a fixed nuclear configuration with a fixed Coulomb potential. In this approximation, each nuclear configuration is associated with a unique electronic ground-state energy. This defines the ground-state Born–Oppenheimer potential energy function $\mathcal{V}(\mathbf{r})$, where the vector \mathbf{r} here specifies only nuclear coordinates and $\mathcal{V}(\mathbf{r})$ has no resemblance to the simple Coulomb potential. In analyzing nuclear motions using this potential function, electronic motions are implicitly assumed to adjust without a time lag. The Born–Oppenheimer approximation breaks down when nuclear motions are fast enough, for example in high-energy collisions, and when small changes in nuclear coordinates lead to large changes in electron distribution, for example in nearly-degenerate states where Jahn–Teller effects become important. Under ordinary conditions in which nuclear kinetic energies are less than electronic kinetic energies and nuclear speeds are accordingly much smaller, the Born–Oppenheimer approximation usually gives an excellent account of molecular behavior.

In most nanomechanical systems, as in most chemical reactions, electron wave functions will transform smoothly with mechanical (i.e., nuclear) motion. Under these conditions, there are no abrupt changes in electron distribution and energy, that is, no electronic transitions. In the absence of electronic transitions, and within the Born–Oppenheimer approximation, molecular dynamics depends only on the Born–Oppenheimer potential. If one knows this potential, or has an adequate approximation to it, then one can analyze molecular dynamics without reference to electronic behavior. (When electronic transitions occur, they place the system on another Born–Oppenheimer potential surface.) The Born–Oppenheimer potential can be used in both classical and quantum models of dynamics.

3.2.3. *Molecular orbital methods*

Practical calculations require further approximations. The most popular approaches lead to a family of techniques known as molecular orbital methods; these make different approximations, yielding computations of widely varying accuracy and cost.

Molecular orbital methods begin with the *independent electron approximation*, in which each electron is treated as moving in the electrostatic potential that would result from the time-average distribution of the other electrons; this neglects *electron correlation* of the sort described above. Together with the Born–Oppenheimer approximation, this reduces the problems of solving a single Schrödinger equation in $3N$ dimensional

configuration space to that of solving N coupled Schrödinger equations in three-dimensional space, where N is now the number of electrons. The coupling is treated by the self-consistent field method: conceptually, each newly-calculated electron wave function results in a new charge distribution, which changes the potential experienced by the other electrons, demanding a new calculation; the resulting iterative process converges toward a set of wave functions, each consistent with the electrostatic potential of the rest (in practice, more efficient iterative procedures are employed). Each single-electron wave function corresponds to a molecular orbital with a characteristic charge distribution and energy. The wave function for the system is then described in terms of determinants based on these orbitals.

In a typical quantum chemistry program, the wave functions are approximated as weighted sums of simple (e.g., “gaussian type”) basis functions, and the programs vary the weighting of each basis function to form a wave function of minimum energy. Increasing the number of basis functions provides increased flexibility in shaping the wave function, enabling construction of more accurate molecular orbitals; the number of basis functions must equal or (far better) exceed the the number of electrons in the molecule. Computations on large molecules are prohibitively expensive and time-consuming, however, because the cost of computing the energy of a single molecular geometry by these methods varies roughly as the fourth power of the number of basis functions, although more recent methods improve on this scaling law.

Molecular orbital computations can be made more accurate by taking electron correlation into account using *configuration interaction* (CI) and related perturbation theory methods. In CI methods, a “configuration” refers not to a spatial configuration of electrons (in the sense of a configuration space) but to a particular pattern of orbital occupancy: that is, to a particular overall wave function and energy. Mixing the ground state wave functions with excited-state wave functions, each calculated with the independent-electron approximation, is equivalent to permitting correlated electron motions and hence reduces the errors of the independent-electron approximation. The set of possible configurations is, however, combinatorially large; practical calculations typically require choosing the most important configurations from a set of millions. The effect of electron correlation on energy varies widely from system to system; it is the source of attractive forces between neutral, non-polar molecules and is particularly important in calculating the energy of bonds far from equilibrium geometries (i.e., during the transition state of a

chemical reaction). Configuration interaction calculations often converge slowly and expensively.

(Calculations using different basis sets or corrections for electron correlation are described as being at different "levels of theory." They might better be described as different "levels of approximation," with high levels converging on the Born–Oppenheimer approximation to the Schrödinger equation.)

The techniques just described (termed *ab initio* methods) make approximations in determining wave functions, but they use no parameters other than fundamental physical constants. Semiempirical molecular orbital methods (such as MNDO, AM1, and PM3) make further approximations in determining wave functions, but they compensate by introducing parameters for different atom-types to fit results to experimental data. These methods typically neglect certain integrals that arise in more complete calculations and treat the compact, high-energy orbitals associated with inner-shell electrons as fixed. Semiempirical methods have far lower computational cost (e.g., by a factor of 50 to 500 for a molecule as small as propane), but vary in their accuracy in a complex manner (e.g., having special problems with hydrogen bonds, or with boron) (Clark 1985). There is much room for cleverness in semiempirical methods, which continue to improve (Stewart 1990).

Molecular orbital methods can be used to calculate parameters such as charge distribution, polarizability, and ionization energies, but their chief application in the present context is the determination of energy as a function of molecular configuration: the Born–Oppenheimer potential. This information suffices to determine stable molecular structures, which correspond to local or global energy minima, but finding these minima requires energy evaluations (often augmented with direct calculations of energy gradients) at many points in configuration space. More complex structures require evaluation at more points, hence computational cost scales with size even more adversely than is suggested above.

The challenge of applying molecular orbital methods to chemical reactions is suggested by the great difference between the magnitude of the total molecular binding energy calculated by these methods (the energy required to separate a molecule into isolated electrons and nuclei) and the energies of chemical interest. A typical total binding energy for a small molecule is on the order of 10^{-15} J, and the energy of a single chemical bond is on the order of 10^{-18} J, but the energy required to rearrange bonds in a typical chemical reaction is less than 10^{-19} J and solution-phase chemical processes are often

critically dependent on energy differences of 10^{-20} J or less. Thus, chemists often require that errors be less than 10^{-4} or 10^{-5} of the total energy, although large errors in total energy can be tolerated so long as energy *differentials* between similar structures are computed accurately enough.

This accuracy can be achieved (for small enough systems) using molecular orbital methods, and such calculations will be of value in the design of reactive devices used in molecular manufacturing systems. Nonetheless, limitations associated with the expense, accuracy, and scalability of molecular orbital methods motivate the continuing popularity of empirically-based models of molecules and molecular transitions.

3.3. Molecular mechanics

3.3.1. *The molecular mechanics approach*

An alternative to performing quantum mechanical calculations of electronic structure is to approximate the Born–Oppenheimer potential directly, in terms of the molecular geometry alone. Molecular mechanics methods (among others) take this approach, which is well suited to most problems of nanomechanical design and modeling.

3.3.1.1. *Overview*

Organic chemists traditionally represent molecular structures with ball-and-stick models in which each ball represents an atom and each stick represents a bond. In quantum mechanical calculations, there is no *a priori* concept of a “chemical bond,” and the *behavior* of bonding emerges afresh in each calculation. Molecular mechanics calculations, in contrast, begin with the concept of bonds and then use them as a basis for modeling the molecular potential energy surface. Molecular mechanics thus builds directly on a familiar and useful visualization of molecular structure.

Molecular mechanics programs calculate equilibrium structures by finding configurations that correspond to local minima of the PES. The PES itself is developed by choosing functional forms and parameters that yield the best achievable fit between predicted and experimentally-measured structures, heats of formation, and vibrational frequencies (experimental structural parameters are typically derived from x-ray crystallography, microwave spectroscopy, and gas-phase electron diffraction; heats of formation from calorimetry; and vibrational frequencies from infrared spectroscopy). Burkert and Allinger (Burkert and Allinger 1982) provide data on the typical accuracy of calculations per-

formed on hydrocarbons using the MM2 model: estimated bond lengths typically match experimental values to within a few times 10^{-13} m ($\sim 0.1\%$), estimated bond angles typically match within about 0.01 rad ($\approx 0.6^\circ$, or $\approx 0.5\%$), and energies within a few mJ. These values are comparable in accuracy to the experimental data itself.

Burkert and Allinger also compare the costs of molecular mechanics calculations to those of *ab initio* quantum mechanical calculations: for a small molecule (propylamine) the cost favors molecular mechanics by a factor of 10^3 . For larger molecules the difference becomes more pronounced: costs for molecular mechanics methods vary at a rate proportional to something between the second and third power of the number of atoms, rather than the fourth power characteristic of molecular orbital methods. As of 1990, molecular mechanics calculations on ten-thousand atom systems have become routine, and systems in the ten-to-hundred-atom range can readily be modeled using personal computers.

3.3.1.2. Limitations and applications

Molecular mechanics systems have, however, been successfully applied to only a narrow range of molecular structures, and most systems are limited to configurations not too far from equilibrium. They use energy functions and parameters tailored to specific, local arrangements of atoms. Fortunately for nanomechanical engineering efforts, the most advanced molecular mechanics methods have been developed to model a class of structures that includes those most suitable for use as nanomechanical components—that is, structures built largely of carbon (typically augmented with one or more of the elements H, B, N, O, F, Si, P, S, and Cl) and held together by strong, directional, covalent bonds. Structures of this sort are the focus of organic chemistry; a subset of these structures comprises most of the molecular machinery of living systems.

Within this set, limitations remain. Aside from the small inaccuracies found in all structures, standard molecular mechanics programs fail dramatically for systems with unusual bonding patterns, and for systems in configurations intermediate between structures with different patterns of bonding. They can model stable structures, even when strained, but cannot model chemical transformations or systems on the verge of such transformations. Computational results must be examined with an eye for such invalidating conditions.

Molecular mechanics programs differ in their intended applications, with some adjusted for speed, and others for accuracy, with some intended for broad classes of

organic structures, and others specialized for biomolecules. In general, systems intended for large biomolecules also place a premium on speed and are not intended for structures with large strain energies; popular examples are AMBER (Weiner, Kollman et al. 1984; Weiner, Kollman et al. 1986) and CHARMM (Brooks, Bruccoleri et al. 1983). A widely-used molecular mechanics system intended for broad classes of structures (including those with large strain energies) is MM2 (Allinger 1977) and the related MMP2, developed chiefly by Norman Allinger and coworkers. Although evolving (and in the process of being superseded by the similar, but more complex and accurate MM3 (Allinger, Yuh et al. 1989)), its basic form has been stable for many years and gives good results for a broad range of structures. After evaluation and with some caveats, the MM2 model will be used as an engineering approximation in much of the present work.

Molecular mechanics programs treat the total potential energy as a sum of terms accounting (*inter alia*) for bond stretching, bending, and torsion, and for van der Waals and electrostatic interactions between non-bonded atoms. To develop a physical intuition for the mechanical behavior of molecular systems, one needs a feel for the nature and magnitude of these forces. Although physical intuition cannot substitute for a more precise analysis, it can be of great value in shaping designs and choosing what to analyze. The relationships and parameters that follow provide a good basis for physical understanding and also enable pocket-calculator estimates of the stiffness, force, and energy associated with simple molecular deformations. In the exploratory phase of design, one must test and modify tentative concepts. This requires quick, accessible estimates of energies, forces, and stiffnesses, and of how they vary. For this, experience shows the value of graphs like those in Figures 3.3–3.8.

3.3.2. The MM2 model

In molecular mechanics, bonds are characterized by the atom-types they join, and these atom-types can depend both on atomic number and bonding environment: for example, carbon atoms that participate only in single bonds are of a different type from those that participate in double or triple bonds. Table 3.1 lists some of the atom-types used by the MM2 program. Since the kcal/mole and kJ/mole units common in chemistry hinder calculations of force, stiffness, acceleration and the like, all energies here (and elsewhere) are given in decimal multiples of SI units and refer to single instances of objects and interactions, rather than to gram-moles of instances.

Table 3.1. Some atom types used in the MM2 program (Allinger 1986) along with (rationalized) MM2 van der Waals parameters. Note that the van der Waals effects of lone pair electrons associated with nitrogen and oxygen atoms are modeled by treating each lone pair of electrons as a pseudo-atom. The full set of atom types described by MM2 is roughly twice as large as this list.

MM2 code	Symbol	Type	ϵ_{vdw} (maJ)	r_{vdw} (10^{-10} m)	Mass (amu)	Mass (10^{-27} kg)
1	C	sp^3	0.357	1.90	12.000	19.925
2	C	sp^2 alkene	0.357	1.94	12.000	19.925
3	C	sp^2 carbonyl	0.357	1.94	12.000	19.925
4	C	sp acetylene	0.357	1.94	12.000	19.925
5	H	hydrogen	0.382*	1.50*	1.008	1.674
6	O	C–O–H, C–O–C	0.406	1.74	15.999	26.565
7	O	carbonyl	0.536	1.74	15.999	26.565
8	N	sp^3	0.447	1.82	14.003	23.251
11	F	fluoride	0.634	1.65	18.998	31.545
12	Cl	chloride	1.950	2.03	34.969	58.064
13	Br	bromide	2.599	2.18	78.918	131.038
14	I	iodide	3.444	2.32	126.900	210.709
15	S	sulfide	1.641	2.11	31.972	53.087
19	Si	silane	1.137	2.25	27.977	46.454
20	LP	lone pair	0.130	1.20	0.000	0.000
21	H	alcohol	0.292	1.20	1.008	1.674
22	C	cyclopropane	0.357	1.90	12.000	19.925
25	P	phosphine	1.365	2.18	30.994	51.464

* see Section 3.3.2.5 for correction factors

3.3.2.1. Bond stretching

Bonds resist stretching and compression, tending toward an equilibrium length. The MM2 expression for the potential energy of bond stretching, \mathcal{V}_s (J), includes a cubic term to account for anharmonicity

$$\mathcal{V}_s = \frac{1}{2} k_s (\ell - \ell_0)^2 \left[1 - k_{\text{cubic}} (\ell - \ell_0) \right] \quad (3.4)$$

In this expression, ℓ_0 (m) is the equilibrium bond length, ℓ is the current bond length, k_s (N/m) is the stretching stiffness, and k_{cubic} (m^{-1}) is the magnitude of the cubic term, for which the MM2 model uses an invariant value of $2 \times 10^{10} \text{ m}^{-1}$. (All MM2 expressions and parameters have here been translated into SI units.) Table 3.2 lists some MM2 bond stretching parameters.

3.3.2.2. Bond-angle bending

Two bonds to a shared atom define a bond angle (four bonds to a shared atom define six such angles). In those systems for which MM2 provides a good approximation, bond angles can be described as having preferred values with displacements countered by restoring forces; this works for trivalent but not pentavalent phosphorus (Burkert and Allinger 1982). In its description of the potential energy of bond bending, \mathcal{V}_θ (J), MM2 includes a sextic term

$$\mathcal{V}_\theta = \frac{1}{2} k_\theta (\theta - \theta_0)^2 \left[1 + k_{\text{sextic}} (\theta - \theta_0)^4 \right] \quad (3.5)$$

In this expression, θ_0 (rad) is the equilibrium bond angle, θ is the current bond angle, k_θ (J/rad^2) is the angular spring constant, and k_{sextic} is the magnitude of the sextic bending term (taken to be 0.754 rad^{-4} for all bonds). Table 3.3 lists some MM2 bond bending parameters. Note that modeling of lone pairs as pseudo-atoms with bond-bending interactions will incorrectly prohibit the inversion of amine groups.

To compare bond bending to bond stretching, one needs a measure of stiffness in terms of spatial (not angular) displacement. To first order, bond bending can be described by a stiffness, $k_{s\perp}$, characterizing displacements perpendicular to a chosen bond, in the plane of the associated angle, with all other bond directions fixed:

$$k_{s\perp} = \frac{k_\theta}{\ell_0^2} \quad (3.6)$$

Table 3.2. MM2 bond stretching parameters for some common bond types (the full set is roughly six times larger) (Allinger 1986)

Bond type (MM2 codes)	k_s (N/m)	ℓ_0 (10^{-10} m)	Bond type
1-5	460	1.113	C-H
1-1	440	1.523	C-C
2-2	960	1.337	C=C
4-4	1560	1.212	C \equiv C
22-22	440	1.501	C-C (cyclopropane)
1-6	536	1.402	C-O
1-8	510	1.438	C-N
3-7	1080	1.208	C=O
1-11	510	1.392	C-F
1-12	323	1.795	C-Cl
1-13	230	1.949	C-Br
1-14	220	2.149	C-I
8-20	610	0.600	N-LP
8-8	560	1.381	N-N
6-20	460	0.600	O-LP
6-21	460	0.942	O-H
6-6	781	1.470	O-O
1-19	297	1.880	C-Si
1-25	291	1.856	C-P
1-15	321	1.815	C-S
19-19	185	2.332	Si-Si

Table 3.3. MM2 bond bending parameters for some common bond types (the full set is roughly ten times larger) (Allinger 1986).

Angle type (MM2 codes)	k_{θ} (aJ/rad ²)	θ_0 (deg)	θ_0 (rad)	Angle type
1-1-1	0.450	109.47	1.911	C-C-C
1-1-5	0.360	109.39	1.909	C-C-H
5-1-5	0.320	109.40	1.909	H-C-H
1-1-11	0.650	109.50	1.911	C-C-F
11-1-11	1.070	107.10	1.869	F-C-F
1-2-1	0.450	117.20	2.046	C-C(sp ²)-C
2-1-2	0.450	109.47	1.911	C(sp ²)-C-C(sp ²)
1-2-2	0.550	121.40	2.119	C-C=C
2-2-5	0.360	120.00	2.094	C=C-H
2-2-2	0.430	120.00	2.094	C=C-C(sp ²)
1-4-4	0.200*	180.00	3.142	C-C≡C
1-3-7	0.460	122.50	2.138	C-C=O
1-6-1	0.770	106.80	1.864	C-O-C
1-8-1	0.630	107.70	1.880	C-N-C
1-1-6	0.700	107.50	1.876	C-C-O
1-1-8	0.570	109.47	1.911	C-C-N
1-6-20	0.350	105.16	1.835	C-O-LP
1-8-20	0.500	109.20	1.906	C-N-LP
20-6-20	0.240	131.00	2.286	LP-O-LP
19-19-19	0.350	111.30	1.943	Si-Si-Si
19-1-19	0.400	115.50	2.016	Si-C-Si
1-19-1	0.480	110.80	1.934	C-Si-C
12-1-12	1.080	111.70	1.950	Cl-C-Cl
1-1-15	0.550	109.00	1.902	C-C-S
1-15-1	0.720	96.30	1.902	C-S-C

* (Allinger and Pathiaseril 1987)

Bond bending stiffnesses are substantially less than bond stretching stiffnesses. For example, in bending of a C–C–C bond angle, $k_{s\perp} = 19.4$ N/m, or about 1/20 the stretching stiffness of a C–C bond (all carbons sp^3). In an sp^3 tetrahedral geometry, bending one bond with respect to the other three is characterized by a stiffness 1.5 times greater than this.

Systems containing sp^2 atoms have a preferred plane and are subject to a restoring force for bending of bonds away from this plane. Table 3.4 lists some force constants for out-of-plane bending. Figure 3.2 illustrates the definition of the in-plane and out-of-plane bending angles; the three atoms surrounding an sp^2 carbon (e.g., in formaldehyde) define three such angles.

Table 3.4. Some MM2 parameters for out-of-plane bending.

Angle type MM2 code ranges	k_θ (aJ/rad ²)
2–(1...9)	0.050
3–(1...9)	0.800
9–(1...4)	0.050
9–(6...9)	0.050

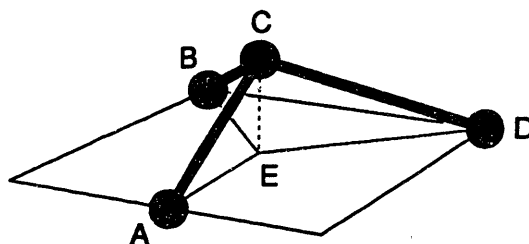


Figure 3.1. Out-of-plane bending geometry in the MM2 model. In calculating angle-bending energies centered on sp^2 atoms (such as C in the above diagram), each angle (such as ACB) is describe in terms of an in-plane component (the angle AEB) and an out-of-plane component (the angle CDE); the line CE is perpendicular to the ABD plane.

3.3.2.3. Bond torsion

Rotation about bonds is not “free rotation” (in the chemist’s technical sense), but can often permit a freely-turning motion in a mechanical engineer’s sense. The MM2 potential describes the variation in energy associated with rotation about a bond in terms of a sum of four-atom torsional potentials, where the torsion angle is defined as illustrated in Figure 3.2. The six hydrogen atoms in ethane define nine such angles. Each four-atom torsional potential takes the form

$$\mathcal{V}_{\omega} = \frac{1}{2} [V_1(1 + \cos \omega) + V_2(1 - \cos 2\omega) + V_3(1 + \cos 3\omega)] \quad (3.7)$$

Table 3.5. MM2 parameters for bond torsion (the full set is roughly 80 times larger).

Torsion type (MM2 codes)	V_1 (maJ)	V_2 (maJ)	V_3 (maJ)	Torsion type
1-1-1-1	1.39	1.88	0.65	C-C-C-C
1-1-1-5	0.00	0.00	1.85	C-C-C-H
5-1-1-5	0.00	0.00	1.65	H-C-C-H
1-1-1-11	0.00	-0.60	6.46	C-C-C-F
1-2-2-1	-0.69	69.47	0.00	C-C=C-C
2-2-2-2	-6.46	55.58	0.00	C(sp ²)-C=C-C(sp ²)
5-2-2-5	0.00	104.21	0.00	H-C=C-H

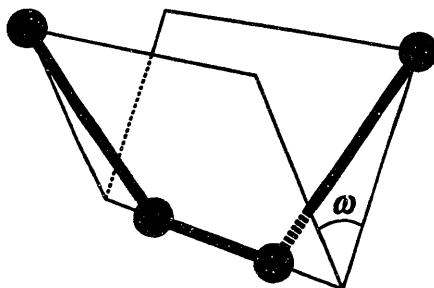


Figure 3.2. Torsion geometry

As with bond bending, torsional restoring forces can (to first order, for configurations at the bottom of a potential well) be described by a stiffness associated with small displacements tangent to the torsional motion. The maximum stiffness associated with displacement of a terminal carbon atom as described by the MM2 C–C–C–C torsion parameters is about 0.36 N/m (considering just four sp^3 carbons). Ordinary torsional stiffness terms (and associated forces and energies) are thus about 1/50 the magnitude of bond bending stiffness terms, or about 1/1000 the magnitude of bond stretching stiffnesses. In nanomechanical structures designed for rigidity, torsional contributions to stiffness will seldom be significant (any structure in which they *were* a substantial source of rigidity would not be very rigid). Double bonds have substantial torsional stiffness, and are an exception to this generalization.

3.3.2.4. *Electrostatics*

Structures containing atoms of substantially different electronegativity (e.g., most non-hydrocarbons) have significant dipole moments. MM2 includes routines for associating fractional charges q with atoms of different types in a given molecular context; the potential energy is then the electrostatic potential

$$V_{\text{charge}} = \frac{q_1 q_2}{4\pi\epsilon_0 d} \quad (3.8)$$

summed over all pairs of atoms. In typical structures, large regions are nearly electroneutral and charge separation occurs chiefly on the scale of individual bond dipoles. An example of a strongly polarized bond is C–F, where the dipole moment is about 4.7×10^{-30} C·m and the fractional charge on the F atom is about $-0.2 e$. Figures 3.5 and 3.6 include a comparison of the magnitude of dipole-dipole and non-bonded interactions for C–F, C–Cl, and C–Br groups.

3.3.2.5. *Non-bonded interactions*

Associated with each atom type (see Table 3.1) is a set of van der Waals parameters describing the attractive and repulsive forces experienced by pairs of uncharged, non-bonded atoms. Physicists typically apply the term “van der Waals force” to the attractive component alone, that is, to the London dispersion force (along with a variety of polar interactions between molecules). This book follows a common usage in computational chemistry, treating polar interactions separately but including the overlap repulsion (a.k.a.

exchange force, hard-core repulsion, or steric repulsion) as part of the van der Waals potential. The MM2 model describes these overall van der Waals interactions with a Buckingham (or *exp-6*) potential

$$\mathcal{V}_{\text{vdw}} = \epsilon_{\text{vdw}} \left[2.48 \times 10^5 e^{-12.5 \frac{d}{d_{\text{vdw0}}}} - 1.924 \left(\frac{d}{d_{\text{vdw0}}} \right)^{-6} \right] \quad (3.9)$$

where d is the separation between the atoms, ϵ_{vdw} for the interaction between atoms 1 and 2 equals $(\epsilon_{\text{vdw1}} + \epsilon_{\text{vdw2}})/2$, and d_{vdw0} equals $r_{\text{vdw1}} + r_{\text{vdw2}}$. This function has a minimum of $-\epsilon_{\text{vdw}}$ at $d = d_{\text{vdw0}}$. The forces between atoms bonded to a common atom (i.e., 1–3 interactions) are included not in the van der Waals energy, but in the bond-bending energy. All other pairs are included in the sum. (The above expression and the parameters in Table 3.1 been converted to SI units and rationalized, making ϵ_{vdw} equal the binding energy at $d = d_{\text{vdw0}}$.)

Equilibrium separations between atoms in different molecules (and larger structures) will typically be smaller than the pairwise equilibrium separations defined by the van der Waals parameters. The short-range nature of the exponential component allows only nearest neighbors to make a significant contribution to the repulsive side of the balance, but the d^{-6} forces have a longer range, allowing many atoms to make a significant contribution on the attractive side (see Sect. 3.5). Equilibrium separations shrink accordingly.

Lack of nucleocentric spherical symmetry in atoms leads to certain complications, however. Calculation of van der Waals interactions for sp^3 nitrogen and oxygen atoms in the MM2 model includes the effects of associated lone-pair pseudo-atoms. The lack of core electrons in hydrogen atoms leads to an unusually large shift of total charge density into the bond; for van der Waals calculations this is modeled by shifting the effective position of the hydrogen atom inward to 0.915 of the full bond length. (Also, describing pairwise van der Waals energies using the above expressions for ϵ_{vdw} and d_{vdw0} is a rough approximation; for the hydrogen– sp^3 -carbon interaction, d_{vdw0} is corrected to 0.982 of the above sum, and ϵ_{vdw} is corrected to 1.011 of the above mean.)

3.3.2.6. Complications and conjugated systems

The MM2 model includes a number of small corrections and special cases. For example, when bond angles are reduced, equilibrium bond lengths increase. This is described in the MM2 potential by a stretch-bend interaction term

Table 3.6. Stretch-bend parameters

Angle type (see below)	$k_{s\theta}$ (10^{-9} N/rad)
X-F-Y	1.20
X-S-Y	2.50
X-F-H	0.90
X-S-H	-4.00

X, Y = first or second row atoms;
F = first row atom; S = second row atom; H = hydrogen.

$$V_{s\theta} = k_{s\theta}(\theta - \theta_0) \left[(\ell - \ell_0)_a + (\ell - \ell_0)_b \right] \quad (3.10)$$

for a C-C-C system, a 0.1 rad reduction in bond angle yields a modest 1.4×10^{-13} m change in the equilibrium lengths of the two associated bonds. A larger (but static) correction to equilibrium bond length can result from the presence of adjacent bonded atoms of differing electronegativity. The extreme case is F, which shortens an adjacent C-C bond by 2.2×10^{-12} m. The MM2 model also includes special-case parameters for three and four membered rings, and for hydrogen bond interactions.

A major complication arises in conjugated double bond systems, in which the model of bonds as entities with properties depending only on their near neighbors breaks down. Where single and double bonds alternate along a chain or ring, delocalization of pi electrons can greatly affect the potential energy function; benzene is a classic example. The MMP2 program (an extension of MM2) deals with conjugated systems by performing a minimal quantum mechanical analysis of the participating pi electrons; it uses the results to estimate the magnitude of fractional bonding between pairs of atoms, and then adjusts the force field parameters (bond lengths, stiffnesses, torsional energies, etc.) accordingly.

3.3.2.7. Notes on MM2 in light of MM3

Known shortcomings of MM2 are discussed in a set of papers on the preliminary version of the MM3 force field (Allinger, Yuh et al. 1989; Lii and Allinger 1989a; Lii

and Allinger 1989b). Aside from those discussed above, the greatest shortcoming of MM2 is its inaccurate predictions of molecular vibrational frequencies; accuracy here was sacrificed to achieve greater accuracy in describing the energy and geometry of equilibrium configurations. Since frequencies depend on molecular stiffnesses, this failure is of significance to the design of molecular machines.

MM3 succeeds in fitting vibrational frequencies accurately while improving accuracy in other areas by virtue of a more complex functional form (e.g., a stretch-torsion interaction, a cubic bending term, a quartic stretching term) and additional parameters. Among the more basic parameters, the ratio of the MM3 to the MM2 values is as follows: For bond stretching, the ratios are C–C = 1.02, and C–H = 1.03; for bond-angle bending, the ratios are C–C–C = 1.49, C–C–H = 1.64, and H–C–H = 1.72. Since MM3 is a better model, it is clear that MM2 stiffness values (especially for the important angle-bending stiffnesses) are substantially lower than those of real molecules. Since greater stiffness is almost always an asset in molecular machinery, this defect in the MM2 will, in most instances, be conservative: it will result in more false-negative evaluations of device feasibility than it will false-positives. The chief exception to this rule will be in structures where angle-bending strain relieves bond stretching: the low angle-bending stiffness of the MM2 model may then result in a false-positive assessment of the stability of a stretched bond.

From a nanomechanical perspective, the other major modification in MM3 (again, indicating a shortcoming in MM2) is its treatment of nonbonded interactions. MM3 uses an exp–6 potential, but with 12.0 in place of 12.5 in the exponent (see Eq. 3.9), and a smaller weighting on the exponential term. The atomic parameters developed thus far have larger atomic radii (by a factor of ~ 1.07) and smaller energy scale factors (by ~ 0.55). The net result is a softer interaction, with lower energies, forces, and stiffnesses in the deep repulsive regime, with typical differences on the order of tens of percent. This difference has mixed effects on nanomechanical systems. Softer interactions will decrease the stiffness of bearings and other devices dependent on nonbonded contacts, but the stiffness of these interactions is sensitive to load and hence is subject to design control. Softer, longer-range interactions in effect make surfaces smoother and will decrease several drag mechanisms found in bearings (Chapter 7). So long as a substantial margin of safety is provided in the design of stiff, non-bonded interfaces, this shortcoming of MM2 should only rarely result in false-positive assessments.

3.3.3. Energy, force, and stiffness under large loads

Nanomechanical engineering and conventional chemistry place different demands on potential energy functions and emphasize different characteristics. Some of these differences are discussed in more detail in Chapter 4, which compares the accuracy demanded from energy calculations by solution-phase and “machine-phase” chemistry. Other differences result from the nanomechanical emphasis on force as a controllable parameter and on stiffness as a determinant of positioning errors; in conventional chemistry, stiffness is of interest chiefly as a determinant of vibrational frequencies in spectroscopy, and force is rarely mentioned. Further, nanomechanical systems will often be used to apply large forces to bonds and to non-bonded interfaces. Although strained organic molecules can experience large bonded and non-bonded forces, potential functions developed for chemistry must be subject to scrutiny before applying them to problems of nanomechanical design involving large loads.

3.3.3.1. Bonds under large tensile loads

For large tensile loads, the MM2 bond stretching function is clearly inadequate: the cubic term eventually results in an unboundedly large repulsive force. Higher-order terms were unimportant in the molecules of interest and were omitted to reduce computational expense. Where stresses and separations are larger, the potential energy of stretching in covalent bonds is commonly modeled using the Morse function

$$\mathcal{V}_{\text{morse}} = D_e \left[\left(1 - e^{-\beta(\ell - \ell_0)} \right)^2 - 1 \right] \quad (3.11)$$

with the associated forces and stiffnesses

$$F_{\text{morse}} = -\frac{\partial \mathcal{V}_{\text{morse}}}{\partial d} = 2\beta D_e \left[e^{-2\beta(\ell - \ell_0)} - e^{-\beta(\ell - \ell_0)} \right] \quad (3.12)$$

$$k_{s, \text{morse}} = \frac{\partial^2 \mathcal{V}_{\text{morse}}}{\partial d^2} = 2\beta^2 D_e \left[2e^{-2\beta(\ell - \ell_0)} - e^{-\beta(\ell - \ell_0)} \right] \quad (3.13)$$

In the above expressions,

$$\beta = \sqrt{\frac{k_s}{2D_e}} \quad (3.14)$$

where the energy D_e represents the difference between the minimum of the potential

Table 3.7. Bond dissociation energies, stiffnesses, lengths and Morse β parameters. References for bond dissociation energies: ^a (Kerr 1990), ^b (McMillen and Golden 1982), ^c (Huheey 1978). Values of ℓ_0 from Table 3.2.

Bond	D_0 (aJ)	D_e (aJ)	k_s (N/m)	β (m ⁻¹)	ℓ_0 (10 ⁻¹⁰ m)	Ref.	Compound
C-H	0.642	0.671	460	1.851	1.113	a	H-C(CH ₃) ₃
C-C	0.545	0.556	440	1.989	1.523	b	CH ₃ -C(CH ₃) ₃
C=C	1.190	1.207	960	1.994	1.337	a	H ₂ C=CH ₂
C≡C	1.594	1.614	1560	2.198	1.212	a	HC≡CH
C-N	0.498	0.509	510	2.238	1.438	b	C ₂ H ₅ -N(CH ₃) ₂
C-O	0.564	0.575	536	2.159	1.402	b	C ₂ H ₅ -OCH ₃
C=O	1.327	1.343	1080	2.005	1.208	c	
C-F	0.876	0.887	510	1.695	1.392	a	F-C ₂ H ₅
N-H	0.631	0.664	610	2.143	1.020	b	H-N(CH ₃) ₂
N-N	0.405	0.417	560	2.592	1.381	a	H ₂ N-N(CH ₃) ₂
O-H	0.725	0.753	460	1.747	0.942	b	H-OC(CH ₃) ₃
O-O	0.259	0.272	781	3.789	1.470	a	(CH ₃) ₃ CO-OC(CH ₃) ₃
C-Si	0.616	0.624	297	1.543	1.880	a	CH ₃ -Si(CH ₃) ₃
C-P	0.439	0.446	291	1.806	1.856	c	
C-S	0.532	0.539	321	1.726	1.815	a	CH ₃ -SCH ₃
C-Cl	0.583	0.591	323	1.653	1.795	b	Cl-CH ₃
C-Br	0.482	0.488	230	1.536	1.949	a	Br-CH ₃
C-I	0.393	0.398	220	1.662	2.149	b	I-CH ₃
Si-Si	0.555	0.559	185	1.286	2.332	a	(CH ₃) ₃ Si-Si(CH ₃) ₃

energy curve and the limiting value at infinite separation. D_e cannot be directly measured, but is the sum of the energy required to break the bond at zero K (D_0), and the zero point vibrational energy. It may be approximated as

$$D_e \approx D_0 + \frac{1}{2} \hbar \sqrt{\frac{k_s}{\mu}}; \quad \mu = \frac{m_1 m_2}{m_1 + m_2} \quad (3.15)$$

where m_1 and m_2 are the masses of the bonded atoms. This expression for the equilibrium dissociation energy incorporates a correction for the zero point energy of the form $\hbar\omega/2$, based on the approximation of a locally harmonic potential and a vibrational frequency like that of a diatomic molecule. In the case of a C–C bond, D_e is about $1.02 D_0$. Table 3.7 lists values for D_e based on the bond stiffness values of the MM2 potential and tabulated values for bond dissociation energies, using the above approximation for the zero point energy after correction to zero K. These values can be substantially modified by the surrounding molecular structure.

Figure 3.3 plots the C–C Morse potential based on the values in Table 3.7 along with the MM2 potential function; Figure 3.4 plots additional Morse potentials for easy reference. Although its form is motivated by quantum mechanical considerations, the Morse function grows increasingly inaccurate far from the equilibrium separation. From a structural perspective, the shape of the Morse function is of interest chiefly within the separation defined by the inflection point, $\ell = \ell_0 + (\ln 2)/\beta$; a bond stretched beyond this length by a position-independent force becomes mechanically unstable. Beyond the inflection point, the stiffness becomes negative, at reaching a most-negative value of $-0.125k_s$.

At larger separations, a more accurate potential is the Lippincott function,

$$V_{\text{lippincott}} = D_e \left[1 - \exp \left(- \frac{k_s \ell_0 (\ell - \ell_0)^2}{2 D_e \ell} \right) \right], \quad \ell \geq \ell_0, \quad (3.16)$$

which better fits data from vibrational spectroscopy (Eggers, Gregory et al. 1964) and accurate *ab-initio* molecular orbital calculations (Brown and Truhlar 1985). Note that its accuracy is poor at separations less than equilibrium. The Lippincott function typically yields a most-negative stiffness of substantially greater magnitude than $-0.125k_s$, and a greater force for bond breakage. Chapter 6 examines bond strengths, using the Morse potential (which underestimates strength); Chapter 8 examines bond breakage processes using the Lippincott potential (which provides a more adverse stiffness estimate).

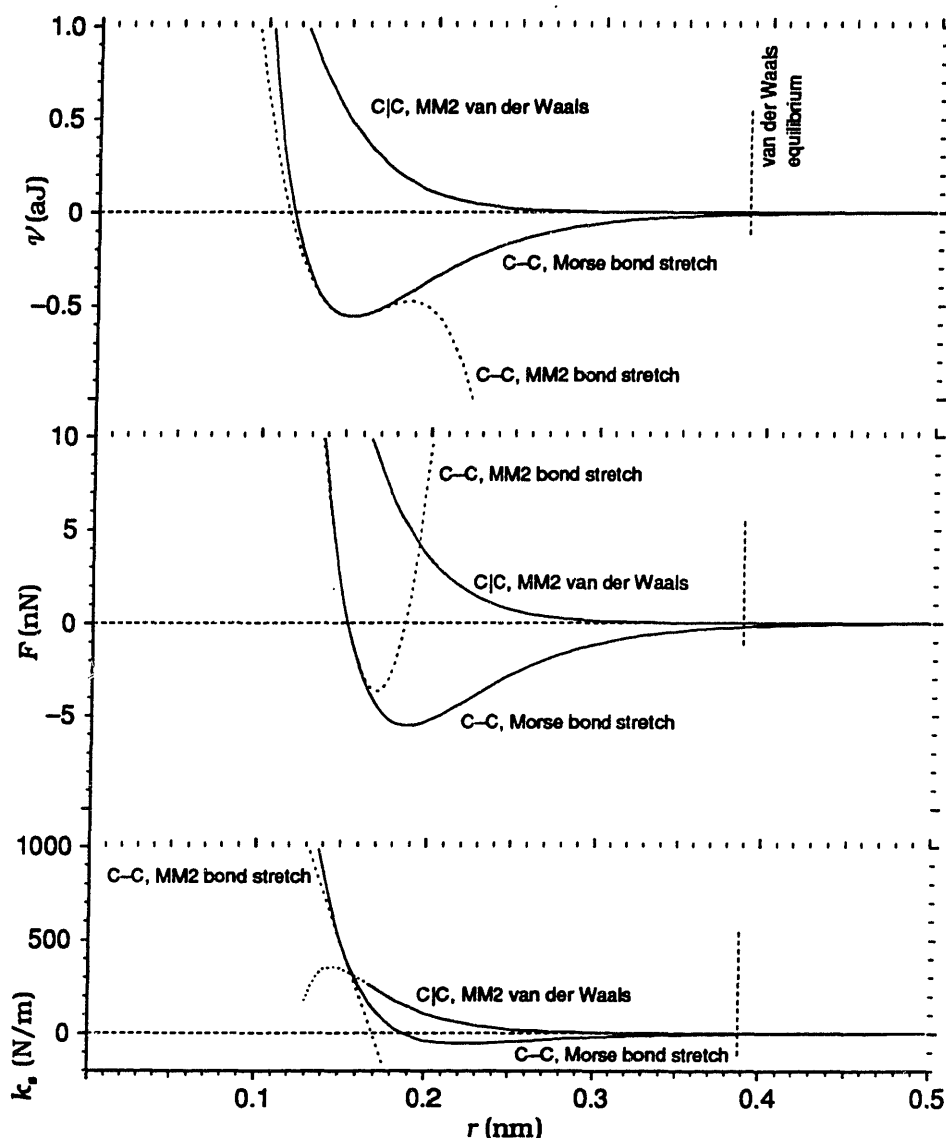


Figure 3.3. Morse and MM2 van der Waals potentials, forces, and stiffnesses for carbon-carbon interactions. The Morse curves are based on parameters from Table 3.7 for single-bonded carbon; a dotted curve compares the MM2 bond stretching potential. The van der Waals curves (vertical bars represent non-bonding contact) are based on parameters from Table 3.1 for sp^2 carbon atoms, which have better-exposed surfaces than do typical sp^3 atoms. Note the breakdown in the MM2 van der Waals model at distances around 0.15 nm (vs. an equilibrium separation of 0.388 nm), with stiffness values suffering at the greatest separations and energies at the least; the dotted extensions represent clearly non-physical regions.

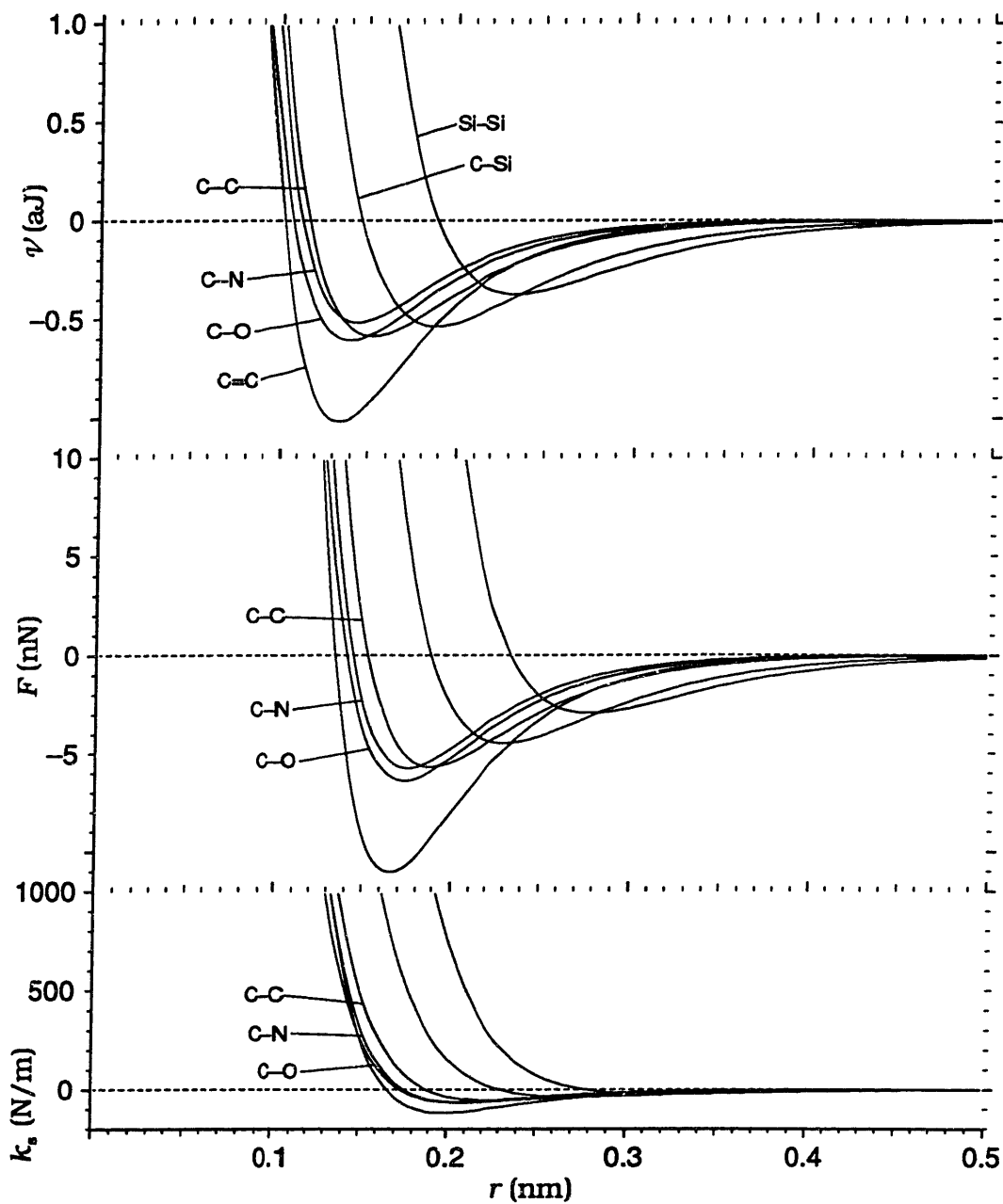


Figure 3.4. Morse potentials, forces, and stiffnesses for a representative sample of bond types of use in structural frameworks. Parameters from Table 3.7.

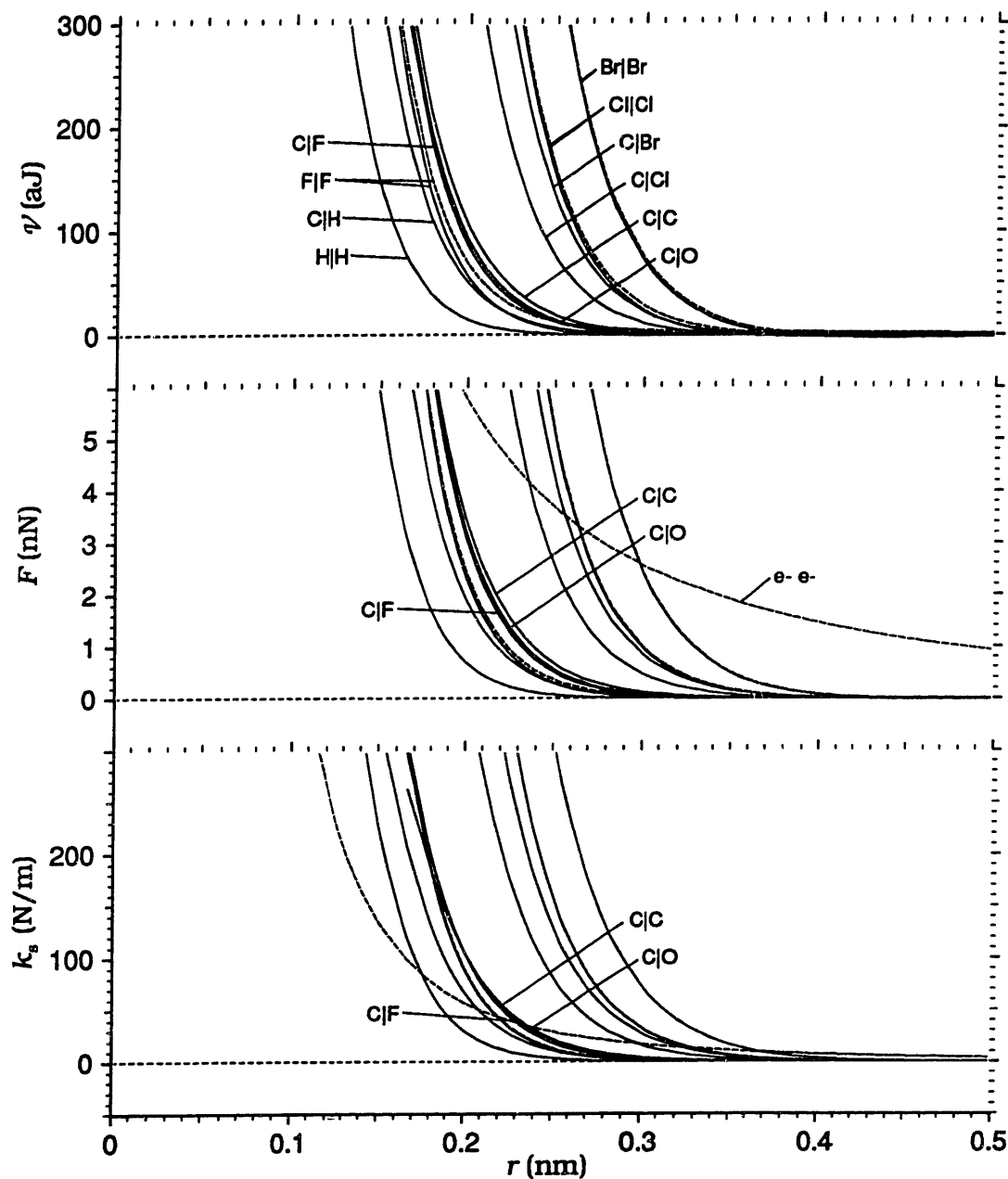


Figure 3.5. MM2 van der Waals potentials, forces, and stiffnesses for a representative sample of pairwise non-bonded interactions (the ClO curves omit the lone-pair contribution). The solid curves for F|F, Cl|Cl, and Br|Br van der Waals interactions are accompanied by dotted curves representing combined van der Waals and electrostatic effects for the collinear dipole systems C-F|F-C, C-Cl|Cl-C, and C-Br|Br-C. Also shown are interactions between two isolated electrons (off scale in the upper graph).

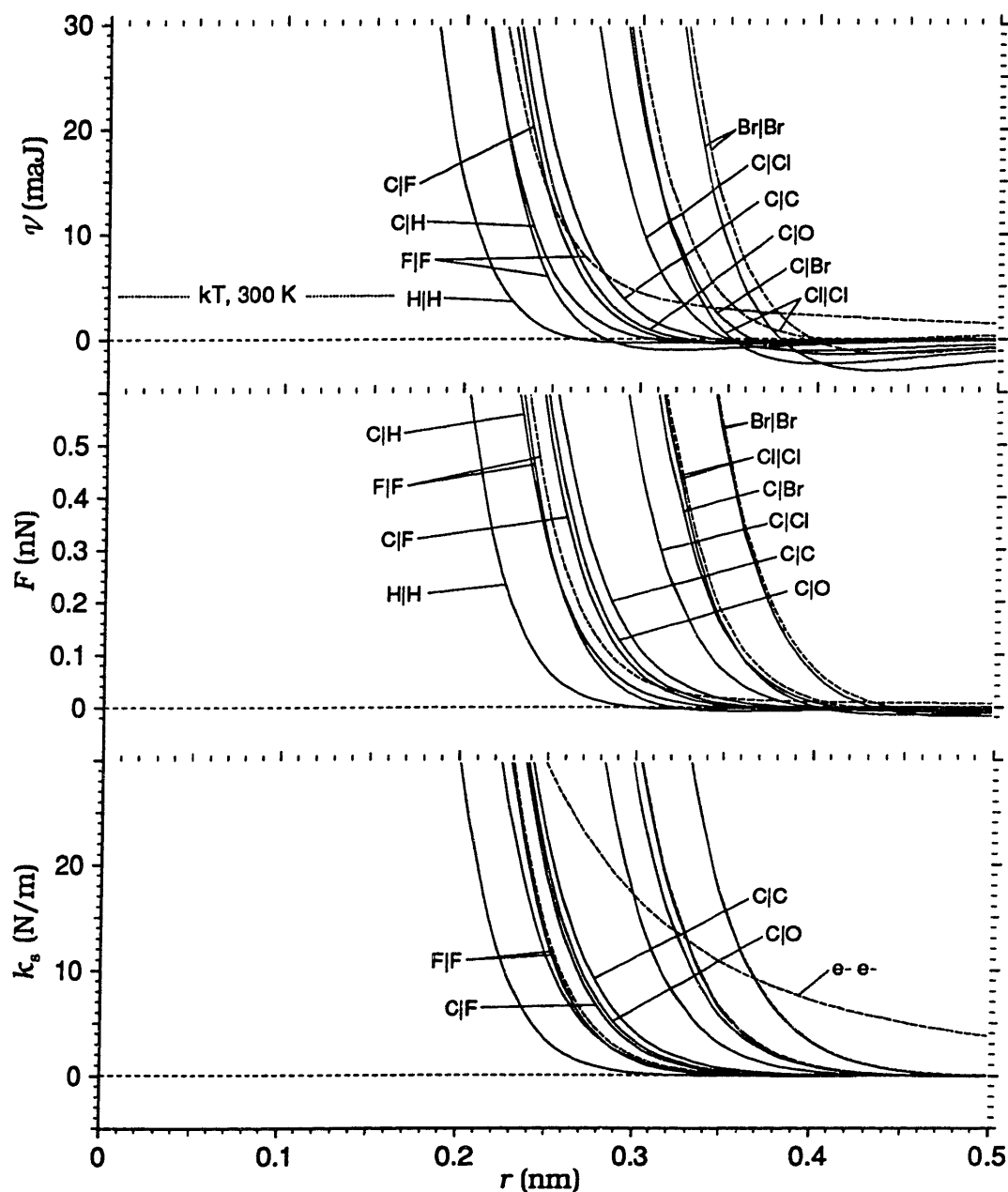


Figure 3.6. Curves as in Figure 3.5, with all vertical scales expanded by a factor of ten to show low-energy, low force interactions. The electron-electron interactions are off scale in the two upper graphs.

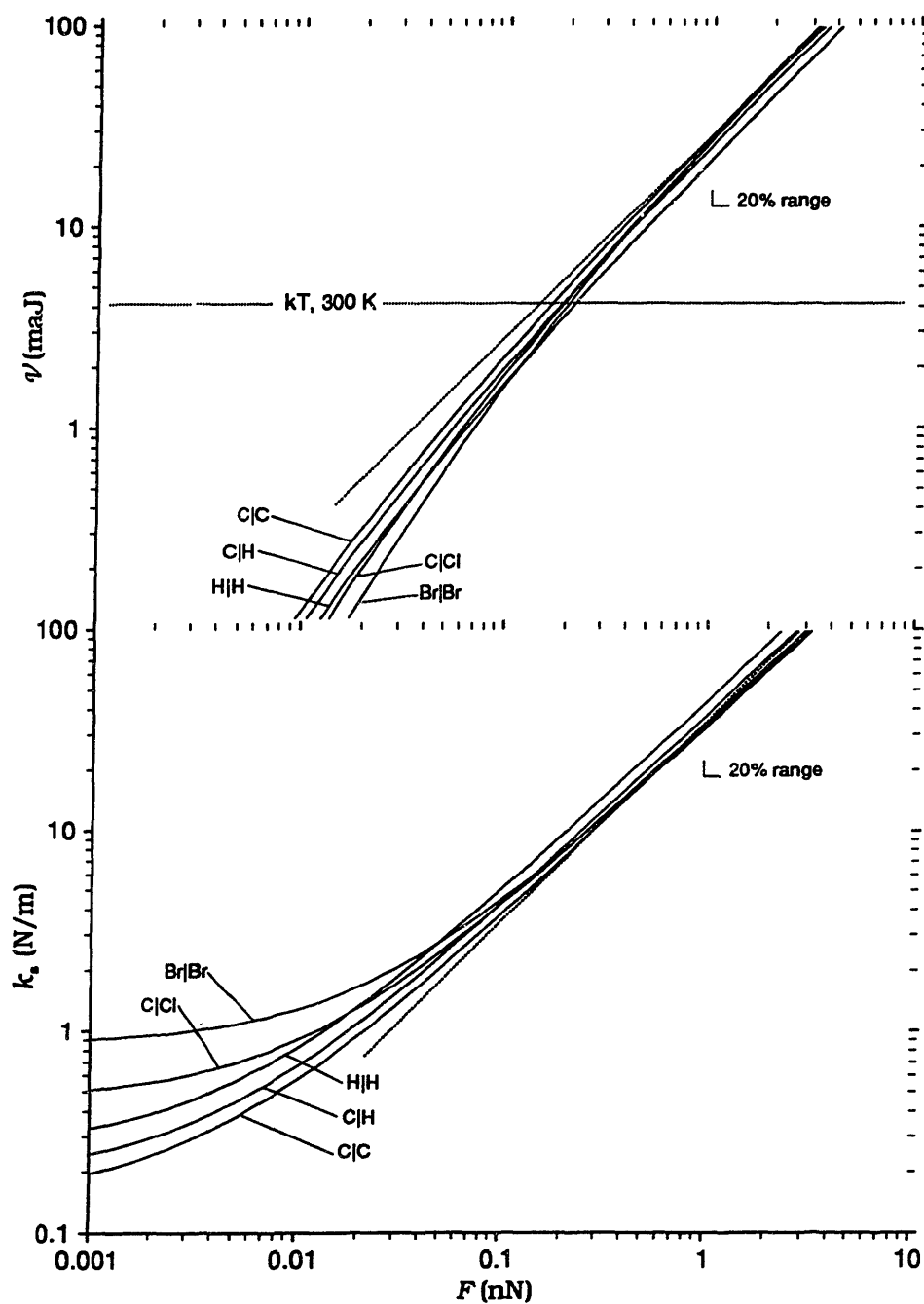


Figure 3.7. MM2 van der Waals potential energies and stiffnesses as a function of the pairwise force for a range of pairwise interactions. The dotted diagonal lines represent Eq. (3.19) (bottom) and Eq. (3.20) (top) for $d_{\text{vdw}0} = 0.36$ nm.

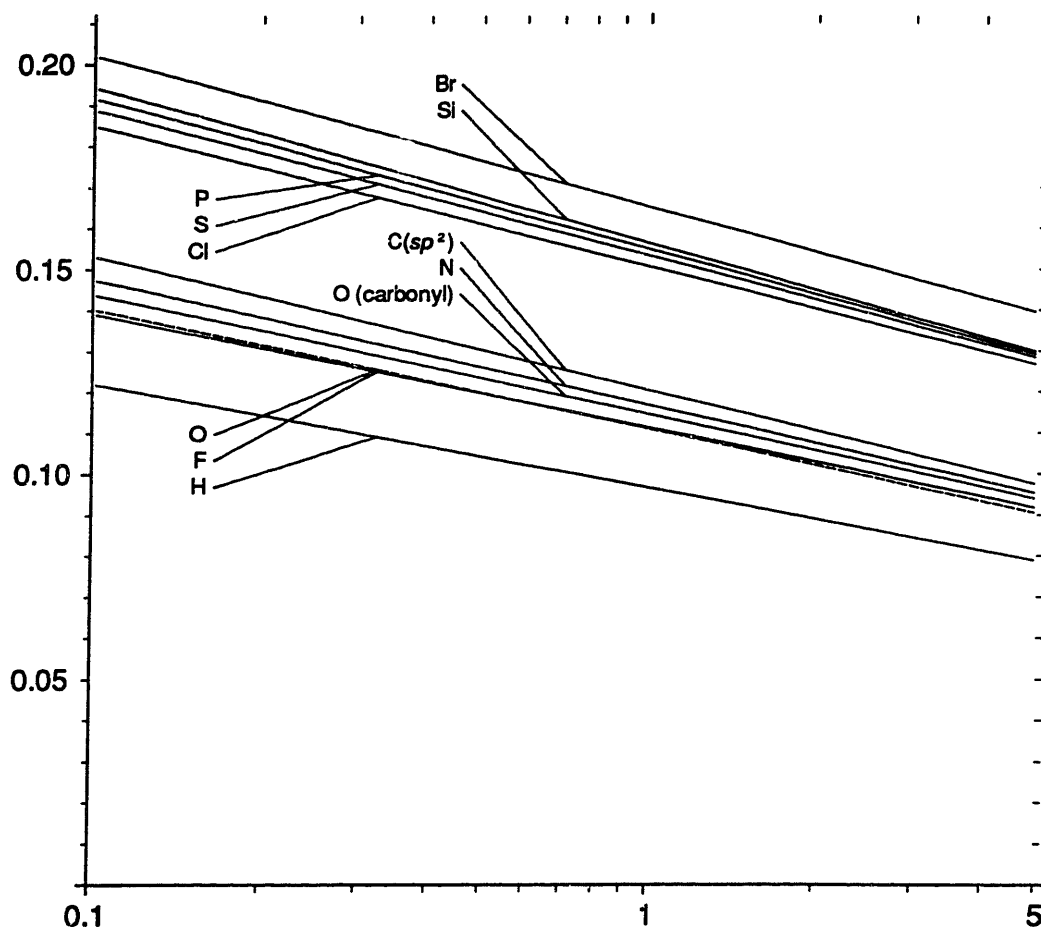


Figure 3.8. Approximate summable van der Waals radii, based on Eq. (3.21): for a given force between two atoms, the separation is approximately the sum of the above radii.

3.3.3.2. *Non-bonded interactions under large compressive loads*

Non-bonded interactions can be divided into many components, of which the MM2 model takes account of three: electrostatic forces, attractive van der Waals forces, and steric repulsion forces. The latter two are represented by the d^{-6} and exponential terms of the exp-6 van der Waals potential, both of which have forms motivated by approximate quantum theories. A more thorough treatment of these forces would include small, attractive d^{-8} , d^{-10} (etc.) terms, three-body interactions, induced dipole effects, different rules of combination for determining the well-depths and equilibrium radii of pairwise interactions, and so forth. Further, modeling interactions within and between molecules in terms of atom-by-atom pairwise potentials of any kind has no theoretical justification and is doubtless inaccurate. Discussions of alternative and more refined models may be found in the literature (Maitland, Rigby et al. 1981; Rigby, Smith et al. 1986). Nonetheless, molecular mechanics potentials including the above three components have proved accurate for a wide range of purposes. Most of the interactions just mentioned have little effect where repulsive forces are dominant, as they often will be at interfaces within nanomechanical devices.

Several popular potential forms are unsuitable for analyzing nanomechanical systems that involve large repulsive energies. The Lennard-Jones 6-12 potential, although simple and time-honored, has a repulsive interaction that lacks theoretical motivation and is unrealistically steep. The Maitland and Smith potential (Maitland, Rigby et al. 1981), although excellent in the low-energy range, is again too steep in the deep repulsive regime.

The MM2 van der Waals expression has two parameters: the equilibrium separation of two atoms and the depth of the attractive well at that point. Well depths are typically on the order of one mJ, but repulsive interaction energies of nanomechanical interest range upward to several hundred mJ. This disparity is cause for concern, as is the behavior of the MM2 potential at small distances, where the exponential repulsion is dominated by the d^{-6} attraction. The repulsive force reaches a maximum at $0.323 d_{vdw0}$ and reverses at sufficiently small separations; this is non-physical. The general realism of the MM2 exp-6 function at intermediate separations and high energies can be tested by comparing its description of noble gas interactions to the results of neutral particle beam collision experiments (Amdur and Jordan 1966), using well-depth and equilibrium separations drawn from other sources (Rigby, Smith et al. 1986). The results suggest that the MM2

potential deviates from the actual potential by only tens of percent down to radii of less than $0.5 d_{\text{vdw}0}$, and thus to energies in the hundreds of mJ. This is good enough for basic nanomechanical engineering work.

As will be detailed in Chapter 5, stiff interactions are necessary where quantum and thermal uncertainties in position are to be minimized. Moving parts will frequently be constrained not by bonds, but by steric repulsion. This makes forces and stiffnesses important. In the MM2 approximation

$$F_{\text{vdw}} = -\frac{\partial \mathcal{V}_{\text{vdw}}}{\partial d} = \epsilon_{\text{vdw}} \left[\frac{3.1 \times 10^6}{d_{\text{vdw}0}} e^{-12.5 \frac{d}{d_{\text{vdw}0}}} - \frac{11.54}{d} \left(\frac{d}{d_{\text{vdw}0}} \right)^{-6} \right] \quad (3.17)$$

and

$$k_{s, \text{vdw}} = \frac{\partial^2 \mathcal{V}_{\text{vdw}}}{\partial d^2} = \epsilon_{\text{vdw}} \left[\frac{3.88 \times 10^7}{d_{\text{vdw}0}^2} e^{-12.5 \frac{d}{d_{\text{vdw}0}}} - \frac{80.81}{d^2} \left(\frac{d}{d_{\text{vdw}0}} \right)^{-6} \right] \quad (3.18)$$

Figure 3.5 plots energy, force, and stiffness as a function of distance for a variety of pairwise nonbonded interactions. In the repulsive regime, stiffness increases with decreasing separation and increasing force; the achievable compressive force will commonly limit the achievable stiffness. Figure 3.7 plots stiffness (and energy) as a function of compressive force for a representative set of pairwise interactions. Where the exponential term dominates, stiffness and force become proportional. For strongly repulsive interactions in the MM2 model

$$k_{s, \text{vdw}} \approx \frac{12.5}{d_{\text{vdw}0}} F_{\text{vdw}} \approx 3.5 \times 10^{10} F_{\text{vdw}} \quad (3.19)$$

A similar expression describes the energy in the strongly repulsive regime

$$\mathcal{V}_{\text{vdw}} \approx 0.08 d_{\text{vdw}0} F_{\text{vdw}} \approx 2.9 \times 10^{-11} F_{\text{vdw}} \quad (3.20)$$

Under these conditions, the relatively wide range of ϵ_{vdw} values (a factor of ~ 10) is of no significance. Only the smaller range of $d_{\text{vdw}0}$ values (a factor of ~ 1.5) affects the stiffness and stored energy resulting from a given compressive load. Note that the form of these relationships makes it a matter of indifference whether a compressive load is concentrated on a single atom or spread over many, so long as each single-atom load is large enough for the approximation to apply.

Chemists define a variety of atomic sizes, including covalent, ionic, and van der Waals radii. These have the property that, under the relevant conditions (covalent bonding, ionic contact, and zero-load van der Waals contact) the distance between atoms of any two types can be approximated as the sum of their radii. For nanomechanical work, it is convenient to define analogous summable van der Waals radii for atoms under mechanical load. The following approximate expression defines this loaded radius as a function of the applied force F

$$r_{\text{loaded}}(F) \approx \frac{r_{\text{vdw}}}{13.6} \ln \left(\frac{2.5 \times 10^6 \epsilon_{\text{vdw}}}{r_{\text{vdw}} F} \right) \quad (3.21)$$

In the range $F = 0.1$ to 5 nN, this approximation yields separations within 4% of the values implied by the MM2 van der Waals model for all pairs of atoms drawn from the list in Table 3.1 save for those including iodine (or the lone-pair pseudoatom). For convenient reference, these functions are graphed in Figure 3.8.

3.4. Potentials for chemical reactions

3.4.1. Relationship to other methods

The molecular mechanics methods discussed above are based on the notion of a molecular structure with well-defined bonds; they can describe structural deformations, but cannot describe transformations that rearrange patterns of bonding. (Consequently they cannot predict chemical instabilities, a topic discussed in Chapter 6.) Potential energy functions for chemical reactions have largely been the subject of separate study. Techniques that combine molecular mechanics potentials for describing large structures with reaction potentials (or with quantum mechanical methods applied to small regions (Singh and Kollman 1986)) will be of use in describing nanomechanisms that make and break bonds.

Bond cleavage and formation present computational challenges for molecular orbital methods. Accurate calculations often require extensive use of CI methods to account for electron correlation effects, raising the cost of computations and preventing computations of useful accuracy on complex systems. Even where computations are feasible at many points on the PES, subsequent analytical studies typically demand that the surface be described by some function fitted to those points. Accordingly, studies of the detailed dynamics of chemical reactions have typically relied on approximate potential energy

surfaces. These are described either by fitting complex functions to quantum mechanical calculations, or by adjusting a few parameters in a fixed functional form to make calculated reaction rates (and their temperature dependence) match experimental data. Only the simplest reactions and PES approximations are described here; a growing literature is concerned with elaborating potentials for polyatomic reaction dynamics (Truhlar and Steckler 1987).

3.4.2. Bond cleavage and radical coupling

The simplest reactions break a single bond and yield two radicals that undergo no subsequent rearrangement. These reactions have potential energy surfaces that are extensions of bond stretching and bond bending. The Morse function Eq. (3.11) can serve as an approximate potential for homolytic reactions (i.e., bond cleavage yielding a pair of radicals rather than a pair of ions). For the design of reactive nanomechanisms, the important region of the Morse function lies between the bottom of the potential well ($\ell = \ell_0$) and the separation at which the stiffness is most negative (which occurs at $\ell = \ell_0 + (\ln 4)/\beta$). Toward the outer end of this range, however, the Lippincott potential appears to give a more accurate account of the potential.

Bond formation by radical coupling (the inverse of homolytic bond cleavage) requires paired, anti-parallel electron spins (termed *singlet* states, for reasons rooted in spectroscopy). With paired spins, this process follows a Morse potential; systems with parallel, unpaired spins (*triplet* states) experience a potential approximated by the repulsive, anti-Morse function

$$V'_{\text{anti-morse}} = \frac{1}{2} D_e \left[\left(1 + e^{-\beta(\ell - \ell_0)} \right)^2 - 1 \right] \quad (3.22)$$

and form no bond. Energetic considerations favor spin pairing and bond formation, but the pairing of initially unpaired spins (an electronic transition between potential surfaces termed *intersystem crossing*) can be slow on a molecular time scale.

3.4.3. Abstraction reactions

More complex reactions make and form bonds simultaneously. The most studied class involves the transfer of a single atom from a molecule to a radical, such as the symmetrical hydrogen abstraction reaction



Reactions of this sort are frequently modeled using the London–Eyring–Polanyi–Sato (LEPS) potential (Sato 1955), or the related extended LEPS potential (Kuntz, Nemeth et al. 1966). The latter is based on the expression

$$V_{\text{LEPS}} = \frac{Q_{AB}}{1+S_{AB}} + \frac{Q_{BC}}{1+S_{BC}} + \frac{Q_{AC}}{1+S_{AC}} - \sqrt{\frac{1}{2} \left[\left(\frac{J_{AB}}{1+S_{AB}} - \frac{J_{BC}}{1+S_{BC}} \right)^2 + \left(\frac{J_{BC}}{1+S_{BC}} - \frac{J_{AC}}{1+S_{AC}} \right)^2 + \left(\frac{J_{AC}}{1+S_{AC}} - \frac{J_{AB}}{1+S_{AB}} \right)^2 \right]} \quad (3.24)$$

which reduces to the LEPS potential when the Sato parameters S_{AB} , S_{BC} , and S_{AC} are all equal, and further reduces to the London equation when all three equal zero. In this expression, the energy Q_{XY} represents the coulomb integral between atoms X and Y, and J_{XY} the exchange integral; these quantities arise in quantum mechanical descriptions of bonding. In the LEPS model, bonding interactions between atoms X and Y are described by the Morse function

$$\frac{Q_{XY} + J_{XY}}{1+S_{XY}} = D_{\text{eXY}} \left[\left(1 - e^{-\beta_{XY}(\ell_{XY} - \ell_{0XY})} \right)^2 - 1 \right] \quad (3.25)$$

and anti-bonding interactions are described by the anti-Morse function

$$\frac{Q_{XY} - J_{XY}}{1-S_{XY}} = \frac{1}{2} D_{\text{eXY}} \left[\left(1 + e^{-\beta_{XY}(\ell_{XY} - \ell_{0XY})} \right)^2 - 1 \right] \quad (3.26)$$

yielding the expressions

$$Q_{XY} = \frac{1}{2} D_{\text{eXY}} \left[\frac{1}{2} (S_{XY} + 3) e^{-2\beta_{XY}(\ell_{XY} - \ell_{0XY})} - (3S_{XY} + 1) e^{-\beta_{XY}(\ell_{XY} - \ell_{0XY})} \right] \quad (3.27)$$

$$J_{XY} = \frac{1}{2} D_{\text{eXY}} \left[\frac{1}{2} (3S_{XY} + 1) e^{-2\beta_{XY}(\ell_{XY} - \ell_{0XY})} - (S_{XY} + 3) e^{-\beta_{XY}(\ell_{XY} - \ell_{0XY})} \right] \quad (3.28)$$

The bond properties of isolated molecules define the pairwise Morse parameters, and for any given choice of Morse and Sato parameters, the LEPS potential can be expressed as a function of the three interatomic distances, ℓ_{AB} , ℓ_{BC} , and ℓ_{AC} . For a collinear system, it can be expressed as a function of two such distances. In the limit as one atom is

removed to infinity while the other pair remains in close proximity, the LEPS potential reduces to the Morse potential for that pair of atoms.

Despite (and because of) their limited flexibility, LEPS and extended LEPS potentials have been the basis for work in molecular reaction dynamics (Levine and Bernstein 1987) and transition state theory (Bérces and Márta 1988). They will be used in Chapter 8 to examine the effects of mechanical forces on atom-transfer reactions. For atom-transfer reactions, a wide range of alternative functional forms has been explored, but improved fits are generally purchased at the expense of greater mathematical complexity (Bérces and Márta 1988).

3.5. Continuum representations of surfaces

Nanomechanical systems will often experience forces dominated by bonding and overlap repulsion, but van der Waals attractions (dispersion forces) between nanometer-scale objects can also be substantial. The modest magnitude and long-range nature of these forces motivates the use of continuum approximations. At small spacings, surface-surface interactions require a model that separately accounts for the first atomic layer, including overlap forces.

3.5.1. Continuum models of van der Waals attraction

3.5.1.1. The Hamaker constant

In terms of the MM2 parameters, the long-range interatomic pairwise potential is

$$\mathcal{V}_{\text{vdw}} = -\frac{C}{d^6}; \quad C = 1.924\epsilon_{\text{vdw}}d_{\text{vdw}0}^6 \quad (3.29)$$

For interactions involving a solid body, the continuum model merges the contributions of individual atoms and describes them in terms of the Hamaker constant, \mathcal{A} . In the simplest description, only non-retarded pairwise interactions are considered and materials are assumed to be separated by vacuum. This yields

$$\mathcal{A}_{12} = \pi^2 C \rho_{a1} \rho_{a2} \quad (3.30)$$

where ρ_{a1} and ρ_{a2} are the atom densities of the interacting bodies. (Where a body contains several atom types, it can be treated as the superposition of several bodies each with a single atom type.) This and the more rigorous Lifshitz model are described in

Israelachvili (Israelachvili 1985), along with the other approximations of this section. Values of the Hamaker constant for condensed media are typically in the range of 40–400 mJ.

For nonpolar insulators of different Hamaker constant interacting across vacuum,

$$\mathcal{A}_{12} \approx \sqrt{\mathcal{A}_{11}\mathcal{A}_{22}} \quad (3.31)$$

(Note that the MM2 model uses the arithmetic mean for calculating pairwise dispersion interactions between unlike atoms; the geometric mean has better theoretical justification, and is adopted in MM3.) For interactions between nonpolar insulators 1 and 2 across a medium 3,

$$\mathcal{A}_{132} \approx \left(\sqrt{\mathcal{A}_{11}} - \sqrt{\mathcal{A}_{33}}\right)\left(\sqrt{\mathcal{A}_{22}} - \sqrt{\mathcal{A}_{33}}\right) \quad (3.32)$$

When one or more phases are polar (e.g., water), an estimate of the Hamaker constant can be derived (Israelachvili 1985) from the Lifshitz theory. In this approximation,

$$\begin{aligned} \mathcal{A}_{132} \approx & \frac{3\hbar\omega}{8\sqrt{2}} \frac{(n_1^2 - n_3^2)(n_2^2 - n_3^2)}{\left(\sqrt{n_1^2 + n_3^2} + \sqrt{n_2^2 + n_3^2}\right)\sqrt{(n_1^2 + n_3^2)(n_2^2 + n_3^2)}} \\ & + \frac{3}{4}kT \left(\frac{\varepsilon_1 - \varepsilon_3}{\varepsilon_1 + \varepsilon_3}\right)\left(\frac{\varepsilon_2 - \varepsilon_3}{\varepsilon_2 + \varepsilon_3}\right) \end{aligned} \quad (3.33)$$

where n is the optical refractive index, ε is the zero-frequency dielectric constant, and ω is the absorption frequency of the medium (or, where frequencies differ, the mean). For a variety of materials, $\hbar\omega \approx 2$ aJ. Table 3.8 lists values of n and ε for several materials, together with the resulting Hamaker constants for interaction between two bodies of identical composition across vacuum.

3.5.1.2. Interactions between objects

Figure 3.9 presents approximate expressions for the attractive potential between objects of various shapes. Note that the potential energy always diverges as the distance $s \rightarrow 0$; a minimum separation $s_{\min} \approx 0.2$ nm is commonly assumed in the literature. This can be rationalized in terms of a ~ 0.1 nm gap between a fictitious “Hamaker surface” (bounding the region of high polarizability in a solid object) and an approximate “overlap surface” (describing the limits of motion imposed by overlap repulsion). The gap arises because polarizability chiefly arises from regions of high electron density within bonds

Table 3.8. Refractive index, zero-frequency dielectric constant, and Hamaker constant (from Eq. 3.33) for several materials; $\hbar\omega$ taken as 2 eV.

Material	n	ϵ	\mathcal{A}_{11} (eV)
polytetrafluoroethylene	1.35	2.1	38
polyethylene	1.52	2.3	76
diamond	2.40	5.5	340
fused silica	1.49	3.8	69
water	1.33	78.5	37
glycerol	1.47	42.5	66
metals (Au, Ag, Cu)*			300–500

* Hamaker constants for metals from (Israelachvili 1985). Values for n , ϵ from (Gray 1972; Lide 1990).

and atomic core regions, but overlap repulsion is substantial at greater distances, in regions of lower electron density. The next section examines flat-surface models good at small separations.

Taking $\mathcal{A} = 400$ eV (a fairly high value) and $s = 0.2$ nm (i.e., contact), the attractive dispersion force between two spheres of $r = 1$ nm is 0.83 nN. For parallel surfaces with the same separation, the force per unit area is 2.7×10^9 N/m², on the order of 1/20 the tensile strength of diamond. Doubling s (perhaps by interposing scattered atomic-scale bumps) would reduce this force to 3.3×10^8 N/m².

3.5.2. Transverse-continuum models of surfaces

Chapters 7 and 10 examine the mechanical properties of sliding, unreactive, atomically-precise interfaces in which van der Waals attraction and steric repulsion are the dominant forces. This requires a model for the interaction energy of two such surfaces at small separations. In the systems of greatest practical interest, the structures of the two surfaces will be out of register (owing to differences in lattice spacing, or to a relative rotation). Neglecting elastic deformations resulting from interfacial forces, this results in a surface-surface potential that can be described in terms of the interactions of layers in which the effects of the constituent atoms have been spread uniformly over the plane. In a

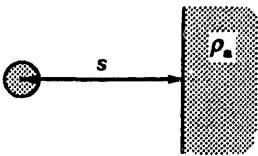
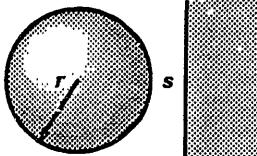
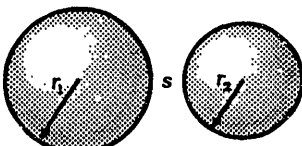
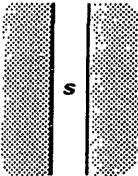
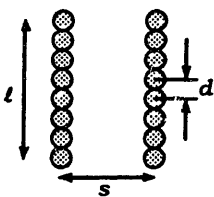
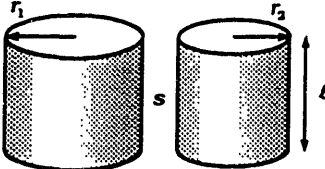
<p>Atom-surface</p>  $\mathcal{E} = -\frac{\pi C \rho_s}{6s^3}$	<p>Sphere-surface</p>  $\mathcal{E} = -\frac{\mathcal{A}r}{6s}; \quad r \gg s$
<p>Two spheres</p>  $\mathcal{E} = -\frac{\mathcal{A}}{6s} \frac{r_1 r_2}{(r_1 + r_2)}; \quad r_1, r_2 \gg s$	<p>Two surfaces</p>  $\mathcal{E} = -\frac{\mathcal{A}S}{12\pi s^3}$
<p>Two parallel chains</p>  $\mathcal{E} = -\frac{3\pi C l}{8d^3 s^3}$	<p>Two cylinders</p>  $\mathcal{E} = -\frac{\mathcal{A}l}{12\sqrt{2}s^{3/2}} \left(\frac{r_1 r_2}{r_1 + r_2} \right)^{1/2}; \quad r_1, r_2 \gg s$

Figure 3.9. The potential energy, \mathcal{V}_{vdw} , of the van der Waals attraction in the non-retarded continuum approximation; after Israelachvili (Israelachvili 1985). In the two-surface expression, S = area. The more complex expressions for geometries with $r \approx s$ appear in Hiemenz (Hiemenz 1986).

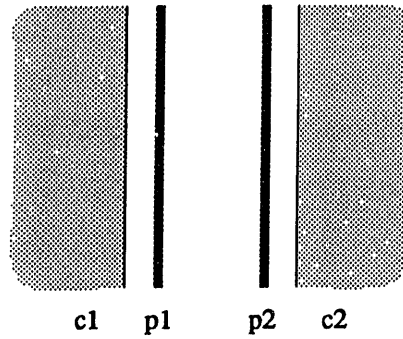


Figure 3.10. Model of two surfaces, showing the topmost atomic layers described as uniform planes and deeper regions approximated as uniform volumes.

further approximation, all but the first layers can be spread uniformly in the third dimension and their overlap repulsions (which are subject to rapid exponential decay) can be neglected, yielding the standard the Hamaker-constant model of the interaction energy for the bulk of each solid (Fig. 3.10.)

With this model, the total surface-interaction energy is the sum of the interactions of the two explicitly-treated planes, of each plane and the opposed continuum, and of the two continua:

$$\mathcal{V}'_{\text{surf}} = \mathcal{V}'_{\text{p1-p2}} + \mathcal{V}'_{\text{p1-c2}} + \mathcal{V}'_{\text{p2-c1}} + \mathcal{V}'_{\text{c1-c2}} \quad (3.34)$$

with obvious generalizations when more layers of atoms are treated explicitly.

The plane-plane interaction energy can be calculated from integrals describing the point-plane interaction energy, using an exp-6 potential. In terms of the density of atoms in the two planes $n_{\text{a,1}}$ and $n_{\text{a,2}}$ (m^{-2}), the energy per unit area is

$$\frac{\mathcal{V}'_{\text{p-p}}}{S} = n_{\text{a,1}} n_{\text{a,2}} \pi \left[\frac{2A}{b^2} e^{-bs_{\text{p-p}}} (bs + 1) - \frac{C_{\text{p-p}}}{2s_{\text{p-p}}^4} \right] \quad (3.35)$$

where C is the same as in Eq. (3.29), and A and b are parameters for the exponential component of the interatomic interaction, when expressed in the form

$$\mathcal{V}'_{\text{rep}} = Ae^{-bs} \quad (3.36)$$

In the MM2 model,

$$A = 2.48 \times 10^5 \varepsilon_{\text{vdw}}; \quad b = \frac{12.5}{d_{\text{vdw}0}} \quad (3.37)$$

For all other interactions, steric repulsion is neglected. The interaction between a plane and a volume is then a small modification of the expression in Figure 3.9,

$$\frac{\mathcal{V}_{\text{p-c}}}{S} = -n_{\text{a},1} \rho_{\text{a},2} \frac{\pi C_{\text{p-c}}}{6s_{\text{p-c}}^3} \quad (3.38)$$

The interaction between the two continuum regions (from Eq. 3.30 and the expression in Figure 3.9) contributes

$$\frac{\mathcal{V}_{\text{c-c}}}{S} = -\rho_{\text{a},1} \rho_{\text{a},2} \frac{\pi C_{\text{c-c}}}{12s_{\text{c-c}}^2} \quad (3.39)$$

The gap d_g between the plane and the surface of the continuum is chosen so as to count the effects of each atom exactly once, in the long-range attraction limit. Figure 3.11 graphs the energy, force, and stiffness resulting from this model for an illustrative set of parameter choices.

3.6. Molecular models and the continuum approximation

In the design process, engineers routinely use abstractions that permit a subsystem to be described without explicit reference to its internal structure and behavior. In electronics, the abstraction of the *digital signal* enables system designers to neglect the detailed properties of transistors; in software engineering, the abstraction of the *bit* is the first in a series of abstractions that enable system designers to neglect the detailed properties of computers. From this perspective, *material properties* and *component properties* are abstractions that enable mechanical engineers to neglect the properties of atoms and bonds.

Molecular mechanics methods can be used to calculate the mechanical properties of diamond-like bulk materials. Where a nanomechanical system incorporates a substantial region of such a material, its mechanical properties can be described in terms of elastic moduli, density, and so forth. Further, where a nanomechanical system includes an extended, regular structure such as a rod or a plate, these can frequently be described in terms of component properties such as stretching stiffness, bending stiffness, and so forth. The mechanical behavior of robust, diamond-like structures will commonly be

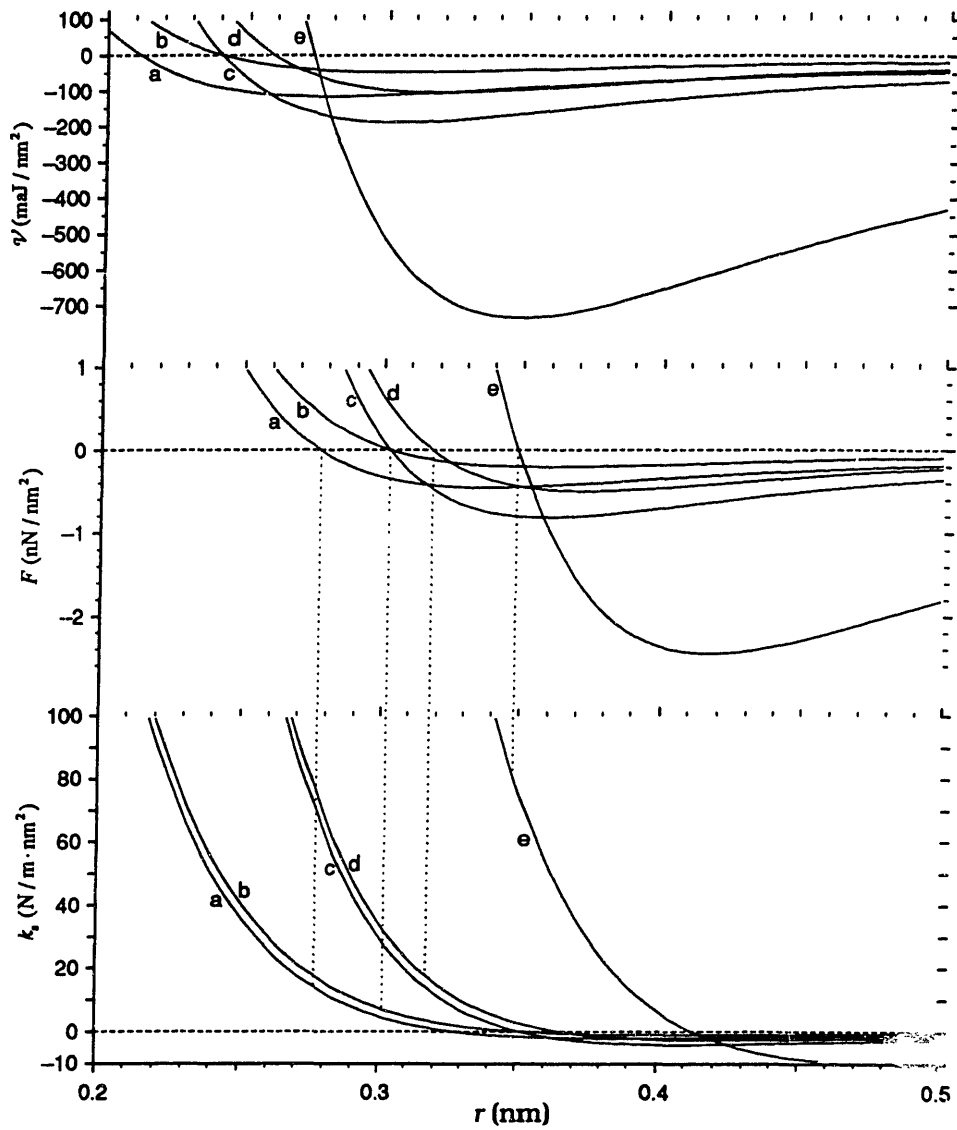


Figure 3.11. Interaction of pairs of surfaces according to Eqs. (3.34–3.39), with the following parameters (identical for both surfaces):

Curve	n_a ($10^{19}/\text{m}^2$)	MM2 type	ρ_a ($10^{29}/\text{m}^3$)	MM2 type	d_g (10^{-9} m)
a	0.9	C (1)	1.5	C (1)	0.09
b	0.9	C (1)	0.75	C (1)	0.09
c	1.8	C (1)	1.5	C (1)	0.09
d	1.8	C (1)	0.75	C (1)	0.09
e	1.4	Si (19)	1.5	C (1)	0.13

well-approximated by linear models, and the atomic details of their interiors will be irrelevant to their external properties, and hence can be abstracted away in the system design process. Note that such approximations are of little use in standard chemistry, where large, stiff, regular structures are rare, and likewise in protein science, where structures are typically flexible and highly inhomogeneous.

For irregular surfaces in contact, there is often no substitute for modeling at the level of interatomic potentials. Nonetheless, Eq. (3.19) and (3.19) provide approximate formulas for the stiffness and energy of repulsive interactions as a function of the applied compressive force. These interactions are relatively insensitive to details of surface structure, within the class of low-polarity organic structures of chief interest here. In the attractive regime, the comparatively long-range nature of van der Waals attractions justifies making a continuum approximation, leading to the formulas presented in Fig. 3.10.

Approximations based on the above observations regarding solids are discussed in Chapter 10. They permit a limited but atomically-motivated return to the continuum approximations discussed in Chapter 2.

3.7. Further reading

This chapter and the next three outline topics that are foundational both to nanotechnology and to conventional chemistry and chemical physics. Since this book is intended for a broad audience, it seems desirable to provide brief bibliographic essays to provide a point of departure for those wishing to explore these topics in more depth. The following cannot be considered a proper review of the very extensive literature, but will suffice to get readers to the proper sections of libraries. It is biased toward general reviews and textbooks that have been used at MIT.

Molecular quantum mechanics

The foundations of molecular quantum mechanics have not changed in decades, and many textbooks teach the subject. Examples include *Atoms and Molecules* (Karplus and Porter 1970) and *Molecular Quantum Mechanics* (Atkins 1970). A useful volume for orientation to the subject is *Quanta: A Handbook of Concepts* (Atkins 1974), a cross-referenced, alphabetically-organized volume with essays, equations, and diagrams on most quantum mechanical phenomena in molecules.

Computational methods in quantum mechanics have advanced greatly over the decades. An excellent introduction to the subject from a user's perspective is *A Handbook of*

Computational Chemistry (Clark 1985); this volume includes concrete examples of the use of popular programs along with less perishable information on techniques and pitfalls in the field. The methods of *ab initio* molecular orbital theory are described in (Hehre, Radom et al. 1986) together with descriptions of numerous computational results.

Potential energy surfaces

A general introduction to molecular mechanics methods is given by *Molecular Mechanics* (Burkert and Allinger 1982). *A Handbook of Computational Chemistry* (Clark 1985) includes descriptions of MM2, although its main focus is on quantum chemistry.

Molecular mechanics methods are evolving rapidly. A common operation in molecular mechanics work is to find an equilibrium point on the PES by a minimization procedure; so-called “Newton methods” have generally required storage space proportional to N^2 (where N is the number of atoms) and a computational time proportional to N^3 , but a new algorithm scales as better than $N^{1.5}$ and N^2 respectively (Ponder and Richards 1987). For engineering work, fast, rough approximations are of considerable value in the initial stages of design and analysis; a method has been reported that gives results similar to MM2 using simpler, faster code based purely on pairwise, interatomic central forces (Saunders and Jarret 1986). Quantum chemistry methods are being used to improve molecular mechanics models, for example in work supported by the Consortium for Research and Development of Potential Energy Functions (Hagler, Maple et al. 1989).

Intermolecular forces are complex, though most are weak enough to be neglected in nanomechanical engineering of the sort considered here. A good introductory book on the topic is *The Forces Between Molecules* (Rigby, Smith et al. 1986); a more advanced text is *Intermolecular Forces: Their Origin and Determination* (Maitland, Rigby et al. 1981). An excellent book with a greater emphasis on condensed matter is *Intermolecular and Surface Forces* (Israelachvili 1985). The broad literature on surface science and surface chemistry contains much that is relevant, but often focuses on surfaces that are unstable and reactive; although these are interesting, they need not be used in nanomechanical devices. (Note that our point of departure has been organic chemistry, which studies structures small enough to be, in effect, all surface.)

Potential energy surfaces for atom-transfer (abstraction) reactions are discussed in (Levine and Bernstein 1987) and at greater length in (Bérces and Márta 1988); both include extensive references to the literature. A review of potential energy surfaces for more complex systems may be found in (Truhlar and Steckler 1987).

Chapter 4

Molecular dynamics

4.1. Models of dynamics

Molecular dynamics is fundamental to molecular machinery and has been widely studied in physical chemistry and chemical physics. Chapters 5–8 all deal with specific aspects of molecular dynamics of importance in a nanomechanical context; the present chapter provides a brief overview of the topic, examining the applicability of various approaches to nanomechanical problems. Section 4.2 reviews methods used in non-statistical descriptions of molecular dynamics, considering both quantum mechanical and classical models for the calculation of system trajectories. Section 4.3 reviews statistical descriptions of molecular dynamics, both classical and quantum mechanical. Section 4.4 returns to the issue of PES approximations, using dynamical principles to examine the differing requirements for accuracy that arise in different applications.

4.2. Non-statistical mechanics

This section outlines some of the non-statistical descriptions of molecular dynamics used in scientific work, commenting on their applicability to nanomechanical engineering problems.

4.2.1. Vibrational motions

Molecular vibration has been extensively studied in connection with infrared spectroscopy; IR vibrational frequencies are a major constraint used in determining parameters for molecular mechanics energy functions. Small displacements of molecules from equilibrium geometry are associated with nearly linear restoring forces. In the resulting harmonic approximation, the vibrational dynamics can be separated into a set of independent normal modes and the total motion of the system treated as a linear superposition of

normal mode displacements. This common approximation is of considerable use in describing nanomechanical systems.

Since both classical and quantum mechanics permit exact solutions for the harmonic oscillator, the time evolution of systems can readily be calculated in the harmonic approximation. In practice, non-linear terms permit energy exchange among vibrational modes, causing relaxation to thermal equilibrium. The equilibrium state, in turn, is best described by the methods of statistical mechanics. A non-statistical description of vibrational motion will be of interest chiefly during (or within a few relaxation times of) the excitation of a vibrational mode by a non-thermal energy source.

4.2.2. Reactions and transition rates

Vibrations involve motion within a potential well; reactions involve transitions between potential wells. Molecular reaction dynamics has been extensively studied by means of crossed molecular beams, and a major application (and test) of potential energy functions for chemical reactions has been the calculation of vibrational states and angular distributions resulting from reactive scattering.

With high-quality beams of simple molecules, one can observe quantal oscillations in the angular distribution of scattered trajectories. These result from interference among alternative collision trajectories that yield indistinguishable outgoing molecular trajectories and states. A broader distribution of initial energies and angles will obliterate these fine features, however, yielding a smooth distribution of product trajectories.

In practice, calculations of molecular reaction dynamics are commonly based on the quasiclassical approximation (Levine and Bernstein 1987). In this approximation, initial molecular states of vibration and rotation and subsequent are described classically, but with initial energies and angular momenta are chosen to match quantum constraints. Quantal uncertainties in position are modeled by calculating many trajectories with randomly chosen vibrational and rotational phase angles. Trajectories are then computed by integrating the classical equations of motion. Finally, quantization of outcomes is modeled by lumping final trajectories into bins in phase space, where each bin corresponds to a permissible product quantum state. These calculations cannot yield quantum interference patterns or resonances, but are accurate enough to be useful in describing the coarse dynamical features of a reactive collision, such as the general distribution of scattering angles and vibrational excitations.

Coarse features from a reaction dynamics perspective are, however, fine features

from the perspective of a chemist or a nanomechanical engineer. Chemical reactions in nanomechanical systems will have little resemblance to reactions in crossed molecular beams. In an extended solid system subject to relatively slow mechanical motions, reactive molecular components do not encounter one another with well-defined energies and momenta; broad, thermal distributions dominate. Likewise, during the course of the encounter, the reactive components do not form an isolated system with locally-conserved energy, momentum, and angular momentum; the reactive system remains coupled to a thermal bath. The notion of a scattering angle is meaningless, and reaction-induced vibrational excitations are quickly thermalized.

In these respects, reactions in nanomechanical systems will resemble the solution-phase reactions familiar to chemists. In such systems, the chief concern is with overall reaction rates, not with the details of trajectories. As a consequence, the detailed shape of the PES (crucial to the details of reactive scattering) is of reduced importance. Reaction rate data only weakly constrains PES properties. Given a few properties of the PES, thermally-activated reaction rates are calculated using classical or quantal transition state theories based on statistical mechanics; these are discussed in Chapter 6.

4.2.3. Generalized trajectories

Even in the absence of reactive transformations, the motions of typical nanomechanical systems cannot be completely described in terms of vibrations. Like protein chains or of molecules in a liquid, these systems typically permit large displacements subject to complex, interacting constraints. Their motions must be described by more general methods.

It is common practice to model the trajectories of such systems by integrating the classical equations of motion, deriving the forces on each atom from approximate potential functions (e.g., molecular mechanics potentials). Where statistical properties are desired, a common approach is to integrate the equations of motion for a substantial time, taking a time-average of the quantities of interest. Time steps are typically $\sim 10^{-15}$ s, and current computers have been used to follow the dynamics of $\sim 10^3$ atom systems for $\sim 10^{-9}$ s. These techniques are suitable for describing the short-term dynamics of nanomechanical components.

From a statistical perspective, thermal motions result in a certain probability density function for the positions of atoms with respect to their surroundings, and quantum effects broaden this probability density function. Quantum effects on molecular trajec-

ries can be approximated by an adaptation of classical molecular mechanics techniques in which each atom is represented by a circular chain of atoms with suitable interactions; this has been applied to large molecules, such as the protein ferrocycytochrome *c* (Zheng, Wong et al. 1988). The effects are, as one would expect, largest for high-frequency vibrations, such as bond stretching, bending, and torsional motions involving hydrogen atoms.

The dynamical behavior of nanomechanical systems can be partitioned into (1) thermally-excited vibrational motions within potential wells, (2) thermally-excited transitions between potential wells, and (3) more general motions, which in nanomachines will frequently be driven by non-thermal energy sources. Motions in category (3) will generally be of low frequency and hence little influenced by quantum effects at ordinary temperatures. They can typically be modeled using classical dynamics based on molecular mechanics potentials. Motions in category (1) can be treated either classically or quantum mechanically; Chapter 5 compares the statistical distributions resulting from classical and quantum models. Chapter 6 examines motions in category (2), again comparing classical and quantum approaches within a statistical framework.

4.3. Statistical mechanics

Statistical mechanics (also known as statistical or molecular thermodynamics) is commonly valued for its ability to relate macroscopic thermodynamic properties to statistical descriptions of the behavior of large numbers of molecules. En route to describing properties of bulk matter, statistical mechanics frequently describes probability distributions for underlying molecular variables, such as position and velocity. In the present context, it is these probabilistic descriptions of the behavior of individual molecular objects that are of primary value.

Statistical mechanics can frequently provide estimates of the statistical behavior of nanomechanical systems without the cost of running a detailed dynamical simulation for long periods of time. Today, it is expensive to do a simulation of 10^3 atoms for as long as 10^{-9} s, yet a nanomechanism that fails even once per millisecond may be unacceptable. Estimating failure rates by observing more than 10^6 expensive simulations would be impractical. Statistical mechanics, in contrast, can provide estimates for the frequencies of extremely rare events based on analytical methods applied to a known PES.

Statistical mechanics is commonly used to calculate quantities such as pressure, entropy, and free energy based on averages taken over many molecules in thermal equi-

librium. There is no fundamental difference, however, between an average computed for many equivalent molecules at an instant of time, an average computed for a single representative molecule over a long period of time, and a mean expected value for a single molecule at a single time. Accordingly, the concepts of pressure, entropy, free energy and the like can be used to reason about (for example) the mean expected efficiency of a single nanomachine. The only caveats regard the accuracy of assuming equilibrium; this is discussed in Section 4.3.4, and the relationship between measurement and equilibrium is discussed in Section 4.3.5.

4.3.1. Detailed dynamics vs. statistical mechanics

By omitting dynamical details, statistical mechanics provides a simplification that can assist both calculation and understanding. In the operation of real nanomechanical systems, the initial conditions will never be known with the precision assumed in classical dynamical models, and seldom with the precision assumed in quantum dynamical models. Instead, the motions and displacements resulting from thermal excitation will be random variables subject to some distribution. Rather than introducing arbitrarily assumptions, statistical mechanics takes these uncertainties as fundamental, yielding inherently probabilistic descriptions of system behavior. The nanomechanical engineer's task, then, is to devise systems in which all probable behaviors (or all but exceedingly improbable behaviors) are compatible with successful system operation.

Even if one were to assume classical, deterministic behavior and nearly-perfect knowledge of initial conditions, a typical nanomechanical system would soon require a statistical description. Consider the trajectory of a particle rebounding from an atom. Because atoms are not flat, a small perturbation in the trajectory will typically cause a particle to rebound at a different angle, leading to a larger perturbation in its next point of impact. In a typical system, trajectories that are initially almost identical will rapidly diverge until they have no similarity. This divergence is characteristic of the phenomenon of chaos. Further, real nanomechanical systems will be in contact with an environment at some non-zero temperature, and the environment will be a constant source of unpredictable thermal excitations.

4.3.2. Basic results in equilibrium statistical mechanics

Statistical mechanics takes its simplest form for systems at thermodynamic equilibrium. Since this is often a good approximation for real systems, some basic results are

worth summarizing.

In quantum statistical mechanics, it is convenient to consider a system that is in thermal equilibrium with a heat bath, yet is assumed to have a set of bath-independent quantum states $i = 0, 1, 2, \dots$. In equilibrium statistical mechanics, a complete description of a system consists of a specification of the probability, $P(i)$, for each state i . This takes the simple form

$$P(i) = \frac{e^{-E(i)/kT}}{\sum_{i=0}^{\infty} e^{-E(i)/kT}} \quad (4.1)$$

where $E(i)$ represents the energy of state i . The probability that the system is in state i is proportional to the Boltzmann factor, $\exp[-E(\text{state})/kT]$, and all states of a given energy are thus equally probable.

A quantity of special importance is the denominator of the above expression,

$$q = \sum_{i=0}^{\infty} e^{-E(i)/kT} \quad (4.2)$$

where q is a temperature-dependent pure number termed the *partition function* of the system (note that its magnitude depends on the choice of the zero of the energy scale). The partition function can be related to the variables of classical thermodynamics. The mean energy of the system is given by an expression involving a constant-volume derivative of the partition function

$$\bar{E} = kT^2 \left(\frac{\partial \ln q}{\partial T} \right)_v \quad (4.3)$$

as is the entropy

$$S = \left[\frac{\partial}{\partial T} (kT \ln q) \right]_v \quad (4.4)$$

The Helmholtz free energy is

$$\mathcal{F} = -kT \ln q \quad (4.5)$$

and the pressure is given by a constant-temperature derivative

$$\begin{aligned}
p &= -\left(\frac{\partial \mathcal{F}}{\partial V}\right)_T \\
&= kT \left(\frac{\partial \ln q}{\partial V}\right)_T
\end{aligned} \tag{4.6}$$

Paralleling the quantum case, in classical statistical mechanics it is common to consider a system that is in thermal contact with a heat bath, yet has a bath-independent energy function, $\mathcal{E}(\text{state})$. For a mechanical system, a state is defined as a point in the phase space defined by the set of position coordinates $q_1, q_2, q_3, \dots, q_n$ and associated momentum coordinates $p_1, p_2, p_3, \dots, p_n$, where n is three times the number of atoms. Here, a complete description consists of a specification of the probability density function (PDF) over phase space. The fundamental result is

$$f_{\text{state}}(\text{state}) = \frac{e^{-\mathcal{E}(\text{state})/kT}}{\int \dots \int_{p, q} e^{-\mathcal{E}/kT} d p_1 d q_1 d p_2 d q_2 \dots d p_n d q_n} \tag{4.7}$$

where the PDF $f_{\text{state}}(\text{state})$ is the probability of occupancy per unit volume of phase space associated with each point in that space.

The denominator of the above expression, together with a factor demanded by the correspondence principle, defines the *classical partition function*

$$q_c = (2\pi\hbar)^{-2n} \int \dots \int_{p, q} e^{-\mathcal{E}/kT} d p_1 d q_1 d p_2 d q_2 \dots d p_n d q_n \tag{4.8}$$

which (in the classical approximation) can be used as the value of the partition function in Eq. (4.7–4.8).

In nanomechanical design, a frequent concern is the probability that a system will be found in a particular configuration at a particular time. Molecular mechanical systems can usually be described in terms of motion on a single potential energy surface, and the total energy can be divided into potential energy and kinetic energy terms

$$\mathcal{E}(\text{state}) = \mathcal{V}(\text{position}) + \mathcal{T}(\text{momenta}) \tag{4.9}$$

With this division, Eq. (4.7) can be factored

$$f_{\text{state}}(\text{state}) = \frac{e^{-\mathcal{V}(\text{position})/kT} e^{-\mathcal{T}(\text{momenta})/kT}}{\int \dots \int_{p, q} (e^{-\mathcal{V}/kT} d q_1 d q_2 \dots d q_n) (e^{-\mathcal{T}/kT} d p_1 d p_2 \dots d p_n)} \tag{4.10}$$

and by integrating over the momentum coordinates of the phase space, a PDF referring the position coordinates alone (the PDF in configuration space) can be obtained

$$f_{\text{position}}(\text{position}) = \frac{e^{-V(\text{position})/kT}}{\int \dots \int_q e^{-V/kT} dq_1 dq_2 \dots dq_n} \quad (4.11)$$

Note that the probability density associated with a configuration is (save for a normalization factor) dependent purely on the potential energy of that configuration. The distribution of momenta is independent of location

$$f_{\text{momenta}}(\text{momenta}) = \frac{e^{-T(\text{momenta})/kT}}{\int \dots \int_p e^{-T/kT} dp_1 dp_2 \dots dp_n} \quad (4.12)$$

and hence the mean kinetic energy of the system is the same in all configurations.

Classical statistical mechanics is frequently a useful approximation. Chapter 5 examines positional PDFs for a variety of elementary nanomechanical systems, comparing the results of quantum and classical models; its results indicate the limits within which classical statistical mechanics yields results adequate for evaluating nanomechanical engineering systems.

4.3.3. The configuration-space picture

Although it yields no new physical information, it can be helpful to regard a classical mechanical system containing n atoms as a moving point in a configuration space of $3n$ dimensions, in which each of the three Cartesian coordinates of each atom corresponds to one dimension. Adding a single “vertical” dimension to represent potential energy yields a potential energy surface. The configuration point can then be imagined as moving over an undulating, frictionless surface—it may oscillate in a potential well, move along a valley, move from well to well through a col between peaks, and so forth.

To make this dynamical picture work out properly, the configuration-space coordinates corresponding an atom must vary in proportional to the Cartesian space coordinates of the atom multiplied by $m^{-1/2}$, where m is the mass of the atom. The kinetic energy of the coordinate point is then an isotropic function, depending only on the square of the speed.

In configuration space, a linear, elastic system corresponds to a point moving in a single potential well (neglecting translational and rotational degrees of freedom). For a two-atom system, this is a one-dimensional parabola. For a non-collinear n -atom system,

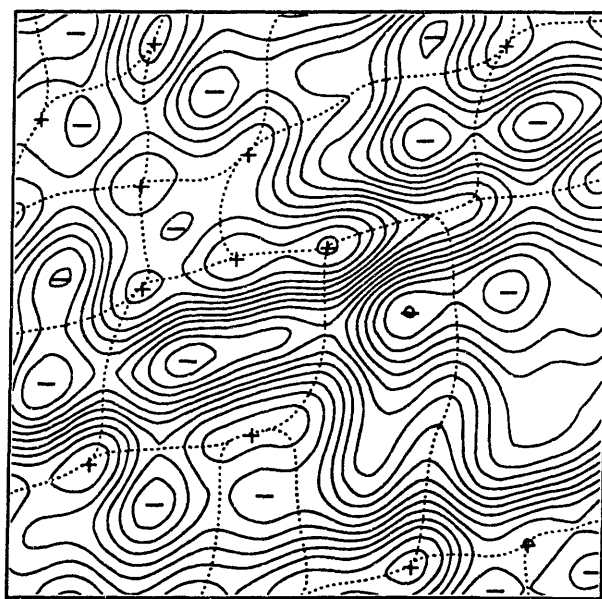
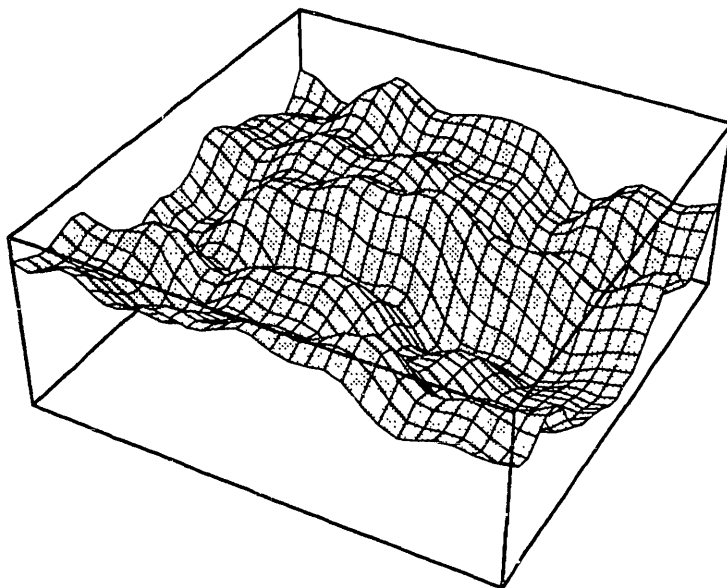


Figure 4.1. Definition of states in terms of potential energy minima. The upper illustration shows an arbitrarily-chosen potential energy surface defined over a two dimensional configuration space; below is the same surface shown as a contour map and partitioned into regions corresponding to local potential energy minima. (Figures 4.2 and 4.3 illustrate configurations and corresponding state-defining minima for *n*-octane in a molecular dynamics simulation.)

the potential well retains a parabolic form along any line through the equilibrium point, and the isopotential surfaces are concentric $3n-6$ dimensional ellipsoids; each of the $3n-6$ axes of an ellipsoid represents the line of motion of a normal mode. A non-linear system might permit (for example) the interchange of two atoms given a sufficiently great thermal excitation; the associated potential surface would have two wells joined by a pass.

For a classical system at thermal equilibrium with a heat bath, statistical mechanics asserts that the probability density of the configuration point is an inverse exponential function of the energy, here represented by the height. Thus, the probability density varies across the configuration-space landscape much as the atmospheric density varies across a real landscape. The configuration-space point is like a gas consisting of a single molecule, with a well defined mean density, flux, and so forth, at every point.

As in an equilibrated atmosphere (unlike Earth's convecting atmosphere), the mean

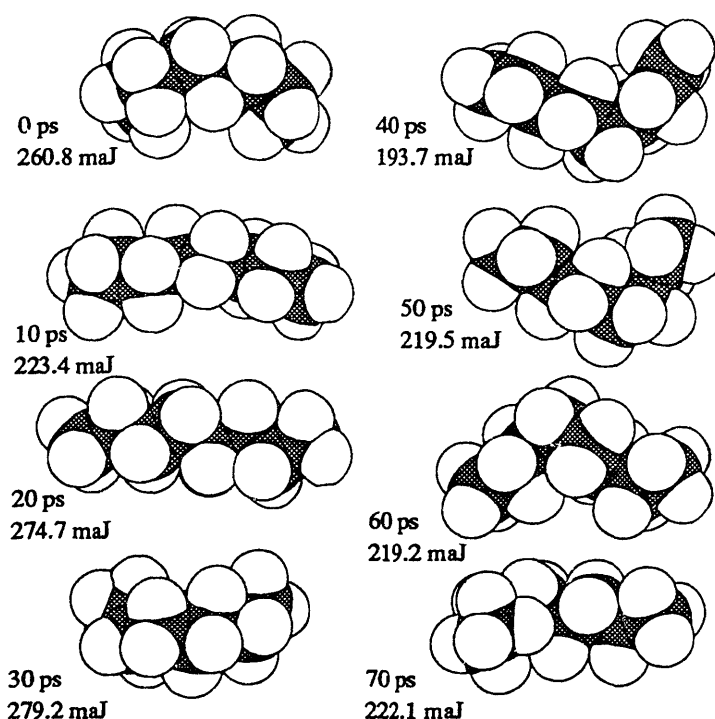


Figure 4.2. Molecular dynamics simulation of *n*-octane at 400 K (MM2/C3D+). The frames illustrate molecular configurations at 10 ps intervals, starting after 10 ps of equilibration at the target energy. Each configuration is labeled with its potential energy relative to the minimum-energy configuration for the molecule. Conformationally-mobile structures of this sort are unsuitable for most nanomechanical applications.

kinetic energy, and hence the temperature, is independent of the height of the land. The equilibrium ratio of the total probability in two connected valleys depends on their effective volumes; these depend on size and altitude, which correspond to the entropy and energy of the associated states. The rate at which probability diffuses through a col between the valleys depends on the height and width of the pass, and on the mean speed and overall probability density of the configuration point. All of these factors appear in transition-state theory (Chapter 6).

In statistical mechanics, the principle of detailed balancing asserts that, at equilibrium, the mean rate of transitions from state A to state B will equal that from state B to state A for all pairs of states. For states defined as regions in configuration space, this has an intuitive interpretation. At equilibrium, gas molecules will cross any arbitrarily-

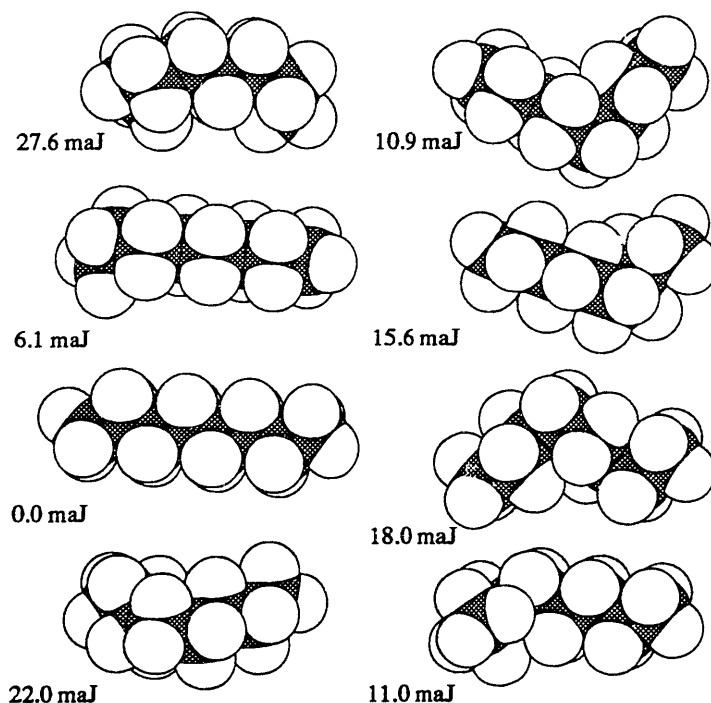


Figure 4.3. Conformers corresponding to the frames in Fig. 4.2. The configurations in Fig. 4.2 correspond to points on a PES like that illustrated in Fig. 4.1; the configurations above correspond to the associated state-defining minima. Note that energy differences between conformers are frequently small. Treating all atoms as distinguishable, the number of distinct conformers for octane is on the order of $3^7 = 2187$ (but this is reduced by excluded-volume effects).

defined surface element at equal rates from both sides; this will likewise be true for the configuration-point gas, and for each surface element of the boundary separating any two states.

The configuration space picture suggests one natural way to define what is meant by “distinct states” of a solid or liquid system (Stillinger and Weber 1984). From each point on the energy landscape, there exists a path of steepest descent, and that path will always end in a point or region of locally-minimum energy. Thus, points correspond to minima, and each local energy minimum can be taken to mark a distinct state, as indicated in

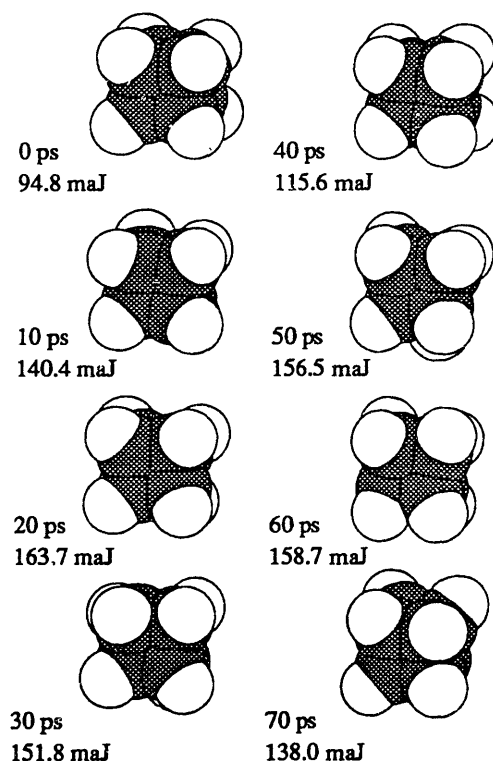


Figure 4.4. Molecular dynamics simulation of pentacyclo[4.2.0.0^{2,5}.0^{3,8}.0^{4,7}]octane (cubane) at 400 K (MM2/C3D+); conditions as in Fig. 4.2. The polycyclic structure of cubane, unlike the open-ended chain of *n*-octane, is representative of structures suitable for nanomechanical systems. It is stiff and lacks alternative conformations, hence the shape is insensitive to small errors in the PES and thermal excitation results in relatively small deformations. (The size and opacity of formal IUPAC names like that given above for cubane grows rapidly with molecular size; no attempt will be made to give formal names to structures like those in Fig. 1.1.)

Figure 4.1. Minima separated by barriers small compared to kT can often be regarded as a single minimum. A nanomechanical system containing a good bearing, for example, will exhibit a chain or loop of minima having essentially the same depth and separated by barriers of negligible height. For all practical purposes, a line or loop of this sort will constitute a single potential well. In a well-designed nanomechanical system, passes between potential wells will either be functional parts of the design, or they will be high enough to block any substantial probability flux.

Even without being able to visualize interconnected, approximately-ellipsoidal potential wells, or ring-shaped valleys in a high-dimensional space, one can get a sense of the strongly-constrained nature of the motion of the configuration point in such systems. A similar description of a chemically-reacting liquid-phase system lacks such well-defined features. Each possible covalent combination of atoms forms a separate valley, and the valleys intertwine in a manner that brings each into contact with many others. Which cols are available—which reactions can occur at appreciable rates—depends on the relative heights and widths of the numerous connections between valleys. A chemist attempting to control the pattern of reactions must exploit small differences in the heights and effective volumes (the energies and entropies) of cols and valleys. In a nanomechanical system, in contrast, a reaction might occur between two groups brought together by a gear-like rotary motion. The available configuration space would consist of two ring-shaped valleys linked by a single col of modest height. Unwanted reactions would be prevented by gross mechanical constraints, not by small differences in energy and entropy.

The configuration-space picture is most useful when applied to a subsystem that is coupled to a larger system that acts as a heat bath. This can be described in terms of the potential surface that would arise if the rest of the system were motionless and fully relaxed for all subsystem configurations, combined with a time-varying perturbing potential representing the effects of thermal vibrations external to the subsystem. In this picture, the landscape vibrates, and total energy is not a constant of motion.

4.3.4. Equilibrium vs. non-equilibrium processes

The relationships cited in Section 4.3.2 describe equilibrium systems, but a functioning nanomechanical systems will seldom be at equilibrium. How useful, then, is equilibrium statistical mechanics?

It is perhaps worth noting that equilibrium statistical mechanics is seldom applied to a true equilibrium system. For matter under ordinary conditions of temperature and density, the equilibrium state is crystalline iron—if one allows equilibration of all nuclear degrees of freedom. In practice, the necessary reactions (e.g., fusion) proceed so slowly that they can be ignored. States (or dynamical domains) that are sufficiently metastable can be treated as stable, in calculations of “equilibrium” properties.

In conventional mechanical engineering, mechanical motion is clearly distinguished from thermal motion. In nanomechanical systems, this distinction can often be drawn with a useful degree of clarity. In mechanical systems of all sizes, it is typical for frictional forces to convert mechanical energy into thermal energy in a spatially inhomogeneous fashion, causing thermal gradients. In estimating the fluctuations resulting from thermal motion, small regions of matter in such systems can then be approximated as being at thermal equilibrium save for deviations associated with temperature gradients, heat flows, and changing temperatures. When these deviations are small, equilibrium models will give good estimates of the statistics of thermally-induced displacements and motions in nanomechanical systems.

A small thermal gradient is one that produces only a small difference in the absolute temperature across the diameter of the system under analysis. For a nanomechanical system with a diameter of 10 nm, a 1% ΔT (at 300 K) would require a gradient of 3×10^8 K/m. Assuming a thermal conductivity of 10 W/m·K, this would yield a thermal power flux of 3×10^9 W/m². This “small” thermal gradient would produce a large (i.e., often unacceptable) temperature difference of 300 K in a modest 10^{-6} m distance.

A small thermal flux is one that produces only a small difference in the equilibrium distribution of thermal vibrations. This equilibrium distribution is characterized by large power fluxes (which cancel, at equilibrium) on the order of $(\text{speed of sound}) \times (\text{thermal energy})/(\text{volume})$. For a typical material at ordinary temperatures, this is on the order of $(3 \times 10^8 \text{ J/m}^3) \times (10^3 \text{ m/s}) = 3 \times 10^{11} \text{ W/m}^2$. One percent of this value would again correspond to a net thermal power flux of $3 \times 10^9 \text{ W/m}^2$. Since thicker layers would result in unacceptable values of ΔT , a “small” power flux of this magnitude would be encountered in a working system only if a $\sim 10^{-6}$ m (or less) thick slab dissipated power at $\sim 10^{15} \text{ W/m}^3$ (or more). Despite the extraordinarily high power-conversion densities possible in small components (see Sec. 2.2.2), the power-dissipation densities for complex, multi-component systems will typically be small compared to this value.

A small rate of change of temperature is one that produces only a small ΔT during the characteristic vibrational relaxation time. In a crystal, a measure of this relaxation time is the phonon mean free path divided by the speed of sound; a typical value might be $(10^{-8} \text{ m})/(10^3 \text{ m/s}) = 10^{-11} \text{ s}$. In a highly inhomogeneous medium (e.g., a typical nanomechanical system), with highly anharmonic interactions (e.g., van der Waals contacts between moving parts), relaxation times will often be shorter. A 1% ΔT (at 300 K) during a single relaxation time would then correspond to a rate of temperature change in excess of $3 \times 10^{11} \text{ K/s}$. For typical volumetric heat capacities ($\sim 10^6 \text{ J/K}\cdot\text{m}^3$), this “small” rate of temperature change would require a large power dissipation (as above, $\sim 10^{15} \text{ W/m}^3$ or more).

Because these “small” thermal gradients, thermal fluxes, and rates of temperature change are all so large, it will often be acceptable to divide motions into mechanical and thermal components, describing the latter in terms of the local temperature and the relationships of equilibrium statistical mechanics. The chief exceptions will be in descriptions of the processes that convert mechanical motion into thermal excitation, where nonequilibrium vibrational motions are generated and then thermalized.

A typical nonequilibrium event is a transition from one potential well to another that is forced to occur at a particular time by an input of mechanical energy. In the configuration-space picture, the configuration point representing the subsystem passes through a col owing to an imposed change in the shape of PES, and is thus injected into the new potential well with an unusually high energy and a somewhat-predictable momentum. Regarding the new well as nearly quadratic, the initial state can be viewed as one with a disequilibrium distribution of excitation of normal modes. Relaxation will then involve two processes: a tendency toward equilibration of the distribution of modal excitation (given an unusually high total energy) *via* anharmonic interactions and a tendency toward equilibration of the expected total energy *via* interactions with a heat bath. For a general anharmonic system in weak contact with a thermal bath, a short-term description of the motion would describe families of trajectories; an intermediate-term description would treat all states of equal energy as equally probable, but would take account of the slowly-decaying excess energy; and the long-term description would be in terms of the statistics of fully equilibrated thermal motion.

4.3.5. Entropy and information

The transition from a detailed dynamical description to a statistical description entails discarding information. State variables that are regarded as having definite values in the dynamical description are regarded as having probability distributions in the statistical description. In quantum statistical mechanics, distinct states can be enumerated, and the entropy

$$S = -k \sum_{\text{states}} P(\text{state}) \ln[P(\text{state})] \quad (4.13)$$

is a weighted measure of the number of possible states. This expression yields the familiar result (the third law of thermodynamics) that a perfect crystal at absolute zero has zero entropy: the structure of the crystal is known and it is in the vibrational ground state with unit probability; $1 \times \ln(1) = 0$, hence $S = 0$. This result is equally true for any completely-specified structure at absolute zero.

“Probability”, however, depends on knowledge. If I flip a coin and peek at it, it may be heads with probability one for me, while remaining heads with probability 1/2 for you. This suggests that the entropy of an irregular solid at absolute zero may be positive if these irregularities are unknown, but zero if they are completely described by some algorithm or external record (the “complete specification” just alluded to). If so, then entropy is not a local property of a physical system.

Studies of entropy and information indicate that this is in fact the case. A structure (such as a polymer with a seemingly random sequence of monomers) can be made to yield more free energy if it is matched against another polymer with a sequence known to be identical—that is, if one has an external record representing “knowledge” of the sequence (Bennett 1982). The Helmholtz free energy of the polymer

$$\mathcal{F} = \mathcal{E} - TS \quad (4.14)$$

is larger, because knowledge of the sequence eliminates many otherwise-possible structures, and thus lowers S .

These issues are intimately related to questions regarding the theoretical energy requirements of computation. In the early 1960s, Landauer observed that compressing the logical state of a computer entails compressing its physical state (Landauer 1961); for example, erasing or overwriting a one-bit storage device with unknown contents entails reducing its possible states (one or zero) to a single state (e.g., zero). Such a transforma-

tion dissipates a quantity of free energy $\mathcal{F} \geq \ln(2)kT$ per bit erased (Sec. 7.6.3).

Theorists have examined devices such as Brownian computers with idealized structures (i.e., hard, rigid parts of arbitrary shape) but subject to realistic dynamics and thermodynamics, as well as computational devices inspired by the molecular machinery of biological systems (Bennett 1982), idealized, deterministic mechanical systems (Fredkin and Toffoli 1982; Toffoli 1981), and quantum mechanical models (Feynman 1985; Likharev 1982). These and related studies indicate that logically-reversible computations—those where the *output* uniquely specifies the *input*—can be performed in a manner approaching thermodynamic reversibility (that is, the dissipation of free energy per logically-reversible operation can be made arbitrarily close to zero). The literature and results in this field have been well reviewed (Bennett 1982; Landauer 1988). They provide an improved understanding of the second law of thermodynamics (and the impossibility of a Maxwell’s Demon), and are of direct relevance to nanomechanical systems.

4.3.6. Uncertainty in nanomechanical systems

In conventional mechanical systems, the positions and shapes of components are never completely known: manufacturing tolerances, measurement inaccuracies, and environmental vibrations ensure this. When systems work despite these uncertainties, it is because the uncertainties are kept within tolerable bounds, either by reducing the uncertainties or by expanding the range of tolerance.

In nanomechanical systems, quantum and statistical mechanics place firm limits on the reduction of uncertainties. For a structure of a given mass and stiffness at equilibrium at a given temperature, positional uncertainties are fixed and irreducible (Chapter 5). Depending on the design of the system, these uncertainties may cause errors at a rate ranging from negligibly low to unacceptably high.

There are strong *qualitative* parallels between the uncertainties of quantum mechanics and those of classical statistical mechanics. In both quantum and statistical mechanics, one begins with a potential energy function, $\mathcal{V}(\mathbf{r})$. In quantum mechanics, the Schrödinger equation ensures that a particle of a given energy has a non-zero (though often vanishingly small) probability of penetrating a potential barrier of any finite height and thickness, and of being found in any region of space. In classical statistical mechanics at finite temperature, Boltzmann’s law yields the same qualitative result by assigning a non-zero (though often vanishingly small) probability to states of arbitrarily high energy. In quantum mechanics, a linear system with a certain mass and stiffness has a positional uncer-

tainty characterized by a gaussian probability distribution; classical mechanics at a finite temperature yields a result of the same form.

Quantum uncertainties measured as the product of the uncertainties in conjugate variables have an irreducible minimum, e.g.,

$$\Delta x \Delta p \geq \frac{\hbar}{2} \quad (4.15)$$

but the uncertainty in either variable can be reduced to an arbitrary degree by a suitable measurement; classical systems permit similar measurements, but present the illusion that the other variable can simultaneously be specified with arbitrary accuracy as well. This difference in the reducibility of uncertainty is, however, irrelevant in the context of equilibrium statistical mechanics. A measurement that reduces uncertainty also reduces entropy (Section 4.3.5) and hence disturbs the equilibrium of the system *from the perspective of the observer*, even in the absence of a physical disturbance. Thus, within the equilibrium statistical mechanical description, uncertainties are irreducible, by definition.

In a nanomechanical system, each component will have many vibrational degrees of freedom, each subject to thermal excitation. Any attempt to use nanomechanical components within a system to represent the results of measurements performed on other nanomechanical components within a system will succeed in encoding only a small fraction of the information needed to represent the total vibrational state of the system. While one can imagine a device that uses measurement to reduce the uncertainty associated with one or a few critical components, the system as a whole will be dominated by components subject to statistical uncertainties that are, in practice, irreducible and that will (as indicated by Section 4.3.4) typically be well-described by equilibrium statistical mechanics. Components and systems of this sort are the focus of the present work.

4.4. PES revisited: accuracy requirements

Both dynamical and statistical mechanical models of molecular behavior depend on potential energy surfaces that are (in all cases of nanomechanical interest) approximations known to deviate from reality. The scientific literature on potential energy surfaces describes efforts to improve the correspondence between experiment and theory, and hence focuses on the imperfections and limitations of existing models. In order to understand the utility of existing models from an engineering perspective, it is useful to consider the sensitivity of different physical phenomena to the existing inaccuracies.

4.5.1. *Physical accuracy*

In chemical physics, experiments are designed to provide stringent tests of theoretical models of molecular systems (including their potential energy surfaces), and theoretical models attempt to predict everything that can be experimentally observed. As discussed in Chapter 3, physicists have made extensive use of molecular beam experiments in which molecules are prepared with precise momenta (and sometimes with control of vibrational states, rotational states, and polarization); they are then allowed to scatter (sometimes with a reactive exchange of atoms) and outcomes are observed and analyzed in terms of scattering angle (etc.)

The quantum interference effects that can be observed in such experiments provide a delicate test of potential energy surface models. Scattering events that involve bond formation and cleavage can traverse energy barriers of over 10^{-19} J, yet the interference effects are sensitive to much smaller energy differences. A characteristic molecular collision time is $>10^{-13}$ s (the time required to travel 10^{-10} m at 10^3 m/s); changing the potential energy along one of the interfering paths by 10^{-21} J or less will shift the phase of that path by a radian and cause a substantial change in the interference pattern. Since different trajectories explore different parts of the configuration space, reproducing interference effects can require accuracy of this magnitude across the entire dynamically-accessible potential energy surface.

4.5.2. *Chemical accuracy*

In solution chemistry, a standard challenge is to predict chemical equilibria, absolute reaction rates, and the relative rates of competing reactions. Synthetic chemistry can be viewed as an engineering discipline aimed at constructing molecules. In this task, rates and equilibria are of central importance: if a reaction equilibrates several species, then the yield of the desired product will depend on the equilibrium concentration ratios; alternatively, if a reaction can proceed along any of several effectively-irreversible paths, then the yield of the desired product will depend on the ratio of the reaction rates. In some reactions, yields are affected by both rates and equilibria.

If entropic factors are equal, then the equilibrium ratio of two species will be an exponential (Boltzmann) function of the difference in potential energy between the species. Likewise, if entropic factors (and certain dynamical factors) are equal, then the ratio of the rates of two competing reactions will be an exponential function of the difference

in potential energy between the two transition states. At 300 K, a difference of 1 mJ changes rates and equilibria by a factor of 1.27, a 10 mJ difference results in a factor of 11, and a 100 mJ difference results in a factor of 3.1×10^{10} . To a practicing chemist, an energy difference of 10 mJ *between* two competing species or transition states which changes the yield of a reaction from 8% to 90% typically matters more than would a 100 mJ shift in all energies (causing no change in the course of the reaction) or a 100 mJ shift in the transition-state energies which slows the reaction-completion time from a microsecond to an hour. In discussions of molecular energies, the phrase “chemical accuracy” is typically taken to imply errors of somewhat less than 10 mJ per molecule in describing the energies of chemical species (potential wells) and transition states (cols). Aside from entropic effects dependent on the breadth of potential wells and cols, reaction rates typically exhibit only modest sensitivity to the shape of the potential energy surface.

Potential energy functions are also used to predict molecular structures; MM2 has good success for a wide range of small organic molecules. A more challenging test is protein modeling, where the shape and stability of the folded protein molecule depend on a delicate balance of free energy terms in which van der Waals interactions, hydrogen bonding, torsional strains, and entropic factors all play crucial roles. The net stability of a folded structure is typically $\sim 50\text{--}100$ mJ, or ~ 0.01 mJ per atom. Although present molecular mechanics potentials are good enough to have found extensive use in protein modeling and design, their errors are large compared to the free energy of folding. Further, energy minimization typically yields structures that differ substantially from those determined by x-ray diffraction, even when the latter are taken as a starting point.

Even for small organic molecules, a slightly-inaccurate molecular mechanics model can predict structures that are totally wrong. Most molecules of concern in organic chemistry can exist in any of a number of conformations, differing by rotations about bonds; a simple example is *n*-octane, Figure 4.2. The relative energies of molecular conformations are sensitive to weak, nonbonded interactions and to torsional energies. Interconformational equilibria, like other chemical equilibria, are strongly altered by energy differences of ~ 10 mJ. Crystal structures (a common source of geometric data) represent one of many possible molecular packing arrangements, again sensitive to weak forces. A small error in relative energies can result in a predicted crystal structure containing the wrong conformation packed in a lattice of the wrong symmetry.

4.5.3. Accurate energies and nanomechanical design

Nanomechanical systems of the sort considered here will be organic structures that resemble (or exceed) proteins in size, and some will be used to perform chemical reactions. It is thus important to consider the sensitivity of nanomechanical designs to errors in potential energy surfaces.

It will be possible to design nanomechanical systems that are exquisitely sensitive to the properties of a potential energy surface. For example, successful operation might be made to depend on a ratio of competing transition rates, as in conventional chemistry, or even on interference phenomena in angle-resolved scattering in crossed molecular beams. In general, any measurable physical property can be made essential to the correct operation of a suitable Rube Goldberg device, and hence any known class of unpredictable discrepancies between model and experiment can be used to design a class of devices that cannot with confidence be predicted to work. The design of such sensitive devices is closely related to good design practice in instrumentation and scientific experimentation, but is the opposite of good design practice in conventional engineering.

In general, the robustness, predictability, durability, and performance of nanomechanical designs will be maximized if they are made of strong, stiff materials. These materials will typically resemble diamond in possessing structures consisting of highly polycyclic, three-dimensional networks of covalent bonds. Predictions of the stability and geometry of rigid, polycyclic structures will be far less sensitive to small errors in potential energy surfaces than are similar predictions for folded protein structures or conformationally-mobile organic molecules. Figure 4.4 illustrates the results of a molecular dynamics simulation of a cyclic octane structure (cubane) under conditions like those of Figure 4.4: only one conformation is available, and deformations are purely vibrational.

Increasing structural stiffness tends to mitigate errors resulting from inaccurate potential energy functions. Good design practice can increase tolerance for the errors that remain. Nonetheless, errors in predicting molecular geometry can pose problems where the resultant discrepancies in shape might interfere with correct operation of a device. Further, even small errors in local geometry can have cumulative effects in large structures. Appendix I discusses how such problems can be overcome through experimentation in standard engineering and through design flexibility in exploratory engineering.

Errors in models of chemical reactions play a different role. If the correct sequence of reactions occurs in building a molecular structure, then the correct structure will result,

regardless of errors in describing reaction rates or transition states. Distinct structures are distinct quantum states (or families of quantum states, allowing for thermal excitations), and the manner of construction has no effect on the nature of the product. In conventional manufacturing, small variations in fabrication steps can have cumulative effects on properties such as product geometry, but no parallel problem arises in molecular manufacturing.

In comparison to conventional chemical synthesis, molecular manufacturing processes based on positional control of synthetic reactions will be less sensitive to small energy differences. Individual reaction steps can be driven to completion either by ensuring that energy differences greatly favor the product state, or by repeating trials until a molecular measurement verifies successful completion. To achieve regio- and stereospecificity, reactions can be guided not by differences in reaction rates and equilibria as in conventional synthesis, but by rigid control of reagent access to different sites. In chemical terms, this can create arbitrarily large ratios of effective reagent concentration between molecular sites that might otherwise be equally reactive. With positional control of synthesis, the main issues are achieving a high enough reaction rate (fast reactions will best exploit the speed of nanomechanical systems) and avoiding (or being able to reverse) unwanted reactions immediately adjacent to the target site. These and related matters are discussed in more detail in Chapter 8.

4.5. Further reading

Molecular dynamics is a broad field with an enormous literature. Basic statistical mechanics is well described in many textbooks, for example (Knox 1971). Trajectory-based molecular dynamics, being more model dependent and computation intensive, is in a greater state of ferment. A good overview of the dynamics of reactive molecular collisions is (Levine and Bernstein 1987), which includes many references.

Chapter 5

Positional uncertainty

5.1. Uncertainty in engineering

In the design of nanomechanical systems, positional uncertainties stemming from thermal excitation and quantum mechanical principles are a fundamental concern. On the scale of conventional mechanical engineering, neither quantum uncertainties nor thermal excitations are significant; the closest macroscopic analogs of these effects arise in systems excited by broad-band noise, yet issues arise there (fatigue as a problem, damping as a solution) that are alien to the molecular domain. Individual bonds are not subject to fatigue, while damping, which degrades mechanical energy to heat, cannot dissipate the vibrations of heat itself.

In an engineering context, problems involving positional uncertainty can frequently be formulated in terms of a probability density function (PDF) for a coordinate describing a part of a system. Typically, a system is designed such that a part should, under specified conditions, occupy a particular position at a particular time to within specified tolerances; these issues are addressed in more detail in the next chapter. Errors occur at some finite rate owing to the finite probability associated with the tail of the PDF that extends beyond the tolerance band. Good approximations to the positional PDFs of typical systems are thus of fundamental value in nanomechanical engineering.

This chapter examines two fundamentally different sources of positional uncertainty, compliance associated with elastic forces and compliance associated with entropic effects. The prototype of the former is the harmonic oscillator; more complex elastic systems include rods subject to thermal excitation of stretching and bending modes. The prototype of the latter is the piston sliding in a gas-filled cylinder; more complex entropic systems include rods subject to length fluctuations resulting from excitation of transverse vibrational modes. In each analysis, the focus is on the exposition of basic principles and

derivation of useful engineering approximations. Graphical summaries are included for ready reference in estimating the relative importance of different effects and the magnitudes of the resulting positional variances.

5.2. Thermally excited harmonic oscillators

Many parts of nanomechanical systems are adequately approximated by linear models, with restoring forces are proportional to displacements. The prototype of such systems is the harmonic oscillator, consisting of a single mass with a single degree of freedom subject to a linear restoring force (measured by the stiffness, k_s). Analytical results for the simple harmonic oscillator can be adapted and extended to systems with multiple degrees of freedom, as is done later in this chapter.

5.2.1. Classical treatment

In classical statistical mechanics, the probability density function for the position coordinate, x , of a particle subject to the potential energy function $\mathcal{V}(x)$ is

$$f_x(x) = \frac{e^{-\mathcal{V}(x)/kT}}{\int_{-\infty}^{\infty} e^{-\mathcal{V}(x)/kT} dx} \quad (5.1)$$

For the harmonic potential,

$$\mathcal{V}(x) = \frac{1}{2} k_s x^2 \quad (5.2)$$

the resulting probability density function is Gaussian:

$$f_x(x) = \frac{e^{-k_s x^2/2kT}}{\int_{-\infty}^{\infty} e^{-k_s x^2/2kT} dx} = \frac{e^{-k_s x^2/2kT}}{\sqrt{2\pi}} \sqrt{\frac{k_s}{kT}} = \frac{e^{-x^2/2\sigma_{\text{class}}^2}}{\sqrt{2\pi}\sigma_{\text{class}}} \quad (5.3)$$

yielding the classical value for the positional variance (= standard deviation squared = mean square displacement):

$$\sigma_{\text{class}}^2 = \frac{kT}{k_s} \quad (5.4)$$

5.2.2. Quantum mechanical treatment

In quantum statistical mechanics, the classical integral over x (more generally, an integral over phase space, yielding a probability density function for both position and momentum) is replaced by a sum and the probability density function is replaced by a probability distribution over a series of states i :

$$P(i) = \frac{e^{-E(i)/kT}}{\sum_{i=0}^{\infty} e^{-E(i)/kT}} \quad (5.5)$$

From elementary quantum mechanics, the vibrational states $n = 0, 1, 2, 3, \dots$ of a harmonic oscillator have energies

$$E(n) = (n + \frac{1}{2})\hbar\omega; \quad \omega = \sqrt{k_s/m} \quad (5.6)$$

The probability of finding the oscillator in vibrational state n ,

$$P(n) = \frac{e^{-(n+\frac{1}{2})\hbar\omega/kT}}{\sum_{n=0}^{\infty} e^{-(n+\frac{1}{2})\hbar\omega/kT}} \quad (5.7)$$

Rearranging and summing the series,

$$P(n) = \frac{e^{-n\hbar\omega/kT}}{\sum_{n=0}^{\infty} (e^{-\hbar\omega/kT})^n} \quad (5.8)$$

$$\sum_{n=0}^{\infty} y^n = 1 + \frac{y}{1-y}, \quad y < 1 \quad (5.9)$$

$$P(n) = e^{-n\hbar\omega/kT} (1 - e^{-\hbar\omega/kT}) \quad (5.10)$$

The variance (mean square displacement) of the oscillator may be derived from its mean energy, making the latter quantity of interest:

$$\bar{E} = \sum_{n=0}^{\infty} P(n)E_n = \sum_{n=0}^{\infty} e^{-n\hbar\omega/kT} (1 - e^{-\hbar\omega/kT}) (n + \frac{1}{2})\hbar\omega \quad (5.11)$$

Rearranging and summing both series,

$$\bar{E} = \hbar\omega(1 - e^{-\hbar\omega/kT}) \left[\frac{1}{2} \sum_{n=0}^{\infty} (e^{-\hbar\omega/kT})^n + \sum_{n=0}^{\infty} n(e^{-\hbar\omega/kT})^n \right] \quad (5.12)$$

$$\sum_{n=0}^{\infty} ny^n = \frac{y}{(1-y)^2}, \quad y < 1 \quad (5.13)$$

$$\bar{E} = \hbar\omega \left(\frac{1}{2} + \frac{1}{e^{\frac{\hbar\omega}{kT}} - 1} \right) \quad (5.14)$$

In a harmonic oscillator, the total energy equals twice the mean potential energy, and the latter is proportional to the mean square displacement:

$$\frac{1}{2}\bar{E} = \bar{V} = \frac{1}{2}k_s\bar{x}^2 = \frac{1}{2}k_s\sigma^2 \quad (5.15)$$

$$\sigma^2 = \frac{\hbar\omega}{k_s} \left(\frac{1}{2} + \frac{1}{e^{\frac{\hbar\omega}{kT}} - 1} \right) \quad (5.16)$$

Describing the frequency in terms of the fundamental mechanical parameters and rearranging yields an equation between dimensionless quantities (see Fig. 5.1):

$$\sigma^2 \frac{\sqrt{k_s m}}{\hbar} = \frac{1}{2} + \frac{1}{e^{\frac{\hbar\omega}{kT}} - 1} \quad (5.17)$$

It is often desirable to determine whether the classical approximation is adequate for describing positional uncertainties in a system of engineering interest. A useful measure is the ratio of the total to the classical variance,

$$\frac{\sigma^2}{\sigma_{\text{class}}^2} = \frac{\hbar\omega}{kT} \left(\frac{1}{2} + \frac{1}{e^{\frac{\hbar\omega}{kT}} - 1} \right) = \frac{\bar{E}}{kT} \quad (5.18)$$

which equals the ratio of the total to the classical energy. This function of the parameter $kT/\hbar\omega$ is graphed in Figure 5.2; where $kT/\hbar\omega$ substantially exceeds unity, the classical variance provides a good approximation.

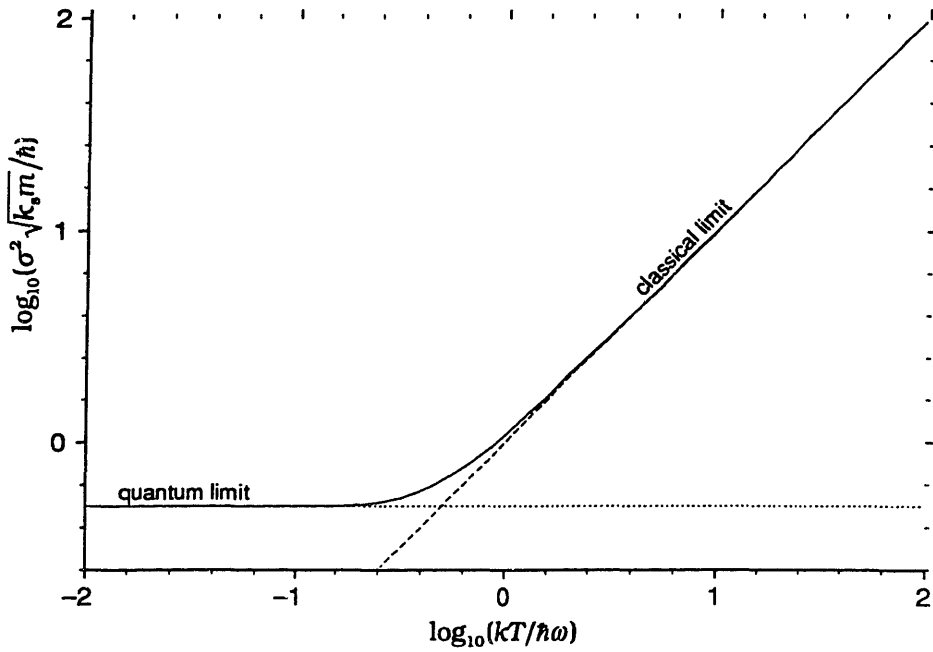


Figure 5.1. A dimensionless measure of variance vs. a dimensionless measure of temperature, Eq. (5.17).

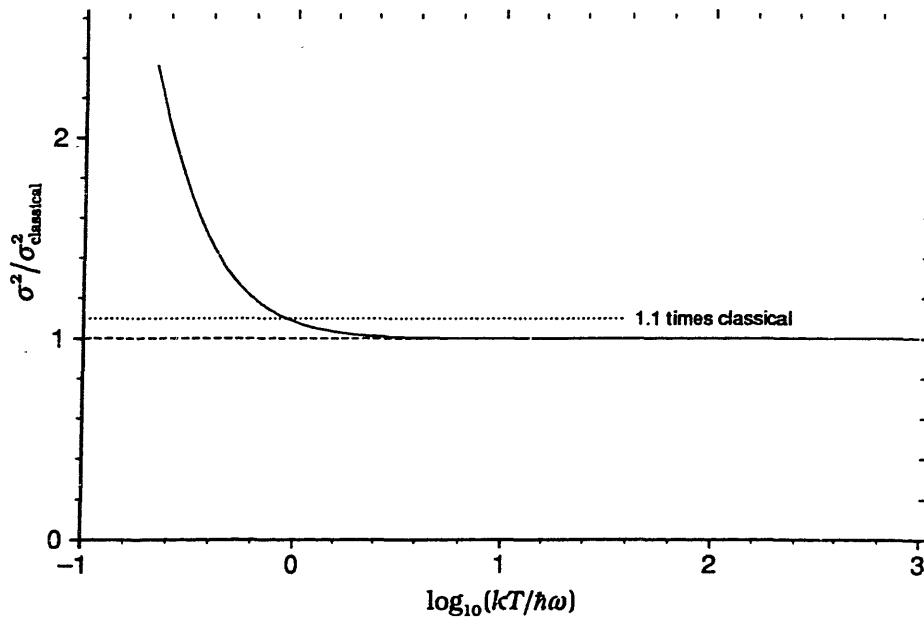


Figure 5.2. The ratio of the actual variance to that predicted by a classical model, vs. a dimensionless measure of temperature, Eq. (5.18).

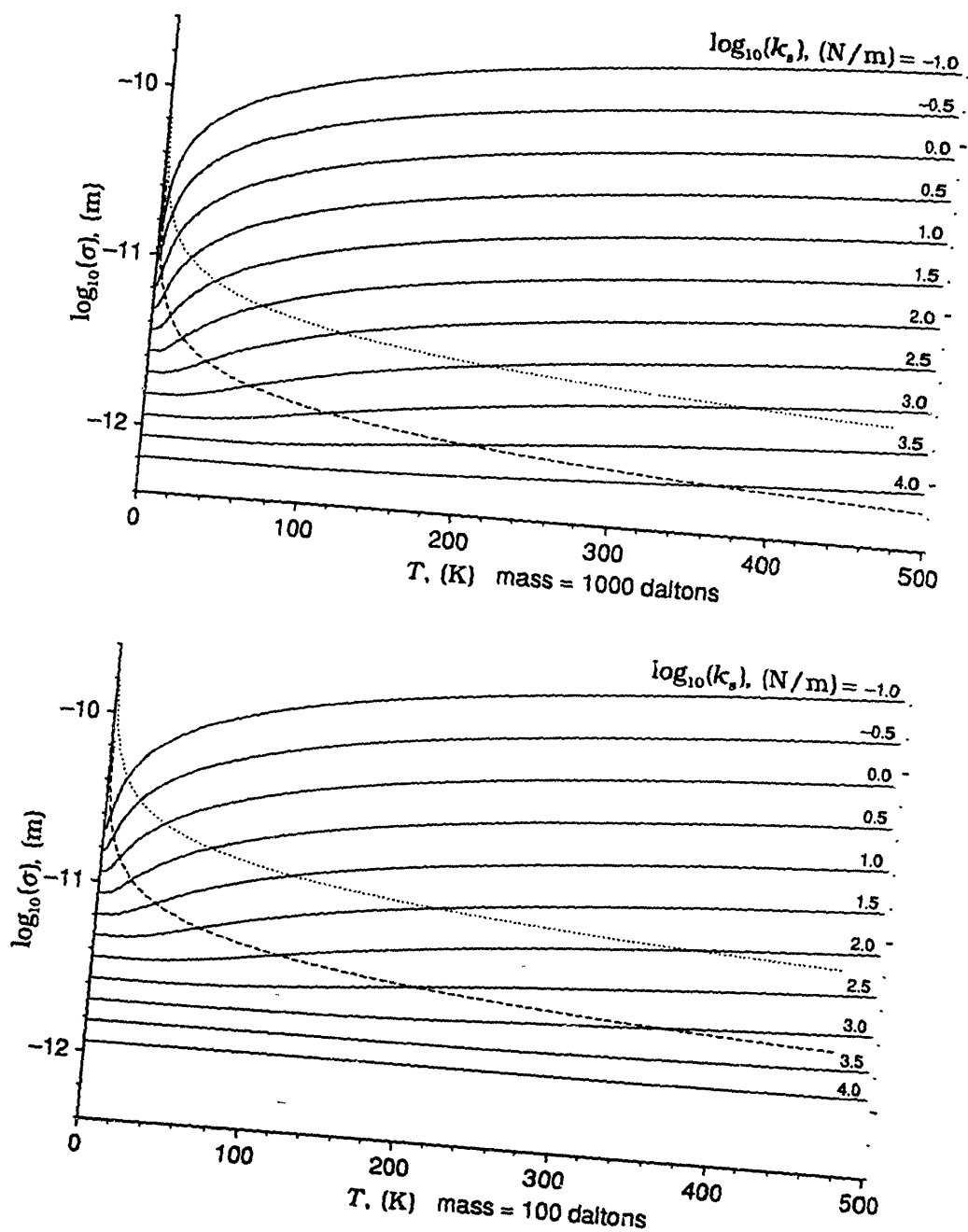
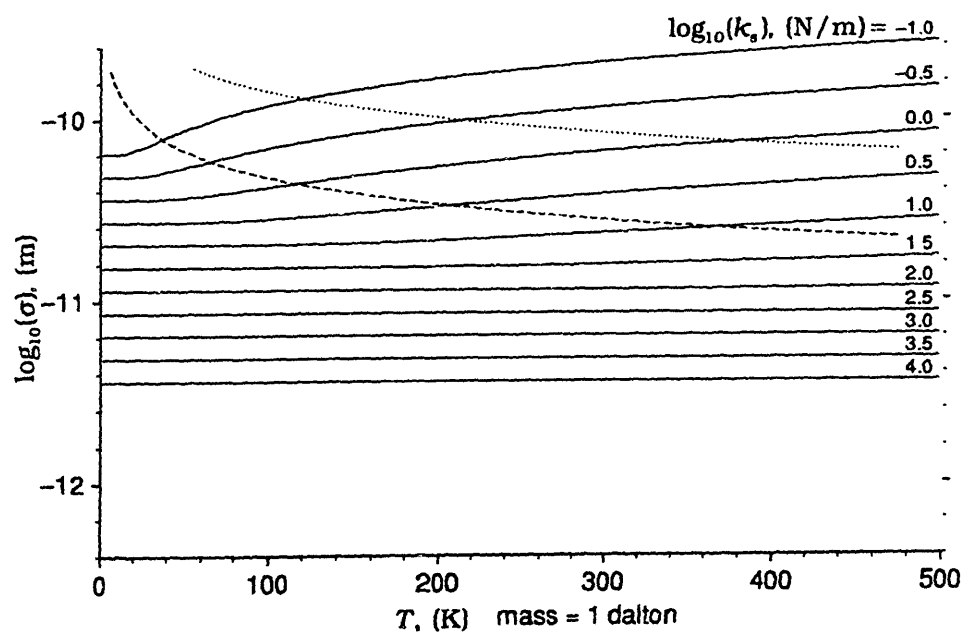
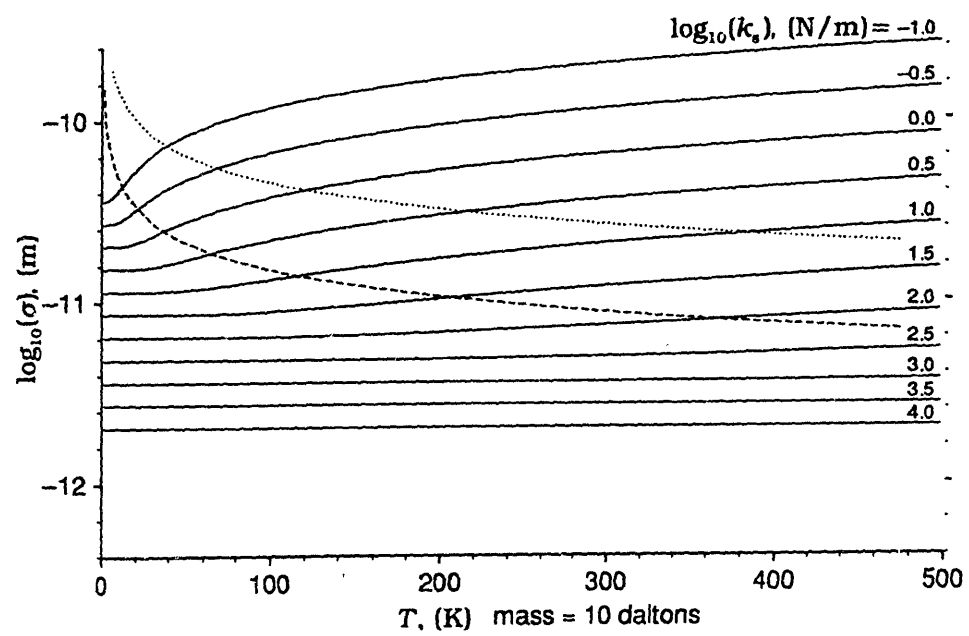


Figure 5.3. This set of graphs plots the logarithm of the rms displacement of harmonic oscillators of varying mass and spring constant as a function of temperature, from Eq. (5.15). Note that the length at the top of the graphs corresponds roughly to an atomic diameter. Above the dashed lines, the classical approximation is accurate to within 10%; above the dotted lines, it is accurate to within 1%.



With less elegance (but more immediately accessible physical significance), these dimensionless quantities can be unfolded into a set of graphs of rms displacement (σ) as a function of temperature, spring constant, and mass (Figure 5.3). For perspective, note that the stiffness of nonbonded interactions at equilibrium separation is typically about 0.1 N/m; that of covalent bond bending, about 30 N/m; and that of covalent bond stretching, about 400 N/m. Small components in nanomechanisms will typically have masses in the hundreds to thousands of daltons.

5.3. Elastic extension of thermally excited rods

A variety of molecular structures resemble rods; these include DNA helicies, microtubules, ladder polymers, and a variety of nanomechanical components. Even flexible molecular chains become rod-like (for purposes of the present analysis) when held in an extended conformation by longitudinal tension or by lateral constraints. More generally, a rod provides the simplest case of an extended object and can serve as a model for a variety of mechanical systems.

A rod clamped at one end and free at the other will undergo both transverse and longitudinal vibrations. This section analyzes the positional variance of the free end resulting from longitudinal vibrational modes; in rods of large aspect ratio, transverse vibrations can make a significant contribution to the longitudinal positional variance of the free end (analyzed in Sec. 5.6). The following first considers a classical model for a continuous elastic rod, then both an approximate and an exact quantum mechanical model for a rod consisting of a series of identical masses and springs, ending with an empirical approximation to this exact model.

5.3.1. Classical continuum treatment

A uniform rod of length ℓ clamped at one end and free at the other will support longitudinal vibrational modes ($n = 0, 1, 2, 3, \dots$) with wavelengths

$$\lambda_n = \frac{4\ell}{2n+1} \quad (5.19)$$

The speed of sound v_s in the rod may be calculated from the linear modulus E_ℓ and linear density ρ_ℓ yielding the modal frequencies ω_n .

$$v_s = \sqrt{\frac{E_\ell}{\rho_\ell}}; \quad \omega_0 = \frac{\pi}{2\ell} \sqrt{\frac{E_\ell}{\rho_\ell}}; \quad \omega_n = \omega_0(2n+1) \quad (5.20)$$

Each longitudinal mode may be regarded as a harmonic oscillator with a certain frequency, effective stiffness, and effective mass. The effective stiffness relates the square of the amplitude (at the rod end, where the variance is to be computed) to the potential energy at maximum displacement during a vibrational cycle, which equals the maximum kinetic energy:

$$\begin{aligned}
 \frac{1}{2} k_n A_n^2 &= \max(\mathcal{V}) = \max(\mathcal{T}) \\
 &= \int_0^\ell \frac{1}{2} \rho_t [v(x)]^2 dx \\
 &= \int_0^\ell \frac{1}{2} \rho_t (\omega_n A_n)^2 \sin^2 \left[\frac{(2n+1)}{2\ell} \pi x \right] dx \\
 &= \frac{1}{4} \rho_t \ell (\omega_n A_n)^2
 \end{aligned} \tag{5.21}$$

This yields the effective stiffness k_n of mode n :

$$k_n = \frac{E_t}{\ell} \frac{\pi^2}{8} (2n+1)^2 \tag{5.22}$$

Combining this with the classical expression for positional variance in a harmonic oscillator as a function of temperature and stiffness, Eq. (5.4), the positional variance of the free end associated with mode n is:

$$\sigma_{n,\text{class}}^2 = kT \frac{\ell}{E_t} \frac{8}{\pi^2} \frac{1}{(2n+1)^2} \tag{5.23}$$

The total variance is the sum of the modal variances:

$$\sigma_{\text{class}}^2 = \sum_{n=0}^{\infty} \sigma_{n,\text{class}}^2 = kT \frac{\ell}{E_t} \frac{8}{\pi^2} \sum_{n=0}^{\infty} \frac{1}{(2n+1)^2} \tag{5.24}$$

Applying the identity

$$\sum_{n=0}^{\infty} \frac{1}{(2n+1)^2} = \frac{\pi^2}{8} \tag{5.25}$$

yields the classical variance in the position of the free end for a rod of uniformly distributed mass and elasticity:

$$\sigma_{\text{class}}^2 = kT \frac{\ell}{E_t} \tag{5.26}$$

This is exactly kT/k_s , where k_s is the stretching stiffness of the rod as a whole. This is, of course, the variance for a simple harmonic oscillator of the same stiffness, and hence of a “rod” having a single mass and a single mode—the opposite extreme from a rod of uniformly distributed mass and elasticity. The general significance of this identity is discussed in Section 5.7.

5.3.2. Quantum mechanical treatment

For a continuous rod, a quantum mechanical treatment yields a divergent series for the positional variance, owing to the zero-point vibrations of an infinity of high-frequency modes; the continuum model is thus unacceptable even as an approximation. The following will work with a more realistic rod model (Fig. 5.4) consisting of a series of N identical springs and masses, supporting N longitudinal vibrational modes. Introduction of complexities such as non-identical masses would alter the detailed dynamics of rod vibrations, but will generally have little effect on the positional variance. More drastic, however, is the assumption that entire planes of atoms perpendicular to the rod axis can be lumped together and treated as single masses, neglecting the degrees of freedom introduced by the physical extent and flexibility of each plane.

A limiting-case analysis illustrates the essential physics. In one limiting case (Fig. 5.5), the atoms in each plane are sufficiently *tightly* coupled to one another, each plane will share a single longitudinal degree of freedom, and the approximation under consideration will be correct (by construction). In the other limiting case, a rod might consist of an *uncoupled* bundle of j component rods, each one atom wide and m atoms long. The total number of modes is now enormously greater, being on the order of m^j , but this is an accounting fiction of no physical significance. If a component rod has a posi-

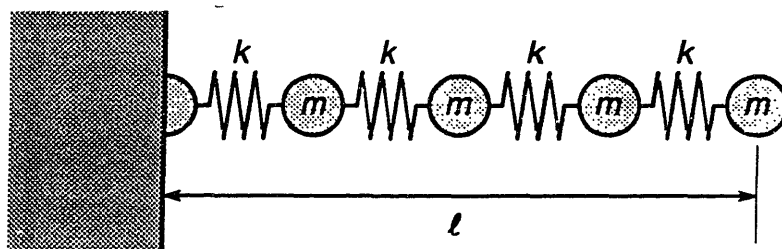


Figure 5.4. Diagram of a discrete rod, showing masses, springs, and the length coordinate (measuring between atomic centers).

tional standard deviation σ_c , then the bundle end position (= the mean position of the component ends) has a standard deviation

$$\sigma_b = \frac{\sigma_c}{\sqrt{j}} \quad (5.27)$$

and the variance of the bundle end position is thus inversely proportional to j . As we will see, this is exactly the variance that results from treating the bundle as a single unit with $N = m$, hence the tightly-coupled and uncoupled results are identical. N may thus be taken as the number of atomic planes along the length of the rod; a conservatively generous estimate (given that real interatomic spacings are greater than 0.1 nm) is

$$N = 10^{10} \ell \quad (5.28)$$

5.3.3. Semi-continuum quantum mechanical approximation

The simplest approximation to a discrete system uses the continuum results but truncates the sum of the modal variances at $n = N - 1$. The continuum expressions for the modal effective stiffnesses and frequencies, together with Eq. (5.13), yield the positional variance associated with mode n in the continuum approximation:

$$\sigma_n^2 \frac{\sqrt{E_t \rho_t}}{\hbar} = \frac{4}{\pi(2n+1)} \left(\frac{1}{2} + \frac{1}{e^{\frac{\hbar \omega_0(2n+1)}{kT}} - 1} \right) \quad (5.29)$$

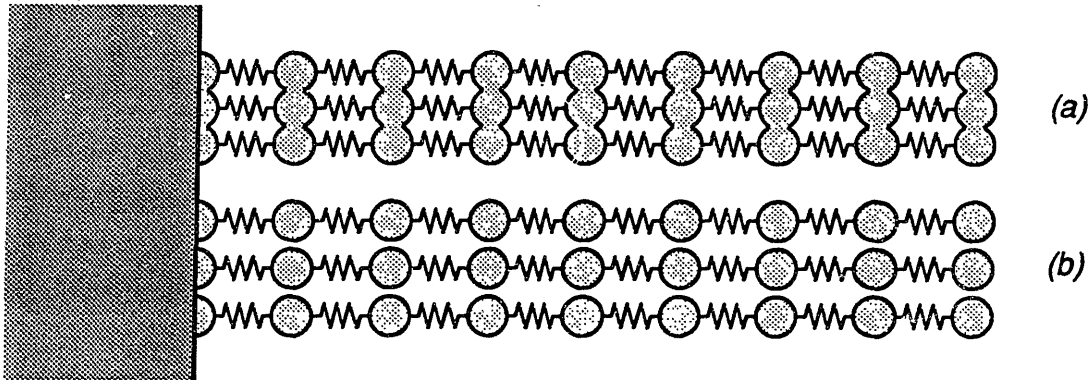


Figure 5.5. Limiting-case rods. Part (b) diagrams a rod in the limiting case in which masses and spring constants are lumped on a plane-by-plane basis; (c) diagrams the limiting case in which the rod is treated as a bundle of decoupled simple rods.

Summing over the N modes in a discrete rod yields the semi-continuum approximation to the total positional variance of the free end of a clamped rod:

$$\sigma^2 \frac{\sqrt{E_t \rho_t}}{\hbar} = \sum_{n=0}^{N-1} \frac{4}{\pi(2n+1)} \left(\frac{1}{2} + \frac{1}{e^{\frac{\hbar \omega_0}{kT}(2n+1)} - 1} \right) \quad (5.30)$$

5.3.4. Exact quantum mechanical treatment

The simple model above has serious defects when vibrational wavelengths approach the interatomic spacing of the rod ($\lambda \approx \ell_\Delta$), where the dispersive nature of the discrete medium becomes important. These effects are large for the single mode of a “rod” with $N = 1$, and continue to be large for a substantial fraction of the modes in rods where $N \gg 1$. Defining the discrete-mass variables in terms of the continuum variables

$$\ell_\Delta = \frac{\ell}{N}; \quad m_\Delta = \frac{\ell \rho_t}{N}; \quad k_s = \frac{E_t N}{\ell} \quad (5.31)$$

Keeping the intuitive definition of the rod length as the distance from the attachment point to the last mass

$$\lambda_n = \frac{4\ell}{(2n+1)} \left(1 + \frac{1}{2N} \right) \quad (5.32)$$

the correct, dispersive relationship of frequency to wavelength (Ashcroft and Mermin 1976) is

$$\omega_n = 2 \sqrt{\frac{k_s}{m_\Delta}} \sin \frac{\pi d}{\lambda_n} = \frac{2}{\ell} \sqrt{\frac{E_t}{\rho_t}} N \sin \left(\frac{2n+1}{2N+1} \frac{\pi}{2} \right) \quad (5.33)$$

The effective mass m_n of a mode n is the sum over the (equal) masses of the square of the local modal amplitude, divided by the square of the modal amplitude at the free end of the rod:

$$m_n = m_\Delta \left[\sin^2 \left(\frac{2n+1}{2N+1} \pi N \right) \right]^{-1} \sum_{i=1}^N \sin^2 \left(\frac{2n+1}{2N+1} \pi i \right) \quad (5.34)$$

which simplifies to

$$m_n = \rho_t \ell \left[\sin^2 \left(\frac{2n+1}{2N+1} \pi N \right) \right]^{-1} \left(\frac{1}{2} + \frac{1}{4N} \right) \quad (5.35)$$

Given Eq. (5.13) and the above values for frequency and effective mass, the modal variance

$$\sigma_n^2 = \frac{\hbar}{m_n \omega_n} \left(\frac{1}{2} + \frac{1}{e^{\frac{\hbar \omega_n}{kT}} - 1} \right) \quad (5.36)$$

which leads to the following dimensionless expression for the exact total variance expressed in terms of ω_0 , the frequency of the fundamental mode given by the continuum approximation:

$$\sigma^2 \frac{\sqrt{E_t \rho_t}}{\hbar} = \frac{2}{2N+1} \sum_{n=0}^{N-1} \frac{\sin^2 \left(\frac{2n+1}{2N+1} \pi N \right)}{\sin \left(\frac{2n+1}{2N+1} \frac{\pi}{2} \right)} \left(\frac{1}{2} + \frac{1}{e^{\left(\frac{\hbar \omega_0}{kT} \right)^4 \pi^{N \sin \left(\frac{2n+1}{2N+1} \frac{\pi}{2} \right)} - 1}} \right) \quad (5.37)$$

This equation is graphed in Figure 5.6; note that the result for $n = 0$, while identical to that given for the simple harmonic oscillator in Figure 5.1, is shifted to the left on this graph by the expression of results in terms of ω_0 (Eq. 5.33) rather than in terms of the true frequency ω .

Note that division of a rod into j component rods with

$$E'_t = E_t / j; \quad \rho'_t = \rho_t / j \quad (5.38)$$

yields

$$\sigma' = \sigma \sqrt{j} \quad (5.39)$$

and the standard deviation of the bundle end position (i.e., of the mean displacement of the component ends) is

$$\sigma_b = \frac{\sqrt{j \sigma'^2}}{j} = \sigma \quad (5.40)$$

5.3.5. An engineering approximation

For engineering purposes, it is advantageous to have a simple analytical expression rather than a sum over a variable number of terms. In seeking such an expression, we can begin with the simple semi-continuum approximation of Eq. (5.30). This sum can be rearranged into two terms

$$\sigma^2 \frac{\sqrt{E_t \rho_t}}{\hbar} = \frac{2}{\pi} \sum_{n=0}^{N-1} (2n+1)^{-1} + \frac{4}{\pi} \sum_{n=0}^{N-1} \left[(2n+1) \left(e^{\frac{\hbar \omega_0}{kT} (2n+1)} - 1 \right) \right]^{-1} \quad (5.41)$$

and the second term, which dominates in the classical limit, can be simplified by taking the classical limit ($\hbar \omega_0 / kT \rightarrow 0$):

$$\sigma^2 \frac{\sqrt{E_t \rho_t}}{\hbar} = \frac{2}{\pi} \sum_{n=0}^{N-1} (2n+1)^{-1} + \frac{4}{\pi} \frac{kT}{\hbar \omega_0} \sum_{n=0}^{N-1} (2n+1)^{-2} \quad (5.42)$$

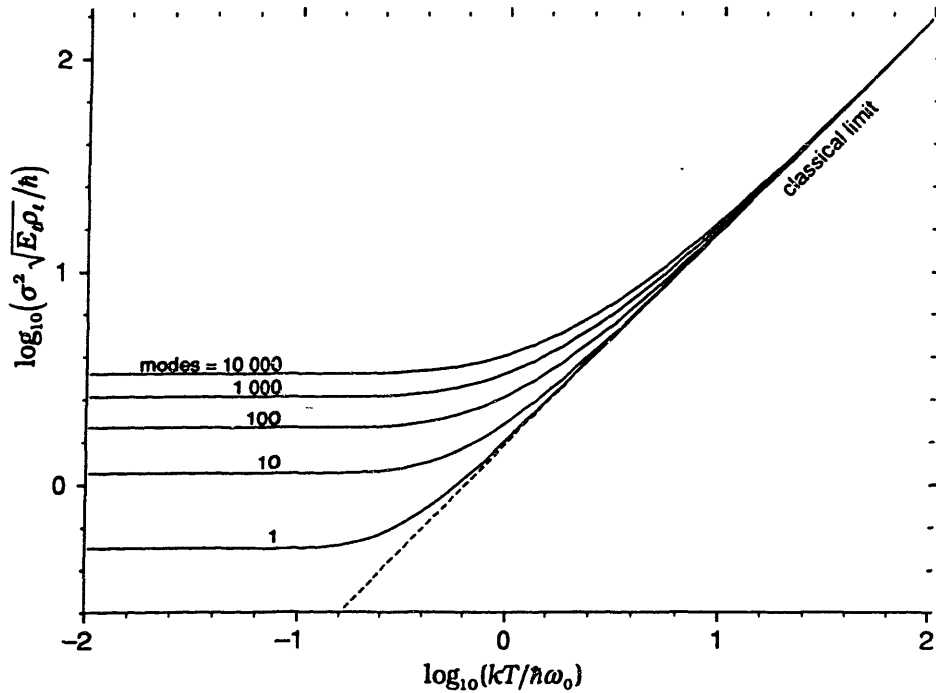


Figure 5.6. A dimensionless measure of variance vs. a dimensionless measure of temperature, for rods supporting varying numbers of longitudinal modes, Eq. (5.37).

The first, logarithmically-divergent sum can be replaced by an integral approximation corrected by a constant. Summing the second series to $N = \infty$ yields a simple expression (also transformed to eliminate ω_0) exhibiting better classical-limit behavior than the original expression:

$$\sigma^2 = \frac{\hbar}{\pi\sqrt{E_t\rho_t}} [0.54 + \log(2N+1)] + kt \frac{\ell}{E_t} \quad (5.43)$$

This expression is accurate in the classical limit ($\hbar\omega_0/kT \rightarrow 0$), and provides a good approximation in the quantum limit ($kT/\hbar\omega_0 \rightarrow 0$). Its deviations are strictly in a conservative direction (overestimating variance), but can amount to tens of percent for values of $kT/\hbar\omega_0 \approx 1$. A similar expression giving a more accurate result in this middle range can be obtained by multiplying the classical term by an empirically-chosen function:

$$\sigma^2 = \frac{\hbar}{\pi\sqrt{E_t\rho_t}} [0.54 + \log(2N+1)] + kt \frac{\ell}{E_t} \exp \left[- \left(0.7 - \frac{0.39}{\sqrt{N}} \right) \frac{\pi\hbar}{2\ell kT} \sqrt{\frac{E_t}{\rho_t}} \right] \quad (5.44)$$

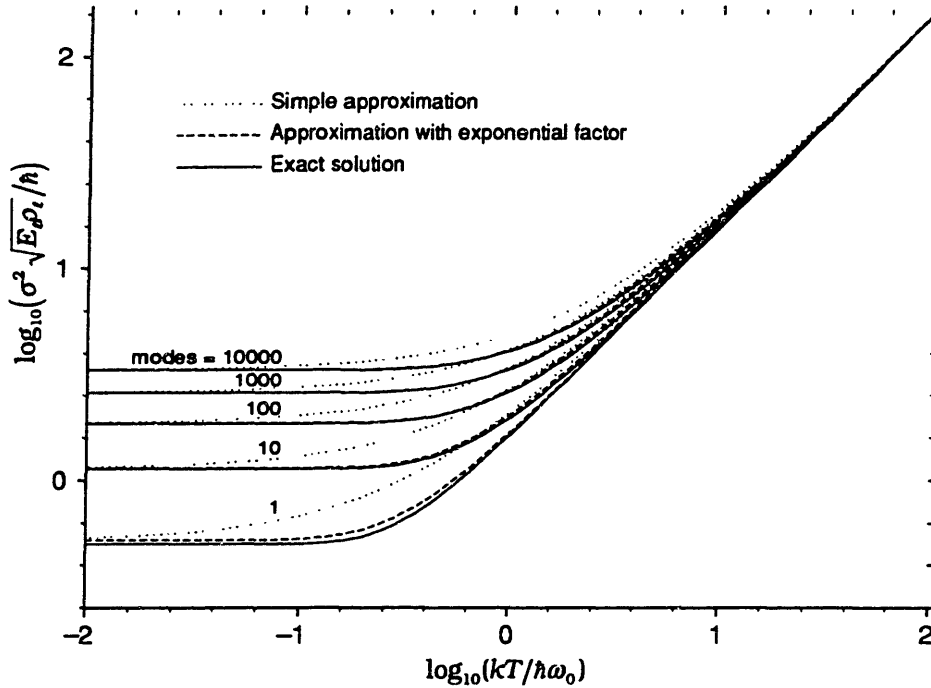


Figure 5.7. A plot of exact variances as in Fig. 5.7, with the approximations of Eq. (5.43) and (5.44) shown for comparison.

Figure 5.7 compares the results of these two approximations to the results given by Eq. (5.37). Both are shown to give conservative estimates of the variance, and the latter equation is shown to give estimates that are within a few percent of the correct values. It will serve as an adequate approximation for most engineering purposes.

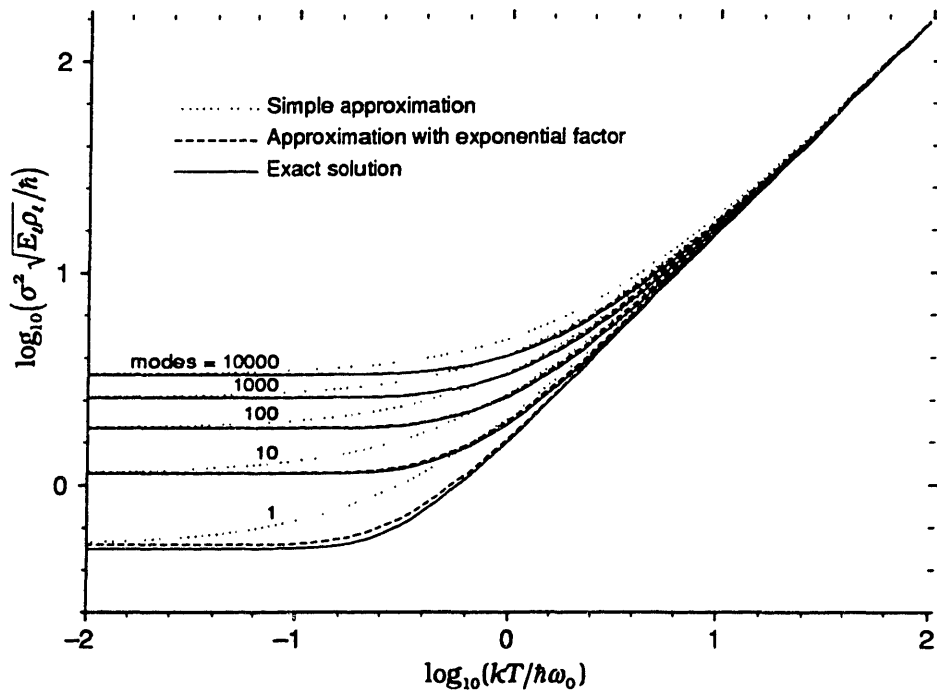


Figure 5.8. Standard deviation in elastic longitudinal displacement for the end of a thermally-excited rod, plotted vs. diameter and length, assuming mechanical properties resembling those of bulk diamond ($\rho = 3500 \text{ kg/m}^3$, $E = 10^{12} \text{ N/m}^2$), $N/\ell = 10^{10}$, and $T = 300 \text{ K}$. Entropic effects, neglected here, are included in Fig. 5.16. Note that quantum effects make a major contribution to the positional uncertainty only for $\ell \leq 1 \text{ nm}$. At sufficiently small dimensions, neglect of structural detail at the atomic level becomes unacceptable even as a rough approximation (e.g., the dark gray region to the left describes nonexistent subatomic diameters); at larger dimensions, this approximation is excellent.

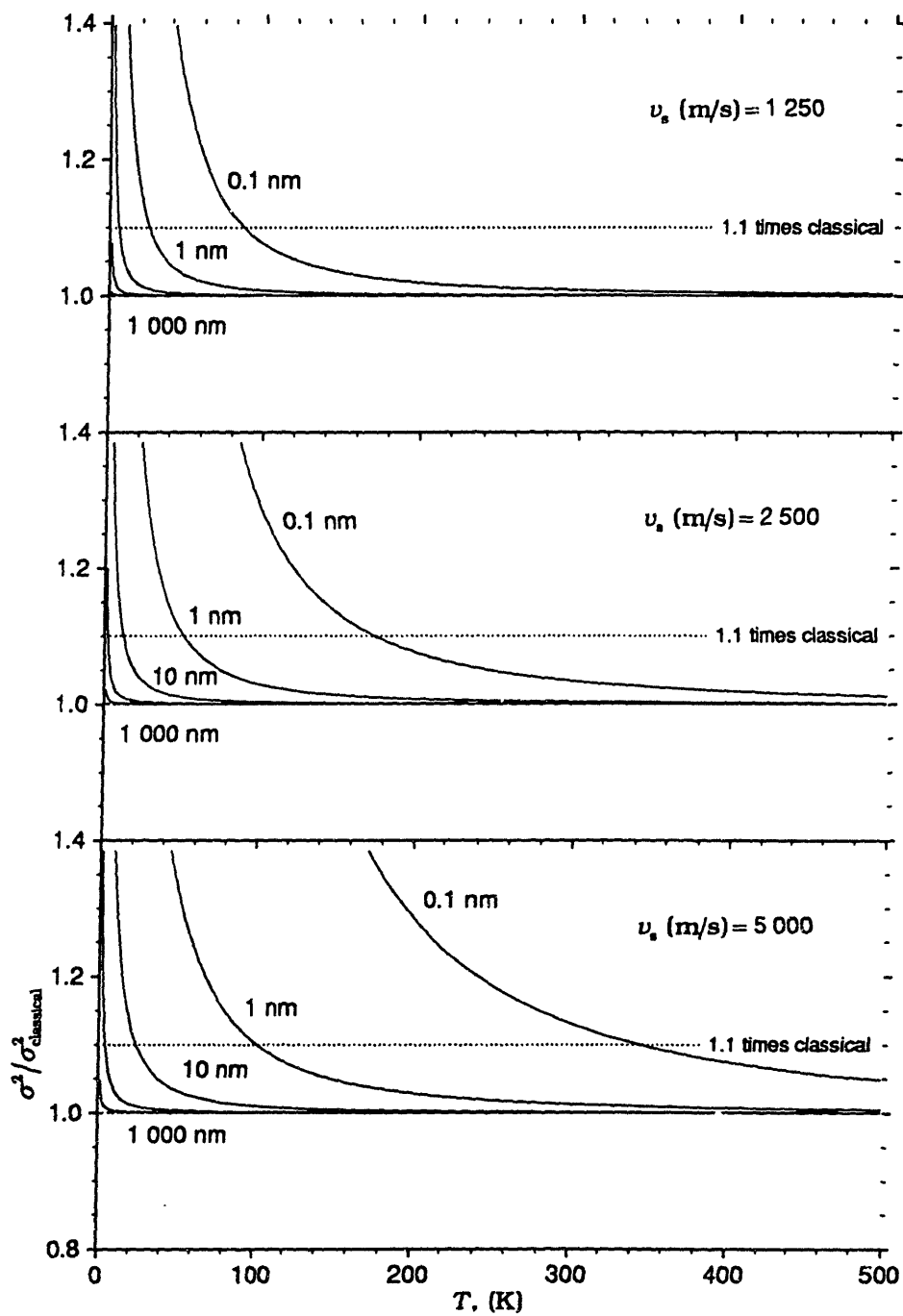
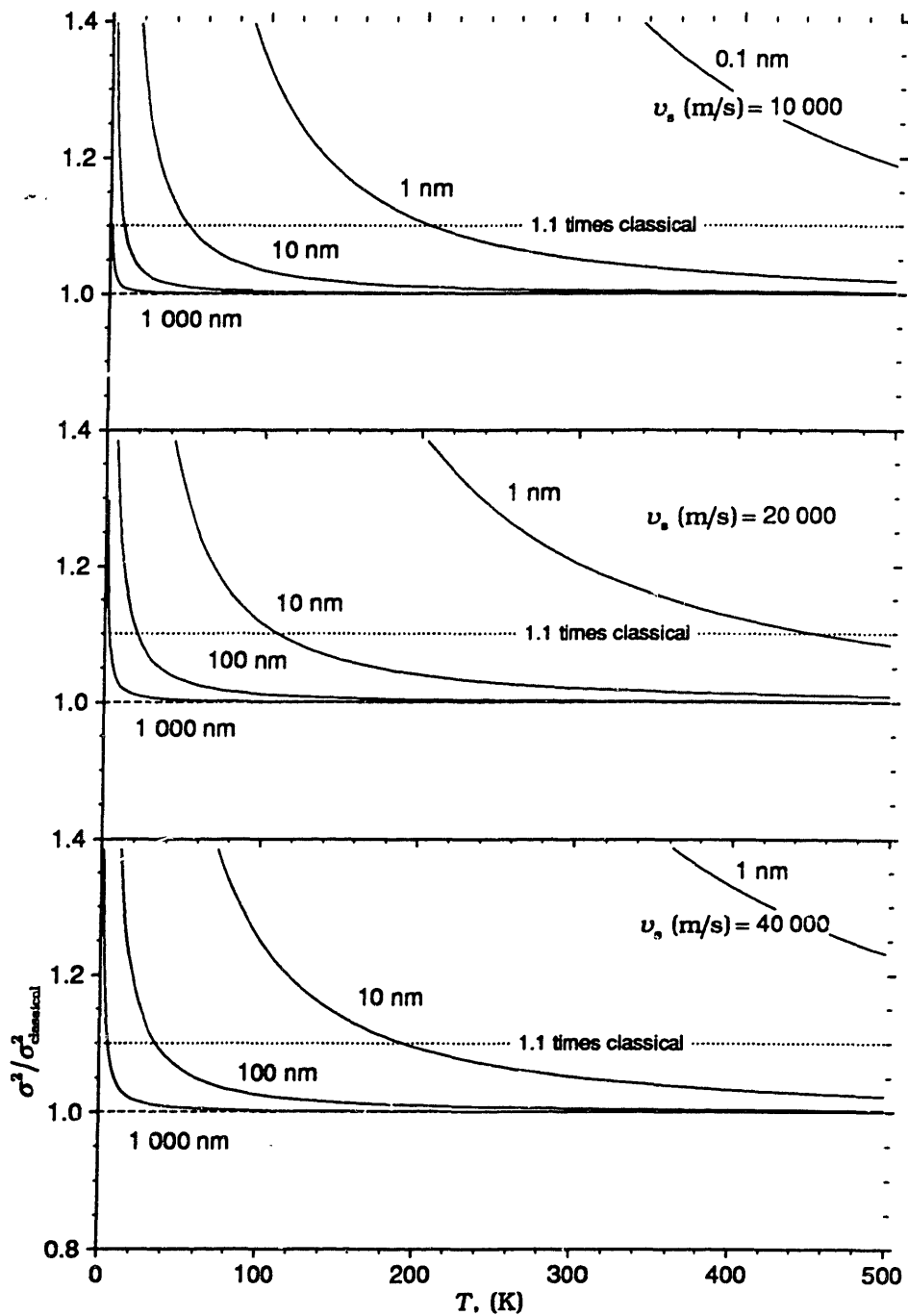


Figure 5.9. A plot of ratios of exact (Eq. 5.37) to classical (Eq. 5.26) variances. The final graph in the sequence assumes an unrealistically high acoustic speed (40 km/s, vs.



~ 17 km/s for diamond), and thus represents an upper bound on the magnitude of quantum effects in real structures under familiar physical conditions.

5.4. Bending of thermally excited rods

The analysis of the transverse positional variance resulting from transverse vibrational modes substantially parallels the longitudinal case. The following first considers a classical model for a continuous elastic rod, then a semi-continuum quantum mechanical model, a fit to the continuum limit of this model, and a conservative approximation to the combined effects of bending and shear deformation.

5.4.1. Classical treatment

Knowledge of modal frequencies and modal effective masses suffices to characterize the system. Following (Timoshenko, Young et al. 1974), the bending vibrations in a plane of symmetry of a uniform rod clamped at one end and free at the other (i.e., of a cantilever beam) will have normal modes with shapes described by

$$y(x) = C_1 \sin \kappa x + C_2 \cos \kappa x + C_3 \sinh \kappa x + C_4 \cosh \kappa x \quad (5.45)$$

where $y(x)$ is the transverse displacement from the equilibrium position at point x . The constants C_1 , C_2 , C_3 , and C_4 are determined by boundary and normalization condition. The boundary conditions require that

$$\cos \mathcal{R}_n \cosh \mathcal{R}_n = -1; \quad \mathcal{R}_n = \kappa_n \ell \approx (n + \frac{1}{2})\pi \quad (5.46)$$

and the modal frequencies are given by the relationship

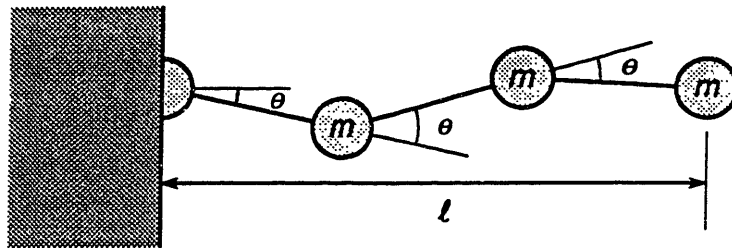


Figure 5.10. A discrete rod, illustrating the angles entering into calculations of bending stiffness effects.

$$\omega_n = \left(\frac{\mathcal{R}_n}{\ell} \right)^2 \sqrt{\frac{k_b}{\rho_t}} \quad (5.47)$$

where k_b is the bending stiffness of the rod in J·m /rad², and the elastic energy per unit length resulting from bending

$$\frac{\partial \mathcal{E}_b}{\partial x} = \frac{1}{2} k_b \left(\frac{\partial^2 y}{\partial x^2} \right)^2 \quad (5.48)$$

The modal effective mass (taking the free-end displacement as the generalized position coordinate) may be found by integrating the square of the local normalized amplitude (requiring $y(\ell) = 1$) with respect to the mass:

$$\begin{aligned} m_n &= \int_0^\ell \frac{\left[\left(\cos \mathcal{R}_n + \cosh \mathcal{R}_n \right) \left(\cos \frac{\mathcal{R}_n}{\ell} x - \cosh \frac{\mathcal{R}_n}{\ell} x \right) - \left(\sin \mathcal{R}_n - \sinh \mathcal{R}_n \right) \left(\sin \frac{\mathcal{R}_n}{\ell} x - \sinh \frac{\mathcal{R}_n}{\ell} x \right) \right]^2}{\left(\sin \mathcal{R}_n \sinh \mathcal{R}_n \right)^2} \rho_t dx \\ &= \frac{1}{4} \rho_t \ell \end{aligned} \quad (5.49)$$

for all values of n . The modal frequency and the modal effective mass define the modal effective transverse stiffness resulting from bending

$$k_{t,b,n} = \frac{k_b}{4\ell^3} \mathcal{R}_n^4 \quad (5.50)$$

which (as in the longitudinal analysis) yields an expression for the total classical transverse variance resulting from bending compliance:

$$\sigma_{t,b,class}^2 = 4kT \frac{\ell^3}{k_b} \sum_{n=0}^{\infty} \frac{1}{\mathcal{R}_n^4} \quad (5.51)$$

Applying the identity

$$\sum_{n=0}^{\infty} \frac{1}{\mathcal{R}_n^4} = \frac{1}{12} \quad (5.52)$$

yields the classical expression for the transverse variance in the position of the free end of a rod of uniformly distributed mass and bending stiffness, neglecting compliance due to shear and the effects of discreteness:

$$\sigma_{t,b, \text{class}}^2 = kT \frac{\ell^3}{3k_b} \quad (5.53)$$

But elementary theory of flexure shows that the bending stiffness of the end of a continuum cantilever rod is

$$k_{t,b} = \frac{3k_b}{\ell^3} \quad (5.54)$$

hence the classical variance is simply $kT/k_{t,b}$. As in the longitudinal case, the vibrational modes prove irrelevant to the classical analysis; knowledge of the overall stiffness suffices.

In bending (unlike stretching) the transition from a continuum to a lumped system changes the overall stiffness (not merely the modal stiffnesses). For a rod considered as a series of masses and angular springs, as shown in Figure 5.10, the rod bending stiffness is related to the inter-mass spacings and angular stiffnesses:

$$k_b = k_\theta \ell_\Delta \quad (5.55)$$

For rods in which the number of segments, N , is finite

$$\frac{1}{k_{t,b}} = \frac{\ell^3}{k_b} \sum_{n=0}^{N-1} \frac{1}{N} \left(1 - \frac{n}{N}\right)^2 \cong \frac{\ell^3}{k_b} \left(\frac{3N+4}{9N-2}\right) \quad (5.56)$$

where the approximation is exact for $N = 1, 2$, and ∞ ; for intermediate values it is always high (i.e., conservative) and in error by less than 1%.

The above analysis takes account of transverse compliance resulting from bending, but rods (of multi-atom width) also exhibit compliance through shear. For typical materials, latter becomes a significant component of the total transverse compliance, k_t , when rod diameter $d \approx \ell$; in vibration, it has significant effects on modal shapes and stiffnesses when $d \approx \lambda$. Figure 5.11 illustrates normal modes due to pure bending (zero shear compliance) and to pure shear (zero bending compliance) for $n = 2$, with the idealization that plane cross sections remain plane despite shear. A standard engineering approximation treats shear as an independent, additive source of compliance, with transverse stiffness effects due to shear taking the same analytical form as longitudinal stiffness effects but substituting the (linear) shear modulus, G_ℓ for the (linear) stretching modulus, E_ℓ . Combining the resulting shear compliance, $1/k_{t,s} = \ell/G_\ell$, with Eq. (5.56) yields the engi-

neering approximation:

$$\sigma_{t, \text{class}}^2 = kT \left[\frac{\ell^3}{k_b} \left(\frac{3N+4}{9N-2} \right) + \frac{\ell}{G_t} \right] \quad (5.57)$$

Aside from quantum corrections, this approximation breaks down when ℓ/d becomes small, yet bending compliance remains important. In this regime, out-of-plane deformation of the free surface becomes a significant source of additional compliance. To treat such low aspect ratio objects as rods is at best a crude approximation.

5.4.2. Semi-continuum quantum mechanical treatment

The above can serve as the basis for a quantum mechanical analysis. As in the earlier analysis of longitudinal vibrations, use of the continuum approximation neglects the dispersive properties of a discrete medium, which become significant as the characteristic wavelengths of normal modes approach interatomic distances. As will be seen, though, zero-point vibrations in the higher transverse modes (unlike those of the higher longitudinal modes) make only a modest contribution to the positional variance. Accordingly, dis-

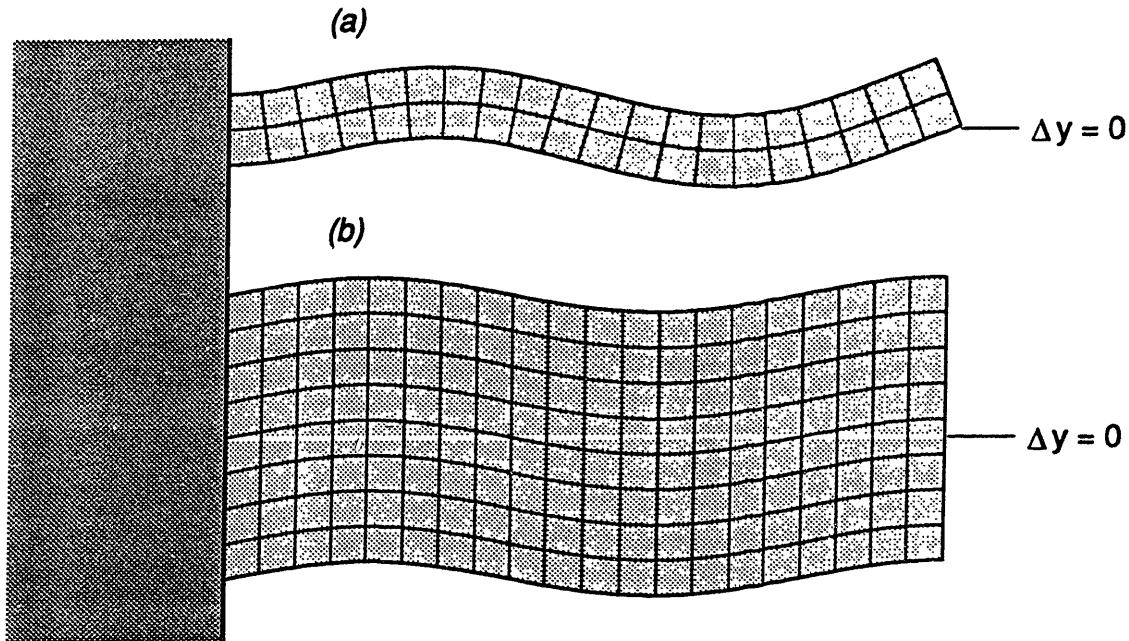


Figure 5.11. Modal deformation in pure bending (a) vs. pure shear (b). Bending dominates in high aspect-ratio rods; shear dominates in low aspect-ratio rods. Both deformations are here shown in idealized form.

person is neglected in the following analysis; this neglect should have little effect for $N \geq 10$.

We begin with a model including only the effects of bending compliance, deferring discussion of shear compliance. (Figure 5.11 illustrates the difference between these forms of compliance.) The above values of modal effective mass and effective stiffness yield the semi-continuum approximation for the transverse bending variance for a uniform rod. In dimensionless terms analogous to those of the longitudinal-mode analysis:

$$\sigma_{t,b}^2 \frac{\sqrt{k_b \rho_l}}{\hbar l} = \sum_{n=0}^{N-1} \frac{4}{\mathcal{R}_n^2} \left(\frac{1}{2} + \frac{1}{e^{\frac{\hbar \omega_0 (\mathcal{R}_n)^2}{kT} - 1}} \right) \quad (5.58)$$

(Fig. 5.12).

5.4.3. Engineering approximations

Expressing the variance directly,

$$\sigma_{t,b}^2 = \frac{\hbar l}{\sqrt{k_b \rho_l}} \sum_{n=0}^{N-1} \frac{4}{\mathcal{R}_n^2} \left(\frac{1}{2} + \frac{1}{e^{\frac{\hbar \mathcal{R}_n^2 \sqrt{k_b}}{kT l^2 \rho_l} - 1}} \right) \quad (5.59)$$

Multiplying out and taking the classical limit of the second term yields

$$\sigma_{t,b}^2 = \frac{\hbar l}{\sqrt{k_b \rho_l}} \sum_{n=0}^{N-1} \frac{2}{\mathcal{R}_n^2} + \sum_{n=0}^{N-1} \frac{4}{\mathcal{R}_n^4} kT \frac{l^3}{k_b} \quad (5.60)$$

In the limit of large N , the second term yields the continuum classical result, and may be replaced by the modified classical expression derived above to take account of the reduced bending stiffness of discrete rods where N is small. For the first term (describing zero-point contributions to the variance) solving the equation that defines values of \mathcal{R}_n yields the convergent series

$$\sum_{n=0}^{\infty} \frac{2}{\mathcal{R}_n^2} = 0.5688 + 0.0908 + 0.0324 + 0.0155 + \dots \approx 0.7588 \quad (5.61)$$

Substituting the infinite series limit yields the approximation

$$\sigma_{t,b}^2 = 0.76 \frac{\hbar \ell}{\sqrt{k_b \rho_t}} + kT \frac{\ell^3}{k_b} \left(\frac{3N+4}{9N-2} \right) \quad (5.62)$$

Given the size of the first term in the above series and the shortcomings of the continuum model for rods of low N , it is useful to consider the case $N = 1$. In a rod consisting of a single mass and a single point of bending, the continuum model overestimates the effective stiffness by a factor of three and underestimates the effective mass by a factor of four, making its estimate of the positional variance conservative by a factor of $(4/3)^{1/2} \approx 1.15$.

This approximation, like that of Eq. (5.43), significantly overestimates the positional variance in the transition region between the quantum and classical limits (Fig. 5.12). A more accurate result for the important case of large N may be obtained by evaluating the limit of the original sum Eq. (5.59) as $N \rightarrow \infty$ and fitting an empirical expression to the results:

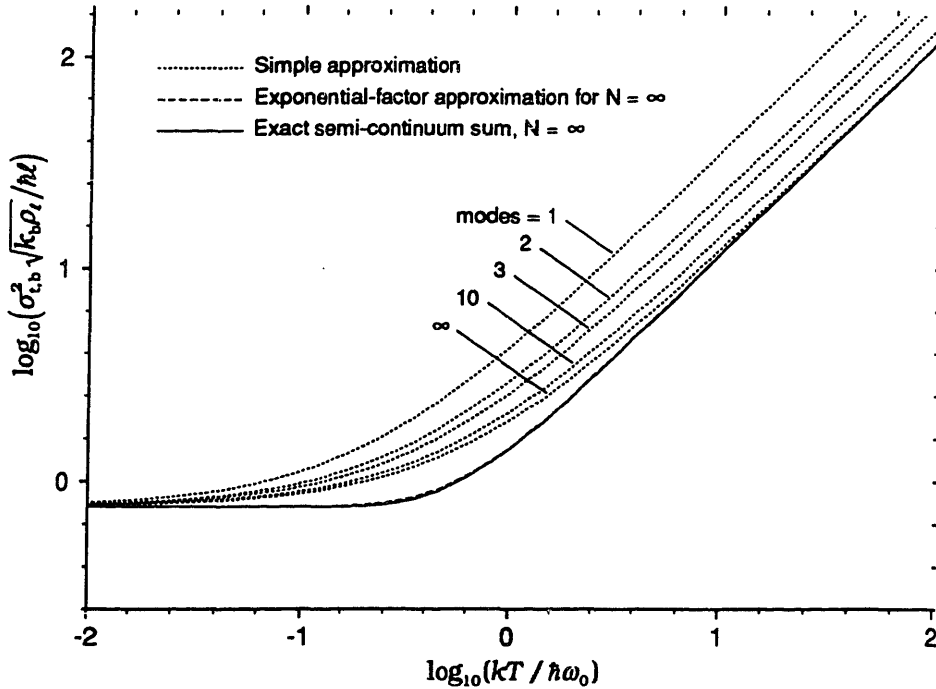


Figure 5.12. Dimensionless transverse variance for rods, neglecting shear compliance. The simple approximation is based on Eq. (5.62); the exponential-factor approximation on Eq. (5.63), and the exact semi-continuum sum on Eq. (5.58).

$$\sigma_{t,b}^2 = 0.76 \frac{\hbar \ell}{\sqrt{k_b \rho_t}} + kT \frac{\ell^3}{3k_b} \exp\left(-\frac{1.97}{\ell^2} \sqrt{\frac{k_b}{\rho_t}} \frac{\hbar}{kT}\right) \quad (5.63)$$

This approximation is always high, but never by more than 1%. Figure 5.12 compares these approximations and Eq. (5.59).

5.4.4. Shear and bending in the quantum limit

The approximations developed for longitudinal positional variance, Eq. (5.43) and Eq. (5.44), have direct analogs for the variance that would result in hypothetical rods having shear but no bending compliance. The more accurate of the two takes the form

$$\sigma_{t,s}^2 = \frac{\hbar}{\pi \sqrt{G_t \rho_t}} [0.54 + \log(2N + 1)] + kt \frac{\ell}{G_t} \exp\left[-\left(0.7 - \frac{0.39}{\sqrt{N}}\right) \frac{\pi \hbar}{2 \ell kT} \sqrt{\frac{G_t}{\rho_t}}\right] \quad (5.64)$$

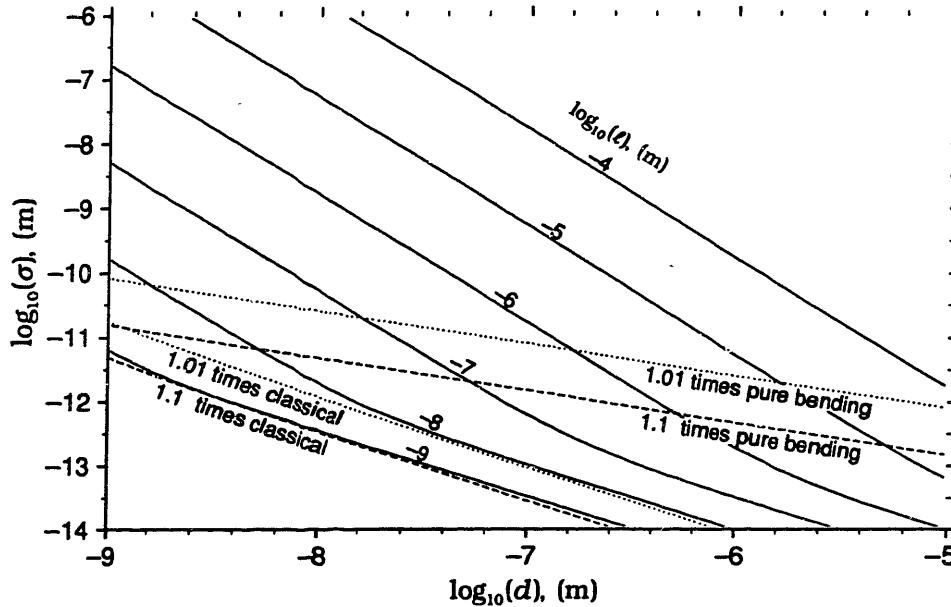


Figure 5.13. Transverse positional standard deviation for rods, including shear compliance. The bending component is based on Eq. (5.59); the shear component is based on Eq. (5.64), assuming mechanical properties resembling those of bulk diamond, as in Fig. 5.8; similar remarks apply.

In the classical limit, the effects of bending and shear compliance are simply additive. In the quantum limit, effective modal masses and frequencies play a role, and a precise analysis would have to include the effects of bending and shear on a mode-by-mode basis. An upper bound on their combined effect can be had more simply. Consider the variance of a harmonic oscillator in the quantum limit

$$\sigma^2 = \frac{\hbar}{2\sqrt{k_s m}} \quad (5.65)$$

If we consider its stiffness to have two sources, k_{s1} and k_{s2} , the variance becomes

$$\sigma^2 = \frac{\hbar}{2\sqrt{m}} \sqrt{\frac{1}{k_{s1}} + \frac{1}{k_{s2}}} \quad (5.66)$$

The expression

$$\sigma_{\text{est}}^2 = \sigma_1^2 + \sigma_2^2 = \frac{\hbar}{2\sqrt{k_{s1} m}} + \frac{\hbar}{2\sqrt{k_{s2} m}} \quad (5.67)$$

will overestimate the actual variance by a factor

$$1 \leq \frac{\sigma_{\text{est}}^2}{\sigma^2} = \frac{\sqrt{k_{s1}/k_{s2}} + 1}{\sqrt{(k_{s1}/k_{s2}) + 1}} \leq \sqrt{2} \quad (5.68)$$

Accordingly, it should be conservative to estimate the variance resulting from a vibrational mode subject to both bending and shear as the sum of the variances of a hypothetical mode constrained purely by bending forces and of one constrained by pure shear forces. (The differences between modal shapes in pure bending and those in pure shear would complicate a more precise analysis.) Thus, treating both classical and quantum variances as additive, expressions for the total transverse variance at the end of a rod take the form

$$\sigma_t^2 = \sigma_{t,b}^2 + \sigma_{t,s}^2 \quad (5.69)$$

with the choice of expressions for the shear and bending contributions depending on the desired accuracy, the magnitude of quantum effects, and the value of N . Figure 5.13 uses Eq. (5.63) and (5.64) to graph the standard deviation of the transverse displacement at room temperature for rods with mechanical properties approximating those of bulk diamond. As can be seen, under these conditions, in the regime where shear and bending

compliance are both important, quantum effects on positional uncertainty are minor for rods of nanometer or greater size; accordingly, Eq. (5.57) provides a good approximation.

5.5. Piston displacement in a gas-filled cylinder

Earlier sections have considered linear, elastic systems in which the motion can be divided into normal modes, treated as independent harmonic oscillators. A different approach is necessary for nonlinear systems in which the displacement of one component affects the range of motion possible to another. Here, the simplest example is not a mass and spring, but a loaded piston in a cylinder containing an ideal gas.

Figure 5.14 illustrates the system and the defining coordinates. The diameter of the cylinder proves irrelevant, and the displacement of the piston is chosen such that a zero displacement corresponds to zero freedom of movement for the gas molecule in the x direction. The assumption of an ideal gas entails ignoring all forces between gas molecules, and accordingly ignoring the reduction in available volume that a molecule experiences as a result of the bulk of other molecules that may be present. Classical positional uncertainty in this system will be analyzed from three (ultimately equivalent) perspectives.

5.5.1. Weighting in terms of potential energy and available states

Let the compressing force on the piston be a constant, F_c . For an empty cylinder, the potential energy of the piston is then $F_c x$, and applying the Boltzmann weighting to this

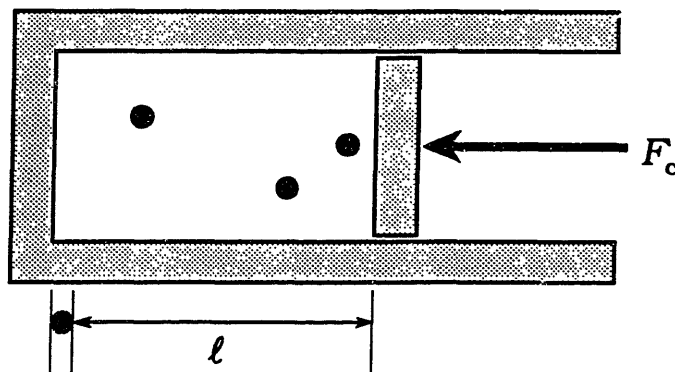


Figure 5.14. Piston, cylinder, and gas molecules. Note that the length coordinate ℓ measures not the distance of the piston from the bottom of the cylinder, but the range of motion available to a gas molecule. A hard sphere, hard surface model is assumed here and in the text.

potential yields the exponential PDF

$$f_x(x) = \frac{e^{-\frac{F_c x}{kT}}}{\int_0^\infty e^{-\frac{F_c x}{kT}} dx} = \frac{F_c}{kT} e^{-\frac{F_c x}{kT}} \quad x \geq 0 \quad (5.70)$$

The addition of N gas molecules does not change the potential energy, but does increase the number of states associated with each piston position by a factor proportion to the space available to each molecule, that is, by a factor of x for each molecule. Introducing this factor to account for the number of states yields the Erlang PDF

$$f_x(x) = \frac{x^N e^{-\frac{F_c x}{kT}}}{\int_0^\infty x^N e^{-\frac{F_c x}{kT}} dx} = \frac{1}{N!} \left(\frac{F_c}{kT} \right)^{N+1} x^N e^{-\frac{F_c x}{kT}} \quad x \geq 0 \quad (5.71)$$

with mean and variance

$$\bar{x} = (N+1) \frac{kT}{F_c}; \quad \sigma_x^2 = (N+1) \left(\frac{kT}{F_c} \right)^2 \quad (5.72)$$

In this approach, the configurational states of the system are treated as known, and the Boltzmann-weighted probabilities are then integrated over these states.

5.5.2. Weighting in terms of a mean-force potential

Alternatively, one can treat the gas as a non-linear spring, as described by the ideal gas equation (here in molecular rather than the more familiar molar units)

$$pV = NkT \quad (5.73)$$

yielding the time-average force due to pressure

$$F_p = -\frac{NkT}{x} \quad (5.74)$$

Since the gas is now treated as a spring external to the piston, its configurational states are now ignored. Treating F_p on the same basis as F_c , the work-energy as a function of position

$$\int F_c - \frac{NkT}{x} dx = F_c x - NkT \log x + C \quad (5.75)$$

hence the Boltzmann-weighted positional PDF is

$$f_x(x) = \frac{e^{\frac{F_c x}{kT} - N \log x + \frac{C}{kT}}}{\int_0^\infty e^{\frac{F_c x}{kT} - N \log x + \frac{C}{kT}} dx} = \frac{x^N e^{\frac{F_c x}{kT}}}{\int_0^\infty x^N e^{\frac{F_c x}{kT}} dx} \quad (5.76)$$

Note the the nonlinearity of the stiffness invalidates the simple relationship

$$\sigma_x^2 = \frac{kT}{k_s} \quad (5.77)$$

Evaluating k_s at the the mean and most probable values of x

$$\frac{kT}{k_{s, \text{mean } x}} = \frac{(N+1)^2}{N} \left(\frac{kT}{F_c} \right)^2; \quad \frac{kT}{k_{s, \text{max-prob } x}} = N \left(\frac{kT}{F_c} \right)^2 \quad (5.78)$$

yields differing results, both of which differ from the true variance Eq. (5.72). Both expressions, however, approach the correct value in the limit of large N , where the standard deviation in position is small enough for a linear approximation to hold for fluctuations of ordinary magnitude.

5.5.3. Weighting in terms of Helmholtz free energy

In a third approach, we begin with the Helmholtz free energy,

$$\mathcal{F} = \mathcal{E} - TS$$

The internal energy \mathcal{E} consists of the potential energy $F_c x$, and the classical kinetic energy contributions from the gas ($3/2 NkT$, assuming a monatomic gas) and the piston ($3kT$, assuming freedom to slide, rotate, and rattle). The translational entropy of an ideal gas (Knox 1971) is

$$S_{\text{trans}} = Nk \left(\frac{5}{2} + \frac{3}{2} \ln \frac{mkT}{2\pi\hbar^2} - \ln \frac{N}{V} \right) = Nk \left(C - \ln \frac{N}{x} \right) \quad (5.79)$$

hence the free energy of the system as a function of piston position is

$$\mathcal{F} = \mathcal{E} - TS = F_c x + \left(3 + \frac{3}{2}N\right)kT - NkT \left(C - \ln \frac{N}{x}\right) \quad (5.80)$$

Taking the Boltzmann-weighted PDF in terms of the free energy once more yields

$$f_x(x) = \frac{e^{\frac{F_c x}{kT} + 3 + \frac{3}{2}N - NC + N \ln \frac{N}{x}}}{\int_0^\infty e^{\frac{F_c x}{kT} + 3 + \frac{3}{2}N - NC + N \ln \frac{N}{x}} dx} = \frac{x^N e^{-\frac{F_c x}{kT}}}{\int_0^\infty x^N e^{-\frac{F_c x}{kT}} dx} \quad (5.81)$$

5.5.4. Comparison and quantum effects

These three perspectives on the same problem are closely related: The variation in available states as a function of piston position considered in the first corresponds to the variation in entropy considered in the third. The second approach considers movement of the piston as doing work on the gas (under reversible, isothermal conditions); work done under these conditions equals the change in \mathcal{F} considered in the third. The choice of perspective in such problems is a matter of convenience.

Physical displacements frequently couple to changes in the range of motion available in some degree of freedom, hence altering its entropy and doing work against an entropic spring. These systems share the basic properties of the piston and cylinder system just considered, including a $(kT)^2$ dependence of positional uncertainty in the classical regime, as opposed to the kT dependence of classical uncertainty stemming from conventional springs. In the quantum limit, systems that would exhibit entropic spring effects at higher temperatures will display a small residual compliance associated with the compression of zero-point probability distributions. This is not an entropic effect, since (in the quantum limit) all modes are consistently in their ground state with an invariant entropy of zero; nonetheless, it is a compressive effect that is a smooth extension of the behavior in the classical regime.

5.6. Longitudinal variance from transverse rod deformation

Transverse vibrations in a rod cause longitudinal shortening by forcing it to deviate from a straight line. The resulting longitudinal-transverse coupling is a source of longitudinal positional variance, and provides another example of an entropic spring.

5.6.1. General approach

Rods sliding in channels can be used to couple mechanical displacements occurring in one location to displacements at a relatively distant location. Thus, it is of interest to consider the longitudinal positional variance of rods in systems where rod motion occurs in a channel that imposes transverse restoring forces (modeled as a stiffness per unit length, k_ℓ) via overlap repulsion, and where boundary conditions may impose a mean tension γ_ℓ on the rod.

The effect of shear compliance on transverse vibration is typically small in systems where longitudinal-transverse coupling is significant. Further, in the approximation that the rod behaves as a piece of elastically linear, isotropic material, shear deformation has no effect on length. Shear is accordingly neglected.

The discrete structure of the rod imposes a limit to the number of modes and modifies the coupling constants as λ_n approaches $2\Delta\ell$ ($= \lambda_{\min}$). The following analysis adopts a semi-continuum model that takes account of the former while neglecting the latter. This approximation can break down in systems where modes with $\lambda \approx \lambda_{\min}$ dominate the variance. Where the restoring force for these modes is dominated by k_ℓ , the approximation is conservative; where the restoring force is dominated by γ_ℓ , the approximation is accurate; where the restoring force is dominated by k_b , the approximation is too low. In the latter case, the maximum correction factor for the variance (for $\lambda = \lambda_{\min}$) is $(\pi/2)^4 \approx 6.09$, falling to 1.52 for $\lambda = 2\lambda_{\min}$ and to 1.02 for $\lambda = 10\lambda_{\min}$.

The non-linearity of overlap repulsions makes the use of the constant stiffness k_ℓ a rough approximation. Since stiffness increases with displacement, this approximation is conservative, underestimating the constraining forces.

Quantum mechanical effects can significantly *decrease* variance resulting from longitudinal-transverse coupling. Their discussion is deferred.

5.6.2. Coupling and variance

The following analysis considers a weighting in terms of potential energy and available states for each of a set of normal-mode deformations, summing the resulting variances to yield the total variance. The use of normal modes here implies nothing about vibrations, but merely provides a convenient set of orthogonal functions with which to describe all possible rod configurations.

For the sinusoidal deformation characteristic of mode n , there exists a constant C_n (dependent on rod parameters) relating the contraction in the length of the rod to the potential energy of the deformation

$$\Delta \ell_n = C_n \mathcal{E}_n \quad (5.82)$$

The classical variance in rod length resulting from modal longitudinal-transverse coupling may be determined from the PDF for the potential energy (which may in turn be derived from the Gaussian PDF for the amplitude).

$$f_{\mathcal{E}_n}(\mathcal{E}_n) = \frac{1}{\sqrt{\pi k T \mathcal{E}_n}} e^{-\frac{\mathcal{E}_n}{kT}} \quad (5.83)$$

and for the contraction in length

$$f_{\Delta \ell_n}(\Delta \ell_n) = \frac{1}{\sqrt{\pi C_n k T \Delta \ell_n}} e^{-\frac{\Delta \ell_n}{C_n k T}} \quad (5.84)$$

From this one can derive the variance in the potential energy

$$\begin{aligned} \sigma_{\mathcal{E}_n}^2 &= \overline{\mathcal{E}_n^2} - (\overline{\mathcal{E}_n})^2 \\ &= \int_0^\infty \frac{\mathcal{E}_n^2}{\sqrt{\pi k T \mathcal{E}_n}} e^{-\frac{\mathcal{E}_n}{kT}} d\mathcal{E}_n - \left[\int_0^\infty \frac{\mathcal{E}_n}{\sqrt{\pi k T \mathcal{E}_n}} e^{-\frac{\mathcal{E}_n}{kT}} d\mathcal{E}_n \right]^2 \\ &= \frac{1}{2} (kT)^2 \end{aligned} \quad (5.85)$$

and hence the variance in length resulting from longitudinal-transverse coupling in mode n .

$$\sigma_{\ell, n}^2 = \frac{1}{2} (C_n k T)^2 \quad (5.86)$$

The mean contraction is

$$\overline{\Delta \ell_n} = \frac{1}{2} C_n k T \quad (5.87)$$

5.6.3. Rods with tension and transverse constraints

Consider a long, continuous rod with a bending modulus k_b and subject to both a mean tension γ_t and a transverse stiffness per unit length k_t . The energy per unit length associated with a sinusoidal deformation may be derived by integrating the contributions from each of these restoring-force terms. For a rod of length ℓ supporting modes with amplitudes A_n and $\lambda = 2\ell/n$ ($n = 1, 2, 3, \dots$),

$$\frac{\mathcal{E}_n}{\ell} = \frac{A_n^2}{4} \left[k_b \left(\frac{n\pi}{\ell} \right)^4 + \gamma_t \left(\frac{n\pi}{\ell} \right)^2 + k_t \right] \quad (5.88)$$

The fractional change in length

$$\frac{\Delta \ell_n}{\ell} = \left(\frac{n\pi A_n}{2\ell} \right)^2 \quad (5.89)$$

hence

$$C_n = \frac{\Delta \ell_n}{\mathcal{E}_n} = \left[k_b \left(\frac{n\pi}{\ell} \right)^2 + \gamma_t + k_t \left(\frac{n\pi}{\ell} \right)^{-2} \right]^{-1} \quad (5.90)$$

and the total (classical) variance resulting from transverse vibrations in one of the two possible polarizations equals the sum of the modal variances

$$\sigma_{t,i}^2 = \sum_{n=1}^N \frac{1}{2} (C_n kT)^2 = \sum_{n=1}^N \frac{1}{2} (kT)^2 \left[k_b \left(\frac{n\pi}{\ell} \right)^2 + \gamma_t + k_t \left(\frac{n\pi}{\ell} \right)^{-2} \right]^{-2}; \quad N = \ell/\Delta \ell \quad (5.91)$$

The total variance will be the sum of contributions from both polarizations; if these are equivalent, the total will be twice the above value.

Stretching the rod compresses the range of motion of the transverse modes, doing work and reducing modal entropies; from a mean-force potential perspective, transverse modes introduce a source of (nonlinear) compliance. This perspective clarifies the nature of the quantum effects: Each transverse mode can be regarded as a harmonic oscillator with a restoring force modulated by the degree of rod extension. In the quantum-mechanical limit, the transverse positional variance is proportional not to the transverse compliance (as in the classical regime)

$$\sigma_{\text{classical limit}}^2 = \frac{kT}{k_s} \quad (5.92)$$

but to its square root

$$\sigma_{\text{quantum limit}}^2 = \frac{\hbar}{2\sqrt{k_s m}} \quad (5.93)$$

hence as quantum effects become significant, mean transverse amplitudes are less easily compressed by increases in restoring forces. Since the longitudinal compliance just described results from compression of transverse modes as a result of increasing tension, it is *lower* in the quantum regime than would be predicted by the classical model. The resulting reduction in longitudinal positional variance is neglected here.

To permit a graphical summary of the (conservative) classical model, the sum over the modes can be expressed in terms of two parameters, k_b/k_ℓ and γ_ℓ/k_ℓ

$$\left(\frac{\sigma_{\ell,i}^2}{\ell} \right) \frac{k_b \sqrt{\gamma_\ell k_\ell}}{(kT)^2} = \lim_{\ell \rightarrow \infty} \sum_{n=1}^{\ell/\Delta\ell} \frac{1}{2\ell} \left(\frac{n\pi}{\ell} \right)^4 \frac{k_b}{k_\ell} \sqrt{\frac{\gamma_\ell}{k_\ell}} \left[\frac{k_b}{k_\ell} \left(\frac{n\pi}{\ell} \right)^4 + \frac{\gamma_\ell}{k_\ell} \left(\frac{n\pi}{\ell} \right)^2 + 1 \right]^{-2} \quad (5.94)$$

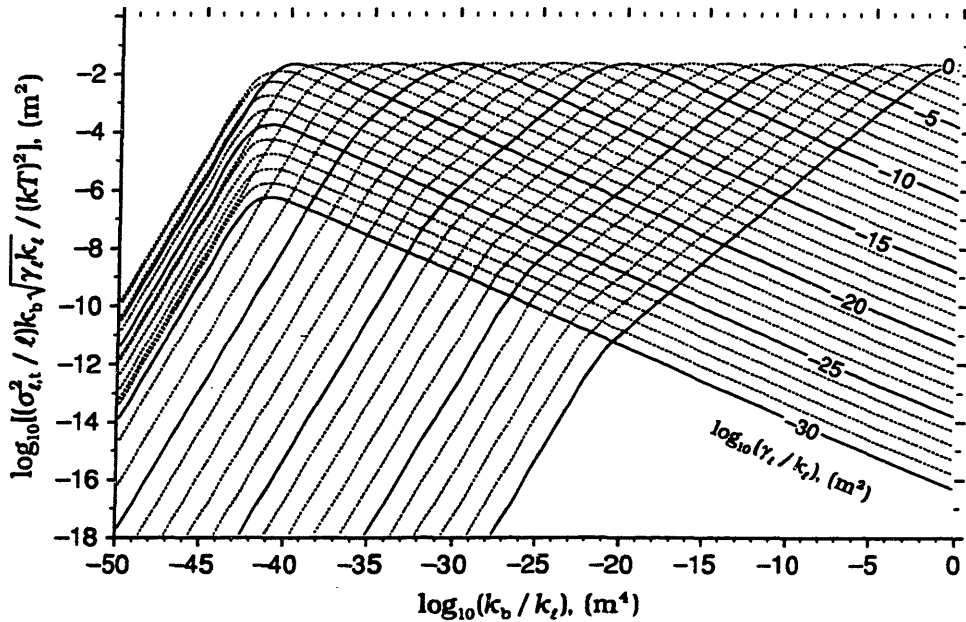


Figure 5.15. A measure of longitudinal variance vs. ratios of restoring forces, based on (5.94).

Since the numerators and denominators of these parameters can both vary over many orders of magnitude, even for sub-micron systems, the ranges covered by Figure 5.15 are large.

5.6.4. Rods with freely-sliding ends and no transverse constraint

The lack of transverse constraint forces and applied tension increases positional variance. For rods of finite length, angular freedom at the ends again increases the positional variance. Thus, these conditions are appropriate for setting upper bounds on the longitudinal positional variance resulting from longitudinal-transverse coupling in rods where the ends are constrained to the axis (or, equivalently, where the coordinate measured is the distance between the ends) but are otherwise free. Under these conditions, Eq. (5.94) simplifies to

$$\sigma_{t,t}^2 = \frac{1}{2} \left(\frac{kT}{k_b} \right)^2 \left(\frac{\ell}{\pi} \right)^4 \sum_{n=1}^N \frac{1}{n^4} \approx \frac{1}{2} \left(\frac{kT}{k_b} \right)^2 \left(\frac{\ell}{\pi} \right)^4 1.082 \quad (5.95)$$

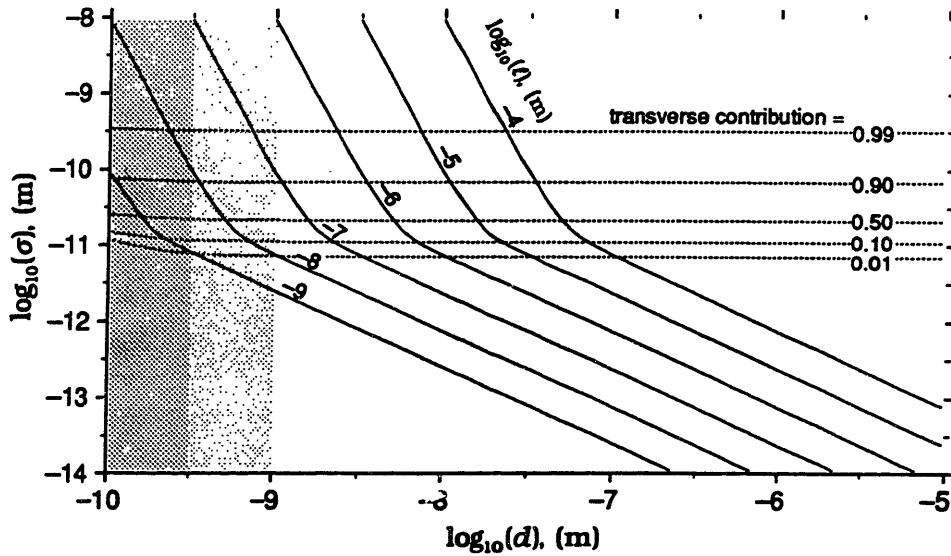


Figure 5.16. Total standard deviation in longitudinal displacement for the end of a thermally-excited rod, including elastic and entropic contributions. Plotted vs. diameter and length, assuming mechanical properties resembling those of bulk diamond (as in Fig. 5.8); similar comments apply.

with the latter expression representing the rapidly-approached limit as N becomes large. Again, contributions for both polarizations must be included. For convenience, Figure 5.16 displays the total longitudinal positional uncertainty resulting from the combined effects of this mechanism (including both equivalent polarizations) and of longitudinal modes as analyzed previously, all at 300K for rods with properties like those of bulk diamond.

It should be noted that the PDF for the value of Δl resulting from longitudinal-transverse coupling involving a single mode is not Gaussian, but a function of the form of Eq. (5.84). The PDF for the case just considered will roughly approximate this, since it is dominated by the contribution of a single mode. Where many modes contribute, however, their sum will approximate a Gaussian; the differences will chiefly affect the tails of the distribution.

5.7. Conclusions

Elastic and entropic springs show distinct behaviors. In elastic systems in the classical regime, positional variance is proportional to compliance and to temperature. Quantum effects increase the variance over the classical value, and become a greater proportion of the whole as stiffnesses increase and as masses and temperatures fall. In entropic systems in the classical regime, positional variance is proportional to the square of the temperature. Quantum effects reduce the variance below the classical value; in the quantum limit of large $\hbar\omega/kT$, all vibrational modes are in the ground state, entropy is zero, and the entropic variance is zero. (Entropic springs are associated with a compliance which, despite the foregoing, does not go to zero in the quantum limit.)

For analyzing the PFDs of elastic springs with respect to some set of coordinates in the classical limit, modal analysis can be bypassed: only the potential energy as a function of these coordinates is of significance. In the quantum regime, an analysis of vibrational modes serves to capture the effects of zero point energy. In the classical regime, an analysis of vibrational modes is a form of entropic bookkeeping; it provides a means of describing all possible system configurations, a step which is useful in analyzing entropic springs but irrelevant in analyzing elastic springs. The simple results Eq. (5.26) and Eq. (5.53) are direct consequences of this.

In the derivation of Eq. (5.26), rod configurations with an end displacement $A_{\text{end}} = \Sigma A_n$ were described as a sum of terms

$$\sum_{n=0}^{\infty} A_n \sin\left(\frac{(2n+1)\pi x}{2\ell}\right) \quad (5.96)$$

with many possible combinations of A_n for a given value of A_{end} . One could equally well describe configurations as a sum of terms

$$A_{\text{end}}x + \sum_{n=0}^{\infty} A'_n \sin\left(\frac{(n+1)\pi x}{\ell}\right) \quad (5.97)$$

This form separates the end displacement from orthogonal rod deformations; the relationship of end position to mean energy, system entropy, etc., would be the same if these deformations were embodied in an entirely separate object. Note that this recasting does not apply in the quantum regime because the deformation $A_{\text{end}}x$ does not correspond to a vibrational mode.

In general, for purely elastic systems in the classical limit, the PDF associated with a set of coordinates can be evaluated by applying a Boltzmann weighting to the potential energy as a function of those coordinates. For linear systems, the result will be a Gaussian distribution (a result which holds in the quantum regime as well). Note that, in the absence of imposed accelerations, inertial mass plays no role in determining equilibrium PDFs for the position coordinates of a system, though it may greatly affect the dynamics of fluctuations. For such systems, variance scales linearly with temperature and compliance; accordingly, variance is inversely proportional to linear dimensions, given uniform scaling of an elastic structure.

Considering quantum effects in elastic systems, it is always conservative to scale up an accurate or conservative variance in proportion to temperature or compliance; it is not conservative to scale down in the same manner. At room temperature, objects made of materials as stiff and light as diamond will have positional uncertainties dominated by classical effects so long as dimensions exceed one nanometer.

For purely entropic systems in the classical regime, variance scales as the square of temperature. It is always conservative to neglect quantum effects, since they reduce entropic variances. Entropic effects in the deformation of structural elements become important as aspect ratios become large. For such systems, the variance can be treated as the sum of entropic and elastic contributions.

Chapter 6

Transitions, errors, and damage

6.1. Overview

In the configuration-space picture, a transition in a nanomechanical system corresponds to a motion of the representative point from one potential well to another. Errors occur when a system executing a pattern of motion transiently enters an incorrect potential well. Damage occurs when a system permanently leaves the set of correct potential wells. Section 6.2 describes standard models used to describe transitions in molecular systems; these are used in later sections to analyze errors and thermomechanical damage.

The analysis of errors (like that of many other processes) is conceptually simpler when subsystems can be considered separately from the system as a whole. Section 6.3 describes methods for modeling subsystems within machines in terms of time-dependent potential energy surfaces, then uses this approach to model error rates in placing molecular objects into potential wells.

The balance of the chapter examines damage mechanisms. These can be divided into categories according to the source of the energy required to move the system over an energy barrier separating working from damaged states. The two broad categories are:

- Internal energy sources, including thermal excitation, mechanical stress, and electromagnetic fields
- External energy sources, including energetic photons and charged particles, and electromagnetic fields

Section 6.4 examines the effects of thermal excitation and mechanical stress, drawing on theoretical models of bond cleavage and experimental models of chemical reactivity. Section 6.5 examines photochemical damage and conditions for avoiding it by means

of opaque shielding; Section 6.6 examines ionizing radiation damage, presenting models based on experimental studies of damage to biological molecular machines. Section 6.7 summarizes the engineering consequences of these damage mechanisms for the reliability of nanomechanical systems, given the assumption that single-point damage always causes device failure.

Owing to scaling laws (Chapter 2), the magnetic fields that can be generated within nanoscale systems will be modest. These will (at most) influence rates and equilibria in electronic phenomena such as triplet-singlet intersystem crossing. In typical molecular structures, even the most intense fields generated in terrestrial physics laboratories will cause no damaging transitions; further discussion of the chemical effects of magnetic fields is accordingly omitted here. Electric fields, in contrast, can cause electrical breakdown and molecular damage. Some constraints resulting from these effects are touched on in Chapter 11.

6.2. Transitions between potential wells

6.2.1. *Transition state theories*

The previous chapter examined the effects of thermal excitation and quantum uncertainty in systems that can be described as points oscillating in a single potential well. This section examines thermal and quantum effects in systems characterized by a PES having two or more potential wells. Within each potential well, the location of the state point will typically be described by a PDF resembling those derived in the previous chapter. This PDF differs, however, in extending over a col leading to another potential well; as a consequence a system in one well has a finite probability per unit time of moving to another. At equilibrium, the PDF places a certain total probability mass in each well that results in equal rates of transition in each direction over each col. *Transition state theories* (TSTs) provide standard approximations for computing these equilibrium transition rates.

Transition state theories find extensive use in studies of chemical kinetics. They are of little value where conditions are far from microscopic equilibrium (for example, during the first vibration-cycle after a molecule has been photochemically excited) but are not impaired by macroscopic disequilibrium (for example, during the first seconds or nanoseconds after two reagents have been mixed). The discussion in Section 4.3.4 of the utility of equilibrium statistical mechanics in non-equilibrium systems is applicable here.

Classical and quantum mechanical TSTs of varying complexity and accuracy have been developed. In the present context, there is little incentive to seek moderate improvements in accuracy at the cost of great increases in complexity. For evaluating nanomechanical designs, it typically suffices to establish either (1) that an undesirable transition does (or does not) occur at a sufficiently low rate in a given structure, or (2) that a desirable transition does (or does not) occur at a sufficiently high rate with a given reagent or other molecular device. Since transition rates at 300 K commonly vary from $> 10^0 \text{ s}^{-1}$ to $< 10^{-50} \text{ s}^{-1}$, an approximate value will frequently suffice to distinguish a workable from an unworkable design. The solution-chemist's concern with 10-fold differences in the rates of competing reactions seldom arises.

6.2.2. Classical transition state theories

Transition state theories (both classical and quantum mechanical) start with a description of the col surrounding the *transition state* of a PES. Within the col between the wells is a saddle point; the surrounding region is locally describable by a quadratic function. With a suitable choice of coordinate axes, the curvature of the surface is negative along one axis (which defines the direction of the *reaction coordinate*) and positive along the other axes. In the most common definition, the transition state corresponds to a surface perpendicular to the reaction coordinate, which passes through the saddle point and divides points associated with one well from those associated with the other. Some versions of transition state theory define different transition surfaces. (Note that this surface has one fewer spatial dimension than the potential energy surface; the next section will speak of *volumes* that correspond to regions of the PES and *transition surfaces* that separate those regions.)

In solution- and gas-phase chemistry, molecules are free to translate and rotate, and molecular number-densities are important variables. In nanomechanical systems, mechanical constraints will typically eliminate both translational and rotational degrees of freedom, and most processes will effectively be unimolecular, making number densities irrelevant to transition rates. Accordingly, the following treatment will be restricted to systems in which the potential wells are localized and distinctly bounded, permitting only vibrational motions within (and transitions between) wells.

6.2.2.1. Standard classical TST

Elementary transition state theory expresses the transition rate, k_{12} (s^{-1}), as

$$k_{12} = \frac{kT}{2\pi\hbar} \frac{q^\ddagger}{q_1} \exp\left(-\frac{\Delta\mathcal{V}^\ddagger}{kT}\right) \quad (6.1)$$

In the classical theory, q_1 is the classical partition function (Section 4.3.2) describing the system when confined to the region of potential well 1, and q^\ddagger is the classical partition function describing a hypothetical system constrained to move along the transition surface. (In calculating the transition-state partition function, the position and momentum coordinates corresponding to motion along the reaction coordinate are dropped from the integral) $\Delta\mathcal{V}^\ddagger$ is the potential energy difference between the the bottom of well 1 and the saddle point of the col.

Although correct within the classical approximation, Eq. (6.1) does not directly suggest a physical visualization that can aid engineering design. Clear expositions of transition state theory can be found (e.g., in (Knox 1971)). Other expositions, however, speak in terms of fictitious concentrations of arbitrarily-defined “activated complexes,” and seem more confusing than enlightening—indeed, to judge by errors in textbooks, they confuse even their authors. Finally, the presence of \hbar in a classical theory suggests that a detour has been taken somewhere along the path from classical concepts to classical consequences. An alternative approach may help to build a more intuitive understanding, at least of the classical model.

6.2.2.2. The probability-gas visualization

In a purely classical model, the physics behind transition state theory can be visualized in terms of the diffusion of a “probability gas” (as introduced in Section 4.3.3). Where the state of the system is uncertain (for example, at equilibrium) motion of the representative point must be treated statistically. Rather than thinking in terms of the behavior of a single point over an indefinitely long period of time, one can instead think in terms of the positions and velocities of an indefinitely large number of points at a single instant. In mass-weighted coordinates, these points behave like an ideal gas of uniform molecular weight and zero collision cross section: they move with the familiar isotropic gaussian velocity distribution at every point in the space, and their density at every point varies in proportion to the Boltzmann factor, $\exp(-\mathcal{V}(\text{position})/kT)$. When the total quan-

tity of gas is taken to be unity, the density of the gas corresponds to the probability density; the diffusion of this gas corresponds to the dynamics described by transition-state theory.

Figure 6.1(a) shows a top view of a two-dimensional square-well potential like that shown in 6.1(b). Since $\exp(-V/kT)$ is a constant within the enclosed region, the distribution of the probability gas will be uniform at equilibrium. Although there are no subsidiary wells, no proper col, and hence no transition state as ordinarily defined, one can divide the region into two nearly circular wells by defining a suitable transition surface,

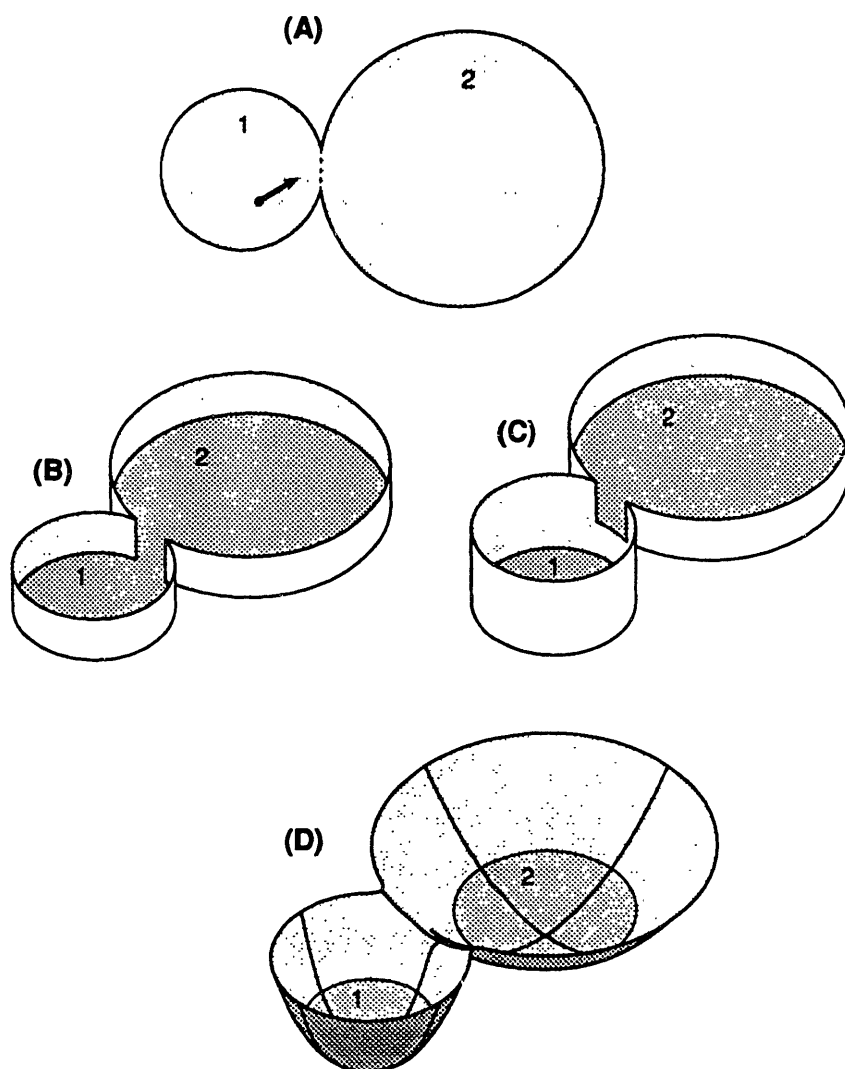


Figure 6.1. Potential energy surfaces illustrating concepts in transition state theory.

shown as the dotted line in 6.1(a). At equilibrium, the region on the right (state 2) will contain more probability gas than the region on the left (state 1): in a chemical system, state 2 would be described as “having a lower free energy owing to entropic factors.”

The rate of transitions from 1 to 2 (given initial occupancy of 1) is easy to calculate. State 1 has a certain volume (area, in two dimension) and the surface between the states has a certain area (width, in two dimensions). The transition rate is simply the volumetric rate at which probability gas exits the region of state 1 through the available patch of transition surface divided by the total volume of gas in state 1:

$$k_{12} = \frac{v_{\text{mean}} A_{\text{trans}}}{V_1} \quad (6.2)$$

where v_{mean} is the mean speed of particles of probability gas along a given direction (that is, the average of the absolute value of the velocity along that direction, divided by two to discount particles traveling in the wrong direction). This expression generalizes to any number of dimensions: An N dimensional volume is always bounded by an $N - 1$ dimensional surface; each patch of surface transmits gas at a rate equaling its area times the mean speed perpendicular to its surface, and the mean speed (in the above sense) is also independent of the number of dimensions. In mass-weighted coordinates,

$$v_{\text{mean}} = \int_0^\infty \frac{v}{\sqrt{2\pi kT}} \exp\left(-\frac{v^2}{2kT}\right) dv = \sqrt{\frac{kT}{2\pi}} \quad (6.3)$$

In a slightly more complicated model, Fig. 6.1(c), one region has a lower potential energy than the other. At equilibrium, the distribution of gas between states 1 and 2 will be determined by the product of their volumes and the Boltzmann factor $\exp(-\mathcal{U}/kT)$ for each region: this product defines a density-weighted volume, or *effective volume*. For suitable choices of V_1/V_2 , $\Delta\mathcal{U}$, and T , state 1 will now be favored and would be described as “having a lower free energy, despite its lower entropy.” The calculation of transition rates from 1 to 2 proceeds as before, but with a Boltzmann factor to account for the lower gas density at the “altitude” of the transition surface between the regions:

$$k_{12} = \frac{v_{\text{mean}}}{V_{\text{state 1}}} A_{\text{col}} \exp\left(-\frac{\Delta\mathcal{U}^\ddagger}{kT}\right) \quad (6.4)$$

This factor reduces the *effective area* of the transition surface relative to the volume of the well. Note that in this example there is no barrier in the reverse direction, hence the

effective area of the transition surface relative to well 2 is unreduced.

Figure 6.1(d) represents a smoother and more realistic PES to which quadratic approximations and classical transition state theory can be applied. In the probability-gas model, the effective volume of each well now must reflect a Boltzmann weighting integrated over the varying floor height, and the effective area of the transition surface (with respect to a well) must likewise be weighted to account for its varying height.

In a general harmonic well, the stiffness k_s may differ for each coordinate. For a given coordinate, the effective width of the well is (without mass weighting)

$$w_{\text{effective}} = \sqrt{2\pi}\sigma = \sqrt{2\pi \frac{kT}{k_s}} \quad (6.5)$$

or, with mass weighting as elsewhere in this section,

$$w_{\text{effective}} = \sqrt{2\pi kT \frac{m}{k_s}} = \sqrt{2\pi kT} \frac{1}{\omega} = \sqrt{\frac{kT}{2\pi}} \frac{1}{\nu} \quad (6.6)$$

allowing expression in terms of the weighting-independent frequency ν . The effective volume of a well is the product of the effective widths in each dimension

$$v_{\text{effective}} = \prod_{i=1}^n \sqrt{\frac{kT}{2\pi}} \frac{1}{\nu_i} = \left(\frac{kT}{2\pi}\right)^{n/2} \prod_{i=1}^n \frac{1}{\nu_i} \quad (6.7)$$

and a similar expression of lower dimensionality yields the effective area of the transition surface. Combining these, the transition rate expression becomes

$$\begin{aligned} k_{12} &= \nu_{\text{mean}} \frac{A_{\text{col}}}{V_{\text{state 1}}} \exp\left(-\frac{\Delta\mathcal{V}^\ddagger}{kT}\right) = \sqrt{\frac{kT}{2\pi}} \frac{\left(\frac{kT}{2\pi}\right)^{(n-1)/2} \prod_{i=1}^{n-1} \frac{1}{\nu_i^\ddagger}}{\left(\frac{kT}{2\pi}\right)^{n/2} \prod_{i=1}^n \frac{1}{\nu_i}} \exp\left(-\frac{\Delta\mathcal{V}^\ddagger}{kT}\right) \\ &= \frac{\prod_{i=1}^n \nu_i}{\prod_{i=1}^{n-1} \nu_i^\ddagger} \exp\left(-\frac{\Delta\mathcal{V}^\ddagger}{kT}\right) \end{aligned} \quad (6.8)$$

By using the expression for the classical partition function of a harmonic oscillator,

$$q_{\text{class}} = \frac{kT}{\hbar\omega} \quad (6.9)$$

and the observation that the partition function for a linear elastic system is the product of the partition functions for each of its normal modes, an expression identical to Eq. (6.8) can be derived from Eq. (6.1). This illustrates the equivalence between the probability-gas model (with its geometric visualization) and the more abstract formalism of classical transition state theory.

Equation (6.8) has a simple interpretation in the special case where the reaction coordinate leading over the col is an extension of one of the normal modes of the well (of frequency ν_{rc}) and the other modal frequencies are identical in both the well and transition state. A system of this sort is effectively one dimensional, and the ratio of products in Eq. (6.8) reduces to ν_{rc} . The representative point can then be described as striking the barrier ν_{rc} times per second, with a success rate in surmounting the barrier of $\exp(-\Delta V^\ddagger/kT)$.

6.2.2.3. Shortcomings and variational theories

Within the classical, equilibrium approximation, the TST described above is exact if the relevant integrals are done with an exact potential energy surface, and *if transitions are defined as motions that cross the transition surface separating the states*. For most practical purposes, however, this is a poor definition.

For some PESs, a typical trajectory from a low energy in one well to a low energy in another (assuming, as always, coupling to a heat bath) will oscillate across the transition surface several times. Further, trajectories may cross the surface and promptly be reflected back into the originating well. These are “transitions” according to the above definition, but not in any practical chemical sense. Accordingly, classical TSTs give an upper bound to the (classical) transition rate of practical interest. The difference is sometimes accommodated by an ad-hoc “transmission factor.”

Variational transition state theories (which lack ad-hoc factors) adjust the geometry of the transition surface to minimize the transition rate. The more sophisticated forms adopt different transition surfaces for trajectories of different energy. These methods yield lower transition rates that remain upper bounds on the rate of practical interest. A review of TSTs, including variational TSTs, can be found in (Bérces and Márta 1988).

Standard transition state theory typically fits exact classical calculations with reasonable accuracy.

6.2.3. Quantum transition state theories

Quantum mechanical corrections to classical transition state theory take three forms: use of a barrier height that accounts for the zero-point energies of the well and transition state, replacement of classical with quantum mechanical partition functions, and use of a correction to account for tunneling through the barrier in the region of the transition state (Bérces and Márta 1988). These can be addressed separately.

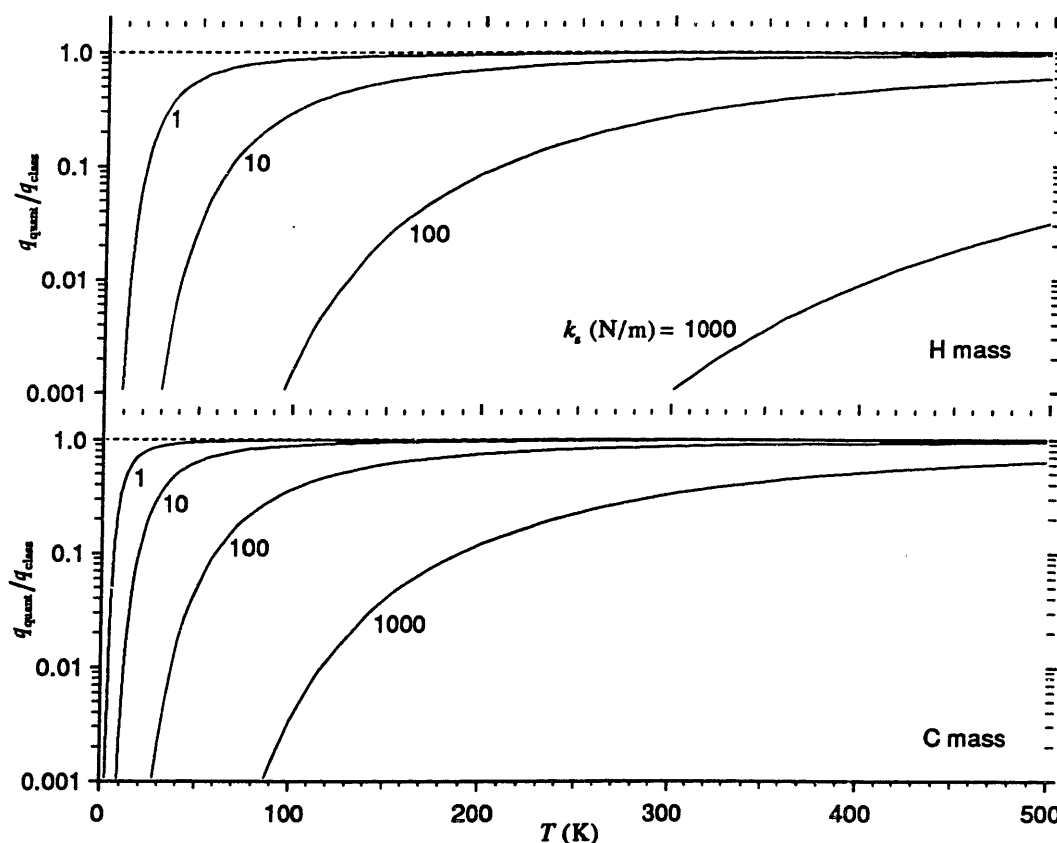


Figure 6.2. Ratios of quantum mechanical to classical partition functions for harmonic oscillators as a function of temperature and effective stiffness for two values of the effective mass at the well minimum. Since the TST expression contains one more modal factor in the denominator than in the numerator, decreases in the above quantum-to-classical ratios tend to increase transition rates.

It is conventional to correct barrier heights by adding a zero-point energy term to the energy of the well minimum. The zero-point energy of a linear elastic system is the sum of the zero point energies of its normal modes; the effective minimum energy of a state is then

$$\nu' = \nu + \sum_{i=1}^n \frac{1}{2} \hbar \omega_i \quad (6.10)$$

A standard form for the quantum-mechanical partition function of the harmonic oscillator,

$$q_{\text{quant,zpe}} = \left[1 - \exp\left(-\frac{\hbar \omega}{kT}\right) \right]^{-1} \quad (6.11)$$

takes the vibrational ground state as the zero of energy and is compatible with corrected barrier heights. Alternatively, one can omit the zero point energy correction of Eq. (6.10) from the barrier height, and instead use a partition function

$$q_{\text{quant}} = \exp\left(-\frac{\hbar \omega}{2kT}\right) \left[1 - \exp\left(-\frac{\hbar \omega}{kT}\right) \right]^{-1} \quad (6.12)$$

which takes the well minimum as the zero of energy. The choice is a matter of notational convenience, and expressions using this form more closely correspond to the classical equations.

The partition function for a linear system having n normal modes is the product of the partition functions of the modes

$$q_{\text{quant}} = \prod_{i=1}^n \exp\left(-\frac{\hbar \omega_i}{2kT}\right) \left[1 - \exp\left(-\frac{\hbar \omega_i}{kT}\right) \right]^{-1} \quad (6.13)$$

Transition state partition functions are again computed by omitting factors associated with motion along the reaction coordinate. Figure 6.2 shows how the ratio of the classical and quantum mechanical partition functions varies with temperature and stiffness.

A standard approximation to the tunneling correction factor, due to Wigner, is

$$\Gamma^* = 1 + \frac{1}{24} \left| \frac{\hbar \omega_{\text{rc}}^*}{kT} \right|^2 \quad (6.14)$$

where

$$\omega_{1c}^{\ddagger} = \sqrt{\frac{k_{s,rc}^{\ddagger}}{m_{eff,rc}}} \quad (6.15)$$

is an imaginary frequency associated with motion along the reaction coordinate (m_{eff} is the effective mass along this coordinate, and the stiffness along the reaction coordinate $k_{s,rc}$ is negative) (Bell 1959; Bigeleisen 1949). Figure 6.3 shows how Γ^* varies with temperature and stiffness. (Section 6.4.4.2 notes the inapplicability of Γ^* at low temperatures.)

Combining the above substitutions and corrections yields the approximate quantum-mechanical TST expression

$$k_{12} = \Gamma^* \frac{kT}{2\pi\hbar} \frac{q_{quant}^{\ddagger}}{q_{quant,l}} \exp\left(-\frac{\Delta\mathcal{V}^{\ddagger}}{kT}\right) \quad (6.16)$$

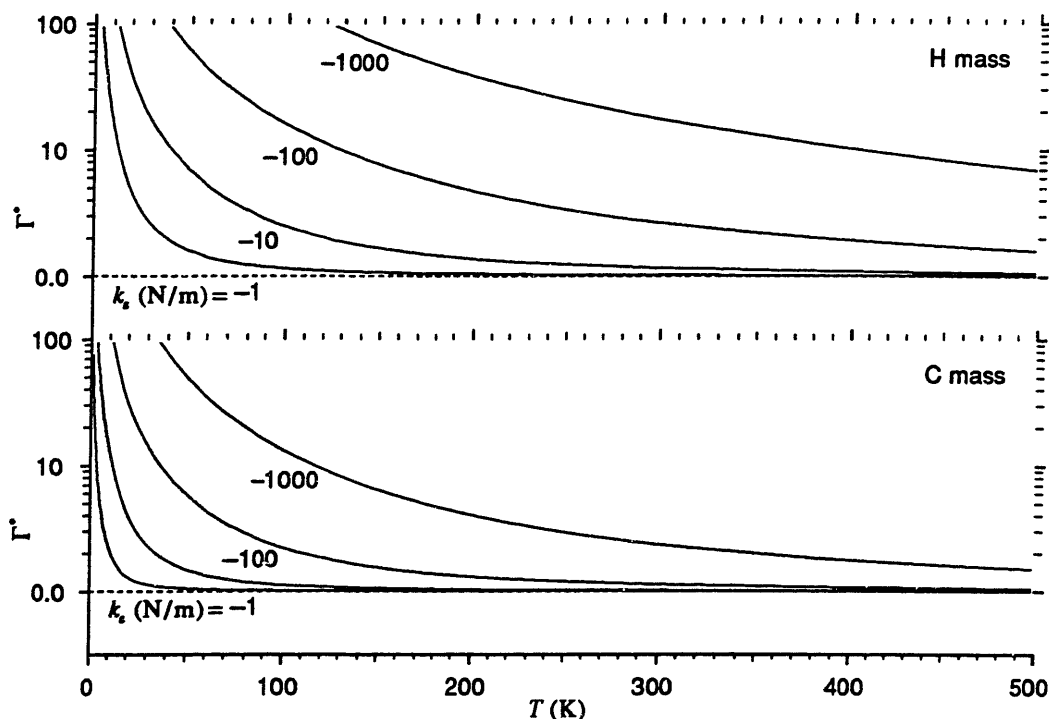


Figure 6.3. The Wigner tunneling correction factor, Eq. (6.14), as a function of temperature and effective stiffness for two values of the effective mass for motion along the reaction coordinate.

More sophisticated quantum-mechanical TSTs are reviewed in (Bérces and Márta 1988).

6.2.4. Tunneling

Tunneling processes dominate the transition rate in all systems at sufficiently low temperatures and in many electronic systems at room temperature. When this occurs, transition state theories based on the Wigner tunneling correction become inadequate and can radically underestimate the transition rate. A different model must be used.

Estimates of tunneling transition rates for a particle in a well bounded by a barrier are derived by multiplying the barrier transmittance, T , by a suitable frequency factor,

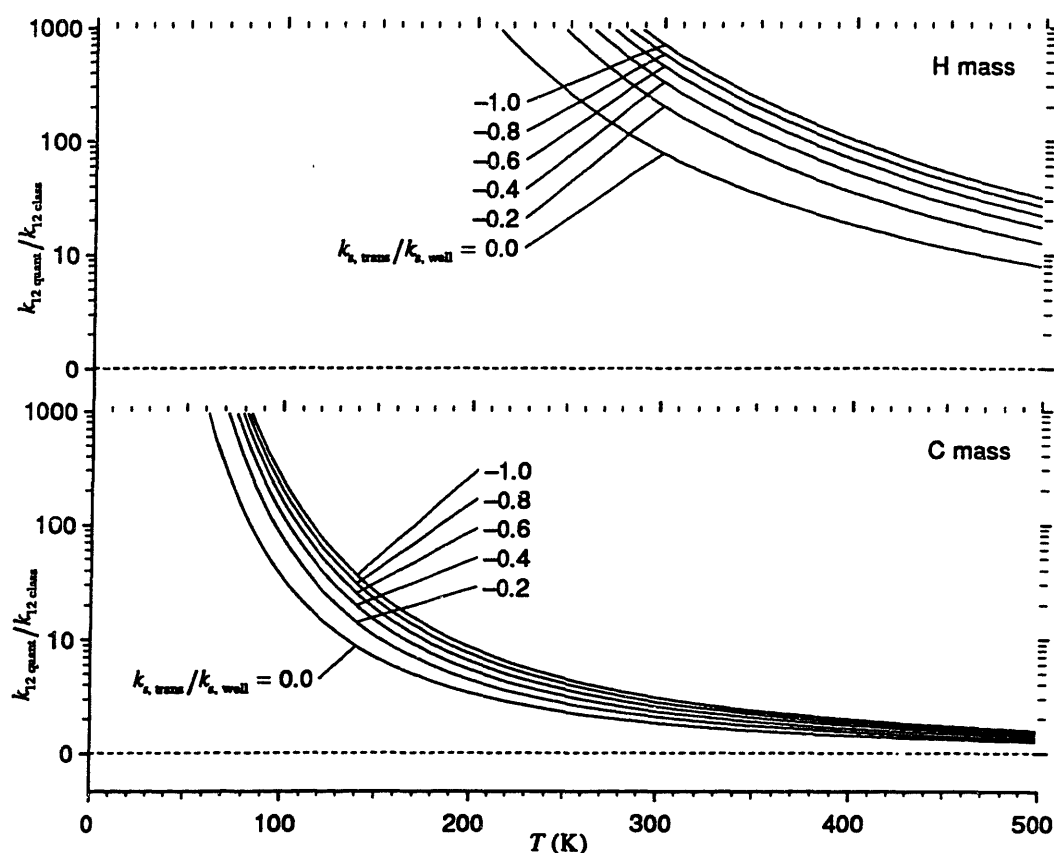


Figure 6.4. The ratio of quantum to classical transition rates, assuming a one-dimensional model with the stated ratios of stiffnesses along the reaction coordinate.¹⁶⁴ The stiffness at the well minimum is 500 N/m (typical of covalent bond stretching, large in most other contexts.). Note that a stiffness ratio of zero implies an infinitely thick barrier and an absence of tunneling.

such as $\omega/2\pi$ (for a harmonic well) or the round-trip traversal time (for a square well). In the classical picture, the particle strikes the barrier with a certain frequency and probability of penetration.

Quantum theory yields exact expressions for T for a variety of cases (such as rectangular barriers) but real systems seldom correspond to one of these. Nanomechanical and electromechanical designs of the sort described in this volume require approximations of broader utility. As with transition state theory, the typical objective will be to discriminate between those systems that do and do not have tunneling rates low enough to be acceptable (e.g., low enough for an insulator to transmit little current or for a stressed bond to have a long expected lifetime). A standard method applicable to many systems of low T exploits the WKB (Wentzel-Kramers-Brillouin) approximation.

6.2.4.1. The WKB tunneling approximation

A useful approximation for the transmittance of a potential barrier $\mathcal{V}(x)$ for a normally-incident particle of energy \mathcal{E} and effective mass m_{eff} is

$$T \approx \exp \left[-2 \int_{x_0}^{x_1} \sqrt{\frac{2m_{\text{eff}}}{\hbar^2} [\mathcal{V}(x) - \mathcal{E}]} \right] \quad (6.17)$$

where the limits of integration define the region in which $\mathcal{V}(x) > \mathcal{E}$ (Fig. 6.5). The approximation requires that the exponential decay of the wave function in the region of negative energy be the dominant phenomenon, making $T \ll 1$, and requires that tunneling occur into a continuum of energy states, such as that characterized by free motion of a particle. Where several barriers occur in close succession, resonance effects can yield complex relationships between energies and transmittances.

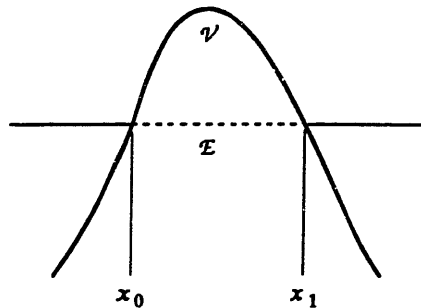


Figure 6.5. Particle total energy and potential energy in tunneling, showing the limits of integration used in the WKB approximation.

6.3. Placement errors

6.3.1. Time-dependent PES models

Section 6.2 implicitly assumes that the PES is unchanging. From a perspective that includes all particles this is always correct, yet it is frequently convenient to consider mechanical subsystems separately, accounting for the effects of other subsystems by imposing changing boundary conditions on the subsystem under analysis. These boundary conditions involve changes in the PES as a function of time. Detailed dynamical simulations (e.g. (Landman, Luedtke et al. 1990)) sometimes model the mechanical motion of an adjacent structure by imposing definite trajectories on boundary atoms, and model thermal equilibration by imposing temperature constraints on additional atoms. The more abstract models discussed in this section describe a time-dependent PES directly; like standard TSTs, they introduce a thermal-equilibrium assumption by means of a basically statistical description.

A variety of processes in nanomechanical systems can be described as attempts to place something in a preferred potential well. In systems that perform measurement, the correct potential well corresponds to a correct measurement of a physical parameter; in nanomechanical computer systems (Chapter 11), the correct potential well corresponds to the correct state of a logic device; in molecular manufacturing (Chapter 12), the correct potential well corresponds to a correctly-bonded product configuration. The potential wells and the trajectories that specify the process of placement will in the general case

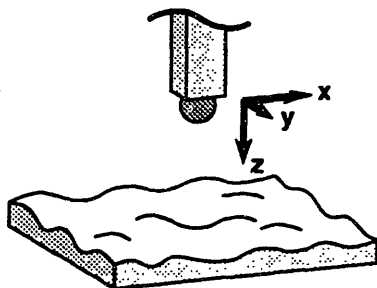


Figure 6.6. Model system consisting of a probe atom positioned by an elastic beam and subject to perturbations both from thermal excitation and from interactions as it encounters the surface below.

involve motions describable only in terms of a PES of high dimensionality; in practice, typical examples of each of the above cases can be described rather well in terms of thermal displacements of an atom (or small group of atoms) in one or two dimensions while subject to mechanical motion in a second or third dimension.

For a concreteness, picture a probe-atom at the tip of an elastic beam descending through empty space toward a surface (Fig. 6.6), described by the coordinates x, y, z , and z' . In this model, z' is the equilibrium value of z in the absence of thermal excitation and perturbing forces from the surface, or alternatively, the limit of the value of z as the stiffness components $k_{sx} = k_{sy} = k_{sz} \rightarrow \infty$. In a nanomechanical system, the value of z' will be determined by the configuration of the rest of the system. The potential $\mathcal{V}(x, y, z)_{z=z'}$, can be regarded as a z' -dependent two-dimensional potential $\mathcal{V}(x, y)$. This potential will typically be harmonic at large distances from the surface, owing to the beam stiffness; as it approaches the surface, corrugations will be superimposed on the harmonic potential, growing in magnitude as the probe grows closer. (At contact, the probe may be forced into a hole by pressure from the arm, or may be attracted to a site by electrostatic interactions, or may be bonded to a reactive group.) We seek a potential of the form $\mathcal{V}'(x, y, z')$ in order to describe time-dependent two-dimensional systems $\mathcal{V}[x, y, z'(t)]$. This is possible if we assume that the probe is mechanically stable in z (i.e., that the z stiffness, summing the beam stiffness and the probe-surface interaction stiffness, remains positive); in the absence of this assumption, transitions involving z excitations in a time-dependent three-dimensional potential must be considered explicitly.

How does \mathcal{V}' differ from \mathcal{V} ? At zero K, it equals

$$\mathcal{V}'_0(x, y, z') = \mathcal{V}(x, y, z_e) + \frac{1}{2}k_{zz}(z' - z_e)^2 \quad (6.18)$$

where z_e is the equilibrium value of z for the given value of x, y , and z' . This is the limiting case of a mean-force or free-energy potential, in which an entropic term accounts for variations in the compression of vibrations in the z direction. Making the harmonic approximation for vibrations about z_e and using the classical vibrational partition function,

$$\mathcal{V}'_0(x, y, z') = \mathcal{V}(x, y, z_e) + \frac{1}{2}k_{zz}(z' - z_e)^2 - kT \left[1 + \ln \left(\frac{kT}{\hbar\omega} \right) \right] \quad (6.19)$$

6.3.2. Error models

A detailed model would use Monte Carlo methods to estimate probabilities, integrating the equations of motion for a system. This method could take account of complex potentials and rapid motions yielding systems far from thermal equilibration, but would be computationally expensive and offer little general insight.

Transition-state theories can provide a basis for a relatively detailed model of placement errors. In this approach, one would define regions corresponding to growing potential wells and moving transition surfaces corresponding to growing barriers. Initial assignments of probabilities to wells would follow the Boltzmann distribution, and integration of transition rates would trace the evolution of those probabilities over time. To each transition rate predicted by a fixed-potential TST would be added a rate corresponding to the product of the probability density along each transition coordinate and the rate of motion of the associated transition surface in the time-dependent potential. This approach would be tractable for many problems.

A simpler model can be constructed starting with the observation that, in most systems of engineering interest, barriers between states are initially absent and eventually so high as to preclude significant transition rates. This leads to consideration of models in which equilibration among regions of configuration space switches from complete to nonexistent, omitting consideration of the narrow range of barrier heights in which transition rates are neither fast nor negligible compared to the characteristic time scale of the placement process. As usual, the goal is not to make a perfect prediction of physical behavior, but to distinguish workable systems, identifying them with a methodology that yields a low rate of false positives without an overwhelming rate of false negatives.

6.3.3. Switched-coupling error models

The probability of a state (or set of states) is effectively frozen (equilibration shut off) when the lowest barrier between it and other states meets the approximate condition

$$v_{\text{TST}} \exp\left(-\frac{\Delta\mathcal{V}''(t)}{kT}\right) \approx \frac{1}{kT} \frac{\partial}{\partial t} \Delta\mathcal{V}''(t) \quad (6.20)$$

assuming that the barrier height $\Delta\mathcal{V}(t)$ and the equilibrium ratio of probabilities are both smoothly increasing. (In this expression, v_{TST} is a frequency factor taken from transition state theory.) This can be termed the time of *kinetic decoupling* for the two wells; a useful

approximation treats this as a discrete time at which equilibration is switched off. A one-dimensional classical model illustrates the consequences of switching equilibration off at different times.

6.3.3.1. The sinusoidal-well model

The simplest time-dependent PES for modeling placement errors combines a fixed, one-dimensional harmonic potential with a time-varying sinusoidal potential

$$\Delta\mathcal{V}'(t) = \frac{k_s}{2}x^2 + A[1 - \cos(2\pi x/d_{\text{err}})]t \quad (6.21)$$

This model captures several basic features of placement errors in nanomechanical systems: the harmonic well is aligned to maximize the probability of finding the system in the target potential well in the center; other wells, corresponding to error states, appear at some distance d_{err} from the target well; well depths grow over time. The bottom panel of Fig. 6.7 illustrates the shape of this potential at varying times for two values of d_{err} .

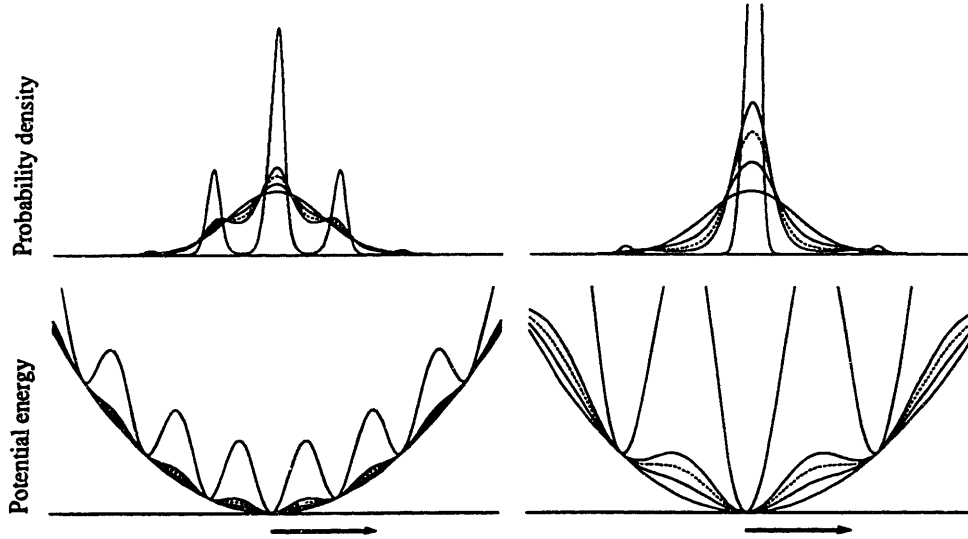


Figure 6.7. The lower panels illustrate time-dependent potentials described by Eq. 6.21; the upper panels illustrate corresponding Boltzmann probability density functions. In both panels, the dotted curves correspond to the potential at t_{crit} (Section 6.3.3.4); other curves are at 0.5, 1.5 and 10 times t_{crit} .

6.3.3.2. The total-equilibrium limit

If equilibration continues until the wells that will eventually hold significant probability are all sharp compared to the initial harmonic well, then the probability mass associated with each is determined only by its depth, which is the height of the harmonic well at a minimum of the sinusoid. In this model, the probability of error is given by the ratio of sums

$$P_{\text{err}} = 2 \sum_1 \exp\left(-\frac{k_s (nd_{\text{err}})^2}{2kT}\right) / \sum_{\infty} \exp\left(-\frac{k_s (nd_{\text{err}})^2}{2kT}\right) \quad (6.22)$$

Figure 6.8 shows the probability of error in the total equilibrium limit, as a function of \mathcal{V}_1 . The generalization of this model is standard chemical equilibrium, in which wells can have varying free energies (that is, depths and effective volumes) and have the corresponding Boltzmann-weighted probabilities of occupancy.

6.3.3.3. The instantaneous-onset limit

If A is sufficiently large, the barriers imposed by the sinusoidal potential will become large in a time short compared to the transit time of a single well. The instantaneous-onset limit falls outside the above model (since probabilities are frozen in their pre-barrier distribution, rather than at a finite barrier height), but its consequences are readily calculated. The probability of an error is the integral

$$P_{\text{err}} = 2 \int_{d_{\text{err}}/2}^{\infty} \sqrt{\frac{kT}{2\pi k_s}} \exp\left(-\frac{k_s x^2}{2kT}\right) dx \quad (6.23)$$

of the initial harmonic Boltzmann distribution outside the location of the peaks that will bound the target well. The upper curve in Figure 6.8 illustrates how this probability varies as a function of \mathcal{V}_1 , the energy difference between the target state and the lowest-energy error state. The generalization of this instantaneous limit in a multidimensional system would start with an ellipsoidal gaussian distribution and impose on this the state boundaries of the final potential surface, taking the integrated probability outside the boundaries of the correct state as the probability of error.

For a nanomechanical system to be well-approximated by the instantaneous limit, its mechanical speeds would typically have to be large compared to its thermal speeds. For

combinations of sufficiently large effective masses and sufficiently high mechanical speeds, this may hold true, yet it is unlikely to be a common case. The instantaneous limit maximizes error rates, hence high speeds (which also increase energy dissipation) will typically be avoided.

6.3.3.4. The worst-case decoupling model

Within the bounds of the kinetic-decoupling model (allowing free equilibration until some time after energy barriers appear), there exists a decoupling time that maximizes the error rate. In the sinusoidal-well model, the first barriers appear at a time

$$t_{\text{crit}} = 0.2332 \frac{d_{\text{gr}}^2}{A} \quad (6.24)$$

then grow in height and move inward toward the origin. Figure 6.7 illustrates two potentials and their associated PDFs, showing t_{crit} . Dividing the PDF at the time and position of the initial barrier appearance yields the dashed curve in Fig. 6.7; varying the time (and resulting PDF and barrier position) to maximize the error probability yields the solid

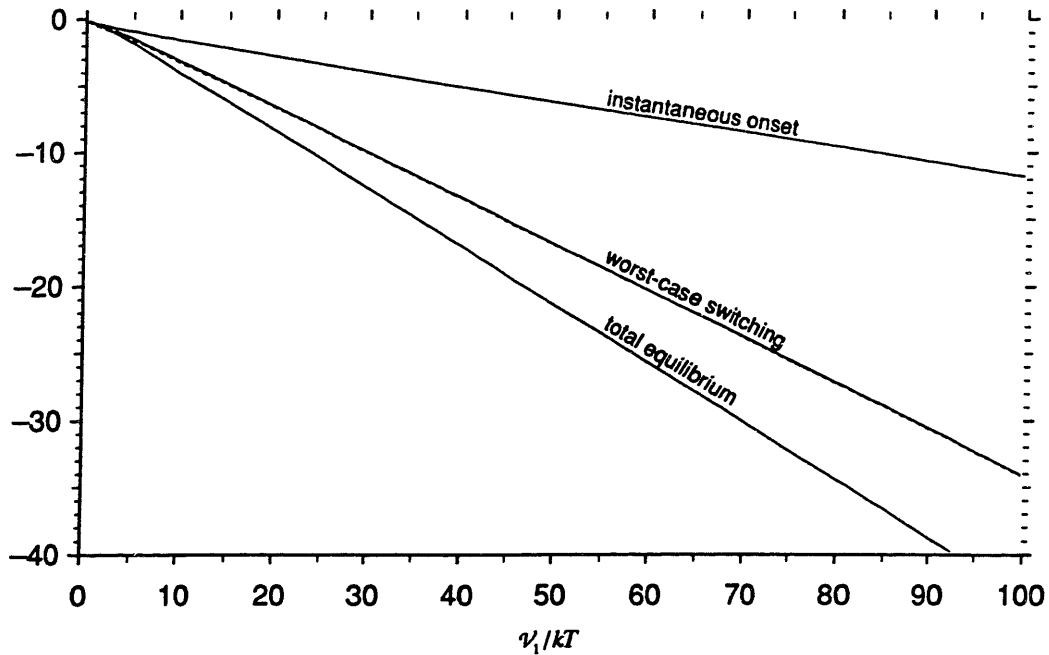


Figure 6.8. Error rates per placement operation resulting from the application of different equilibration assumptions to the sinusoidal-well model.¹⁷¹

curve slightly above it.

This model, which permits barrier crossing only so long as it increases errors, yet gives error-rates far lower than the instantaneous-onset model, seems a good choice for making conservative estimates of error rates in a variety of circumstances. The conservatism breaks down when error rates are large, that is, when $d_{\text{err}} < \sim 2(kT/k_s)^{1/2}$, but this is outside the usual range of interest. A more common failure would seem to be excessive conservatism in realistic, non-sinusoidal systems where a large well may appear *after* being cut off by a high barrier: it will then assign an unrealistically high probability to that well. With the notion of state boundaries that change over time, the worst-case switching model readily generalizes to multidimensional systems.

6.4. Thermomechanical damage

6.4.1. Overview

Thermal excitation, sometimes aided by mechanical stress, can cause permanent damage to devices built with molecular precision. Damaged states (as distinct from transiently disturbed states) occur when a transition occurs into a stable potential well corresponding to a non-functional device structure. Causes of damage include:

- Reactions like those between small molecules
- Rearrangements like those within small molecules
- Reactions like those at solid surfaces
- Rearrangements like those within solid objects
- Bond cleavage accelerated by mechanical stress

The first four points can be approached via analogies to known chemical systems. Several of the following sections follow this strategy, building on empirical knowledge and theoretical scaling relationships that describe how transition rates vary with temperature. Here (as in Chapter 8) it seems wise to begin by comparing the general conditions of solution-phase and machine-phase chemical processes, as an aid to forming the proper generalizations from present chemical experience.

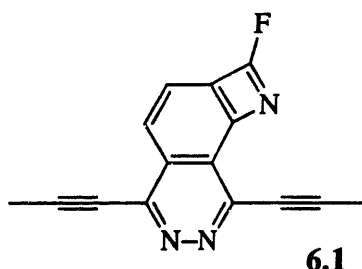
Throughout this section, a transition lifetime of 10^{20} s will be taken as a standard of comparison. As will be seen in the following, this lifetime suffices to make damage resulting from the associated transitions negligible compared to that from ionizing radiation, assuming as always that complete device failure will result from a single damaging transition.

6.4.2. Machine- vs. solution-phase stability

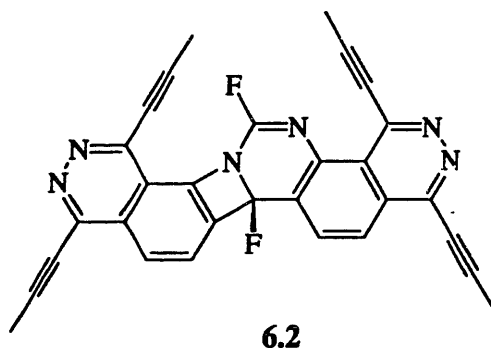
Machine-phase chemistry comprises systems in which all potentially reactive moieties follow controlled trajectories. Although borderline cases can be imagined, the contrast is clear between systems of small, diffusing molecules in the liquid phase and systems of molecular machinery as envisioned here. In solution-phase chemistry, potentially reactive moieties encounter each other in positions and orientations constrained by only local molecular geometry, not by control of molecular trajectories. In the solution phase, access to a reactive moiety may be blocked by attached hydrocarbon chains; in the machine phase, even an exposed reactive moiety may never encounter another molecule.

6.4.2.1. Examples

The stability of a structure often depends on its physical environment. For example, a molecular structure including the fragment:

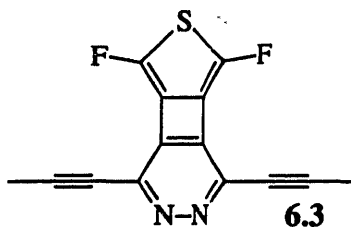


would be unstable in solution, readily forming dimeric products such as:

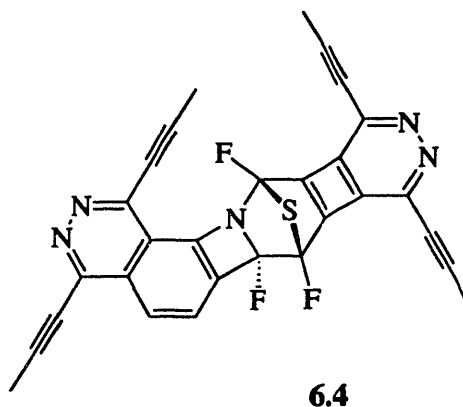


among others. The formation of dimers, however, requires that two molecules of the same kind encounter one another; this is inevitable in solution, but not in molecular machines. Structure 6.1 has been proposed for a role in a nanomechanical computer (Drexler 1988) in which such encounters would not occur.

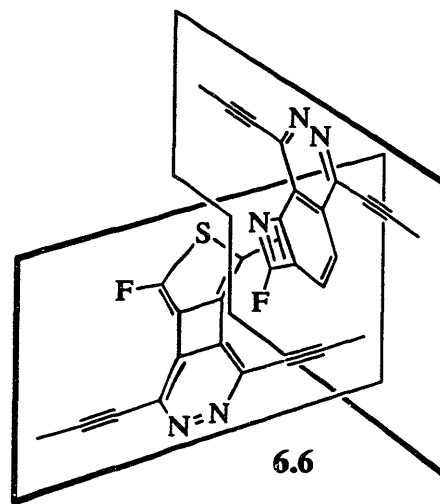
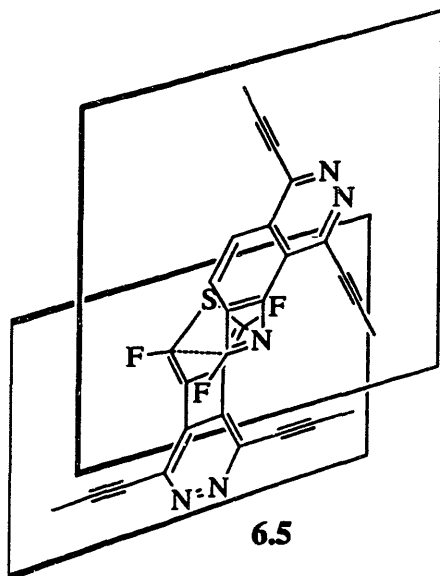
In that design, however, **6.1** would encounter other structures including the fragment:



which in solution might yield the Diels-Alder adduct:



and a variety of other products. To approach the Diels-Alder transition state, the rings would have to encounter one another in a nearly parallel orientation (**6.5**), but in the application proposed, a surrounding matrix would mechanically constrain the rings, forcing them to approach in a perpendicular orientation (**6.6**),



thereby precluding this reaction.

6.4.2.2. *Machine-phase damage mechanisms*

To generalize from these examples:

- Structures that are unstable in all solution environments owing to bimolecular self-reactions can fail to react in mechanical systems owing to a lack of collisions.
- Structures that react on collision in all solution environments can fail to react on collision in mechanical systems owing to constraints on orientation.

In solution chemistry, exposed moieties always encounter solvent molecules and frequently encounter dissolved oxygen, dissolved water, trace contaminants, and container materials, each potentially reactive. In machine-phase chemistry, potentially-reactive moieties are assumed to encounter other structures only by design. In some nanomachines (Chapters 8 and 12), certain structures and encounter conditions will be designed to foster reactions; reagent devices of this sort are not considered here. More typically, structures and encounter conditions will be designed to discourage reactions, and a reaction will constitute damage.

In non-reagent nanomechanical systems, contacts between structures serve mechanical functions: bearing surfaces, cam riders, gear teeth and the like are typical examples. A theme in later chapters will be the stability of the resulting interfaces. Where no mechanical function is served, no contact is necessary, and bimolecular reactions can be excluded from possibility.

Nanomechanical systems will, however, be subject to increased reactivity owing to mechanical stresses not found in solution chemistry. These stresses can be used to facilitate desired reactions Chapter 8, but can also facilitate damage. Stresses are subject to engineering control, and estimation of their chemical effects is important to engineering design. These stress-induced reactions, however, are special cases of unimolecular damage processes involving unimolecular reactions of sorts familiar in conventional chemistry. These reactions will be little affected by control of trajectories.

6.4.2.3. *Design and unimolecular stability*

To provide a thorough discussion of unimolecular reactions would require a thorough discussion of much of chemistry: impossible in a single book and further from possibility in a section of a chapter. Even restricting the discussion to organic molecules would require inclusion of much of organic chemistry, since the ends of long, flexible molecules can react with each other in a manner essentially the same as two separate molecules.

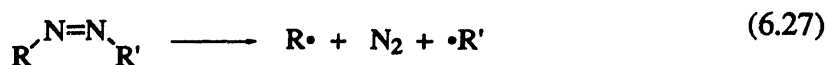
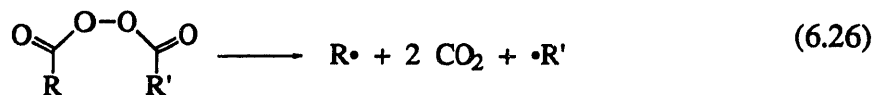
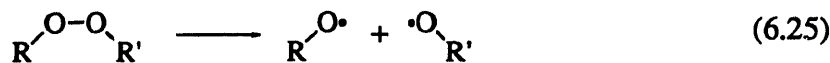
The pursuit of engineering objectives, however, greatly restricts the range of structures to be considered. For example, these objectives typically favor designs in which structures are either rigid, or are kept in an extended conformation by tension or by matrix constraints. Designs with these characteristics will exclude this class of reactions.

Likewise, the pursuit of engineering objectives will generally favor the selection of structures with strong, stable patterns of bonding. Errors in discriminating between stable and unstable structures during the early phases of nanomechanical engineering will be eliminated by more thorough analysis and (eventually) experimental testing. The aim of the following is to outline some principles of importance in (1) generating reasonable designs and (2) using the chemical literature to evaluate specific designs in more detail.

With regard to the latter, a possible source of cognitive bias is worth noting: In solution chemistry, the focus of attention is on the active reagent molecules, not on the comparatively inert solvent molecules. Indeed, in a typical textbook of organic chemistry, roughly half the chemical species appear to the left of an arrow representing a transformation to another species, $A \rightarrow B$; if A were stable under the specified reaction conditions ($A \rightarrow A$), it would seldom be worth mentioning. For those only slightly acquainted with chemistry, this consistent focus can create the illusion that significant reactivity is nearly ubiquitous. In nanomechanical systems, however, many structures play a role more like that of solvents: these structures may move with respect to one another (and perhaps transport reagents) without necessarily being reactive. If need be, the interactions between these nanomechanical surfaces could be made to more closely resemble those among (e.g.) hexane, benzene, and methyl ether than those among (e.g.) sodium hydroxide, bromoethane, hydrochloric acid, and cyclopentadiene. Reactivity will depend on specifics of surface structure and interaction geometry, both of which will eventually be subject to engineering control guided by experimentation.

6.4.3. Thermal bond cleavage

Direct cleavage of a bond by thermal excitation, forming a pair of radicals, is usually slow at ordinary temperatures. Among the major exceptions are organic peroxides, diacyl peroxides, and azo compounds, which find use as initiators of radical reactions:



6.4.3.1. Thermal bond cleavage rates

The rate of bond cleavage, $k_{\text{cl}} \text{ s}^{-1}$, cannot be estimated using standard transition state theory, for molecules separating into gas-phase radicals: the effective transition state is at infinity, and the transition-state frequencies are zero. Variational theories can give sensible answers, but a rough approximation (low by a small factor) can be had using a one-dimensional theory. In this,

$$k_{\text{cl}} = \frac{kT}{2\pi\hbar} \frac{1}{q_{\text{long}}} \exp\left(-\frac{\Delta\mathcal{V}}{kT}\right) \quad (6.28)$$

where q_{long} is the partition function for longitudinal vibrations of the bond,

$$q_{\text{long}} = \exp\left(-\frac{\hbar}{kT} \sqrt{\frac{k_s}{\mu}}\right) \left[1 - \exp\left(-\frac{\hbar}{kT} \sqrt{\frac{k_s}{\mu}}\right)\right]^{-1} \quad (6.29)$$

approximating the effective mass of the system as the reduced mass

$$\mu = \frac{m_1 m_2}{m_1 + m_2} \quad (6.30)$$

and the effective stiffness as the bond stiffness, k_s . The tunneling correction Γ^* is unity owing to the effectively infinite barrier thickness. Rate increases owing to the increasing effective area of the escape channel with increasing bond length are neglected. Figure 6.9 shows values for the characteristic bond-cleavage time, $T_{\text{cl}} = (k_{\text{cl}})^{-1}$, based on Eq. (6.28). As will be seen in Section 6.7, there is seldom reason to seek bond lifetimes $> 10^{20} \text{ s}$; of the bonds selected for listing in Table 3.7, only the O—O bond of organic peroxides falls short of this lifetime in the absence of strain or other adverse influences on bond stability.

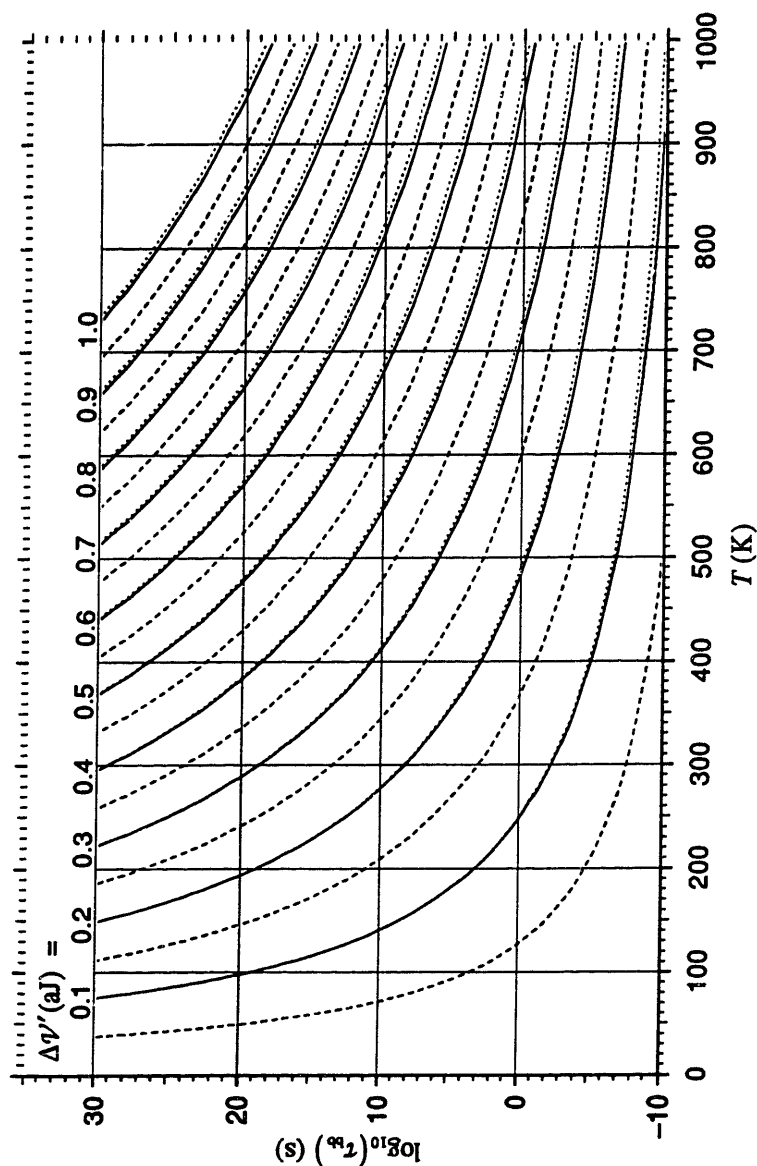
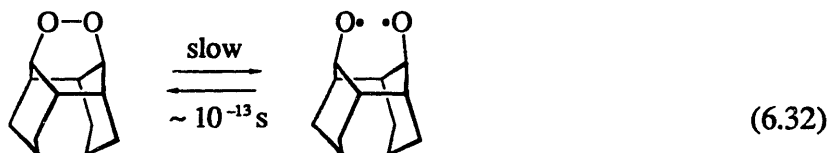


Figure 6.9. Values of the characteristic time for bond breakage, τ_{cl} , as a function of temperature and the bond energy $\Delta\mathcal{V}'$. The solid and dashed lines represent systems in which the partition function q_{long} is unity (corresponding to the high-frequency limit); the dotted lines just above the solid lines represent systems in which this partition function is calculated for $\omega = 7 \times 10^{13}$ rad/s, a relatively low frequency like that of a Si–Si bond.

6.4.3.2. Rate modifications in liquid and solid media

In liquids, the observed rate of thermal bond cleavage is substantially reduced by the dynamic “cage effect”: radicals surrounded by solvent molecules are confined for several molecular vibration times, permitting opportunities for immediate (geminate) recombination. In solids, geminate recombination is far more likely than in liquids owing to the near-absence of diffusion; they are trapped in solid cages. In the interior of a rigid diamondoid solid, and in the absence of intense stress, separation of radicals will frequently require the simultaneous cleavage of additional bonds, multiplying the effective magnitude of the barrier height and drastically reducing failure rates.

Attachment to a rigid structure can have dramatic effects even for surface moieties. For example, while an ordinary organic peroxide would be subject to irreversible cleavage in solution (once separated, the two radicals are unlikely to recombine), Eq. (6.31), an analogous peroxide structure attached to a rigid structure would (in the absence of neighboring reactive moieties) reliably recombine on a 10^{-13} s time scale Eq. (6.32).



At interior sites, a physical model for thermally-activated structural damage is self-diffusion in diamond-like crystals of germanium and silicon, believed to occur largely through a vacancy-hopping mechanism (Reiss and Fuller 1959). This process requires the cleavage of bonds, as would a damage mechanism in a diamondoid solid; unlike an atom in the interior of a well-designed diamondoid solid, however, an atom adjacent to a vacancy has available to it both room into which to move, and unsatisfied valencies with which to bond. Despite these facilitating structural features, the activation energies (which approximate the barrier energies discussed above) for vacancy diffusion in germanium and silicon (Lide 1990) have been estimated to be 0.503 and 0.745 aJ, about 1.6 and 2.0 times their respective bond energies.

A better physical model for thermal damage in the interior of a stable diamondoid solid might be the creation of a new vacancy within a covalent crystal. For this process, activation energies are much higher.

6.4.4. Thermomechanical bond cleavage

Tensile stress destabilizes bonds, increasing the rate of thermal cleavage and sometimes opening a tunneling path to cleavage. Angular strains are common in organic chemistry, being displayed in three-membered rings. Tensile strains are less common in organic chemistry but will frequently be important in nanomechanical systems, where engineering performance often depends on the imposed stress, providing an incentive to push toward the allowable limits of stress.

6.4.4.1. Thermomechanical bond cleavage rates

Cleavage rates for bonds under simple tensile stress at ordinary temperatures can be approximated with a one-dimensional quantum TST like that in Section 6.2.3 above, but with frequencies and barrier heights calculated with the aid of a potential energy function describing bond extension (a model extended to include shear would be of value). Combining the Morse function Eq. (3.11) with the potential energy resulting from an applied force F yields

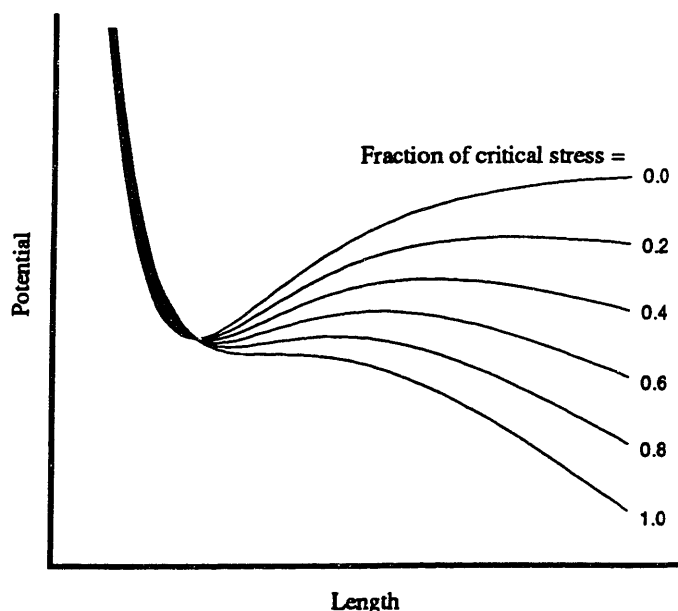


Figure 6.10. Morse curves modified by tensile stresses ranging from 0.0 to 1.0 times the critical stress for barrier elimination (in the absence of zero-point energy).

$$\mathcal{V}_{\text{stressed}} = D_e \left[\left(1 - e^{-\beta(\ell - \ell_0)} \right)^2 - 1 \right] - F(\ell - \ell_0) \quad (6.33)$$

The Morse function underestimates the energy (overestimates the well depth) in the high extension region, underestimating the energy gradient and the tensile strength of the bond (Sec. 3.3.3.1); it is thus conservative for engineering analyses where bond stability is required for a workable design. The above expression assumes that the applied force is independent of the displacement. Where applied forces are associated with large positive stiffnesses (as will be true for bonds inside diamondoid structures) this analysis, by ignoring the “solid cage effect,” will grossly overestimate cleavage rates.

Manipulating the above expression yields the classical barrier height (Kauzman and Eyring 1940)

$$\Delta\mathcal{V} = D_e \sqrt{1 - 2F/\beta D_e} + \frac{F}{\beta} \ln \left(\frac{1 - \sqrt{1 - 2F/\beta D_e}}{1 + \sqrt{1 - 2F/\beta D_e}} \right) \quad (6.34)$$

and frequencies at the well minimum and barrier maximum yielding a one-dimensional partition function

$$q_{\text{quant}} = \exp \left(-\frac{\hbar\omega}{kT} \right) \left[1 - \exp \left(-\frac{\hbar\omega}{kT} \right) \right]^{-1}, \text{ where} \quad (6.35)$$

$$\omega = \sqrt{\frac{D_e\beta}{\mu} \left(\beta - \frac{2F}{D_e} + \sqrt{\beta^2 - \frac{2\beta F}{D_e}} \right)},$$

and an imaginary frequency for calculation of the Wigner tunneling correction, Eq. (6.14)

$$\omega_{\text{tc}}^{\dagger} = \sqrt{\frac{D_e\beta}{\mu} \left(\beta - 2F/D_e - \sqrt{\beta^2 - 2\beta F/D_e} \right)} \quad (6.36)$$

The zero-point energy estimate uses the harmonic approximation, yielding a (conservative) underestimate of the barrier height. All factors incorporating bond frequencies neglect the increase in effective mass and effective stiffness resulting from atoms not directly participating in the bond.

6.4.4.2. Tunneling cleavage rates

At sufficiently low temperatures, the Wigner tunneling correction becomes inadequate and methods like those discussed in Section 6.2.4 must be applied. At zero K, only an estimate of tunneling from the ground state is necessary; the model used here makes the above assumptions regarding the potential energy function and effective mass, approximating the transition rate as the product of the vibrational frequency in the well and the WKB transmission probability, Eq (6.17). Intermediate cryogenic temperatures would require a more complex treatment, giving intermediate values of the bond lifetime.

6.4.4.3. Allowable stresses in covalent bonds

Figure 6.11 graphs the characteristic bond cleavage time, τ_{cl}^{-1} , vs. applied stress for a variety of bond types at 0, 300, and 500 K, with bond stiffnesses and dissociation energies from Table 3.7. It shows that the criterion $\tau_{bb} > 10^{20}$ s is met at 300 K is so long as the stress is $\leq \sim 1.2$ nN/bond. Replacing each stressed C–C bond by a pair of C–C bonds (modeled by doubling the mass, energy, and stiffness) yields the curve second-farthest to the right; the threshold stress for this system is ~ 6 nN, substantially more than twice the single-bond threshold stress.

Allowable stresses are strongly temperature dependent: Although this volume is concerned with systems at room temperature, the range of workable structures and stresses would be substantially broader at cryogenic temperatures. Conversely, as temperatures increase, workable designs will become increasingly constrained, ultimately dwindling to the empty set.

These calculations can be compared to the estimated theoretical tensile strength of diamond, $\sim 1.9 \times 10^{11}$ N/m² (Field 1979). Diamond normally cleaves along the {111} plane, which has 1.8×10^{19} bonds/m²; the strength just cited corresponds to ~ 10.6 nN/bond, roughly twice the theoretical limiting strength implied by the more conservative calculations of this section. The theoretical shear strength has been calculated (Field 1979) to be $\sim 1.2 \times 10^{11}$ N/m², or 6.7 nN/bond. At ~ 1300 K, diamond undergoes plastic flow under mean shear stresses of only 0.18 nN/bond (Brookes, Howes et al. 1988), but this process depends on the breakage of bonds in dislocations (where stresses are far higher than the mean) and occurs at temperatures that rapidly cleave C–C bonds in normal hydrocarbons.

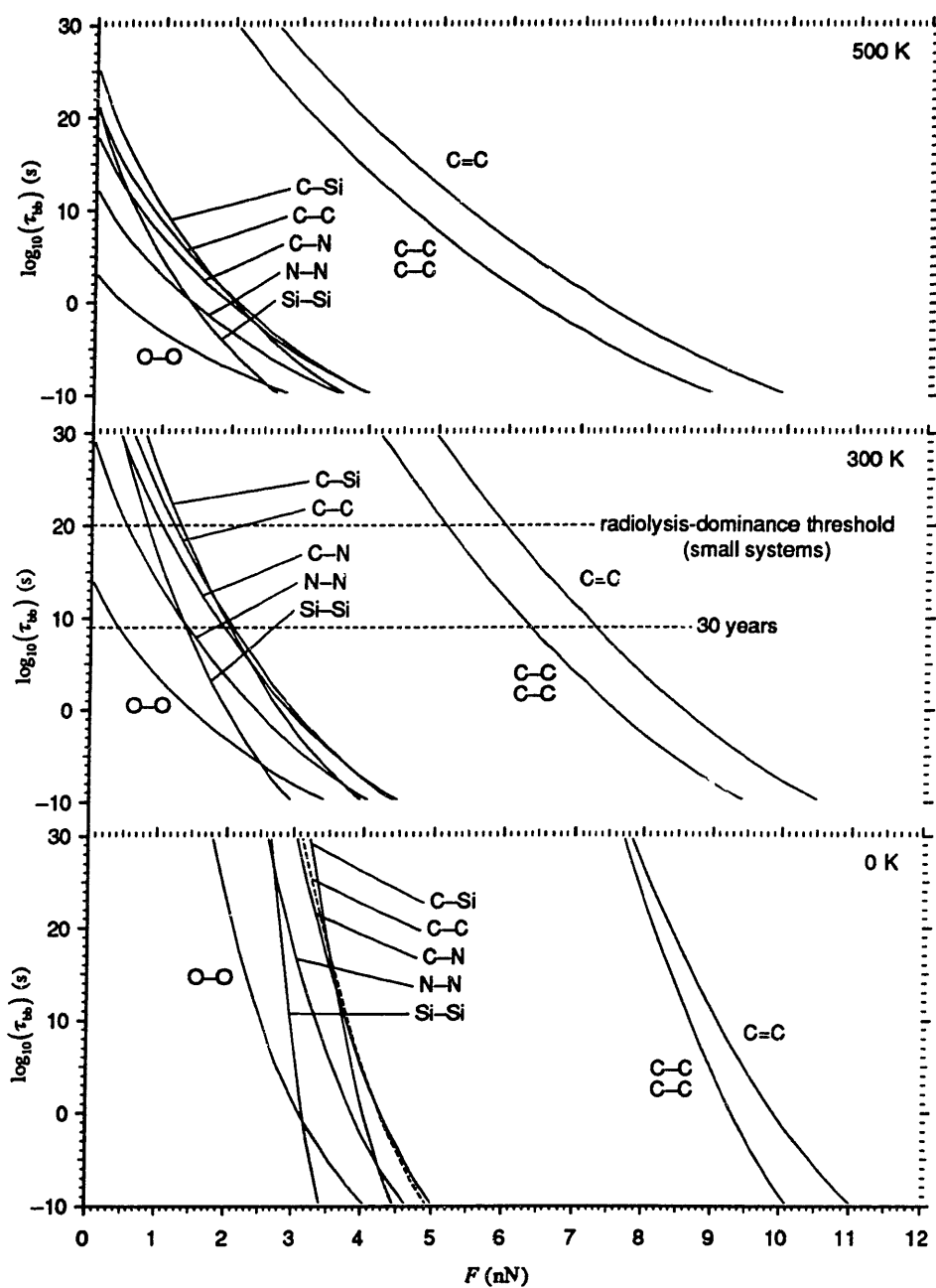


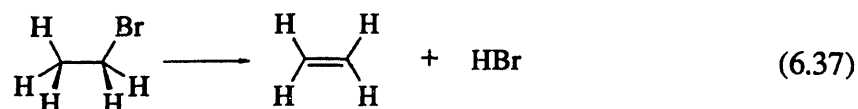
Figure 6.11. Characteristic bond cleavage time vs. applied stress at 0, 300, and 500 K. Calculations at 300 and 500 K are based on quantum transition state theory; those at zero K are based on the WKB tunneling approximation (see text). The second curve from the right represents a pair of single C–C bonds mechanically constrained to cleave in a concerted process.

6.4.5. Other chemical damage mechanisms

6.4.5.1. Elementary, non-elementary, and solvent-dependent reactions

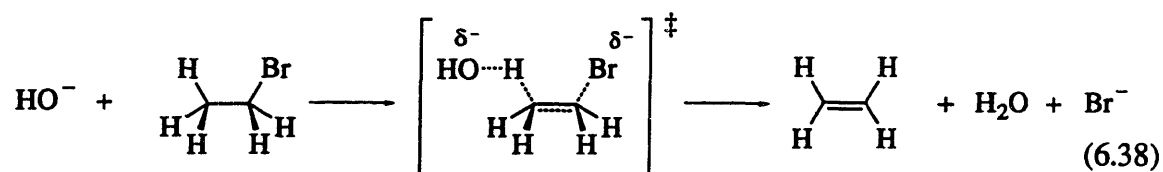
As suggested by Section 6.4.2.3, unimolecular reactions provide important models for damage mechanisms of significance in nanomechanical systems. Common unimolecular reactions include elimination (loss of part of a molecule) and rearrangements (structural transformations that leave overall composition unchanged). Often, however, such transformations occur by non-unimolecular processes, or by unimolecular processes that are solvent-dependent.

A simple example is an elimination reaction sometimes written as

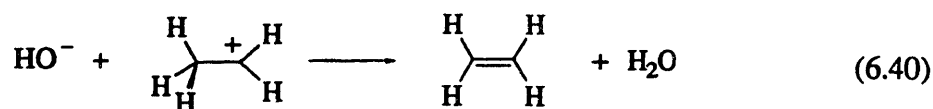
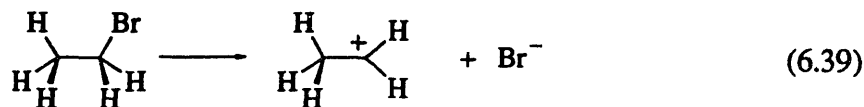


This would not occur at an appreciable rate among molecules isolated in vacuum, because the reaction does not occur as shown. The above equation represents a reaction (or compound reaction), but not an elementary reaction, which is to say a single molecular transformation. Not all reactions are elementary, and mistaking a compound reaction for an elementary reaction can lead to mistaken conclusions.

The steps yielding ethylene in the above reaction can proceed by at least two distinct mechanisms comprising three elementary reactions. The first mechanism is a single-step E2 process, such as:



The second is a two-step E1 process, such as:



Neither reaction mechanism actually yields the HBr molecule shown in the first equation, and both require the participation of other molecules. The 2 in E2 refers to the requirement for participation of two molecules in the slow, rate-limiting step; the 1 in E1 indicates that the rate-limiting step is unimolecular. Even the first step of the E1 reaction is solvent-dependent: the separation of a positive and negative ion from 0.3 nm to 3 nm in vacuum requires ~ 700 mJ of energy, but in a medium with the dielectric constant of water, only ~ 9 mJ. For similar reasons, Na^+ and Cl^- ions readily leave a salt crystal in water, causing swift dissolution despite the lack of any significant tendency for salt to evaporate into the vacuum-like emptiness of air. In general, processes generating free ions will be rare in vacuum at 300 K.

6.4.5.2. Evidence from pyrolysis

The TST-derived relationships described in Section 6.4.3 can be used to estimate unimolecular reaction rates at 300 K from reaction rates at higher temperatures. The partition functions for such reactions will typically be small and can be approximated as unity; further, comparing the reaction rates of a single molecular species at two temperatures causes a partial cancellation of the errors in this approximation. Equation (6.28) implies that a species requiring ≥ 1 s at ≥ 750 K (≈ 480 C) for thorough pyrolysis will have a transformation lifetime $\tau_{\text{trans}} \geq 10^{20}$ s/molecule at 300 K. Thermal exposure times of 1 to 100 s are common in pyrolysis experiments in organic chemistry (Brown 1980).

These experiments (discussed in Section 6.4.5.4), like the polymer degradation experiments discussed in the next section, give evidence regarding the suitability of various classes of structures for use as major components of long-lifetime nanomechanical systems. As should be expected, some are adequate and others are too unstable.

6.4.5.3. Polymer pyrolysis

In the pyrolysis of polymers, 50% weight loss during 30 min at 610 K (≈ 340 C) suggests a lifetime against fragmentation reactions of $\geq 10^{20}$ s/monomer at 300 K. Table 6.1 lists the characteristic temperatures for a variety of polymers; data regarding heat-resistant polymers can be found in (Critchley, Knight et al. 1983). Note that many exceed 340 C despite the availability of bimolecular processes for degradation; polyvinylchloride, for example, is thought to undergo elimination of HCl via a radical chain reaction mechanism. In the analysis of nanomechanical system failures, it will be assumed that the first chemical transformation is sufficient to cause failure (Section 6.7.1.1), making chain

reactions irrelevant. Polymer degradation studies are accordingly better for establishing lower bounds on achievable stability than they are at establishing upper bounds.

These empirical observations can also be criticized for being sensitive to the cleavage of polymer backbones, but perhaps not to rearrangements of side-chains. Rearrangements are better examined in small molecules.

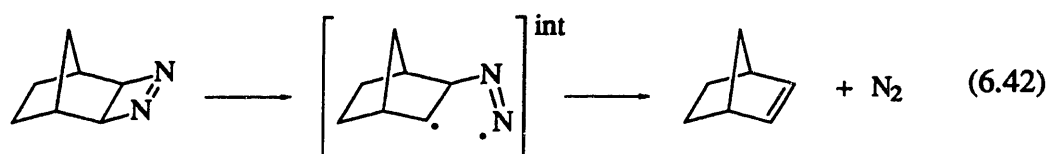
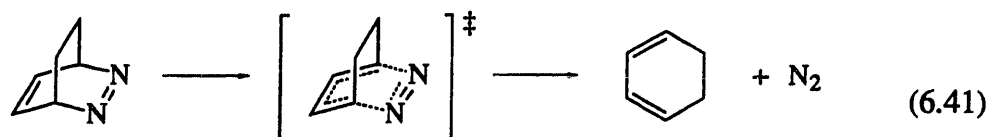
6.4.5.4. Unimolecular elimination, fragmentation, and rearrangement

The rapid pyrolysis of small organic molecules in the gas phase provides a useful model for long-term damage to machine-phase systems at room temperature. These processes can provide a guide to the relationship between stability and local molecular structure. This section provides a few examples and limited generalizations regarding molecular stability.

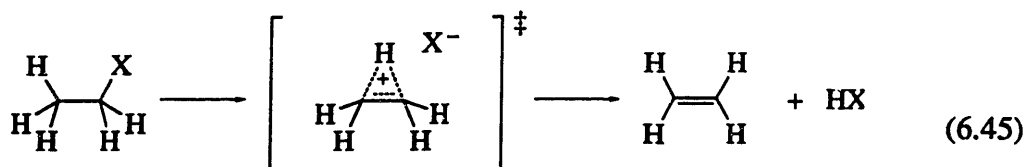
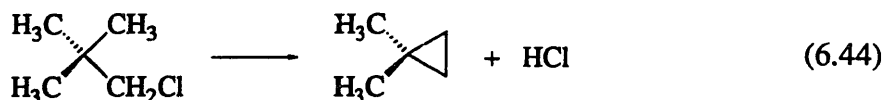
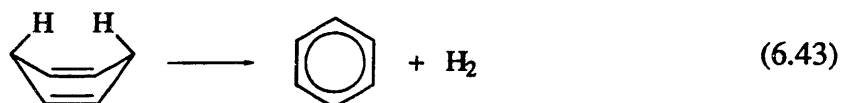
Many elimination reactions would proceed at unacceptably high rates at room temperature. In reactions such as

Table 6.1. Temperatures for 30-minute volatilization half-lives in the absence of air (Schnabel 1981).

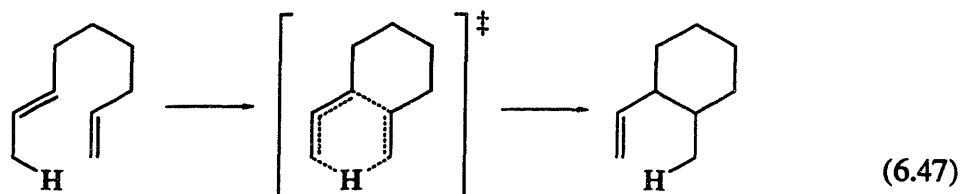
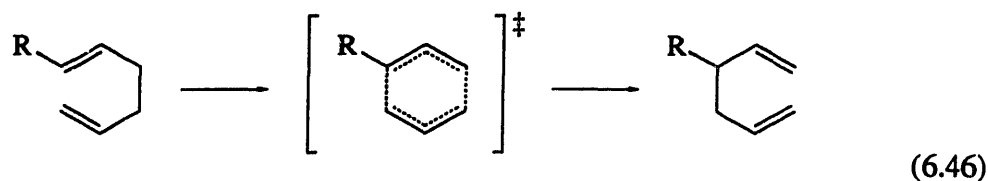
Polymer	T_{vol} (C)
Polytetrafluoroethylene	510
Polybutadiene	410
Polypropylene	400
Polyacrylonitrile	390
Polystyrene	360
Polyisobutene	350
Poly(ethylene oxide)	350
Polymethylacrylate	330
Polymethylmethacrylate	330
Polyvinylacetate	270
Polyvinylchloride	260



bonds rearrange to make the extremely stable N_2 molecule. Reaction (6.41) proceeds quickly at 200 K reaction (6.42) at does so only at ≥ 370 K (the difference in reaction mechanism stems from differences in orbital symmetry) (Carey and Sundberg 1983). Other significant processes of this sort yield CO , CO_2 and H_2 . Under pyrolytic conditions, hydrogen halides can be eliminated from alkyl halides (possibly as shown in Eq. (6.37), or in a surface reaction).



An extensive review of eliminations under gas-phase pyrolytic conditions appears in (Brown 1980) (from which (6.43), (6.43), and (6.43) are taken). Chapter 7 of (Carey and Sundberg 1983) provides a briefer overview of elimination reactions and discusses some classes of unimolecular rearrangements of chemical interest. Examples include Cope rearrangements and intramolecular ene reactions, such as:



(Brown 1980) surveys reactions occurring during gas-phase pyrolysis, describing hundreds of examples of reactions from numerous classes. It is possible to draw some conclusions from this body of data, even in the absence of generalizations that neatly categorize molecules according to stability. Brown's stated selection criteria favor the inclusion of reactions that are (1) useful in organic synthesis and (2) proceed at temperatures > 350 C. The former criterion favors the inclusion of molecules having a specific reactivity, that is, a specific weakness relative to more inert molecular structures (such as aliphatic hydrocarbons of low strain). Further, achieving specific reactivity typically favors the selection of the lowest reaction temperatures at which reaction rates are adequate for practical work. Of the structures in (Brown 1980), selected for exhibiting useful pyrolytic reactions, the fraction having listed reaction temperatures > 500 C (suggesting adequate stability for nanomechanical use at room temperature) is roughly half.

For pyrolytic reactions in general, the structural characteristics favoring reaction temperatures ≤ 480 C are too complex to enumerate, but include inherently-weak covalent bonds, resonant stabilization of reaction products, high strain, and others. A generalization does hold for the set of unimolecular rearrangements reviewed in (Brown 1980): listed reaction temperatures ≤ 480 C do not occur unless the molecule either (1) contains two or more unsaturated bonds, or (2) contains both a strained, three-membered ring and one or more unsaturated bonds, or (3) contains a strained, four membered ring that includes a single unsaturated bond. Many molecules meeting one or more of these conditions nonetheless have listed reaction temperatures > 480 C. Exceptions to this generalization can be found elsewhere in the literature.

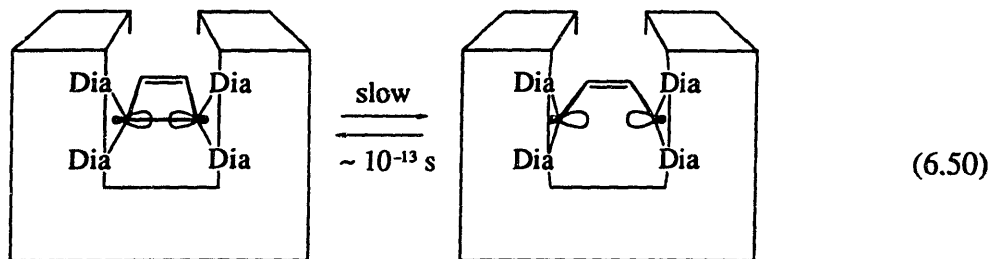
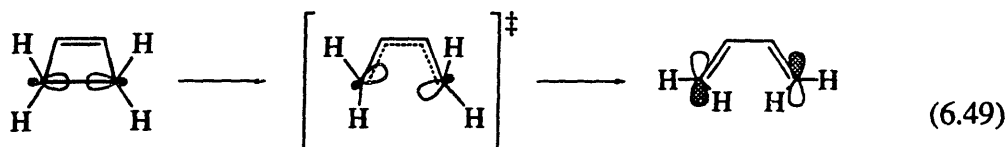
It should be noted that surface-catalyzed and non-unimolecular gas-phase reaction pathways may occur in some of the systems reviewed in (Brown 1980). Removing these pathways can only improve stability. (Rearrangements occurring under a variety of condi-

tions are reviewed in the volumes of (Thyagarajan 1968–1971).)

The thermal rearrangement of hydrocarbons, which has been practiced on a large scale in the petroleum industry to increase the octane rating of gasoline, is relevant to estimating the failure rates of molecules having no special destabilizing characteristics. The temperatures applied are in the 780–840 K (~ 500–570 C) range, with exposure times of many seconds; this implies $\tau_{\text{trans}} > 10^{20}$ s at 300 K.

The study of pyrolytic reactions in small molecules provides useful indications of stability in nanomechanical systems at 300 K, but fails to account for two expected differences. The first is the presence of mechanically-imposed tensile and shear stresses: these will be destabilizing and are at best only partially modeled by the methods of Section 6.4.4. Since activation energies for rearrangements frequently fall within the range of activation energies for the cleavage of structurally-useful bonds, it seem safe to assume that failure due to rearrangement under moderate applied stresses will in these instances meet the $\tau_{\text{trans}} > 10^{20}$ s at 300 K criterion.

The second difference is that potentially unstable substructures found in nanomechanical systems will typically be embedded in rigid, polycyclic matrices. As we have seen, thermal bond cleavage will often be strongly inhibited by the rigid cage effect. Likewise, rearrangements typically involve atomic displacements that may be feasible (e.g.) for a carbon atom in a methylene group, but not for a carbon atom in a diamond-like structure.



A useful exercise is to examine a list of rearrangement reactions and observe how few could occur if most of the bonds to hydrogen were instead bonds to carbon atoms in

a surrounding matrix of diamond-like rigidity. Diamond itself is stable to 1800 K in an inert atmosphere ($\tau_{\text{trans}} > 10^{85}$ s at 300 K), despite the well-known energetic advantage of rearrangement to graphite.

For the nanomechanical design process, it would be useful to have an automated classification system capable of reliably labeling structures as adequately stable, inadequately stable, or of unknown adequacy, with the set of structures labeled as adequately stable being large enough to permit effective nanomechanical design. Such a design tool could in substantial measure be validated using the existing data on small-molecule pyrolysis. Lacking a tool of this sort, one must proceed using informal chemical reasoning (drawing on the accumulated generalizations of organic chemistry) and analogy based on model compounds (drawing on the yet more massive accumulation of data in organic chemistry), supplemented by computational methods. The strategy pursued in the present work is to favor structures in which strong bonds form frameworks of diamond-like rigidity that lack substructures of known instability. This strategy is more likely to exclude workable designs (yielding false negatives) than to include unworkable designs (yielding false positives). As development progresses, both classes of error will be reduced.

6.4.5.5. Stability of reagent devices

Molecular manufacturing systems will include devices that play the role of reagents in organic synthesis. These devices will be nanomechanical components that include moieties that will exhibit high reactivity under certain conditions (as discussed in Chapter 8). These moieties evidently cannot be designed for high stability in a general sense, hence their stability against unwanted transformations will require special attention on a case-by-case basis.

Some general observations are possible: Reagent moieties will make up only a small fraction of the mass of a molecular manufacturing system as presently conceived, falling within the range of 10^{-3} to 10^{-6} ; this somewhat reduces their quantitative significance as targets for damage. Further, they will be subject to rapid cycles of synthesis and use, rather than serving as structural elements that must be stable for the life of the system; this potentially reduces their qualitative significance as targets for damage. Finally, the reagent devices used in a molecular manufacturing system will be selected from a wide range of candidate structures, including analogs of familiar reagents, catalysts, and reactive intermediates; this generates a large set of options with varying stability characteristics.

The stability of reagent devices will depend on how they are designed, which will depend in part on their applications. Chapter 8 discusses reagent devices in the context of mechanochemical operations, giving special attention to the stability and use of structures that would in solution chemistry be regarded as non-isolable reactive intermediates.

6.4.6. The stability of surfaces

Nanoscale devices will have a high surface to volume ratio; concern is occasionally expressed regarding the stability of these surfaces. It is well known in materials science that surface diffusion occurs more readily than diffusion in bulk materials, that surfaces frequently undergo spontaneous reconstruction to form arrangements unlike those in the interior, that surfaces are associated with modified electronic properties in semiconductors, and so forth. In judging the relevance of these concerns, two considerations are central: the chemical nature of the surfaces being considered, and the nature of the structures to which they are being compared.

Metals frequently exhibit high surface reactivity, diffusion, or both, and metals are not here proposed for use as nanomechanical components. Section 6.5 discusses the use of aluminum films for photochemical shielding, but films of the sort required are familiar in present technology. Chapter 11 discusses metals as conductors in electromechanical systems, but even here, alternative conductors are feasible. Certain metal surfaces are doubtless stable enough for some roles, but metals play a peripheral role in the present work.

Unlike metals, semiconductors are covalently-bonded solids with a basic similarity to the structures considered for use in nanomachinery. Clean semiconductors in vacuum consistently exhibit high surface reactivity and generally undergo reconstruction. Their reactivity and reconstruction, however, are associated with clean surfaces that (from a chemist's perspective) would in the absence of reaction or reconstruction consist of a dense array of free radicals. Nanomachines do not require the use of such surfaces. Polished diamond undergoes surface reconstruction (Kubiak and Kolasinsky 1989), but only after heating to 1275 K has removed the hydrogen from its surface (Hamza, Kubiak et al. 1988). Until then, it is an unusually stable hydrocarbon. (Carbon frameworks built to match this reconstruction may be useful in nanomechanical systems.)

The idea that surfaces present special problems arises from a perspective that takes the bulk phase as the norm. The present work, however, draws chiefly on concepts and models based on experimental results regarding the chemistry of small molecules. These

molecules are, in effect, all surface. Accordingly, the concepts and models need no modification to make them applicable to surfaces of the sort contemplated here. From this perspective, surfaces are the norm, and it is (for example) the special stability of the diamond *interior* relative to surface-dominated hydrocarbons that has been worthy of remark.

6.5. Photochemical damage

6.5.1. Energetic photons

Photochemical processes can excite molecules to energies that are effectively unavailable in equilibrium systems at ambient temperatures. Photons at visible and ultraviolet wavelengths (for example, in sunlight) have energies characteristic of far higher temperatures (for example, the ~ 5800 K of the solar photosphere) and they deliver that energy to a single molecular site.

Where photochemical damage in the ambient terrestrial environment is concerned, the electromagnetic spectrum has traditionally been divided as shown in Table 6.2.

Ultraviolet exposure is limited by atmospheric opacity: From the visible to the UV-B, the atmosphere transmits solar radiation; within the UV-C, absorption by oxygen and ozone effectively block solar radiation; at energies beyond the UV-C (the vacuum ultraviolet range), air becomes opaque. Consequently, both solar and local radiation sources will be effective in the visible to UV-B range, only local sources will be effective in the UV-C, and exposure in the vacuum ultraviolet will (within the atmosphere) typically be negligible. At yet higher energies, the UV spectrum merges into the X-ray spectrum and pho-

Table 6.2. Wavelengths, frequencies, and energies

Name	Wavelength range (nm)	Maximum frequency (Hz)	Maximum energy (aJ)
Visible	700 – 400	7.5×10^{14}	0.50
UV-A	400 – 320	9.4×10^{14}	0.62
UV-B	320 – 280	1.1×10^{15}	0.71
UV-C	280 – 200	1.5×10^{15}	1.00

tons become both penetrating and ionizing; the resulting damage is discussed in Section 6.6.

6.5.2. Overview of photochemical processes

Photochemical processes at a given wavelength begin with the absorption of a photon, which requires a chemical species able to absorb at that wavelength. Ordinary alkanes, alcohols, ethers and amines absorb at wavelengths < 230 nm (Robinson 1974), deep in the UV-C. In the absence of oxygen and other photochemically-sensitive molecules, lack of absorption renders these substances photochemically stable (by ordinary standards) under ambient UV conditions. Systems of π electrons characteristically absorb at longer wavelengths (organic dyes, which absorb at visible wavelengths, contain large conjugated π systems). Quantum mechanical selection rules constrain electronic transitions and absorption cross sections. Multiphoton absorption processes relax constraints on photon energies but require intensities seldom encountered in the absence of lasers.

For a process to do permanent damage to a nanomachine, bonds must be rearranged or cleaved, or a charge must be displaced and trapped. For a given molecular structure, each process has an energy threshold. Charge displacement (photoionization) in typical organic molecules requires energies in the vacuum ultraviolet range. To produce an electron-hole pair within diamond requires a photon of $\lambda < 230$ nm.

Energy thresholds for rearrangements vary greatly with molecular structure and can be quite low. In laboratory photochemistry, rearrangements typically involve conjugated π systems. Broad classes of structures resist low-energy rearrangement.

Bond cleavage requires an energy equal to or greater than the dissociation energy of the bond; in practice, it typically requires significantly more energy, since the photochemical process that overcomes the negative potential energy of the bond also deposits energy in vibrational and electronic excitations. Many bonds of interest in constructing nanomechanical systems—including carbon-carbon bonds—are subject to cleavage at UV-A and UV-B energies.

Photochemical bond cleavage is more common in low-pressure gases than in condensed matter, where competing processes more efficiently dissipate excitation energy. Further, in condensed matter, photochemical (like thermal) bond cleavage is frequently reversed by geminate recombination. Bond cleavage can be characterized by quantum yield, the ratio of bonds cleaved to photons absorbed. For carbonyl-rich polymers, such

as poly(methyl-isopropyl ketone), quantum yields are as high as 0.22 at $\lambda = 253.7$ nm; for polystyrene, the quantum yield is 9×10^{-5} (Ranby and Rabek 1975).

In diffusive chemistry, the consequences of photochemical reactions are typically complex. Free radicals can initiate chain reactions involving a variety of reactive species, and all species have unconstrained opportunities for collision. In a structured solid phase, the opportunity to build more constrained systems will enable the construction of systems with simpler behavior, including reduced sensitivity to photochemical damage.

6.5.3. Design for photochemical stability

In designing nanomechanical systems, photochemical damage can be limited or avoided in three alternative ways:

- (1) by keeping the entire system away from UV light,
- (2) by providing the system with a UV-opaque surface layer, or
- (3) by requiring that all components be photochemically stable.

Approach (1) is simple and will be workable for many purposes, but it will limit the nature of operating environments. Approach (2) is analyzed in the next section. Approach (3) will be preferable for some applications, but adding the constraint of photochemical stability will substantially increase the complexity of molecular systems design, and will eliminate a variety of device designs that would otherwise be attractive. The discussion in this book assumes the use of approach (1) or (2), thereby avoiding photochemical stability constraints.

It is nonetheless worth briefly considering how one might design systems within the constraints of approach (3). One methodology would focus on absorption processes, attempting to avoid all use of structures that absorb in the UV-A and UV-B bands, but this may not prove practical. An alternative methodology would examine all potential modes of absorption and attempt to ensure that the absorbing structures can tolerate the resulting photochemical excitation. Tolerance for excitation will be more achievable in a structured solid phase than in a liquid phase, owing to greater control of reaction opportunities and radical recombination processes. One approach would be to seek structures in which mechanical constraints will force swift recombination of any cleaved bond: complications include photoexcitation to triplet states (delaying recombination in a manner that is absent in thermal processes) and rearrangements that trap the system in a ground-state potential well other than that desired. In summary, it may prove possible to develop a library of nanomechanical components that meet the stringent conditions required for

UV-B exposure tolerance, but for the present it is easier to assume the use of shielding.

6.5.4. Photochemical shielding

Most metals block UV radiation of ordinary wavelengths. The optical transmittance of a metal layer is determined (Gray 1972) by its index of refraction, n , its extinction coefficient, k , and (secondarily) by the indexes of refraction of the adjacent media, n_1 and n_2 . Table 6.3 lists values of n and k for aluminum at UV wavelengths. When an absorbing layer is thick enough, one can neglect interference among reflected waves within the layer. In this single-path approximation, appropriate for shielding calculations, the transmittance is given by:

$$T = \frac{16n_1n_2(n^2 + k^2)}{\left[(n + n_1)^2 + k^2\right]\left[(n + n_2)^2 + k^2\right]} \exp\left(\frac{-4\pi kd}{\lambda}\right) \quad (6.51)$$

where d is the thickness of the metal layer. Figure 6.12 graphs the transmittance of aluminum layers as a function of their thickness, at several UV wavelengths. Note that trans-

Table 6.3. UV optical properties of aluminum (Gray 1972)

Wavelength (nm)	n	k
120	0.057	1.15
160	0.080	1.73
200	0.110	2.20
320	0.280	3.56
400	0.400	4.45195

mittance is nearly independent of wavelength from the UV-A to the edge of the vacuum ultraviolet.

A typical shield thickness is readily estimated. Assume that the mean time between (photochemical) failures is required to be 30 years ($\approx 10^9$ s) for a device with an area of one square micron exposed to terrestrial sunlight. Assume that the mean power density at effective wavelengths is 5 W/m^2 , that all shield-penetrating photons are absorbed in the device, that cleavage of one bond causes the device to fail, and that the quantum efficiency of bond cleavage is 10^{-2} . With these assumptions, the mean photon flux is $\sim 10^{19}/\text{m}^2\cdot\text{s}$, the 30-year dose to the device surface is 10^{16} photons, and the transmittance must be limited to 10^{-14} or less, which can be achieved with a shield thickness of just under 250 nm. A shield of this thickness would impose a large volumetric penalty on a system with an overall diameter of 2μ , but only a modest penalty on system with a diameter of 10μ .

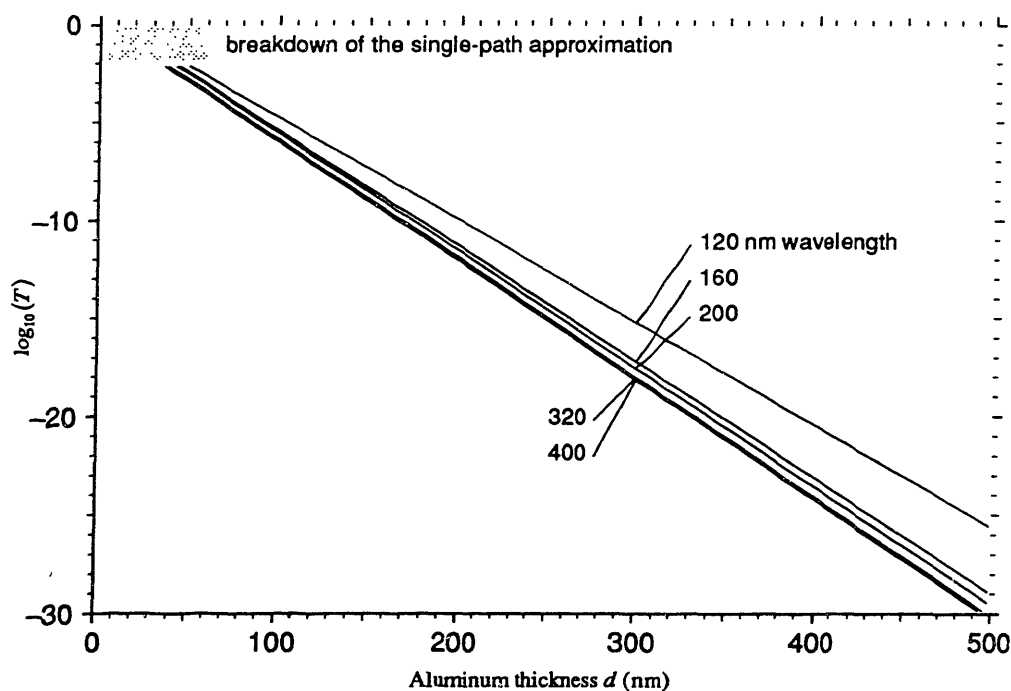


Figure 6.12. Transmittance (fraction of optical power transmitted) vs. thickness of aluminum for various wavelengths, $n_1 = n_2 = 1.5$. Values computed from Eq. (6.51) and Table 6.1.

6.6. Radiation damage

6.6.1. Radiation and radiation dosage

Forms of ionizing radiation include high-energy photons and charged particles. At a nanometer scale, most of the damage done by these forms of radiation is mediated by high-energy secondary electrons. Experiment shows that the energy carried by charged particles (including secondary electrons) is chiefly deposited in chains of excitation events with spacings ranging from a few 100 nm at MeV energies, to ~ 10 nm at 5 keV, to a virtual continuum at low energies; the local chemical effects are like those of UV radiation in the 10–30 eV range (Williams 1972).

Ionizing radiation is measured both in rads (1 rad = 100 ergs absorbed per gram = 10^{-2} J/kg) and in roentgens (defined in terms of ionization produced in air by X- and gamma radiation; 1 roentgen deposits ~ 87 ergs in a gram of air). For many forms of ionizing radiation impinging upon light-element targets, the quantity of energy absorbed by a small volume of matter is roughly proportional to its mass. Background radiation in the terrestrial environment seldom exceeds 0.5 rad/year.

6.6.2. Classical radiation target theory

Existing experimental evidence relates radiation dosage to damage in molecular machinery, where that molecular machinery takes the form of proteins acting as enzymes. Classical radiation target theory (Lea 1946) is a standard technique for estimating the molecular mass of enzymes from the loss of enzymatic activity as a function of radiation dose applied to dried enzyme in vacuum (Beauregard and Potier 1985; Kepner and Macey 1968). In the target theory model, a single hit suffices to inactivate an enzyme molecule, and the probability that a molecule will be hit is proportional its mass (for radiation doses small enough to make multiple hits improbable). Studies of the inactivation of enzymes of known molecular weight indicate that roughly 10.6 aJ of absorbed radiation is required to produce one inactivating hit (Kepner and Macey 1968), yielding $\sim 10^{15}$ hits/kg-rad.

How good is this relationship for estimating the rate of destruction of nanomechanical systems owing to radiation damage? As a rough guide, it should be fairly reliable because it rests directly on experimental evidence. It will, however, surely be inaccurate because of the physical differences between the targets. Proteins can tolerate small

changes in side-chain structure at many sites without loss of function (as shown by protein engineering experience (Bowie, Reidhaar-Olson et al. 1990; Ponder and Richards 1987)); this suggests a significant ability to absorb structure-altering events while remaining active, which in turn suggests that nanomachines of more tightly-constrained design will be *more* sensitive to radiation damage. Weighing on the other side, however, is the greater radiation tolerance of polycyclic structures: where geminate recombination of radicals is strongly favored by mechanical constraints, only a small fraction of excitation events will lead to permanent structure alteration. This suggests that typical nanomachines will be *less* sensitive to radiation damage than are proteins, which readily suffer chain scission. In summary, both proteins and nanomachines will exhibit some tolerance for bond cleavage, with nanomachines having a greater fraction of sites at which bond cleavage will cause no permanent alteration, and proteins having a greater fraction of sites at which permanent alterations can be tolerated. On the whole, it seems reasonable to assume that nanomachines, like proteins, will suffer $\sim 10^{15}$ inactivating hits/kg-rad. For this estimate to err by being substantially too low would require that enzymes have an implausibly-large probability of surviving random structural damage. In this model, the probability that an initially functional device has not been destroyed by radiation damage is

$$P_{\text{functional}} = e^{-10^{15} D m} \quad (6.52)$$

where D is the dose in rads and m the device mass in kilograms.

6.6.3. Effects of track structure

Aside from structural differences, typical nanomechanical systems will be far larger than typical enzymes. This affects scaling relationships because radiation hits are distributed not at random, but along particle tracks. In the range of sizes for which the diameter of a mechanism is large compared to the spacing of excitation events along a typical particle track (while remaining small compared to the length of the track), damage becomes proportional not to mass but to area, and hence scales not as m but as $m^{2/3}$. In typical radiation environments, devices of ordinary density with a diameter of 100 nm or greater should benefit from this favorable breakdown in linearity. A rough model representing this shift from m to $m^{2/3}$ scaling in the vicinity of 100 nm diameters is

$$P_{\text{functional}} \approx \exp \left\{ - \left(10^{15} D m \right)^{-1} - \left[10^{11} D (m/\rho)^{2/3} \right]^{-1} \right\}^{-1} \quad (6.53)$$

where ρ is the density of the device in kg/m^3 . This model (graphed in Fig. 6.13) neglects a variety of factors affecting track structure, uses an arbitrary functional form to represent the transition between regimes, and should be taken only as a rough guide to damage rates for large devices.

6.6.4. Radiation shielding

To shield against ionizing radiation typically requires macroscopic thicknesses of dense material, ranging from a fraction of a millimeter for medium-energy alpha particles

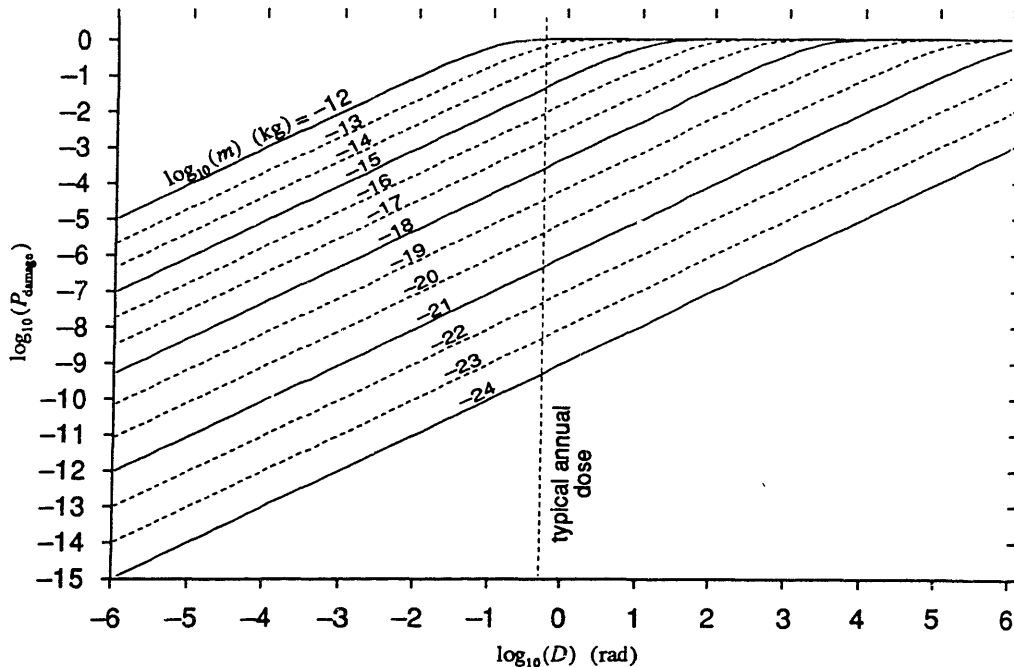


Figure 6.13. Probability of device failure owing to ionizing radiation damage vs. radiation dose, according to Eq. (6.53), for devices of differing masses; assumes failure after a single hit and a device density equaling that of water. Background radiation in the terrestrial environment seldom delivers more than 0.5 rad/year, marked as “annual background.” Note that at the assumed density, the graphed curves correspond to device volumes ranging from 1 nm^3 to $10^{12} \text{ nm}^3 = (10 \mu)^3$.

to meters for gamma and cosmic rays. Earth's atmosphere has an areal mass density of $\sim 10^4 \text{ kg/m}^2$, yet showers of secondary particles from cosmic rays deliver a significant annual dose at sea level.

Any non-zero exposure to ionizing radiation will adversely affect either reliability or performance in nanoscale systems, yet even with extensive macroscale engineering, radiation exposure seems inescapable. Thick layers of non-radioactive shielding materials, although able to reduce radiation exposure by large factors, cannot prevent sporadic showers of secondary particles resulting from high-energy, cosmic-ray neutrinos. These particles can penetrate planetary thicknesses of rock, and the resulting secondary showers have been observed in deep mines, traveling upward.

6.7. Device and system lifetimes

6.7.1. Device lifetimes

6.7.1.1. The single-point failure assumption

Under the conditions of machine-phase chemistry, some sources of molecular damage (e.g., reactions with water and free oxygen) are excluded, and most others are subject to control during the design phase. Damage rates can accordingly be quite low. A single chemical transformation, however, may cause failure. Since a typical device may consist of 10^6 to 10^{12} atoms, damage sensitivity can be quite high.

Components of sufficient size can avoid this sensitivity: macroscopic machines are proof of this. Use of large-scale components, however, would sacrifice many of the advantages that scaling laws provide to systems of minimal size, while components of intermediate scale would require a complex analysis of failure modes. Accordingly, to simplify analysis and ensure conservatism, the present work assumes that devices will fail if they experience a single chemical transformation anywhere in their structures. This can be termed the *single-point failure assumption*.

6.7.1.2. Choosing reliability criteria

What will be the main causes of chemical transformation and resulting failure? By a suitable choice of design, almost any mechanism could be made dominant. By permitting exposure to photochemically active UV wavelengths, photochemical damage could be made dominant; by using organic peroxides or highly strained bonds as structural elements, bond cleavage could be made dominant; other choices could make rearrangements

or interfacial reactions dominant. All of these damage mechanisms, however, are subject to rate laws that are exponentially dependent on parameters that are subject to engineering control. Under ambient terrestrial conditions, all can be reduced to low levels by careful design, augmented by experimental testing and redesign if necessary.

Ionizing radiation is less subject to control. Although (imperfect) shielding is possible, it would be awkward in many applications; major reductions in radiation exposure require shielding thicknesses measured in meters, in contrast to the hundreds of nanometers that suffice at UV-A, B, and C wavelengths. The present work assumes that ambient terrestrial radiation sets a practical lower bound on molecular damage rates.

This practical lower bound can be used to choose design criteria for other damage rates. A variety of engineering objectives would be served by the use of structures of marginal stability; examples include high speeds (associated with high bond stresses), compact designs (associated with exploitation of relaxed design rules on bond types and geometries), and so forth. Reliability, however, is itself a major engineering objective. Given the assumption of a fixed, structure-independent failure rate resulting from ionizing radiation, it is reasonable to constrain the controllable damage mechanisms to be comparatively small. This can be termed the *radiation-damage dominance criterion*.

In the terrestrial environment, background radiation rarely exceeds ~ 0.5 rad/year, which corresponds to $\sim 1.5 \times 10^7$ hits/kg·s (Section 6.6). The single-point failure assumption implies that a hit may be equated to a cleaved bond or any other damaging transition. If we demand that all other damage rates be an order of magnitude lower than radiation damage rates, then an acceptable mean failure rate per bond in small structures (100 nm diameter or less) is $\sim 10^{-20} \text{ s}^{-1}$, given typical numbers of bonds per kilogram. The 250 nm photochemical shielding thickness calculated in Section 6.5.4 meets this criterion in full sunlight at Earth's surface. For thermally activated processes, Eq. (6.28) (with $q_{\text{long}} = 1$) yields acceptable failure rates so long as all barrier heights (\approx activation energies) are greater than ~ 0.313 aJ at 300 K, or ~ 0.366 aJ at 350 K.

In the absence of strong stresses or other destabilizing influences, most covalent bonds of interest (in particular, those between C and H, N, O, F, Si, P, S, and Cl) substantially exceed these thresholds (Table 3.7); all of those mentioned yield cleavage rates below 10^{-33} s^{-1} . If most bonds in a system have failure rates this low, then a smaller population can have higher failure rates while still meeting the 10^{-20} s^{-1} criterion for the mean failure rate. If this population is 1% of the total, then its barrier heights (by the above method) need only exceed 0.294 aJ at 300 K and 0.344 aJ at 350 K. For perspec-

tive, the former condition is *almost* met by organic peroxides, which are normally considered quite reactive.

6.7.1.3. Assumptions and damage rates

The single-point failure assumption and the radiation-damage dominance criterion are both subject to criticism and improvement, but neither is a hypothesis. The former is a conservative rule of calculation; it can be relaxed when better models are available. The latter is a design objective; it can be discarded whenever there is something to be gained by doing so. Both are stated in order to guide design and analysis within the context of the present work.

So long as devices meet the above 10^{-20} failure/bond-s criterion, the uncorrelated radiation-damage lifetime model Eq. (6.52) can with reasonable accuracy be treated as the overall device lifetime model. Systems meeting the slightly more stringent 10^{-22} failure/bond-s criterion can be described by Eq. (6.53) over the range of device volumes graphed in Figure 6.13.

6.7.2. System lifetimes

For a given level of reliability, finite damage rates and the single-point failure assumption set upper bounds to device size. Figure 6.13 indicates that devices of $10^9 \text{ nm}^3 (= 1 \mu^3)$ will have annual failure rates of several percent if they meet the radiation-damage dominance criterion. For some applications, this will be unacceptable.

Where device failures cannot be excluded and systems must be reliable, standard engineering practice resorts to redundancy. Although the structure of radiation tracks violates the assumption of random device failure and introduced geometrical complications, the essential effects of redundancy can be illustrated by a simple model. Assume that a system of mass m_{ref} is divided into devices each of mass m_{dev} . The effects of redundancy can be modeled by assuming that each device is one of a set of n , constructed such that system failure will occur only if there exists at least one set in which all devices have failed. The total number of sets is $m_{\text{ref}}/m_{\text{dev}}$, and the total mass of the system $m_{\text{tot}} = n \cdot m_{\text{ref}}$. Assuming uncorrelated failures described by Eq. (6.52),

$$P_{\text{functional}} = \left\{ 1 - \left[1 - \exp(-10^{15} D m_{\text{dev}}) \right]^n \right\}^{\frac{m_{\text{ref}}}{m_{\text{dev}}}} \quad (6.54)$$

Choosing a value of D such that $10^{15} D m_{\text{ref}} = 1$ (implying a 0.63 probability of failure in the absence of redundancy), $m_{\text{ref}}/m_{\text{dev}} = 1000$, and $n = 3$, the probability of failure ($1 - P_{\text{functional}}$) $\approx 10^{-6}$. To reach a probability of failure ≈ 0.5 requires increasing D by a factor of 100.

Where indefinitely prolonged system life is required, the standard engineering answer is a combination of redundancy and replacement of damaged components. This will be feasible in nanoscale systems, but at the cost of substantial increases in system complexity.

Chapter 7

Energy dissipation

7.1. Overview

Energy dissipation will limit the performance of nanomechanical systems, particularly when they are aggregated to form macroscopic systems. Energy dissipation limits the feasible rate of operations owing to limits on cooling capacity, (e.g.) in massively parallel computational systems. It could likewise reduce the attractiveness of nanomechanical systems relative to alternatives, (e.g.) in carrying out chemical transformations that are simple enough to be performed by diffusive, solution-phase chemistry.

Energy dissipation will seldom impose *qualitative* limits, that is, constraints on the *kinds* of operations that can be performed on a nanoscale, as opposed to constraints on the speed and efficiency with which they can be performed. In studies of the potential of molecular nanotechnology, energy dissipation is often important to estimate, yet seldom crucial to estimate precisely. Moderate overestimates will yield conservative estimates of system performance, while seldom falsely implying that a system is infeasible. The present chapter surveys those mechanisms of energy dissipation that seem significant, estimating or bounding their magnitudes. Mechanisms of energy dissipation specific to metals are ignored since the systems under consideration are chiefly dielectric; mechanisms occurring during plastic deformation are ignored (except as analogies) since under the rules adopted in Chapter 6 any degree of plastic deformation is counted as fatal damage.

The mechanisms of importance are of several fundamentally different types. These are:

- *Acoustic radiation from forced oscillations*, which carries mechanical energy to remote regions where it is thermalized,
- *Phonon scattering*, in which mechanical motions disturb phonon distributions by

reflection,

- *Thermoelastic effects and phonon viscosity*, in which elastic deformations disturb phonon distributions via anharmonic effects,
- *Compression of potential wells*, in which nonisothermal processes result in thermodynamic irreversibility, and
- *Transitions among time-dependent potential wells*, in which the merging of initially separate wells dissipates free energy by a combination of free expansion and forced oscillation.

Different mechanisms result in power dissipation rates that scale differently with speed. For acoustic radiation from an oscillating force, $P \propto v^2$; for radiation from an oscillating torque, $P \propto v^4$. Phonon scattering, thermoelastic and phonon viscosity effects, and nonisothermal compression of potential wells all (to a good approximation) exhibit $P \propto v^2$. Transitions among time-dependent potential wells, in contrast, are better described by $P \propto v$. Of these mechanisms, all but the last exhibit a speed dependent energy dissipation per operation (or per unit displacement) which approaches zero as $v \rightarrow 0$.

Note that qualifiers such as “typically,” “frequently,” “many systems,” and the like are used throughout this chapter when directing attention to situations and parameters characteristic of systems of practical interest. They thus reflect the results of design and analysis from Part II of the present work. Parameter values used in sample calculations are usually chosen to yield significant but moderate energy dissipation.

7.2. Radiation from forced oscillations

7.2.1. Overview

Time-varying forces in an extended material system can excite mechanical vibrations that are eventually thermalized. The energy dissipated in this fashion is distinct from the energy transiently stored in local elastic deformations, which is (unless subject to other dissipation mechanisms) recovered in the course of a cycle. This mode of energy dissipation is affected by the structure of the system in a substantial region surrounding the device in question. In estimating magnitudes, it is natural to begin by modeling the structure as a uniform elastic medium and the vibrations as acoustic waves in that medium. At the frequencies and energies of interest here, quantum effects will be small.

The accuracy of this approximation will vary, but it will be good when the wavelength of the acoustic radiation is long compared to the scale of inhomogeneities in the mechanical system. Many systems considered in later chapters have structures characterized by an extended matrix or housing that supports numerous nanoscale moving parts. To estimate dissipation due to acoustic radiation, it is reasonable to treat such a system as uniform on a scale of tens of nanometers or more. If the structural matrix includes roughly 1/10 the total mass and is of diamond-like stiffness, then the speed of sound across the system will be $\sim (1/10)^{1/2}$ times the speed of sound in diamond, or ~ 5000 m/s. For $\lambda = 100$ nm, $\omega \approx 3 \times 10^{11}$ rad/s. At higher frequencies, the acoustic radiation model should still yield results of the right order, so long as estimates are based on mean properties of the structure within a wavelength of the device. (Phonon scattering processes will depend on material properties on a nanometer scale.)

Treating dissipation as simple acoustic radiation still leaves a complex problem. Only a few cases will be treated here, and then by approximation. Acoustic radiation in fluids is commonly described; expressions for the power radiated by a pulsating sphere and a piston in a wall appear in (Gray 1972); (Nabarro 1987) adapts an expression for a pulsating cylinder in a fluid to describe analogous radiation losses in a solid. Of more interest in the present context is radiation resulting from a sinusoidally-varying force, torque or pressure at a point (or small region) in an elastic medium. General expressions for radiation from a time-varying force applied at a point within a solid medium appear in the literature (e.g., (Hudson 1980)). Sections 7.2.3, 7.2.4, and 7.2.5 derive a simple approximations for a sinusoidally-varying force, torque, and pressure. Section 7.2.6 discusses radiation from traveling disturbances in a medium, taking dislocations in crystals as a model.

7.2.2. Acoustic waves and the equal-speed approximation

An isotropic elastic medium will support transverse waves of velocity

$$v_{s,t} = \sqrt{\frac{G}{\rho}} \quad (7.1)$$

where G is the shear modulus and ρ the density, together with longitudinal waves of velocity

$$v_{s,l} = \sqrt{\frac{E}{\rho} \frac{1-\nu}{(1+\nu)(1-2\nu)}} \quad (7.2)$$

where E is Young's modulus and ν is Poisson's ratio, which (save in unusual structures of negative ν (Lakes 1987)) falls in the range $0 \leq \nu \leq 0.5$. In an isotropic medium,

$$E = 2G(1 + \nu) \quad (7.3)$$

hence $\nu_{s,\ell} > \nu_{s,t}$. For diamond, $\nu \approx 0.1$, and $\nu_{s,\ell} \approx 1.5\nu_{s,t}$.

Given the approximations involved in treating a nanomechanical system as a uniform medium, it is not unreasonable to add the approximation that $\nu_{s,\ell} = \nu_{s,t}$ for waves radiated from the origin. This equal-speed approximation in effect assumes anisotropic elastic properties that simplify the mathematics, rather than making the mathematics fit an arbitrarily-assumed isotropy. The equal-speed approximation also underlies the standard Debye model of heat capacity and the phonon distribution. It is used extensively in the phonon-drag models of Section 7.3.

7.2.3. Oscillating force at a point

Many mechanical systems will cause disturbances that can be approximated by an oscillating force applied to a point. Among these are unbalanced rotors, reciprocating power-driven mechanisms, and vibrating, elastically-restrained masses.

7.2.3.1. A model

With the above equal-speed approximation, propagating disturbances can be a function of radius alone: the restoring forces between uniformly-displaced spherical shells will then uniform over each shell, leading to uniform accelerations and continued spherical uniformity of displacements. The linearized dynamical equation has the form

$$\frac{\partial}{\partial r} \left(4\pi r^2 M \frac{\partial}{\partial r} y(r, t) \right) = 4\pi r^2 \rho \frac{\partial^2}{\partial t^2} y(r, t) \quad (7.4)$$

where the function $y(r, t)$ specifies a displacement along the line of the force, and M is a modulus of elasticity, uniform in magnitude but differing in nature from the axis aligned with the force (where it is equivalent to $E[1 - \nu]/[1 + \nu][1 - 2\nu]$) to the plane perpendicular to that axis (where it is equivalent to G).

The oscillating force of amplitude F_{\max} is introduced through the boundary condition

$$4\pi r^2 M \frac{\partial}{\partial r} y(r, t) \Big|_{r=0} = F_{\max} \sin(\omega t) \quad (7.5)$$

and solutions corresponding to outbound waves are required. These constraints yield

$$y(r, t) = -\frac{F_{\max}}{4\pi Mr} \sin \omega \left(t - r \sqrt{\frac{\rho}{M}} \right) \quad (7.6)$$

The instantaneous power at a given radius is the force transmitted times the velocity

$$P = \left(4\pi r^2 M \frac{\partial}{\partial r} y(r, t) \right) \frac{\partial}{\partial t} y(r, t) \quad (7.7)$$

which has a time-average value equaling the (isotropic) mean radiated power

$$P_{\text{rad}} = F_{\max}^2 \omega^2 \sqrt{\rho} \frac{1}{8\pi M^{3/2}} \quad (7.8)$$

7.2.3.2. Damping of an embedded harmonic oscillator

A harmonic oscillator like that in Figure 7.1 will be damped by acoustic radiation. At a given amplitude, the force is related to the energy and effective stiffness by

$$F_{\max} = \sqrt{2k_s \mathcal{E}} \quad (7.9)$$

Equating the net radiated power to the time-average value derived above yields an

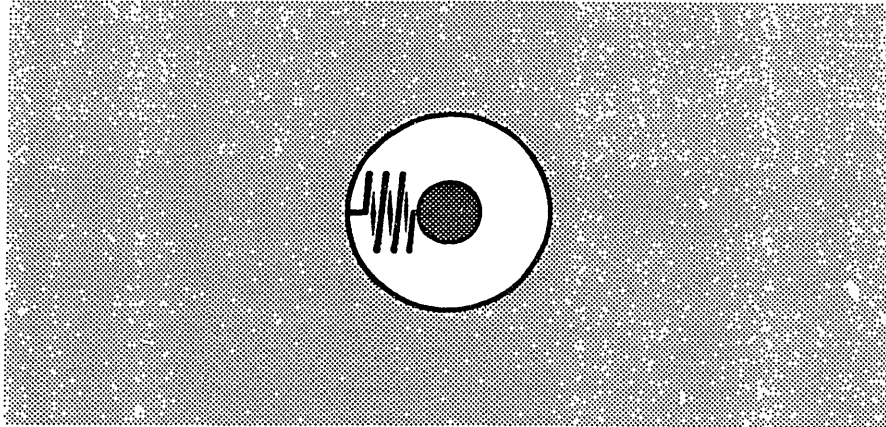


Figure 7.1. Model of a mechanical harmonic oscillator embedded in a medium. The oscillator can be treated as a point source of force so long as its dimensions are small compared to the wavelength of the sound emitted at its characteristic frequency.

exponential decay of the oscillation energy with a time constant (in seconds) of

$$\tau_{\text{osc}} \approx \frac{4\pi m M^{3/2}}{k_s^2 \sqrt{\rho}} \quad (7.10)$$

Alternatively, the fractional energy loss per cycle can be expressed as

$$f \approx \frac{1}{2} \sqrt{\frac{\rho}{m}} \left(\frac{k_s}{M} \right)^{3/2} \quad (7.11)$$

for $f \ll 1$.

Many systems of low stiffness will be constrained by nonbonded interactions with strong anharmonicity. In the limiting case, the stiffness at equilibrium is small, and the oscillation can be viewed as a series collisions with bounding walls. Energy loss is then better modeled using thermal accommodation coefficients (Section 7.5.1).

7.2.4. Oscillating torque at a point

Torsional harmonic oscillators are directly analogous to the linear oscillators just discussed, and can be modeled as sinusoidally varying torques applied to a point. Further, the potential energy of an imperfect bearing will vary with the rotational angle, causing a varying torque. For a bearing in uniform rotation, the resulting torque can be treated as a sum of sinusoidally varying components, each resulting in acoustic radiation.

7.2.4.1. A model

An oscillating torque in a uniform, isotropic medium radiates pure shear waves, hence such media serve as a convenient approximation for real systems. The analysis roughly parallels that given above. Again, spherical shells can be treated as undergoing rigid motion (here rotation rather than displacement), reducing the problem to a linearized equation with a single spatial dimension. The linearized dynamical equation is

$$\frac{\partial}{\partial r} \left(\frac{8}{3} \pi r^4 G \frac{\partial}{\partial r} y_\theta(r, t) \right) = \frac{8}{3} \pi r^4 \rho \frac{\partial^2}{\partial t^2} y_\theta(r, t) \quad (7.12)$$

where the function $y_\theta(r, t)$ specifies an angular displacement about the axis of the applied torque, which sets the boundary condition at the origin:

$$\left. \frac{8}{3} \pi r^4 G \frac{\partial}{\partial r} y_{\theta}(r, t) \right|_{r=0} = T_{\max} \cos(\omega t) \quad (7.13)$$

Together with the requirement for outbound waves, this yields the solution

$$y_{\theta}(r, t) = \frac{T_{\max} \omega \sqrt{\rho}}{8 \pi G^{3/2}} \left[\frac{1}{r^2} \sin \omega \left(t - r \sqrt{\frac{\rho}{G}} \right) + \frac{1}{r^3 \omega} \sqrt{\frac{G}{\rho}} \cos \omega \left(t - r \sqrt{\frac{\rho}{G}} \right) \right] \quad (7.14)$$

(which includes a near-field component). This solution implies a time-average radiated power

$$P_{\text{rad}} = T_{\max}^2 \omega^4 \frac{\rho^{3/2}}{48 \pi G^{5/2}} \quad (7.15)$$

which is steeply dependent on frequency.

7.2.4.2. Damping of an embedded torsional harmonic oscillator

Paralleling the above development, a torsional harmonic oscillator characterized by an angular spring constant k_{θ} (J/rad²) and a moment of inertia I (kg·m²) has a time constant for radiative decay of oscillation energy

$$\tau_{\text{osc}} \approx \frac{24 \pi I^2 G^{5/2}}{k_{\theta}^3 \rho^{3/2}} \quad (7.16)$$

and a fractional energy loss per cycle

$$f \approx \frac{1}{12} \left(\frac{\rho}{I} \right)^{3/2} \left(\frac{k_{\theta}}{G} \right)^{5/2} \quad (7.17)$$

7.2.5. Oscillating pressure in a volume

A component sliding through a channel with corrugated walls will exert a varying pressure on its surroundings. The force applied in one direction is balanced by the force applied in the opposite direction, distinguishing this from the case described in Section 7.2.3. This and related systems can be modeled as a sinusoidally varying pressure in a spherical cavity.

7.2.5.1. A model

An oscillating pressure in a spherical cavity in an isotropic, homogeneous medium will radiate spherical, longitudinal waves. In the near field, however, hoop stresses transverse to the wave can play a dominant role in the balance of forces. The materials of greatest engineering interest have low values of ν ; for example, diamond has $\nu \approx 0.2$ (Field 1979); the analysis will be simplified and render somewhat more conservative by assuming $\nu = 0$ and treating the effective modulus M as equal to Young's modulus E . With these approximations, the linearized dynamical equation is

$$\frac{\partial}{\partial r} \left(4\pi r^2 M \frac{\partial}{\partial r} y(r, t) \right) - 8\pi M y(r, t) = 4\pi r^2 \rho \frac{\partial^2}{\partial t^2} y(r, t) \quad (7.18)$$

where the function $y(r, t)$ specifies a radial displacement. Because the forces resulting from the applied pressure are attenuated by the containing layers of the medium in a way impossible for forces or torques, the a radius of application r_0 must be defined for the applied pressure p and the associated force $F (= \pi r^2 p)$. The boundary condition imposed by the oscillating force is then

$$4\pi r^2 M \frac{\partial}{\partial r} y(r, t) \Big|_{r=r_0} = F_{\max} \sin(\omega t) \quad (7.19)$$

Together with the requirement for outbound waves, this yields the solution

$$y(r, t) = \frac{F_{\max}}{4\pi M} \left(\frac{\rho \omega^2 r_0^2}{M} + \frac{4M}{\rho \omega^2 r_0^2} \right)^{-1/2} \left[\frac{1}{r} \sin \omega \left(t - r \sqrt{\frac{\rho}{M}} \right) - \frac{1}{\omega r^2} \sqrt{\frac{M}{\rho}} \cos \omega \left(t - r \sqrt{\frac{\rho}{M}} \right) \right] \quad (7.20)$$

This again includes a near-field component. This solution implies a time-average radiated power

$$P_{\text{rad}} = F_{\max}^2 \omega^2 \sqrt{\rho} \frac{1}{16\pi M^{3/2}} \left(\frac{\rho \omega^2 r_0^2}{2M} + \frac{2M}{\rho \omega^2 r_0^2} \right)^{-1} \quad (7.21)$$

The trailing factor (in parenthesis) strongly reduces the radiated power when the radius of the driven region is small compared to the radiated wavelength.

7.2.6. Moving disturbances

7.2.6.1. Dislocations as a model

Dislocations provide a model for nanometer-scale mechanical disturbances moving through a medium, exhibiting many distinct energy dissipation mechanisms. The major role of dislocation motion in determining the strength of bulk materials has encouraged extensive analysis and experimentation; recent reviews include (Nabarro 1987) and (Alshits and Indenbom 1986). Several of the following sections draw directly or indirectly on this body of analysis.

7.2.6.2. Subsonic disturbances

Source of moving mechanical disturbance include objects sliding or rolling on a surface and alignment bands (Section 7.3.5.1) in sliding interfaces. At any given point, the motion imposed by a moving disturbance takes the form of an imposed oscillation. Nonetheless, in a uniform environment, a uniform disturbance moving at a uniform, subsonic speed will radiate no acoustic power. In a real system, inhomogeneities and variations in speed and in force as a function of time will lead to forced-oscillation radiation, but these mechanisms can be considered separately.

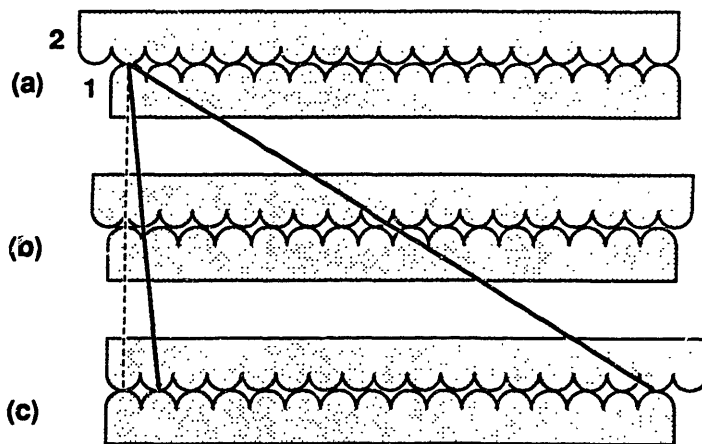


Figure 7.2. The motion of bands of atomic alignment as two surfaces with differing row spacings slide over each other. Panels (a), (b), and (c) represent three successive positions, the arrows trace the motion of an atom in surface 1 (left) and of the alignment band between the surfaces (right).

7.2.6.3. Supersonic disturbances

Material motions of subsonic speed can lead to supersonic patterns of disturbance. The chief mechanism of interest here is the the motion of bands of atomic alignment (closely analogous to dislocations) in sliding interfaces.

Figure 7.2 illustrates the geometry for two rows of atomic bumps, with wave vectors κ_1 and κ_2 (rad/m). Panels (a), (b), and (c) show three successive configurations as surface 2 moves over surface 1: the arrow to the left shows the motion of an atom in surface 2; the arrow to the right shows the motion of an alignment band. Taking v as the velocity of 2 with respect to 1, and d_2 as the interatomic spacing of surface 2, it can be seen that the spatial frequency of the alignment bands is $|\kappa_{\text{bands}}| = |\kappa_2 - \kappa_1|$, and that the velocity ratio R is

$$R = \frac{v_{\text{bands}}}{v} = \frac{|\kappa_1|}{|\kappa_2 - \kappa_1|} \quad (7.22)$$

which can attain arbitrarily high values as $|\kappa_1 - \kappa_2| \rightarrow 0$.

More generally, each surface can be viewed as having many sets of rows, with sets being described by wave vectors that need not be collinear with each other, or with the sliding velocity vector. Interpreting κ_1 and κ_2 as vectors with signs chosen to minimize $|\kappa_2 - \kappa_1|$, each pair of opposed row-sets defines a set of alignment bands in which $\kappa_{\text{bands}} = \kappa_2 - \kappa_1$. From a geometrical construction, it can be seen that

$$R = \frac{v_{\text{bands}}}{v} = \left| (\kappa_2 - \kappa_1) \frac{|\kappa_1|}{|\kappa_2 - \kappa_1|^2} \cos \alpha + \frac{\kappa_1 \times (\kappa_2 \times \kappa_1)}{|\kappa_1 \times (\kappa_2 \times \kappa_1)|} \sin \alpha \right| < \frac{|\kappa_1|}{|\kappa_2 - \kappa_1|} + 1 \quad (7.23)$$

where α measures the angle between the velocity vector v and the vector κ_1 .

In the limiting case, $|\kappa_2 - \kappa_1| = 0$, $R = \infty$, and the interface as a whole periodically enters and leaves the aligned state, radiating sound like an oscillating piston, Eq. (7.24). (This limiting case also sets an upper bound to the power dissipation of supersonic alignment bands.) Nanomechanical bearings of several kinds contain sliding interfaces (Chapter 10). The present work adopts the design constraint that the alignment-band speeds remain subsonic, thereby avoiding this mode of energy dissipation.

The above limiting case can be modeled as a compliant interface in which a sinusoidal variation in the equilibrium separation occurs at a frequency $\omega = \kappa v$. The time-

average radiated power is then

$$P_{\text{rad}} = A^2 \omega^2 \sqrt{M\rho} \left(M\rho \frac{\omega^2}{k_a^2} + 4 \right)^{-1} S \quad (7.24)$$

where S is the area of the interface, A is the amplitude of the variation in equilibrium separation (the limit of the actual amplitude as $\omega \rightarrow 0$), both media are assumed identical, and power radiated from both sides of the surface is included. Typically, unless ω is unusually high or k_a unusually low, the approximation

$$P_{\text{rad}} \approx A^2 \omega^2 \sqrt{M\rho} \frac{1}{4} S \quad (7.25)$$

is accurate (it is always conservative). For the not-atypical values $M = 10^{11} \text{ N/m}^2$, $\rho = 2000 \text{ kg/m}^3$, $k_a = 2 \times 10^{10} \text{ rad/m}$, and $A = 0.05 \text{ nm}$, the radiated acoustic power is $\approx 4 \times 10^6 \text{ W/m}^2$ at a speed of 1 m/s, and $\approx 4 \times 10^2 \text{ W/m}^2$ at a speed of 1 cm/s. Again, most sliding interfaces need not be subject to losses by this mechanism.

7.2.6.3. *Non-adiabatic processes*

J. Soreff notes that, if one surface of a sliding interface is modeled as an array of atomic-scale harmonic oscillators, these will be exposed to mechanical perturbations resulting from the passage of atomic features on the other surface and will be subject to excitation at some rate due to adiabatic processes (that is, “adiabatic” in the quantum-mechanical rather than the thermal sense). The probability of an encounter resulting in an excitation (from first-order perturbation theory (Kogan and Galitskiy 1963)) is proportional to a ratio of the perturbing energy to the oscillator quantal energy ($\hbar\omega$), a quantity typically of order unity, times the factor $\exp(-2\omega\tau)$, where τ is the characteristic time of the perturbation. Since $\omega\tau \approx v_s/v$, Soreff observes that $\exp(-2\omega\tau)$ will typically be extremely small. For example, in a material with $v_s = 10^4 \text{ m/s}$, a system with v as high as 10^2 m/s will have a transition probability on the order of 10^{-85} . In a typical system, a single excitation would dissipate on the order of 10^{-21} J .

7.3. Phonons and phonon scattering

7.3.1. Phonon momentum and pressure

Thermal phonons in a crystal resemble blackbody radiation in a cavity, and both resemble a gas. As discussed further in Sections 7.3.3, 7.4.1, and 7.4.2, the phonon gas is responsible for energy dissipation by mechanisms analogous to those in ordinary gases. Here, we consider drag resulting from scattered and reflected phonons.

In calculating drag due to scattering, phonons can be treated (Lothe 1962) as having a momentum equal to their quasi-momentum (i.e., crystal momentum), of magnitude

$$|p| = \hbar \kappa = \frac{\hbar \omega}{v_s} = \frac{\mathcal{E}}{v_s} \quad (7.26)$$

where κ is, in the following, the *magnitude* of the wave vector (in rad/m) and v_s is the speed of sound (here again approximated as constant for all frequencies and modes). With the substitution of c for v_s , the above expression also describes the momentum of photons in vacuum.

A phonon-reflecting surface in an isotropic medium with a energy density \mathcal{E} experiences a pressure

$$p = \frac{\mathcal{E}}{3} \quad (7.27)$$

Note that this pressure will be exerted on a (hypothetical) surface able to move with respect to the medium, but *not* on features, such as a free surface of a crystal, that can move only by virtue of elastic deformation of the medium. Accordingly, phonon pressure makes no contribution to the thermal coefficient of expansion, which for an ideal harmonic crystal is zero (Ashcroft and Mermin 1976).

7.3.2. The Debye model of the phonon energy density

Phonon scattering drag depends on the phonon energy density and (generally) on the energy distribution vs. wave vector. The commonly-used Debye model of the distribution (discussed at greater length in (Ashcroft and Mermin 1976)) assumes that all waves propagate at a uniform speed v_s . It gives the total phonon energy density as an integral over a spherical volume in κ -space as

$$\varepsilon = \frac{3\hbar v_s}{2\pi^2} \int_0^{\kappa_D} \kappa^3 \left[\exp\left(\frac{\hbar \kappa v_s}{kT}\right) - 1 \right]^{-1} d\kappa \quad (7.28)$$

where κ_D (the Debye radius) is a function of n , the atomic number density (m^{-3}):

$$\kappa_D = (6\pi n)^{1/3} \quad (7.29)$$

The Debye temperature

$$T_D = \frac{\hbar \kappa_D v_s}{k} \quad (7.30)$$

is a measure of the temperature at which the highest-frequency modes of a solid are excited. For $T \ll T_D$, $\varepsilon \propto T^4$, as in blackbody radiation. To yield the correct value of the phonon energy density for $T \ll T_D$, v_s in the above expressions must (in an isotropic material) be taken as

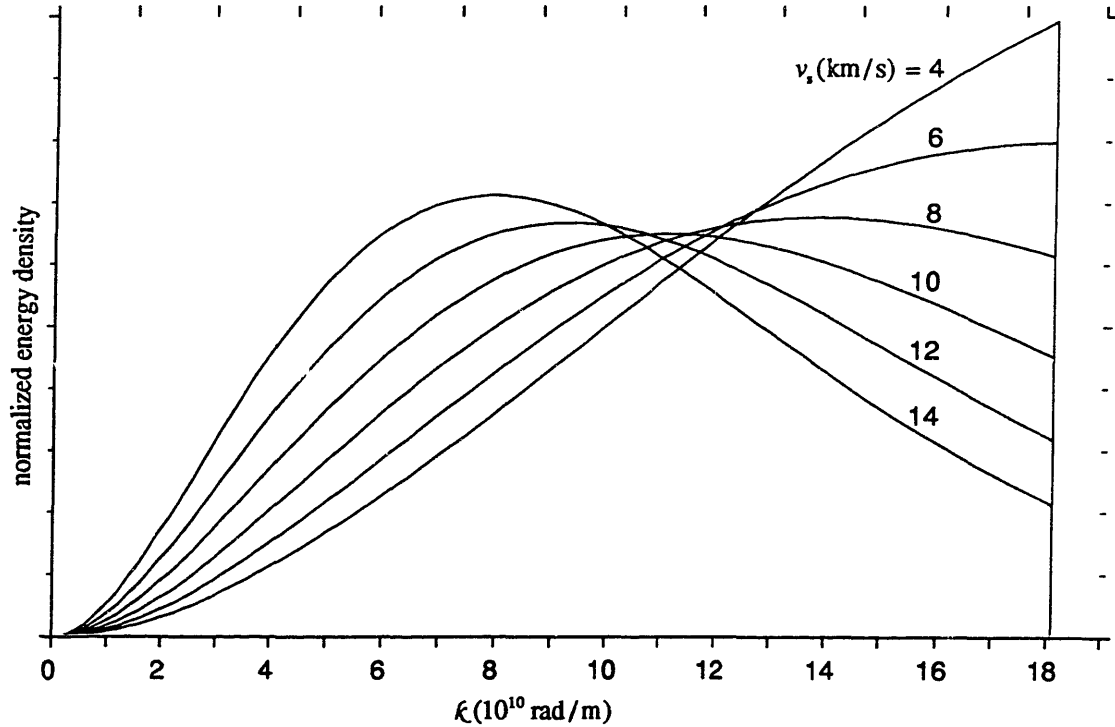


Figure 7.3. Phonon energy density per unit interval of κ in the Debye model for $n = 100/\text{nm}^3$ and $T = 300 \text{ K}$, normalized to constant total energy. The maximum value of $\kappa = \kappa_D = 1.24 \times 10^{10} \text{ rad/m}$.

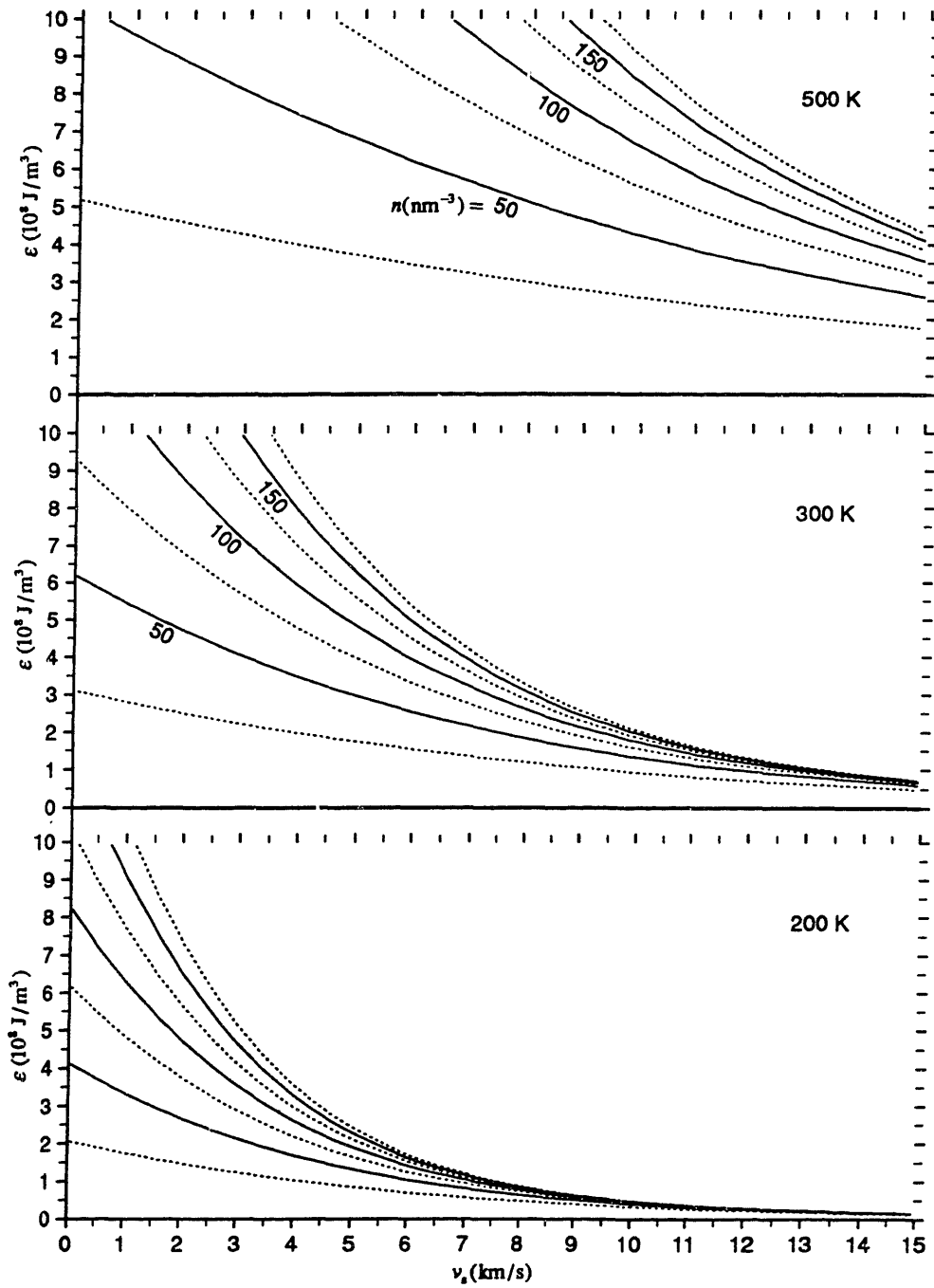


Figure 7.4. Phonon energy density in the Debye model vs. the *effective* speed of sound, v_s and the atomic number density, n . For perspective, in diamond the effective speed $v_s \approx 13.8 \text{ km/s}$ and $n \approx 176/\text{nm}^3$.

$$v_s = \left(\frac{1}{3} v_{s,\ell}^{-3} + \frac{2}{3} v_{s,t}^{-3} \right)^{-1/3} \quad (7.31)$$

which has a maximum value of $1.084v_{s,t}$ in the limiting case of $E = 2G$. For a hypothetical isotropic substance with $v_{s,t}$ and $v_{s,\ell}$ equal to those of diamond along an axis of cubic symmetry, the above relations give $v_s = 1.38 \times 10^4$ m/s and $T_D = 1570$ K (vs. a value for diamond itself of 2230 K (Gray 1972)).

The Debye model has several shortcomings in describing real crystals, to say nothing of nanomechanical systems treated as continuous media; in the present context, its chief defects arise from its neglect of waves of low group velocity. Near the limiting value of k , acoustic dispersion (ignored in the Debye model) results in group velocities approaching zero. The Debye model also neglects so-called optical modes of vibration, which typically have low group velocities. These effect of these shortcomings can be significant, but is typically small for $T \ll T_D$ (Alshits and Indenbom 1986).

7.3.3. Phonon scattering drag

A scattering center moving with velocity v will experience drag from the “phonon wind” resulting from its velocity (Alshits and Indenbom 1986); this can be treated as analogous to scattering of photons in a vacuum. In the simplest situation, a scattering center has both an isotropic cross section in the rest frame and an isotropic emission pattern in its own frame. Drag can then be calculated from of the anisotropic momentum distributions and encounter frequencies of phonons seen in the moving frame.

Phonons approaching at an angle θ from the forward direction will be doppler shifted, changing their frequency and energy by a factor $(1 + v\cos\theta/v_s)$; the rate of encounter for phonons from this direction will changed by the same factor. Phonons approaching from the side will experience an aberrational shift in direction through an angle $-v\sin\theta/v_s$ (for $v \ll v_s$). Combining these factors, discarding terms of order v^2/v_s^2 and higher, and integrating yields the phonon-scattering drag and power dissipation for the scattering center:

$$F_{\text{drag}} = -\frac{4}{3} \epsilon \sigma_{\text{therm}} \frac{v}{v_s}, \quad P_{\text{drag}} = \frac{4}{3} \epsilon \sigma_{\text{therm}} \frac{v^2}{v_s} \quad (7.32)$$

where σ_{therm} is a thermally-weighted scattering cross section (in m^2) derived from a wave-vector dependent scattering cross section $\sigma(k)$. For the Debye model of the phonon

distribution,

$$\sigma_{\text{therm}} = \int_0^{k_D} \sigma(k) k^3 \left[\exp\left(\frac{\hbar k v_s}{kT}\right) - 1 \right]^{-1} dk / \int_0^{k_D} k^3 \left[\exp\left(\frac{\hbar k v_s}{kT}\right) - 1 \right]^{-1} dk \quad (7.33)$$

7.3.4. Scattering from harmonic oscillators

A variety of nanomechanical components can be treated as moving scattering centers. A roller bearing moving across a surface, a follower moving over a cam, an object sliding in a tube: each results in the motion of a small region of contact with respect to a medium. The effect of the contact can be modeled as an embedded harmonic oscillator of the sort described in Section 7.2.3.2, excited by incident phonons and radiating to a degree that can be approximated by Eq. (7.8).

In the limit of large mass and low stiffness, the motion of the mass will be small and the oscillating force will be proportional to the stiffness, making $\sigma \propto k_s^2$. In the limit of low mass and large stiffness, the deformation of the spring will be small and the force will be proportional to the mass, making $\sigma \propto m^2$. In general, far from resonance,

$$\sigma \approx \frac{1}{2\pi} \left(\frac{k_s}{M} \right)^2 \left(\frac{k_s}{m\omega^2} + 1 \right)^{-2} \quad (7.34)$$

Resonant scattering occurs when $k v_s \approx [k_s/m]^{1/2}$, with resonant cross sections limited by radiation damping. Values of σ_{therm} can be estimated by numerical integration of the damped harmonic oscillator cross section over the Debye phonon distribution. The results for representative values of material parameters at 300 K are graphed in Figure 7.5, for oscillators with isotropic effective stiffnesses and masses. For oscillators with differing values along three principle axes, the cross section is the mean of those that would be exhibited by three isotropic oscillators with these values.

A sliding scattering center will typically be coupled to the medium by nonbonded interactions. The relationships summarized in Fig. 3.7 suggest that, regardless of how low the equilibrium stiffness may be, thermal excitation of small scattering centers at 300 K will frequently explore regions in which the stiffness is of the order of 10 N/m. Thus, anharmonicity and thermal excitation will place an effective lower bound on the effective mean stiffness, and hence on the phonon scattering cross section.

It is useful to examine the magnitude of drag for a typical case. A sliding contact with a stiffness of ~ 30 N/m will have $\sigma \approx 10^{-20}$ m² in a moderately stiff medium. With $v_s \approx 10^4$ m/s and $\varepsilon \approx 2 \times 10^8$ J/m³, its power dissipation Eq. (7.32) will be $\sim 3 \times 10^{-16}$ W at 1 m/s and $\sim 3 \times 10^{-20}$ W at 1 cm/s, or (equivalently) 3×10^{-25} and 3×10^{-27} J/nm traveled.

7.3.5. Scattering from alignment bands in bearings

7.3.5.1. Alignment bands in bearings

Sliding interfaces will contain alignment bands that are closely analogous to dislocations. Phonon-scattering drag plays a major role in dislocation dynamics and has accordingly received substantial attention (Alshits and Indenbom 1986; Lothe 1962; Nabarro 1987). Causes of scattering include both the mechanical inhomogeneity of the dislocation core and “flutter,” in which phonons excite oscillations in dislocations, which then re-radiate power.

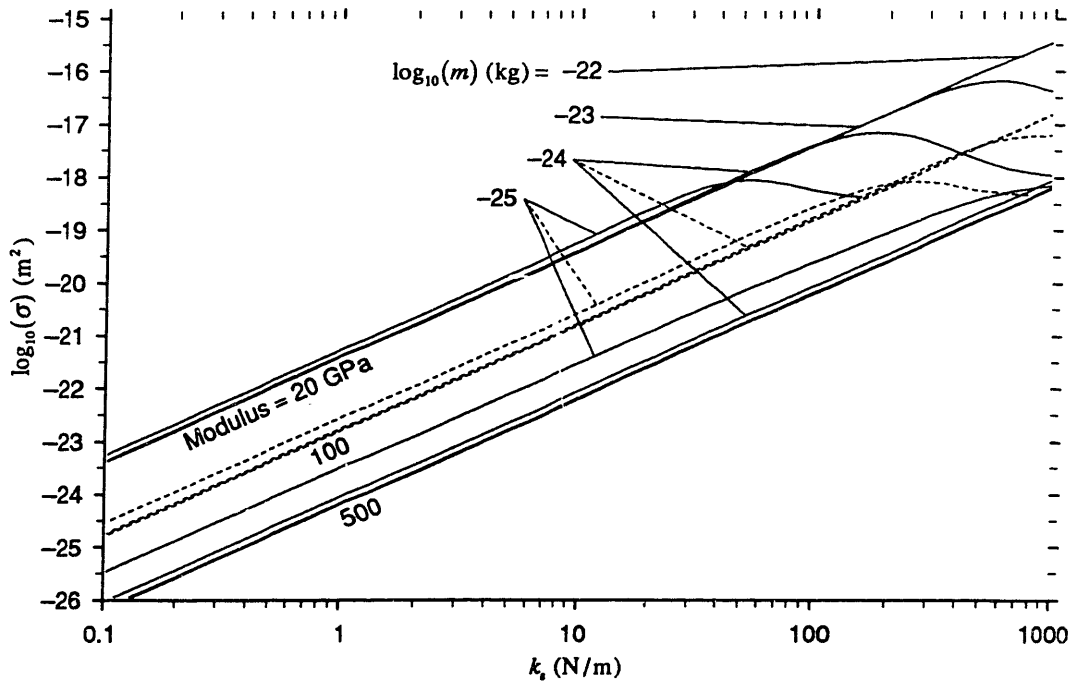


Figure 7.5. Values of σ_{therm} at 300 K for a range of values of m and k_s and three values of the modulus M . For perspective, 10^{-25} kg is the approximate mass of 5 carbon atoms, and the Young's modulus of diamond is ~ 1000 GPa ($= 10^{12}$ N/m²). The assumed density is 2000 kg/m³ (vs. ~ 3500 for diamond), with $n = 100/\text{nm}^3$.

In typical materials, dislocations are narrow, causing severe disruption of crystalline alignment over ~ 5 atomic spacings (Lothe 1962), and inducing local variations in stress that are significant compared to the modulus E . Alignment bands in sliding-interface bearings, however, will typically be broad and will typically induce only small variations in stress. These variations in stress, however, can yield significant variations in the stiffness of the interface owing to the strong anharmonicity of nonbonded interactions. Alignment bands are sufficiently similar to dislocations that analogs of both flutter scattering and inhomogeneity scattering will occur, yet are different enough to invalidate the approximations that have been used in modeling dislocations. Suitable approximations are developed in Sections 7.3.5.2–7.3.5.5.

7.3.5.2. Common features of the models

Alignment bands in interfaces can be modeled as sinusoidally-varying disturbances moving at a speed v_{bands} (related to the sliding speed v by Eq. (7.22) or (7.23), in two limiting cases). The nature of the disturbance varies with the mode of scattering. Both flutter scattering and stiffness scattering are here described by approximate models, intended to provide only upper bounds.

In many systems of engineering interest, the shear stresses transmitted across the interface will be small compared to the normal stresses, and the shear stiffness will likewise be small compared to the normal stiffness. Shear stresses and stiffnesses will accordingly be neglected in the following discussion, although their treatment would be entirely analogous.

As noted by J. Soreff, the assumption that $v_{s,l} = v_{s,t}$ permits waves to be resolved into components with x , y , and z polarizations, where the polarization axes may be chosen for convenience and without regard to the direction of propagation. For scattering from band stiffness variations, only polarizations perpendicular to the interface are relevant; for flutter scattering, only polarizations parallel to the interface and oriented in the band-shifting direction are relevant.

7.3.5.3. Band-stiffness scattering

The interface can be treated as a compliant sheet with a stiffness per unit area k_a and a transmission coefficient (Section 7.3.5.5.) equaling T_{trans} (this coefficient includes the factor of $1/3$ resulting from the effectiveness of only one of three polarizations). The mechanical inhomogeneity of the alignment bands will typically be well-approximated by

a sinusoidal variation in stiffness of amplitude $\Delta k_a/2$, which (in interfaces with small values of T_{trans}) will cause variations in the transmission coefficient on the order of $\Delta T_{\text{trans}} = 1.7T_{\text{trans}}\Delta k_a/k_a$.

Specular reflection and simple transmission of phonons makes no contribution to the drag, but a fraction of incident phonons $\leq \Delta T_{\text{trans}}$ will be subject to diffractive scattering from the bands owing to the spatial variation in the transmission (and accordingly, reflection) coefficient. For normally-incident phonons of $\kappa \gg \kappa_{\text{bands}}$, the diffraction angle will be small, and for $\kappa < \kappa_{\text{bands}}$, it will be zero. In these cases, the scattering contribution to drag will be relatively small or nonexistent. (Owing to the angular variation in the transmission coefficient, the actual results will be strongly influenced by the diffraction of phonons at grazing incidence.) The present estimate will nonetheless assume isotropic scattering for all κ tending to overestimate the drag.

The incident power on a single side of a surface is $v_s \epsilon/4$, and hence the average collision cross section for a flat surface of area S (counting both sides and all angles of incidence) is $S/2$. (The quantity $\epsilon/4$ will appear frequently in normalization expressions and can be termed the *effective energy density*.) Combining these factors yields the bound

$$P_{\text{drag}} < 0.85 \epsilon T_{\text{trans}} \frac{\Delta k_a}{k_a} R^2 \frac{v^2}{v_s} S \quad (7.35)$$

where v is the sliding speed of the interface and a factor of order unity (analogous to the 4/3 in Eq. (7.32)) has been dropped. (Note that this and similar expressions do not hold when the formal value of $T_{\text{trans}} \approx 0$.)

As discussed in Chapter 10, $\Delta k_a/k_a$ can be made small in properly-designed bearings of certain classes, and values of T_{trans} (Section 7.3.5.5.) can easily be less than 10^{-3} . A not-atypical value for R is 10. For these values, with $\Delta k_a/k_a \approx 0.1$, $v_2 = 10^4$ m/s, and $\epsilon = 2 \times 10^8$ J/m³, P_{drag} from this mechanism will be bounded by ~ 200 W/m² at $v = 1$ m/s, and ~ 0.02 W/m² at 1 cm/s. The latter values correspond to 2×10^{-25} and 2×10^{-27} J/nm² per nanometer of travel.

7.3.5.4. Band-flutter scattering

Alignment bands also cause deformations of the equilibrium shape of each surface of the interface with amplitude A and spatial frequency κ_{bands} ; this results in sinusoidally-varying slopes with a maximum magnitude of $A\kappa_{\text{bands}}$. A shear wave of the proper polarization will cause the bands to move by a distance that is a multiple R of the particle dis-

placements caused by the shear wave itself. The ratio of the amplitude of the equilibrium displacement of the interface to the amplitude of the incident shear wave is $A\kappa_{\text{bands}}R$. These displacements are like those imposed by an incident wave of perpendicular polarization and scaled amplitude; after this transformation, the interface can again be regarded as a moving diffraction grating.

Taking the mean square value of the slope over the interface introduces a factor of 1/2 in the scattered power; consideration of radiation from both surfaces introduces a compensating factor of 2. The time-reversal of an equilibrium system is an equilibrium system, hence in the limit of slow band motion, power scattered from parallel to perpendicular polarizations by band flutter must equal power scattered from perpendicular to parallel. This introduces a further factor of 2 in the drag expression.

With the bounding approximations made above, the analysis proceeds essentially as before, yielding

$$P_{\text{drag}} < \epsilon T_{\text{trans}} (A\kappa_{\text{bands}}R)^2 R^2 \frac{v^2}{v_s} S \quad (7.36)$$

By Eq. (7.22), the product $\kappa_{\text{bands}}R = \kappa_1 = 2\pi/d$, where d is the spacing of atomic rows in either surface. (In the general case described by Eq. (7.23), this remains a reasonable approximation.) This yields the expression

$$P_{\text{drag}} < \epsilon T_{\text{trans}} \left(\frac{2\pi A}{d} \right)^2 R^2 \frac{v^2}{v_s} S \quad (7.37)$$

As with stiffness variations, A/d can be made small in properly-designed bearings of certain classes; a not-atypical value will be 10^{-2} . For values of other variables as assumed above, the drag power from this mechanism will be bounded by $\sim 10 \text{ W/m}^2$ at 1 m/s.

7.3.5.5. The transmission coefficient

As seen in Sections 7.3.5.3 and 7.3.5.4, the phonon transmission coefficient T_{trans} greatly affects drag at sliding interfaces. In a simple one-dimensional model of longitudinal vibrations propagating along a rod with a linear modulus E_l (N) interrupted by spring of stiffness k_s , the transmission coefficient is

$$T_{\text{trans,rod}} = \left[1 + \left(\frac{E_t \kappa}{2k_s} \right)^2 \right]^{-1} \quad (7.38)$$

Where κ is the spatial frequency (rad/m).

A detailed analysis by J. Soreff shows that in a medium in which all speeds of sound are equal, the transmission coefficient at a planar interface takes the same form,

$$T_{\text{trans,perp}} = \left[1 + \left(\frac{M \kappa_z}{2k_s} \right)^2 \right]^{-1} \quad (7.39)$$

in which κ_z is the z-component of the wave vector of an incident wave of perpendicular polarization, and M is the single modulus. The overall mean power transmission coefficient can then be estimated by an integral over one hemisphere of the allowed volume of κ -space, weighting contributions from different wave vectors in accord with the Debye model of the distribution of phonon energy and including a factor of 1/3 to account for the transmission of incident power in only one of the three possible polarizations:

$$T_{\text{trans}} = \frac{\frac{4}{3} \int_0^1 \frac{\kappa'^3}{\exp(\kappa'/T') - 1} \int_0^{\pi/2} \frac{\sin(2\theta)}{(d' \kappa' \cos \theta)^2 + 4} d\theta d\kappa'}{\int_0^1 \frac{\kappa'^3}{\exp(\kappa'/T') - 1} d\kappa'} \quad (7.40)$$

where $T' = T/T_D$, and d' is a dimensionless measure of the stiffness of the interface, $d' = \kappa_D M/k_s$. Values of T_{trans} are plotted in Figure 7.6 with respect to $d_n = n^{-1/3} M/k_s$ ($\propto d'$) for a range of values of interest in the present context. The parameter d_n can be interpreted as the thickness of a slab of the medium, in atomic layers (assuming a simple cubic lattice), that has a stiffness per unit area equaling that of the interface itself.

The above expression does not lend itself to easy evaluation or use in analytical models. A reasonable engineering approximation is

$$T_{\text{trans}} \approx \frac{z}{1+3z}; \quad z = \frac{0.6}{d_n^{1.7}} \left(1 + \frac{0.075}{T'^{1.8}} \right) \quad (7.41)$$

or

$$T_{\text{trans}} \approx z, \quad z \ll 1. \quad (7.42)$$

The former expression is plotted in Figure 7.6. Both expressions consistently overestimate transmission (and hence drag), yielding conservative results for most purposes; they are chosen to give a good fit for parameters in the anticipated region of engineering interest, rather than being chosen to exhibit correct behavior at the simple physical limits (e.g., $T' \rightarrow 0$).

7.3.5.6. Curved interfaces and dissimilar media

In the above model, when d_n is large and T' is not small, normally-incident phonons are reflected almost perfectly and grazing-incidence phonons make a large contribution to T_{trans} . The efficient transmission of grazing-incidence phonons results from a resonant process that depends on (1) the prolonged interaction associated with grazing-incidence collisions and (2) the matching acoustic speeds of the media on either side of the interface. In the limit of small angles, the transmission probability approaches unity.

Curved interfaces will disrupt this process, reducing T_{trans} . A difference in acoustic speed between the two media will do likewise, causing the grazing-incidence transmis-

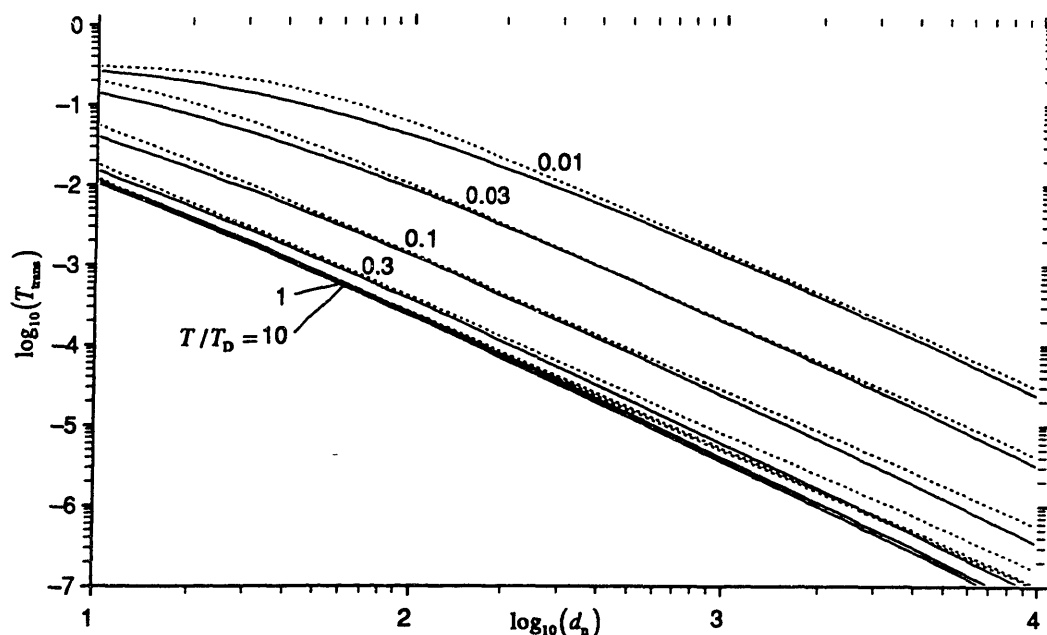


Figure 7.6. The transmission coefficient for a compliant interface, Eq. (7.40), plotted for several values of a dimensionless measure of temperature, T/T_D , and a dimensionless measure of interface compliance, d_n (see text). Dotted curves represent the approximation given by Eq. (7.41).

sion probability to fall to zero on one side (owing to total internal reflection) and to approach zero in the limit of small angles on the other side (as shown by the angular variation of reflectivity in analogous optical systems). Sharply curved interfaces will be common in bearings, and differences in acoustic speed will likewise be common; indeed, such differences can be a design objective. In critical applications, practically-significant differences can be ensured even between chemically-identical structures by building them from different isotopes (e.g., C^{12} vs. C^{13}). An analysis taking account of curved interfaces or differentiated media would be desirable, but the above values provide an upper bound on transmission-dependent drag processes. This suffices for present purposes.

7.3.6. Shear-reflection drag

The asymmetry of doppler shifts for phonons transmitted through a sliding interface shows that a thermal distribution is transformed into a non-thermal distribution, implying an increase in the free energy of the phonons at the expense of the energy of sliding. Analyzing this energy loss mechanism, however, is difficult. A study by J. Soreff of sound propagation through the model interface of Section 7.3.5.5 yields an expression for the wave-vector resolved transmission coefficient of an interface as a function of the Mach number, M_s , and the dimensionless measures of interfacial stiffness and phonon spatial frequency described in Section 7.3.5.5:

$$T_{\text{trans,perp}} = 4r_z \left[(1+r_z)^2 + (d'\kappa_z \sin \psi \cos \phi)^2 \right]^{-1} \quad (7.43)$$

where

$$r_z = \frac{\sqrt{\sin^2 \psi \cos^2 \phi - 2M_s \cos \psi + M_s^2 \cos^2 \psi}}{\sin \psi \cos \phi} \quad (7.44)$$

In this expression, the coordinates are chosen such that sliding motion occurs along the x axis with the z axis perpendicular to the interface; ψ measures the angle between the wave vector and the x axis, and ϕ measures the angle between the $x-\kappa'$ plane and the $x-z$ plane. The square root in Eq. (7.44) is to be taken as positive, with formally imaginary values taken as zero. Soreff observes that phonons crossing the interface experience no change in the component of their wave vector parallel to the interface and hence no change in that component of momentum; only velocity-dependent asymmetries in the

transmission coefficient lead to momentum transfer and hence to drag. (Soreff has verified the correctness of equating crystal momentum and ordinary momentum in this instance by a detailed analysis of transverse forces in the phonon-deformed interface.)

An expression for the drag per unit area per unit phonon effective energy density takes the form of the ratio of integrals in Eq. (7.45). This closely resembles the expression for the transmission coefficient save for a change of coordinates, use of the velocity-dependent expression for transmission, and the introduction of a $\cos(\psi)$ factor in the numerator to account for contributions to the x momentum. At low Mach numbers, contributions from the leading and trailing regions of k -space nearly cancel, and are associated with rapidly-varying length scales; these characteristics make analysis difficult. Soreff reports that a variety of approaches for developing analytical approximations or bounds fail to yield useful results.

On physical grounds, one expects that at low Mach numbers the drag will be approximately proportional to the energy density, to the zero-velocity transmission coefficient, and to the Mach number itself. This encourages consideration of the expression

$$D_{sr} = \frac{1}{M_s T_{trans}} \frac{-\frac{8}{3\pi} \int_0^1 \frac{k'^3}{\exp(k'/T') - 1} \int_0^{\pi/2} \int_0^\pi \frac{r_z \sin^2 \psi \cos \psi \cos \phi}{(1 + r_z)^2 + (d' k'_z \sin \psi \cos \phi)^2} d\psi d\phi dk'}{\int_0^1 \frac{k'^3}{\exp(k'/T') - 1} dk'} \quad (7.45)$$

in hopes that D_{st} is a slowly-varying quantity. Numerical investigation indicates that D_{st} is of order unity and does indeed vary only moderately across a range of parameters in which the drag varies by more than seven orders of magnitude. Use of the approximation $D_{st} = 1$ appears conservative for the systems of interest in the present context, frequently overestimating drag by a factor of 10 or more.

The expression for the power dissipation from shear-reflection drag includes a factor of two to account for phonons approaching interface from each side:

$$P_{drag} \approx \frac{\epsilon}{2} T_{trans} M_s D_{sr} v S = \frac{\epsilon}{2} T_{trans} D_{sr} \frac{v^2}{v_s} S \quad (7.46)$$

The magnitude of drag from this mechanism relative to those described previously will vary with the design parameters. For bearings in which alignment-band drag has been minimized, it will frequently be dominant. With the assumptions of Section 7.3.5.3,

the drag power is $\sim 10 \text{ W/m}^2$ at 1 m/s.

7.4. Thermoelastic damping and phonon viscosity

The phonon gas causes energy loss by mechanisms analogous to those occurring in the compression and shear of ordinary gases. These mechanisms are termed *thermoelastic damping* and *phonon viscosity*.

7.4.1. Thermoelastic damping

When a typical solid is compressed, the energy of its normal modes increases. In the absence of equilibration, phonon energies likewise increase and the solid becomes hotter. Since this process involves changes in the dimensions of the solid rather than motions with respect to the lattice, no work is done against the pressure of the phonon gas directly (as in the compression of an ordinary gas or a photon gas). Instead, phonon energies increase with compression as a result of the anharmonicity of the solid, which in turn arises because interatomic potentials become stiffer as distances shrink. A widely-used measure of anharmonicity is the Grüneisen number,

$$\gamma_G = \frac{\beta K}{C_{\text{vol}}} \quad (7.47)$$

where β is the volume coefficient of thermal expansion (K^{-1}), K is the bulk modulus (N/m^2), and C_{vol} is the heat capacity per unit volume ($\text{J/K}\cdot\text{m}^3$). (C_{vol} equals the molar heat capacity at constant volume divided by the molar volume.) Values of γ_G for many ordinary materials fall in the range 1.5 to 2.5 and have little temperature dependence near 300 K. For diamond, tabulated values are $\beta = 3.5 \times 10^{-6} \text{ K}^{-1}$, $K \approx 4.4 \times 10^{11} \text{ N/m}^2$, and $C_{\text{vol}} \approx 1.7 \times 10^6 \text{ J/K}\cdot\text{m}^3$ at 300 K, yielding $\gamma_G \approx 0.9$.

Thermoelastic damping arises from the difference between the adiabatic and the isothermal work of compression. Starting with an expression for small values of ΔV and ΔT (Lothe 1962), and applying thermodynamic identities,

$$\begin{aligned} \Delta W &= \frac{1}{2} (K_{\text{adiabatic}} - K) \frac{\Delta V^2}{V} \\ &= \frac{1}{2} \gamma_G^2 T C_{\text{vol}} \frac{\Delta V^2}{V} \\ &= \frac{1}{2} \beta^2 \frac{T}{C_{\text{vol}}} \Delta p^2 V \end{aligned} \quad (7.48)$$

A worst-case thermodynamic cycle would involve adiabatic compression of a volume (increasing the temperature) followed by nonequilibrium cooling (producing entropy), followed by adiabatic expansion and nonisothermal warming, resulting in an overall energy dissipation of $2\Delta W$, from Eq. (7.48). For diamond, this amounts to $\sim 2.2 \times 10^{-24} (\Delta p)^2 \text{ J/nm}^3 \cdot \text{cycle}$, where the p is here measured in $\text{nN/nm}^2 (= \text{GPa})$.

Thermoelastic damping falls to zero if the cycle is either adiabatic or isothermal, and nanomechanical systems will frequently approach the latter limit. The ratio of the energy dissipated in a nearly-isothermal cycle to that dissipated in the above worst-case cycle equals the ratio of the mean displacement-weighted temperature increments. A component undergoing smooth mechanical cycling with a period t_{cycle} and a characteristic time for thermal equilibration τ_{therm} will experience a temperature rise during the cycle on the order of $\tau_{\text{therm}}/t_{\text{cycle}}$ times that of the adiabatic case, implying

$$\Delta W_{\text{cycle}} \approx 2\beta^2 \frac{T}{C_{\text{vol}}} (\Delta p)^2 \frac{\tau_{\text{therm}}}{t_{\text{cycle}}} V \quad (7.49)$$

For a component in good thermal contact with its environment, τ_{therm} will be of the order of

$$\tau_{\text{therm}} \approx \max \left(\frac{C_{\text{vol}}}{K_T} \ell^2, \frac{\ell}{v_s} \right) \quad (7.50)$$

Table 7.1. Values of the volumetric thermal coefficient of expansion, β , for a variety of strong solids in the neighborhood of 300 K (Gray 1972).

Material	β (10^{-6} K^{-1})
Diamond	3.5
Silicon	7.5
Silicon carbide	11.1
Sapphire	15.6
Quartz	36.0
Silica (vitreous)	1.2

where K_T is the thermal conductivity (W/m·K) and ℓ is a characteristic dimension. Values of K_T for glasses and nonporous ceramics are typically in the range 1–10 W/m·K, with the value for diamond being ~ 700 (Gray 1972). For $K_T = 10$, $\ell = 10$ nm, and $C_{vol} = 2 \times 10^6$, $\tau_{therm} \approx 10^{-11}$ s; ΔW_{cycle} is accordingly multiplied by a factor of $\sim 10^{-2}$ at 1 GHz and $\sim 10^{-5}$ at 1 MHz relative to the values given by the worst-case expression, Eq. (7.48).

7.4.2. Phonon viscosity

Shear deformation causes no volume change and hence no thermoelastic losses. Shear does, however, cause compression along one axis and extension along another: phonons traveling along one axis are increased in energy; those along the other, reduced. Within a factor of 3/2, an analogy between this and the thermoelastic effect, Eq. (7.48), yields an estimate of the difference between the adiabatic and isothermal shear moduli (Lothe 1962), resulting in an effective viscosity

$$\eta_{phonon} = \tau_{relax} \frac{3}{2} \gamma_g^2 T C_{vol} \quad (7.51)$$

The analysis of the energy dissipation proceeds essentially as for thermoelastic damping, but with the substitution of the phonon relaxation time τ_{relax} for $\tau_{thermal}$, and the shear stress γ for the pressure p , yielding

$$\Delta W_{cycle} \approx \frac{3}{2} \beta^2 \frac{T}{C_{vol}} (\Delta \gamma)^2 \frac{\tau_{relax}}{t_{cycle}} V \quad (7.52)$$

The time τ_{relax} measures the rate of equilibration of phonon energy between different directions in the solid, which can be accomplished by elastic scattering such as that occurring at the boundaries of a solid body or at internal inhomogeneities. In nanomechanical systems, scattering will typically limit phonon mean free paths to nanometer distances, resulting in values of $\tau_{relax} \approx 10^{-13}$ s for $v_s \approx 10^4$ m/s. Because of this small time constant, phonon viscosity losses will typically be small compared to thermoelastic losses save in systems undergoing very high frequency motion or nearly pure shear.

7.4.3. Application to moving parts and alignment bands

Alignment bands, like sliding and rolling components, impose moving regions of stress on the surrounding medium. These regions can be characterized by their volume V

$\approx \ell^3$ (for contact regions) or $\approx d\ell^2$ (for bands extending over a distance d), where ℓ is a measure of the scale of the region (e.g., the wavelength of the alignment bands). For motions of velocity v , $\tau_{\text{cycle}} \approx \ell/v$. This leads to an estimate of the magnitude of the thermoelastic drag of a set of bands:

$$P_{\text{drag}} \approx \beta^2 \frac{T}{K_T} \ell (\Delta p)^2 R^2 v^2 S \quad (7.53)$$

assuming that phonon mean free paths are shorter than ℓ . With $\beta = 3.5 \times 10^{-6} \text{ K}^{-1}$, $K_T = 10 \text{ J/m} \cdot \text{K}$, $\ell = 10 \text{ nm}$, $R = 10$, $T = 300 \text{ K}$, and $\Delta p = 10^8 \text{ N/m}^2$, $P_{\text{drag}} \approx 4 \text{ W/m}^2$ at $v = 1 \text{ m/s}$, or 0.04 W/m^2 at 1 cm/s . As with stiffness and displacement, good design can in many instances yield very low values of Δp .

The above estimates of thermoelastic dissipation have used a value of β appropriate for diamond. Table 7.1 lists values for a variety of other strong solids that can serve as models for the materials of nanomechanical components. The large difference between SiO_2 as quartz and as vitreous silica indicates that β is sensitive to patterns of bonding and hence will be subject to substantial design control in the products of molecular manufacturing, including nanomechanical components.

7.5. Compression of square and harmonic potential wells

7.5.1. Square well compression

Many nanomechanical systems will contain components that move over a relatively flat potential energy surface within a bounded region of variable size. Thermodynamically, the motion of the component within this region is like the motion of a gas molecule in a container; changes in the size of the region correspond to compression and expansion. At any finite speed, compression will be nonisothermal, heating the gas, raising its pressure, and so increasing the work of compression and causing energy dissipation. The better the thermal contact between the component and its environment, the lower the dissipation. A conservative model assumes contact only between the “gas molecule” component and two moving pistons (Figure 7.7), treating the component as a one-dimensional gas consisting of one molecule.

7.5.1.1. Accommodation coefficients

Thermal contact between a gas and a solid is usually described by a thermal accommodation coefficient α that measures the extent to which the excess energy of an impinging gas molecule is lost in a single collision with a wall:

$$(T_1 - T_s)\alpha = T_1 - T_2 \quad (7.54)$$

where the temperature of the incident molecules is T_1 , that of the surface is T_s , and that of the outbound molecules is T_2 . Separate accommodation coefficients can be defined for the energy of translation, rotation, and vibration. The most accurate measurements have been made for monatomic gases, in which all the energy is translational. As just defined, α is a function of three temperatures; in the limit as T_1 , T_2 , and T_s become equal, α becomes a function of a single temperature. A value of the latter sort (an equilibrium accommodation coefficient) can be used with reasonable accuracy so long as none of the three temperatures differs greatly from the reference temperature. Theory and experimental data for thermal accommodation coefficients are reviewed in (Goodman 1980; Goodman and Wachman 1976; Saxena and Joshi 1989).

Save for light gases (helium, neon) impinging on a clean surface with massive atoms (tungsten), tabulated values of α typically range from 0.25 to ~ 1.0 . Surface contamination generally increases accommodation; stable structures with similar effects could be provided in many systems.

7.5.1.2. A square-well temperature-increment model

A simple model for the temperature rise assumes that the statistics of the velocity of a moving component of mass m are those of a system at equilibrium at some temperature T_{rest} , with an rms velocity (along the axis) of $v_{g,\text{rms}} = (kT_g/m)^{1/2}$. If each piston moves at

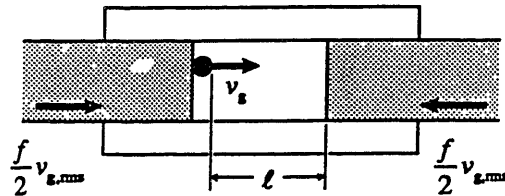


Figure 7.7. Two-piston model of a compressed square well; cylinder walls are assumed adiabatic.

a speed $(f/2)(kT_g/m)^{1/2}$, then

$$v_{1, \text{rms}} = v_{g, \text{rms}} \left(1 + \frac{f}{\sqrt{\pi}} + \frac{f^2}{4} \right), \quad v_{2, \text{rms}} = v_{g, \text{rms}} \left(1 - \frac{f}{\sqrt{\pi}} + \frac{f^2}{4} \right) \quad (7.55)$$

where $v_{1, \text{rms}}$ and $v_{2, \text{rms}}$ are measured in the frame of reference of the scattering piston (pistons are assumed to interact only with particles moving in their direction in the rest frame, neglecting molecules overtaken; this approximation is good for small f). Since mean square velocity is proportional to temperature, Eq. (7.54) can be converted to velocity terms, yielding the condition

$$\left(v_{1, \text{rms}}^2 - \frac{kT_s}{m} \right) \alpha = v_{1, \text{rms}}^2 - v_{2, \text{rms}}^2 \quad (7.56)$$

where kT/m is the mean square thermal speed, and hence

$$R_{\text{temp}} = \frac{T_g}{T_s} = \left[1 + \frac{f^2}{4} - f \sqrt{\frac{2}{\pi}} \left(\frac{2}{\alpha} - 1 \right) \right]^{-1} \quad (7.57)$$

A useful approximation, good for small f and moderate to large α , is

$$R_{\text{temp}} \approx 1 + f \sqrt{\frac{2}{\pi}} \left(\frac{2}{\alpha} - 1 \right) \quad (7.58)$$

7.5.1.3. Energy losses

The cases of greatest interest in the present context are those in which $f \ll 1$, and $T_g \approx T_s$. The work done in isothermally compressing a freely-moving particle from a range of motion ℓ_1 to a range ℓ_2 is

$$W = - \int_{\ell_1}^{\ell_2} \frac{kT}{\ell} d\ell = kT \ln \frac{\ell_1}{\ell_2} \quad (7.59)$$

and for a system undergoing compression at a uniform speed, resulting in a constant value of ΔT_{comp} , the free energy lost (with the above approximations) is

$$\begin{aligned} \Delta W &\approx k\Delta T_{\text{comp}} \ln \frac{\ell_1}{\ell_2} = kT_s f \sqrt{\frac{2}{\pi}} \left(\frac{2}{\alpha} - 1 \right) \ln \frac{\ell_1}{\ell_2} \\ &\approx \sqrt{\frac{2mkT_s}{\pi}} \left(\frac{2}{\alpha} - 1 \right) \ln \frac{\ell_1}{\ell_2} v_{\text{total}} \end{aligned} \quad (7.60)$$

where v_{total} is the speed of one piston with respect to the other. For a sliding component with $m = 2 \times 10^{-25}$ kg, a compression ratio of 10, $\alpha = 0.5$, and $T_s = 300$ K, the energy lost is $\sim 1.6 \times 10^{-22}$ J at 1 m/s, and $\sim 1.6 \times 10^{-24}$ J at 1 cm/s. So long as f remains small, the above expressions are equally applicable to nonisothermal expansion losses.

7.5.1.4. Large molecules

For moving components substantially more massive than ordinary gas molecules, however the literature values of α for the latter offer only a poor guide. As molecular motions become slow, collisions become more nearly elastic and energy transfer decreases, but as molecules become large, the slowing of their *free* motion is offset by the effect of their increased van der Waals attraction energy (Goodman 1980): the final approach to a surface is accelerated, and the loss of a portion of the resulting increment in kinetic energy can result in a negative total energy relative to the free state. This results in adsorption, complete thermal accommodation and (in the present context) elimination of further nonisothermal-compression losses until the pistons press the molecule from both sides. (Energy losses resulting from the fall into the van der Waals potential well can be described within the framework of Section 7.6.)

7.5.2. Harmonic well compression

Nanomechanical systems will sometimes contain components confined to approximately-harmonic wells of time-varying stiffness. This is, for example, a reasonable description of the final compression of a single-molecule gas when it is subject to repulsive forces from both pistons. Compression corresponds to an increase in k_s , reducing the effective volume (Section 6.2.2.2) available to the oscillator. Thermal exchange with the medium in these instances can be modeled as acoustic radiation from a harmonic oscillator with an energy equaling the excess thermal energy; near equilibrium, the same coefficients will describe the absorption of energy by an oscillator undergoing expansion.

7.5.2.1. A harmonic-well temperature-increment model

Assume that the compression process is slow compared to the vibrational period, and that the temperature increment ΔT_{comp} is small compared to the equilibrium temperature T . The total work done in compressing the system by increasing the stiffness from $k_{s,1}$ to $k_{s,2}$ is

$$W = kT \ln \sqrt{\frac{k_{s,2}}{k_{s,1}}} \quad (7.61)$$

Equating the derivative of W with respect to k_s to the net radiated acoustic power Eq. (7.8) associated with ΔT_{comp} yields the expression

$$\Delta T_{\text{comp}} = \frac{2\pi m M^{3/2}}{\sqrt{\rho}} T \frac{\partial k_s}{\partial t} k_s^{-3} \quad (7.62)$$

where the constants are as defined in Section 7.2.2.

7.5.2.1. Energy dissipation models

The energy dissipated is the integral of the difference in work resulting from ΔT_{comp} , or

$$\Delta W = \int_{k_{s,1}}^{k_{s,2}} \frac{k \Delta T_{\text{comp}}}{2k_s} dk_s = \frac{\pi m M^{3/2}}{3\sqrt{\rho}} k T \frac{\partial k_s}{\partial t} (k_{s,1}^{-3} - k_{s,2}^{-3}) \quad (7.63)$$

assuming that the stiffness increases linearly with time. In systems where k_s results from nonbonded repulsions, Eq. (3.19) implies that $k_s \approx 3.5 \times 10^{10} F_{\text{load}}$ (N/m); a roughly linear increase of F_{load} will not be uncommon.

The value of this expression is strongly sensitive to the value of $k_{s,1}$, the stiffness at the onset of compression. This will frequently be on the order of the stiffness of an unloaded nonbonded contact between two objects. A model for this, in turn, is the contact between two planes in solid graphite. With an interlayer spacing of 0.335 nm and a modulus of 1.0×10^{10} N/m² (Kelly 1973), the stiffness of this contact is $\sim 3 \times 10^{19}$ N/m³, or 30 N/m per nm². The contact area between a blocky component and a surface will typically be of the order of

$$S \approx (m/\rho_c)^{2/3} \quad (7.64)$$

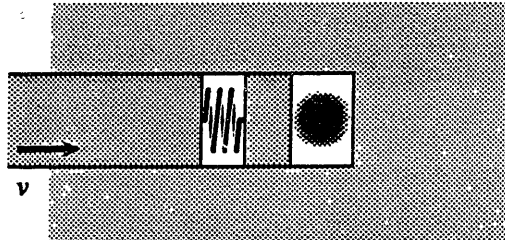


Figure 7.8. Model of compression by an elastic system.

and hence

$$k_{s,1} \approx 3 \times 10^{19} (m/\rho_c)^{2/3} \quad (7.65)$$

In a typical situation (Fig. 7.8), a system with a finite stiffness external to the interfaces under consideration, $k_{s,\text{ext}}$, will be loaded by a steady displacement at a rate v_{ext} . Using Eq. (7.63) and assuming substantial compression ratios,

$$\Delta W \approx kT \frac{\pi m M^{3/2}}{3\sqrt{\rho}} 3.5 \times 10^{10} \frac{k_{s,\text{ext}} v_{\text{ext}}}{k_{s,1}^3} \quad (7.66)$$

where the approximation assumes $k_{s,\text{ext}}$ to be substantially less than $k_{s,1}$. Assuming $T = 300$ K, $m = 10^{-24}$ kg, $M = 5 \times 10^{11}$ N/m², $\rho = \rho_c = 2000$ kg/m³, $k_{s,\text{ext}} = 10$ N/m, and $k_{s,1} = 19$ N/m, $\Delta W \approx 2 \times 10^{-21}$ J at $v_{\text{ext}} = 1$ m/s, and $\approx 2 \times 10^{-23}$ J at 1 cm/s.

These estimates indicate magnitudes of energy dissipation large relative to those identified thus far. Of the parameters subject to design control, the most significant is M ; the value assumed above is for a diamond-like material, which strongly suppresses radiative coupling and increases ΔT . In the immediate neighborhood of a component undergoing fast compression, a better choice would be a structure with properties like those of an engineering polymer, $M \approx 3 \times 10^9$ N/m², $\rho \approx 1000$ kg/m³. This decreases ΔW by a factor $\approx 7 \times 10^{-4}$, yielding $\approx 1 \times 10^{-24}$ J at $v = 1$ m/s, and $\approx 1 \times 10^{-26}$ J at 1 cm/s.

It would be of interest to test the predictions of the above models, and those in Section 7.5.1, against direct molecular dynamics simulations. Studies along these lines could reduce uncertainties regarding accommodation coefficients and velocity distributions in the square-well regime and could probe energy losses in the transition to a harmonic well compression regime. For this purpose, the routines for integrating the equations of motion must conserve energy with considerable accuracy.

7.5.3. Multidimensional systems

Sections 7.5.1 and 7.5.2 assume one-dimensional motion of the compressed component. Real systems will have two additional degrees of freedom in translation and further degrees of freedom in torsion and internal vibration. In general, excitation of a single translational degree of freedom by nonisothermal compression will lead to rapid equilibration with the other degrees of freedom, approaching equipartition of energy. Each of these degrees of freedom will open new channels for equilibration with the environment, reducing ΔT_{comp} ; typically, some will have higher frequencies and more effective radia-

tive coupling than the single mode considered above. Consequently, energy dissipation will be lower than that implied by a one-dimensional model.

7.5.3.1. Multidimensional square-wells

In a simple gas-in-cylinder model, the one-dimensional approximation is equivalent to assuming that the walls of the cylinder are adiabatic, permitting energy loss only to the pistons. A realistic, multidimensional model would include equilibration through collisions with the walls.

7.5.3.2. Multidimensional harmonic wells

In a realistic, approximately-harmonic well, longitudinal vibrations will (at a minimum) be coupled to transverse vibrations, providing two additional vibrational modes for equilibration through acoustic radiation. Further, vibrational frequencies in these transverse modes can often be made substantially higher than those in the longitudinal mode: a component can be tightly clamped against transverse motion while performing a function requiring relatively low longitudinal stiffness. This can greatly reduce ΔT_{comp} .

7.5.3.3. Mixed systems

A component may be subject to strong harmonic constraints against transverse motion, yet move in an approximately square-well potential in longitudinal motion. Again the modes will be coupled, and again the transverse modes will provide relatively effective coupling to the heat bath of the surrounding medium.

7.6. Transitions among time-dependent wells

7.6.1. Overview

Each of the energy dissipation mechanisms discussed in Sections 7.2–7.5 approaches zero energy per operation (or per unit displacement) as the speed of the system approaches zero. The merging of potential wells, in contrast, will usually dissipate energy no matter how slowly it is performed. Since well-merging can dissipate on the order of kT or more per cycle, it can impose a significant energy cost on systems operation.

7.6.2. Energy dissipation in merging wells

Energy dissipation in merging wells occurs when transitions between states move a disequilibrium system closer to equilibrium: the dissipation is equal to the associated

reduction in free energy. In analyzing time-dependent wells, we begin by describing wells as distinct states linked by cols which are divided by transition surfaces. For concreteness, we will consider the case of systems with two wells linked by a single col, using the harmonic approximations made in Section 6.2.

In the cases of interest, transition rates are low compared to vibration times, and the partition function of each state can be evaluated separately under the assumption of internal equilibrium. Losses owing to nonisothermal compression within a well can be estimated by adapting the methods of Section 7.5.2; this mechanism can be treated as independent given either that the transition rate is negligible, or that the temperature rise has little effect on the transition rate. The latter condition is satisfied if

$$\frac{\Delta T_{\text{comp}}}{T} \ll \frac{kT}{\Delta \mathcal{V}^\ddagger} \quad (7.67)$$

7.6.2.1. Free energy

The free energy for a system of two wells can conveniently be written as the sum of the free energy of each well (assuming unit probability of occupancy), weighted by the actual probabilities of occupancy, plus an entropic term reflecting the uncertainty regarding which well is occupied:

$$\mathcal{F}_{1,2} = P_1 \mathcal{F}_1 + (1 - P_1) \mathcal{F}_2 + kT [P_1 \ln(P_1) + (1 - P_1) \ln(1 - P_1)] \quad (7.68)$$

or, equivalently,

$$\mathcal{F}_{1,2} = -kT \left[P_1 \ln \frac{q_1}{P_1} + (1 - P_1) \ln \frac{q_2}{1 - P_1} \right] \quad (7.69)$$

where q_1 and q_2 are the partition functions of the wells considered as isolated, occupied systems.

7.6.2.2. Free-energy changes

For a fixed value of P_1 , the free energy of the two-well system, $\mathcal{F}_{1,2}$, depends on the shape of the PES. This changes over time, but so long as equilibrium is maintained, $\Delta \mathcal{F}_{1,2}$ exactly equals the work done on the system by whatever mechanisms are responsible for the change in the PES. The total free energy is then conserved, and no dissipation results.

For a fixed PES, a finite change ΔP_1 results in a change in free energy without work being done by (or on) the external mechanisms. Accordingly, $\Delta \mathcal{F}_{1,2}$ can only be negative and corresponds to an increment of energy dissipation. The rate of decrease of free energy, holding the PES constant, is thus

$$\left(\frac{\partial \mathcal{F}_{1,2}}{\partial t} \right)_{\text{PES}} = \frac{\partial P_1}{\partial t} \frac{\partial \mathcal{F}_{1,2}}{\partial P_1} = -R_{12} \left(\mathcal{F}_1 - \mathcal{F}_2 + kT \ln \frac{P_1}{1-P_1} \right) = -R_{12} kT \ln \frac{P_1 q_2}{(1-P_1) q_1} \quad (7.70)$$

Note that at equilibrium

$$\mathcal{F}_1 - \mathcal{F}_2 = -kT \ln \frac{P_{1,\text{eq}}}{1-P_{1,\text{eq}}} \quad (7.71)$$

hence

$$P_{1,\text{equil}} = \exp \left(-\frac{\mathcal{F}_1 - \mathcal{F}_2}{kT} \right) \left[\exp \left(-\frac{\mathcal{F}_1 - \mathcal{F}_2}{kT} \right) + 1 \right]^{-1} \quad (7.72)$$

and the flow of probability from well to well occurs without dissipation.

The net rate of transitions from 1 to 2, R_{12} can be expressed in terms of the unidirectional transitions rates k_{12} and k_{21} calculated from an appropriate version of transition state theory:

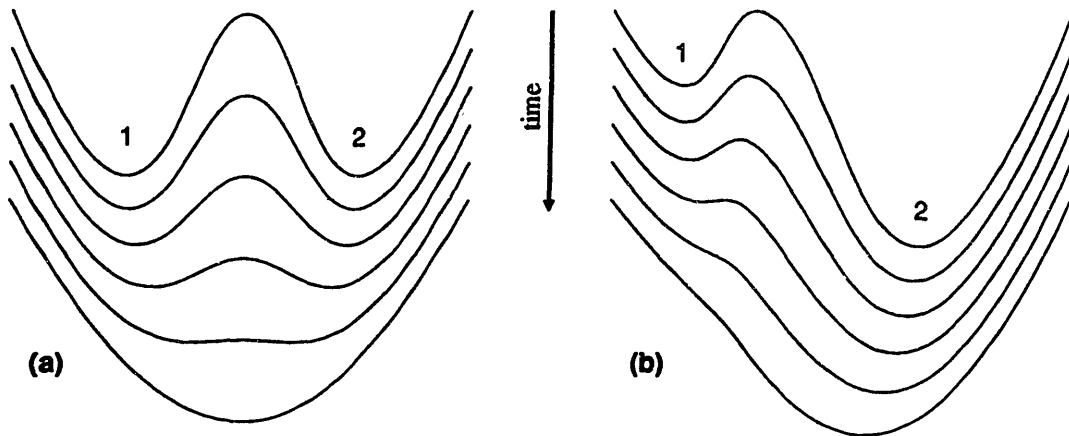


Figure 7.9. Merging of symmetrical (a) and asymmetrical (b) potential wells, showing potentials at a series of times.

$$R_{12} = -\frac{\partial P_1}{\partial t} = P_1 k_{12} - (1 - P_1) k_{21} \quad (7.73)$$

The total energy dissipation in a process between times t_1 and t_2 is

$$\Delta \mathcal{F}_{\text{diss}} = -\int_{t_1}^{t_2} R_{12}(t) \left(\mathcal{F}_1(t) - \mathcal{F}_2(t) + kT \ln \frac{P_1(t)}{1 - P_1(t)} \right) dt, \quad P_1(t) = P_{1,\text{init}} - \int_{t_1}^t R_{12}(\tau) d\tau \quad (7.74)$$

7.6.2.3. Switched-coupling models

As in calculations of error rates, one useful approximation treats the transition rate as switching between rapid and negligible. As two wells merge and the barrier between them shrinks, energy dissipation will typically be low when transition rates are negligible, and low again when rates are high enough to ensure the maintenance of near-equilibrium values of P_1 ; between these regimes will be a period in which transition rates first become significant and then establish and maintain near-equilibrium. The switched coupling model approximates that period as a single time, t_{cpl} , hence

$$\Delta \mathcal{F}_{\text{diss}} = \mathcal{F}_{1,2}(t_{\text{cpl}}, P_{1,\text{equil}}) - \mathcal{F}_{1,2}(t_{\text{cpl}}, P_{1,\text{init}}) \quad (7.75)$$

where the equilibrium probability $P_{1,\text{equil}}$ is given by Eq. (7.72) with values of \mathcal{F}_1 and \mathcal{F}_2 evaluated at t_{equil} . A value of t_{cpl} can always be chosen such that Eq. (7.75) yields the same value as Eq. (7.74); in practice, simple rules for choosing t_{cpl} will yield approximations to this value.

7.6.3. Free expansion and symmetrical well merging

A simple and important case is the symmetrical merging of two identical wells (e.g., Fig. 7.9(a)), where one is initially occupied with $P_1 = 1$. At all times, $\mathcal{F}_1 = \mathcal{F}_2$, the dissipation resulting from transitions is purely entropic, and the switched-coupling model yields

$$\Delta \mathcal{F}_{\text{diss}} = -kT \ln 2 \quad (7.76)$$

which is simply $-TS$, where S is the entropy increase resulting from a volume-doubling free expansion of a single-molecule gas; this equals the work required to restore the initial condition by isothermally compressing a single-molecule gas to half its initial vol-

ume. In computational systems, the symmetrical merging of two wells corresponds to the erasure of one bit of information; information erasure is the only abstract computational process that *of necessity* results in an increase in entropy and a corresponding loss of free energy (Bennett 1982; Landauer 1988).

7.6.4. Asymmetrical well merging

Simple approximations suffice when merging wells are sufficiently different in free energy, as in Fig. 7.9(b). If $(\mathcal{F}_1 - \mathcal{F}_2) \gg kT$ during the time of equilibration, and $P_{1,\text{init}} = 1$, then

$$\Delta\mathcal{F}_{\text{diss}} \approx \mathcal{F}_2(t_{\text{equil}}) - \mathcal{F}_1(t_{\text{equil}}) \quad (7.77)$$

In this case, equilibration changes P_1 from 1 to $\sim \exp[-(\mathcal{F}_1 - \mathcal{F}_2)/kT] \approx 0$, hence the entropy of uncertain well-occupancy is almost unchanged. If the difference in free energy between the wells chiefly results from a difference in potential energy (rather than within-well entropy), then the system can be viewed as falling into a deeper well, with subsequent thermalization of the energy of the fall. Processes of this sort will occur, for example, when an elementary step in a chemical reaction is exothermic, or when a mechanical process places an elastic component under load, then allows it to slip past an obstacle. $\Delta\mathcal{F}_{\text{diss}}$ can be large compared to kT .

If $(\mathcal{F}_1 - \mathcal{F}_2) \gg kT$ during the time of equilibration, and $P_{1,\text{init}} = 0$, then

$$\Delta\mathcal{F}_{\text{diss}} \approx -kT \exp\left(-\frac{\mathcal{F}_1 - \mathcal{F}_2}{kT}\right) \quad (7.78)$$

In this case, $\Delta\mathcal{F}_{\text{diss}} \ll kT$, and is dominated by the entropy of free expansion into the small effective volume of the upper well.

7.6.5. Optimal well merging under uncertainty

The value of $\Delta\mathcal{F}_{12} = \mathcal{F}_1 - \mathcal{F}_2$ for a pair of merging wells will often be subject to design control, since well depths can be modulated by a variety of influences. If so, and if the value of P_1 is known, then the energy dissipation resulting from well merging can be kept near zero by ensuring that $\Delta\mathcal{F}_{12}$ nearly satisfies Eq. (7.72) at $t = t_{\text{cpl}}$ in the switched coupling model, or during the critical period surrounding t_{cpl} in the model described by Eq. (7.74). Note that in a cyclic process where $P_1 \neq 0$ or 1, the associated uncertainty indicates that some other step in the cycle has increased the entropy and dissipated a cor-

responding increment of free energy; optimal well merging merely avoids imposing an additional loss.

7.7. Conclusion

A variety of energy dissipation mechanisms operate in nanomechanical systems, each strongly influenced by design parameters. Most have a magnitude per operation or per unit displacement that is proportional to the speed of the system; merging of potential wells (including information erasure in computational systems) is the chief exception, since the energy per operation is typically speed-independent, or nearly so. With the exception of the last, each of these mechanisms results in energy dissipation that is small compared to kT for nanoscale systems with reasonable choices of physical parameters (one criterion of a reasonable choice, of course, is that it result in acceptable energy dissipation). The approximations and bounds described and developed in this chapter will be applied and occasionally elaborated in the chapters of Part II.

Chapter 8

Mechanosynthesis

8.1. Overview

Mechanosynthesis and nanomechanical design are interdependent subjects. Advanced mechanosynthesis will employ advanced nanomachines, but advanced nanomachines will themselves be products of advanced mechanosynthesis. This circular relationship must be broken both in exposition and (eventually) in technology development. In development, the circle can be broken by using either conventional synthesis or non-eutaxic mechanosynthesis to construct first-generation nanomachines, as discussed in Chapter 13. In the following exposition, the circle is broken as follows:

(1) This chapter shows that a wide range of diamondoid structures can be made, provided that accurately-controlled mechanical motions can somehow be imposed as a boundary condition in molecular systems;

(2) Part II then shows how diamondoid structures can serve as components of nanomachines able to power and guide accurately-controlled mechanical motions, including motions of the sort assumed in this chapter.

8.1.1. Mechanochemistry: terms and concepts

The term “mechanochemistry,” coined by Ostwald to describe “a branch of chemistry dealing with the chemical and physico-chemical changes of substances of all states of aggregation due to the influence of mechanical energy” (Heinicke 1984); it has more recently been used to describe both the conversion of mechanical energy to chemical energy in polymers (Parker 1984) and the conversion of chemical energy to mechanical

energy in biological molecular motors; proponents of the first use (Casale and Porter 1978) urge that processes of the second sort be termed “chemomechanical” in accord with the direction of energy flow. As used here, “mechanochemistry” refers to processes in which mechanical motions *control* chemical reactions, typically providing activation energy or otherwise modulating the reaction conditions; this is consistent with Ostwald’s broad definition. Since mechanochemical systems can approach thermodynamic reversibility (Sec. 8.5), the direction of conversion between chemical and mechanical energy cannot be a defining characteristic. The present usage is thus related to but substantially distinct from those just cited. The term “mechanosynthesis” is here introduced to describe mechanochemical processes that perform steps in the construction a complex molecular product.

Mechanochemistry, in present sense, has features in common with reactions in organic crystals (McBride, Segmuller et al. 1986) and in enzymatic active sites (Creighton 1984). In both crystals and enzymes, the environment surrounding a molecule can induce substantial stresses on it, or on a reaction intermediate, and can constrain reaction pathways so as to produce a specific product. In crystals, however, the effect of these forces is typically to limit motion (McBride, Segmuller et al. 1986). Mechanochemical systems of the sort contemplated here, in contrast, will characteristically transport reactive moieties over substantial distances, induce substantial displacements in the course of a reaction, and supplement local thermal and molecular potential energies with work done by external devices.

Mechanochemistry will play several roles in the systems described in Part II. These include the transformation of small molecules, generation of mechanical power, preparation of bound reactive moieties, and mechanosynthesis of complex structures using such moieties. Although the discussion in the present chapter centers on the latter two roles, the principles described are of general applicability.

8.1.2. Scope and approach

Eutaxic mechanochemistry, like solution-phase chemistry, will embrace an enormous range of possible reactions. This potential breadth places the field beyond any hope of thorough exploration, and beyond any possibility of summarizing in the space of a chapter. Nonetheless, an attempt must be made to delineate the general capabilities and limitations of mechanochemical processes. How is this problem to be approached? In the absence of opportunities for immediate experimentation, two options suggest themselves:

The first approach relies on detailed theoretical modeling of mechanochemical processes. This approach yields quantitative descriptions of the sort familiar to those schooled in physics and engineering; several examples of this sort will be presented in Section 8.5. Theoretical modeling suffers from several difficulties and limitations, however. Chemical reactions cannot be modeled using a molecular mechanics PES, but require either a specialized PES (e.g., Sections 8.5.3, 8.5.4) or the use of general quantum mechanical methods to calculate the energies of a set of molecular configurations. The latter approach is typically either expensive or inaccurate, save for small structures (note, however, that the reduced sensitivity of mechanochemical processes to small differences in energy reduces accuracy requirements). Computational quantum chemistry can eventually provide an extensive and sufficiently-accurate picture of the capabilities of mechano-synthesis, but it cannot now provide a rapid overview of those capabilities.

The second approach begins with a survey of the known capabilities of solution-phase organic synthesis, then examines how the conditions of mechanochemistry both add capabilities and impose constraints relative to this model. The resulting picture of the capabilities of mechanosynthesis will thus be linked to experimental results in chemistry, rather than being directly founded on theory. This picture will contain fewer quantitative arguments than a physicist or an engineer may expect. However, if we think that chemists know what they are doing (as their achievements amply demonstrate), then we have reason to assume that their methodologies and modes of explanation are well-suited to their subject matter. Organic synthesis is a notoriously qualitative field (it is “very little dependent on mathematical calculations,” (Hendrickson 1990)); much of the discussion in Sections 8.3 and 8.4, taking organic synthesis as its point of departure, accordingly represents a substantial departure from the style of exposition in the surrounding chapters. Like much of chemistry, it relies heavily on argument by example, reasoning from model systems and experimental results, supplemented by the general principles of chemical kinetics and thermodynamics.

The following sections approach this topic from several directions: Section 8.2 provides a brief overview of the nature and achievements of solution-phase organic synthesis, offering a perspective from which to judge the expected capabilities of mechano-synthesis. Section 8.3 compares solution-phase synthesis and mechanosynthesis, examining the limitations and strengths of the latter with respect to the former. This section develops quantitative criteria for reaction reliability in light of the anticipated requirements of molecular manufacturing, and provides a preliminary discussion of the role of

mechanical force in chemical reactions. Section 8.4 provides an overview of several classes of reactive species from a mechanosynthetic perspective; Section 8.5 then examines a variety of mechanochemical processes in more detail, focusing in particular on the exemplary processes of tensile bond cleavage and hydrogen abstraction. Finally, Section 8.6 presents several approaches to the mechanosynthesis of diamondoid structures, building both on the preceding sections and on results related to current techniques for low-pressure diamond synthesis.

8.2. Perspectives on solution-phase organic synthesis

8.2.1. The scale and scope of chemistry

The advent of organic synthesis is commonly dated from the 1828 synthesis of urea by Wöhler, a time when “chemists were just beginning to speculate about the arrangement of atoms within organic molecules” (Brooke 1985). By the late 1800s, the basic concepts of atoms and molecular geometry were well established, synthesis was subject to rational planning, and synthetic organic chemicals had become a major industry.

One year’s output of *Chemical Abstracts* now occupies ~ 2 m of shelf space, containing abstracts of ~ 5×10^5 articles; the volumes published to date occupy ~ 30 m of shelf space, with the indexes occupying ~ 30 m more. After 110 years of publication, *Beilsteins Handbuch der Organischen Chemie* occupies ~ 20 m of shelf space with over 300 volumes. The chemists’ view of the current capabilities of organic synthesis is suggested by the observation that “The problem of synthesis design is simply that the number of possible synthetic routes to any target molecule of interest is enormous” (Hendrickson 1990) (this author estimates tens of millions of paths for a typical five-step synthesis, with vastly differing yields), and by the barely-cautious warning that “The idea persists that we can synthesize anything. The major flaw in this view is that it fails to recognize effectiveness and practicality” (Trost 1985).

During the first century of organic synthesis, chemists proceeded with no guidance from quantum mechanics, relying instead on a growing set of concepts regarding molecular geometry and thermodynamics, together with rules of thumb for estimating the energy differences between similar molecular configurations and the rate differences between similar reactions. These concepts and rules grew from efforts to rationalize experimental results; they are still being extended and remain central to organic synthesis to this day. Many reactions are known to yield particular results; somewhat fewer can be described in

mechanistic detail (i.e., which atoms go where as the system passes through its transition state); far fewer have been modeled at the level of potential energy surfaces and molecular dynamics. Detailed theory thus describes only a fraction of the synthetic operations available to the modern chemist.

8.2.2. The prominence of qualitative results in organic synthesis

Organic synthesis is fundamentally qualitative, in much the same sense that topology (as distinct from, say, analytic geometry) is fundamentally qualitative. In synthesis, the chemist seeks to construct molecules with a particular pattern of bonding that can be described (give or take discrete, stereochemical differences) in topological terms. The discovery that substances of type 1 can be transformed into substances of type 2 is inherently qualitative and often of great value. The development of such useful species and processes as Grignard reagents, Wittig reactions, and Sharpless epoxidations are examples of contributions of this kind.

In describing an organic synthesis, numbers are used to specify temperatures, quantities of material, and characteristics such as melting points, refractive indexes, and so forth, but these numbers play a peripheral role. One can frequently examine several chapters of a monograph on chemistry while encountering no numbers save for dimensionless integers, such as those used in specifying molecular structures and charge states. To be useful in experimental work, most measurements need only discriminate among different molecular species, and accurate discrimination need not require great precision. Even theoretical contributions of great fame and value often take the form of non-quantitative, geometric rules (some of which are subject to violation); examples include Markovnikoff's rule for the regioselectivity of addition to alkenes, the Alder rule for the production of isomers in Diels–Alder reactions, and the Woodward–Hoffmann rules regarding orbital symmetry in cycloaddition processes.

Returning to the picture of potential energy surfaces in configuration space, the discovery of a new kind of reaction corresponds to the discovery of a set of similar cols between potential wells corresponding to molecular species. In standard chemistry, useful cols must be low enough to be thermally accessible at moderate temperatures *and* other cols must be relatively inaccessible (or harmless to traverse). The quantitative results of greatest interest in organic synthesis relate to the height and width of the cols (which determine reaction kinetics) and the height and width of the potential wells (which determine reaction equilibria).

8.2.3. A survey of synthetic achievements

No simple, first-principles argument leads from elementary facts regarding molecular physics to conclusions regarding the range of structures that can be made by diffusive, solution-phase chemical processes. Accordingly, one must examine the actual capabilities that chemists have developed. These include the synthesis of:

- Highly strained molecules, such as cyclopropane rings **8.1**, with 60° bond angles, rather than the optimal 109.5° ; cubane **8.2**, with a total strain energy of ~ 1.15 aJ (Eaton and Castaldi 1985), about twice the typical C–C bond energy; and [1.1.1]-propellane **8.3** (Wiberg and Walker 1982), in which the bonding configuration at two of the carbon atoms is inverted relative to the normal tetrahedral geometry.



8.1

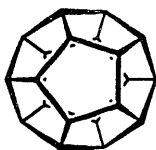


8.2

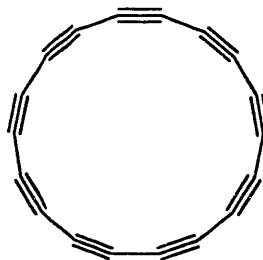


8.3

- Highly symmetrical molecules, such as dodecahedrane **8.4** (Ternansky, Balogh et al. 1982) and cyclo[18]carbon **8.5** (Diederich, Rubin et al. 1989).

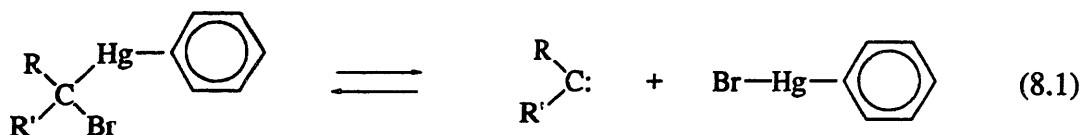


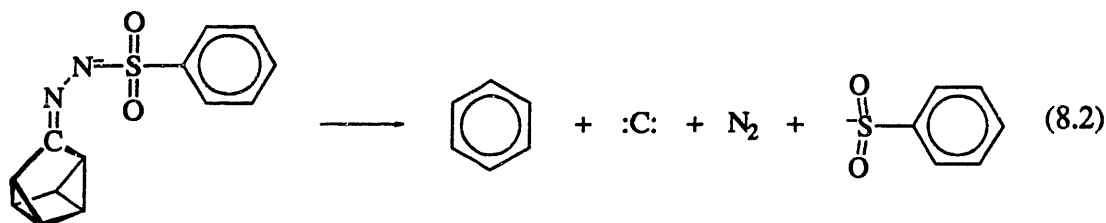
8.4



8.5

- Molecules that spontaneously decompose to free highly-reactive species, such as carbenes (species containing divalent carbon), Eq. (8.1) and even free carbon atoms (Shevlin and Wolf 1970), Eq. (8.2). (These species ordinarily exist only briefly before reacting.)





- Over six hundred organic structures, containing a few tens to a few hundred atoms, identical to molecules synthesized by molecular machinery in living systems (Corey and Cheng 1989).

- Chains of a hundred or more monomers, joined in a precise sequence to build biologically-active protein and DNA molecules containing on the order of 1000 precisely-bonded atoms (Caruthers 1985; Kent 1988).

- Molecules designed to self-assemble into larger structures (Cram 1988; Lehn 1988; Rebek 1987), and even to bind other molecules temporarily to catalyze the formation of copies of themselves (Tjivikua, Ballester et al. 1990).

8.3. Solution-phase synthesis and mechanosynthesis

8.3.1. Analytical approach

This section uses the demonstrated mechanisms and capabilities of solution-phase organic synthesis as a basis for understanding the capabilities of mechanosynthesis. It proceeds by examining differences:

- The solution-phase conditions and unit operations *excluded* by mechanosynthetic constraints.
- The new conditions and unit operations *provided* by mechanosynthesis.
- The additional system-level capabilities implied by the above differences.
- The additional constraints that must be satisfied to exploit these system-level capabilities.

Taken together with the known capabilities of diffusive organic synthesis, this set of differences gives substantial insight into the capabilities of mechanosynthesis.

8.3.1.1. A terminological note: exothermic vs. exoergic

In solution-phase chemistry, when molecular potential energy is transformed into mechanical energy, this promptly appears as heat. Reactions that reduce potential energy are accordingly termed “exothermic”; those that increase it, “endothermic.” In mechanochemical systems, however, potential energy released by a reaction can often be stored elsewhere in the mechanical system, either as potential energy or as kinetic energy in the form of orderly motion. Accordingly, the more general terms “exoergic” and “endoergic” are appropriate. It is generally appropriate to focus on changes in free energy, but inasmuch as entropy will seldom change greatly during the reactions under consideration, free energy differences and potential energy differences can be roughly equated.

8.3.1.2. Mechanochemical parameters

A repeated theme in the following sections will be reaction rates and reaction reliability in mechanochemical systems. Useful variables will include the actuation time t_{act} (s), during which motions carry the system toward and then away from a reactive geometry; the transformation time t_{trans} (s), during which a reactive geometry exists and a reaction may occur; the reaction rate k_{react} (s^{-1}), which measures the probability per unit time of a reactive transformation during t_{trans} , given that a reaction has not yet occurred, the intersystem crossing rate k_{isc} , which measures the rate of the electronic intersystem crossing “reaction”; and the error rate k_{err} , which measures the rate of unwanted reactions. These parameters will be related to the probability of an error in a mechanochemical operation, P_{err} . It will generally be assumed that $t_{\text{act}} \approx 10^{-6}$ s, $t_{\text{trans}} \approx 10^{-7}$ s, and that each contribution to P_{err} is to be $\leq 10^{-12}$ (see Sections 8.3.4.3–8.3.4.6).

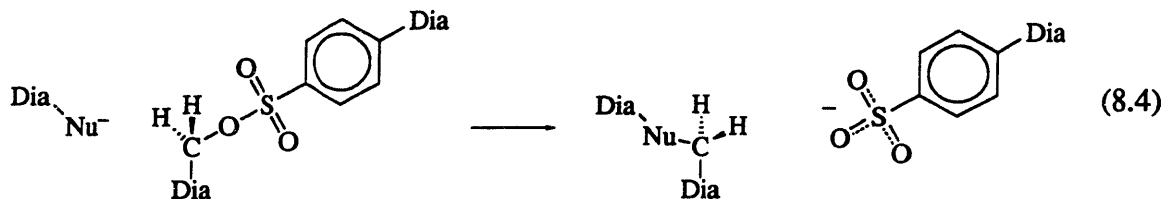
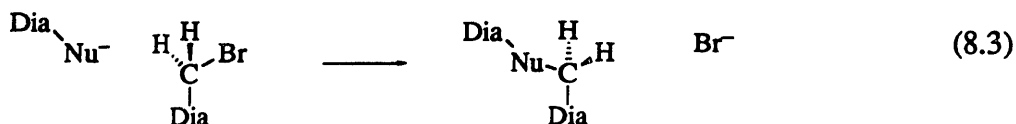
8.3.2. Basic constraints imposed by mechanosynthesis

8.3.2.1. Loss of natural parallelism

Solution-phase synthesis typically yields macroscopic quantities of product. In replacing a macroscopic volume of reagent solution with a single device, the number of reacting entities is reduced by a factor on the rough order of 10^{-23} . To compensate for this difference will require the use of large numbers of mechanosynthetic devices, each operating at a high frequency. Some consequences of this constraint are discussed in Section 8.3.4.3.

8.3.2.2. Limitations on reagents and products

Mechanical control of reactions will typically require that reagent moieties be bound to extended “handle” structures; control of products will typically require that they, too, be bound, excluding the use or production of small, freely-moving species. This constraint is less severe than it might seem, however, because free and bound species can often participate in similar reactions. For example, although an S_N2 reaction producing a free ion, Eq. (8.3), would be excluded, (8.4) would be permissible.



In reality, an ionic process like (8.3) would require a favorable electrostatic environment, and that environment could be used to bind the product ion. In water, small ions (from Li^+ to I^-) are found in solvent cages, and are bound with energies ≥ 330 mJ relative to vacuum (Bockris and Reddy 1970a). In condensed-phase acid reactions “ H^+ ” is always bound. Reactions characteristic of small free radicals (e.g., H , F , OH) can frequently be made to proceed by concerted mechanisms in which the radicals are never free. Binding of such species typically lessens reactivity, but Section 8.3.3 describes several effective techniques for speeding reactions in mechanochemical systems. Purely steric confinement will sometimes be feasible, and will have little effect on the electronic structure of the reagent.

The requirement for handles (of any kind) places significant steric constraints on feasible reactions. Small, diffusing species will be able to reach many sites (e.g., deep in pockets on a surface) that will be inaccessible to a mechanically-positioned reagent moiety. In planning the synthesis of an object, geometric constraints will thus be of increased importance. Finally, for reaction processes to be reliable, reagents must have adequate stability against unimolecular decomposition reactions (Sec. 8.3.4.6).

8.3.2.3. Lack of true solvation

Fully eutaxic systems must lack true solvation, since solvents are not eutaxic. As discussed in Section 8.3.3.2, however, solvent-like eutaxic environments will typically be superior to true solvents in promoting reaction speed and specificity. Mechanical transport can likewise substitute for diffusive transport.

8.3.2.4. Inutility of true photochemistry

Strictly speaking, photochemical reactions require photons, but the relevant wavelengths are measured in hundreds of nanometers, and a cubic-wavelength block accordingly contains millions of cubic nanometers. This mismatch relative to the molecular scale will limit the utility of photochemistry in mechanosynthesis (but note Sec. 8.3.3.4).

8.3.2.5. Limitation to moderate temperatures

Mechanosynthetic systems will be sensitive to damage from thermomechanical degradation (Sec. 6.4); the single-point failure assumption (Sec. 6.7) is adopted here. With careful design, reliable operation at substantially elevated temperatures may be feasible. Nonetheless, temperatures like those used in melt-processing of typical ceramics are almost surely infeasible. The present work will assume that temperatures are limited to near 300 K, thereby eliminating from consideration a large class of reactions that require elevated temperatures to proceed at an acceptable rate. This constraint could be partially evaded by exploiting highly localized, non-thermal kinetic energy sources to provide activation energy (e.g., impacts driven by stored elastic energy), but this approach will not be pursued here.

8.3.3. Basic capabilities provided by mechanosynthesis

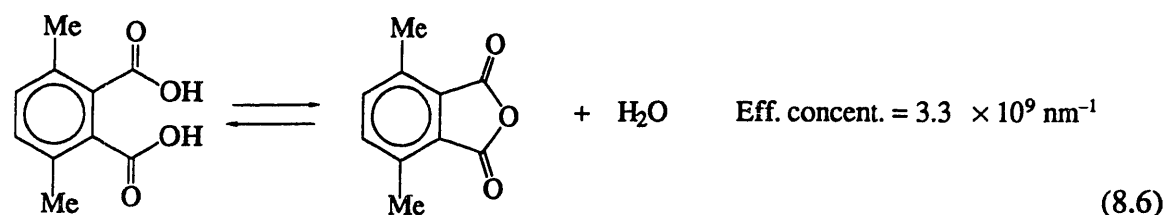
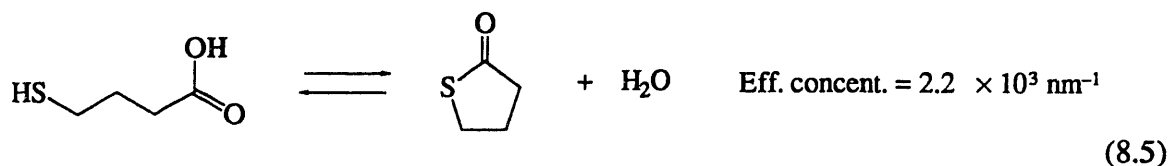
8.3.3.1. Large effective concentrations

In a diffusive reaction $A + B \rightarrow C$, the reaction frequency of molecules of type A is (by definition) directly proportional to the *effective concentration* of molecules of type B. In a typical solution system, the effective concentration equals the actual concentration, often stated in bulk chemistry in moles per liter (M), but here described by a molecular number density (nm^{-3}). The concentration of water molecules in liquid water is $\sim 55 \text{ M}$, or $\sim 33 \text{ nm}^{-3}$.

Mechanochemical processes strongly resemble unimolecular processes. Unimolecular reactions, in turn, are often analogous to bimolecular reactions, save for the pres-

ence of a covalent connection between the reacting moieties. Given the unimolecular reaction frequency and the rate constant for an analogous bimolecular reaction, one can compute the effective concentration of a moiety B with respect to a moiety A in the unimolecular process by computing the (hypothetical) concentration of B that would subject A to the same reaction frequency in the bimolecular process.

Experiment shows that effective concentrations for reactions between moieties on small molecules are not limited to $\sim 10 \text{ nm}^{-3}$ (typical of immersion in a liquid); they typically range from $\sim 10^3$ to 10^{10} nm^{-3} (Creighton 1984). Two examples cited by Creighton are:



The occurrence of large magnitudes can readily be understood: In the probability-gas picture of classical transition state theory (Sec. 6.2.2), the rate of transitions is proportional to the density of the gas. For concreteness and ease of visualization, consider a unimolecular reaction in which a single atom is transferred (e.g., hydrogen abstraction). The effective concentration of hydrogen at the bottom of the potential well defining its initial state is just the reciprocal of the effective volume; using Eq. (6.5) and a stiffness of 20 N/m in each of three dimensions (vs. $\sim 30 \text{ N/m}$ typical of bond bending) yields a number density of $\sim 2 \times 10^4 \text{ nm}^{-3}$; this local density is unrelated to (and far larger than) the mean number density of nuclei in solid hydrogen at ordinary pressures. Further, many reactions require a relatively precise orientation of the reagent moieties; structural constraints in unimolecular reactions can greatly increase the concentration of properly oriented moieties, multiplying the effective concentration by a large factor. (Unimolecular reactions can also be accelerated by inter-moiety strain; this can, however be viewed as a violation of the unimolecular-bimolecular analogy that motivates the notion of effective concentration, and is more closely analogous to the piezochemical effects described in Sec.

8.3.3.3.)

Binding of reagents so as to ensure high effective concentrations is a one mechanism of enzymatic catalysis (Creighton 1984). Mechanochemical processes can likewise present reagent moieties to one another in favorable orientations and positions, resulting in high effective concentrations. Relative to bimolecular reactions in solution with a not-atypical concentration of $\sim 1 \text{ nm}^{-3}$, this mechanism should routinely yield rate accelerations on the order of 10^4 or more.

8.3.3.2. *Eutaxic "solvation"*

Where reactions involve substantial changes in charge separation, solvent effects can alter reaction rates by factors in excess of 10^5 . Relative to vacuum or nonpolar environments, a polar solvent greatly decreases the energetic cost of creating dipoles or separated ions by orienting its molecules in a manner that decreases electrostatic field strengths. The feasible reduction in free energy is, however, lessened by the entropic cost of orienting the solvent molecules. A preorganized structure (Cram 1986), in contrast, can equal the electrostatic effects of a good solvent without imposing entropic costs on the transition state.

Indeed, a preorganized environment can yield larger electrostatic effects. The solvent structure at the transition state of a solution-phase reaction is minimizes the free energy of the transition-state system as a whole. A very different organization of solvent molecules would minimize the free energy of the transition-state reactants alone; this organization would typically be more highly ordered (imposing additional entropic costs) and more highly polarized (imposing substantial energetic costs). Preorganized environments can be constructed to mimic this hypothetical, high-free-energy solvent state, paying the free-energy cost in advance and thereby out-performing a good solvent in promoting the desired reaction (a principle exploited by enzymes).

The magnitude of these effects can be roughly estimated. Ionic reactions will typically involve charge displacements on the order one electron-charge and one bond length, or $\sim 2.5 \times 10^{-29} \text{ C}\cdot\text{m}$. Fields of $2 \times 10^9 \text{ V/m}$ (e.g., 1 V across 0.5 nm) are easily achieved in a polarized environment; contact potentials in metallic systems can produce fields of this magnitude. The energy difference associated with the above field and charge displacement, $\sim 50 \text{ mJ}$, gives a rough measure of the effect that local electrostatics can have on the energies of transition and product states in ionic reactions. This energy suffices to change rates and equilibria by factors of $> 10^5$ at 300 K. The electrostatic energy differ-

ence between two orientations of an HCN molecule adjacent to a pair of oriented HCN molecules (Fig. 8.1) is ~ 110 mJ, an energy differential sufficient to change a rate or equilibrium by a factor $> 10^{10}$.

Enzymatic active sites often resemble preorganized solvent structures, reducing the energy of a transition state. Mechanochemical systems based on a more general class of molecular structures will be able to equal or exceed enzymes in this regard. The requirement that all structures be well-bound, prohibiting true solvation (Sec. 8.3.2.3) does not appear to be a significant sacrifice.

As enzymes show, molecular flexibility can be achieved without freely-moving small molecules. In a mechanochemical context, however, it will frequently be desirable to provide relatively rigid support for a reagent structure; one strategy is to surround the structure with a nonbonded, “solvent-like,” but rigid shell.

8.3.3.3. Mechanical forces

The term “piezochemistry” is in general use to describe solution-phase chemical processes in which mechanical pressure modifies chemical reactivity. Derived from the Greek “piezein” (“to press”), the term is here adopted to refer to a wider range of machine-phase chemical processes in which time-dependent mechanical forces (not necessarily a homogeneous, isotropic, slowly-varying pressure) modify chemical reactivity.

In conventional piezochemistry, pressures accessible in relatively non-exotic laboratory equipment (e.g., 0.1–2 GPa) frequently have substantial effects on reaction rates and equilibria. Transition states can be characterized by an activation volume, ΔV^\ddagger , defined by Eyring in terms of a constant-temperature partial derivative:



Figure 8.1. Dipole-dipole interaction energies. In (a), the two parallel HCN molecules (modeling bound nitrile moieties) have been held fixed at a separation of 0.3 nm while the energy of the third is minimized. In (b), the orientation of the third has been reversed. The difference in the MM2/C3D electrostatic (and total) energies is ~ 110 mJ.

$$\Delta V^\ddagger = -kT \left(\frac{\partial}{\partial p} \ln k_{\text{react}} \right)_T \quad (8.7)$$

where k_{react} is the rate constant for the reaction (Sec. 6.2.1) and p is the pressure. Typical values of ΔV^\ddagger are in the range of -0.01 to -0.10 nm^3 (values can be positive, e.g., in fragmentation reactions) (Isaacs and George 1987; Jenner 1985). Reactions frequently have negative volumes of activation because transition states typically combine bond-making with bond-breaking, and because the shortening of nonbonded distances by partial bond formation exceeds the lengthening of bonded distances by partial bond breakage. The relatively compact transition state is then favored by increased pressure. (In polar, solution-phase systems, however, electrostrictive effects associated with solvent reorganization may be dominant, hence charge-neutralization reactions can also have positive volumes of activation.)

In the (poor) approximation that ΔV^\ddagger is independent of pressure, the rate enhancement is an exponential function of pressure. In reality, increasing pressures alter ΔV^\ddagger by altering the potential energy surface. The above relationship and moderate-pressure values give only rough guide to the magnitudes of the effects that can be expected in higher-pressure piezochemical processes. With this assumption, a pressure of 2 GPa applied to a reacting system with $\Delta V^\ddagger = 0.02 \text{ nm}^3$ results in a rate increase of 1.6×10^4 at 300 K.

Nanomechanical mechanosynthetic devices can be built of diamond and diamond-like structures, and in the continuum approximation, stress is a scale-independent parameter. Accordingly, mechanosynthetic devices will be able to achieve pressures equaling those in macroscopic diamond-anvil pressure cells. These reach $\geq 550 \text{ GPa}$ (Mao, Wu et al. 1989; Xu, Mao et al. 1986), corresponding to a $\geq 30 \text{ nN}$ compressive load per bond in the (111) diamond plane (neglecting the change in areal bond density with pressure). The effects of such pressures on bonding are substantial: at $\sim 150 \text{ GPa}$, H_2 becomes metallic (Hemley and Mao 1990), as does CsI at $\sim 110 \text{ GPa}$ (Mao, Wu et al. 1989); xenon has likewise been metallized. Even a more modest pressure of 50 GPa ($\sim 3 \text{ nN/bond}$) will have striking chemical effects. In the constant ΔV^\ddagger approximation, $p = 50 \text{ GPa}$ and $\Delta V^\ddagger = 0.01 \text{ nm}^3$ would yield a 500 mJ reduction in activation energy (and a physically unrealistic $> 10^{52}$ speedup); effects in real systems will be different, but large. The change in free energy resulting from high pressures can far exceed the change resulting from high temperatures (within the conventional laboratory range).

Mechanically applied energy (unlike random thermal vibration) is subject to precise control. The above discussion has been cast in terms of pressure and volume because it draws on experiments and theories applicable to solution-phase processes. In mechanochemistry, however, force and displacement are the more useful variables, and (for a particular mode of displacement), and an activation length

$$\Delta \ell^\ddagger = -kT \left(\frac{\partial}{\partial F} \ln k_{\text{react}} \right)_T \quad (8.8)$$

can be defined in terms of the applied force F . A typical magnitude for $\Delta \ell^\ddagger$ will be ~ 0.1 nm, and feasible compressive loads will often extend to > 5 nN, which would once again yield a characteristic energy of ~ 500 mJ. (More realistic estimates are developed below.)

Mechanical instabilities can limit compressive loads: for example, if a single-atom tip is pressed against another atom, there will be a tendency to slip sideways (“down off the hill”) unless this is resisted by an adequate transverse stiffness. The negative stiffness associated with this instability has a magnitude

$$k_{s,\text{instab}} = -\frac{F_{\text{compr}}}{r_1 + r_2} \quad (8.9)$$

in the (conservative) model of hard spheres with radii r_1 and r_2 . A typical transverse stiffness $k_{s\perp}$ for an unloaded tip is 20 N/m (characteristic of angle bending for a single bond), and a typical atomic radius for a non-bonded contact under substantial loads is ~ 0.1 nm (Fig. 3.8), hence the condition

$$k_{s,\text{instab}} + k_{s\perp} > 0 \quad (8.10)$$

permits $F_{\text{compr}} \approx 4$ nN. A load of this magnitude suffices to store > 100 mJ of potential energy in overlap repulsion between two unreactive atoms (Fig. 3.7), and would store additional energy in the more-compliant deformation modes that lead toward a chemical reaction.

Unlike forces resulting from hydrostatic pressure, forces applied by mechanochemical devices can be highly anisotropic and inhomogeneous on a molecular scale: large loads (including tension, shear, and torsion) can be applied to specific atoms and bonds in a controlled manner. As noted in Section 6.4.4, these forces can cleave otherwise-stable bonds. Further, steric difficulties can be reduced by molecular compression and deforma-

tion in conventional piezochemistry (Jenner 1985). Under mechanochemical conditions larger effects can be obtained, which can significantly offset steric difficulties posed by the mechanosynthetic requirement for bound reagents (for comparison, at 4 GPa the atomic number density in liquid cyclohexane is increased by a factor of ~ 1.5 (Gray 1972)). In general, the availability of controlled forces of bond-breaking magnitude will permit piezochemical modulation of reactions far exceeding that seen in solution-phase chemistry or in the comparatively low-strength, low-stiffness environment of an enzymatic active site. Section 8.5 discusses piezochemical effects in further detail.

8.3.3.4. *Localized electrochemistry, "photochemistry"*

Mechanochemical systems will be able to exploit non-mechanical energy sources, for example, through electrochemistry and energy transfer via electronic excitations. Both of these mechanisms can be controlled more precisely in a mechanochemical environment than in solution or solution-surface interface systems.

Electrochemistry finds significant use in organic synthesis (Kyriacou 1981). Electrostatic potentials and tunneling rates can vary sharply on a molecular scale (Bockris and Reddy 1970b), resulting in molecular-scale localization of electrochemical activity. Accordingly, electrochemical processes are well-suited to exploitation in a mechanochemical context; they are also subject to modulation by piezochemical means (Swaddle 1986). In electrochemical cells, pyridine can tolerate an electrode potential of 3.3 V without reaction, and tetrahydrofuran can tolerate -3.2 V, both with respect to a (catalytically active) platinum electrode (Kyriacou 1981); these potentials correspond to energy differences with a magnitude > 500 mJ per unit charge. In field-ion microscopes (which provide one model for an electrode surface), electric fields can reach ~ 50 V/nm (Nanis 1984). Electrochemical processes, despite their undoubted utility, are generally neglected in this volume.

Direct photochemistry suffers from problems of localization (which do not preclude its use), but photochemical effects can often be achieved by nonphotochemical means. Photochemical processes begin with the electronic excitation of a molecule by a photon, but this energy can often migrate from molecule to molecule as a discrete "exciton" before inducing a chemical reaction, and this process is highly sensitive to molecular structures and positions. Accordingly, the transfer of excitons can provide a better-controlled means for achieving photochemical ends; potential applications are, however, neglected in this volume.

8.3.3.5. Broadened options for catalysis

The structural requirements for mechanochemical reagents are satisfied by many catalytic structures: some are parts of solid surfaces already, and others (e.g., many homogeneous transition metal catalysts) have analogues that can be covalently anchored to a larger structure. The remarks of Section 8.3.2.2 apply to small catalytic species (e.g., hydrogen and hydroxide ions).

Aside from regeneration treatments (which are infrequent on a molecular time scale), conventional catalysts operate under steady-state conditions. In a typical catalytic cycle, reagents are bound to form a complex, the complex rearranges, and a product departs, all in the same medium at constant pressure, temperature, and so forth. If any transition state in this sequence of steps is too high in energy, its inaccessibility will block the reaction. If any intermediate state is too low in energy, its stability will block the reaction. If any feasible alternative reaction (with any reagent or contaminant in the diffusing mixture) leads to a stable complex, the catalyst will be poisoned. Many successful catalysts have been developed, but the above conditions are stringent, requiring a delicately-balanced energy profile across a sequence of steps (Crabtree 1987).

The range of feasible catalytic processes will be broadened by the opportunities for control in mechanochemical processes. The elementary reaction in a catalytic cycle can occur under distinct conditions, lessening the requirement for delicate compromises to avoid large energy barriers or wells. As will be discussed in Section 8.5.10.3, mechanochemical catalysts can be subjected to a variety of manipulations that modulate the energy of bonds and transition states, typically by many times kT . Finally, comprehensive control of the molecular environment will enable the designer to prevent many unwanted reactions (Sec. 8.3.3.6), permitting the use of more-reactive species (Sec. 8.3.3.7).

8.3.3.6. Avoidance of competing reactions

In diffusive synthesis, achieving 95% yield in each of a long series of steps is typically considered excellent. At the end of a 100-step process, however, the net product would be $\sim 0.6\%$; and at the end of a 2000-step process, $\sim 10^{-43}\%$. A million tons of starting reagents would then be expected to yield zero molecules of the desired product structure. A reaction with $\Delta\mathcal{F} \leq -115$ mJ at 300 K will, at equilibrium, leave less than 10^{-12} of the starting molecules unreacted. An energy difference of this magnitude is not uncommon, and a series of reactions with this yield would permit over 10^{10} sequential

steps with high overall yield, in the absence of side reactions.

The complexity of the structures that can be built up by diffusive synthesis is limited not by an inability to add molecular fragments to a structure, but by the difficulty of avoiding mistaken additions. This problem is substantial, even in 100-atom structures. Moreover, as structures grow larger and more complex they tend to have increasing numbers of functional groups having similar or identical electronic and steric properties (on a local scale). Reliably directing a conventional reagent to a specific functional group becomes increasingly difficult, and ultimately impossible.*

In a well-designed eutaxic mechanochemical system, unplanned molecular encounters will not occur and *most* unwanted reactions will accordingly be precluded. One class of exceptions consists of reactions analogous to unimolecular fragmentation and rearrangement; these instabilities have been discussed in Section 6.4, and will be discussed further in connection with reagent moieties in Sections 8.4 and 8.5. The other class of exceptions consists of reactions that occur in place of desired reactions; these can be termed *misreactions*.

A typical mechanosynthetic step will involve the mechanically-guided approach of a reagent moiety to a target structure, followed by its reaction at a site on that structure. In general, *unguided* reactions would be possible at several alternative sites, each separated from the target site by some distance (properly, a distance in configuration space). At one extreme, the alternative sites will be separated by a distance sufficient to make an unwanted encounter in the guided case energetically infeasible (e.g., requiring that the mechanical system either break a strong bond or undergo an elastic deformation with a large energy cost). At the other extreme, the potential energy surface will be such that passage through a single transition state leads to a branching valley, and then to two distinct potential wells, only one of which corresponds to the desired product; in this circumstance, unwanted reactions would be unavoidable. In intermediate cases, transition states leading to desired and undesired products will be separated by intermediate distances,

* Note, however, that convergent, "structure-directed" synthesis strategies (Ashton, Isaacs et al. 1989) can substantially loosen this constraint by combining larger fragments in a manner analogous to biological self-assembly; reactions then depend on distinctive, larger-scale properties of the structure. The above argument thus sets no firm limit on the capabilities of diffusive synthesis.

and the mechanical stiffness of the guiding mechanism will impose a significant energy cost on the unwanted transition state, relative to the unguided case.

Considering the approach of a reagent moiety to a target structure in three-dimensional space, a reaction pathway can be characterized by the trajectory of some atom in the moiety (such as an atom that participates in the formation of a bond to the surface). Inspection of familiar chemical reactions concurs with elementary expectations in suggesting that reaction pathways leading to alternative products are commonly characterized by trajectories that differ by a bond length ($d_t \approx 0.15$ nm) or more at the competing transition states. This is not universally true, but it is the norm; accordingly, avoidance of the exceptional situations will impose only a modest constraint on the available chemistry. Where reagents are selective, or have reactivities with a strong orientational dependence, alternative trajectories may be separated by far greater distances.

A not-atypical stiffness for the displacement of an atom in a mechanically-guided reagent moiety (measured with respect to a reasonably rigid or well-supported target structure) will be 20 N/m (comparable to the ~ 30 N/m transverse stiffness of a carbon atom with respect to an sp^3 carbon). Note that the mechanical constraints on a reactive moiety can in many instances include not only the stiffness of the covalent framework, but forces resulting from a closely-packed, chemically solvent-like surrounding structure that substantially blocks motion in undesired directions (see Fig. 8.2). Using the worst-case decoupling model of Section 6.3.3.4 to estimate error rates with non-sinusoidal potentials, the condition $P_{\text{err}} \leq 10^{-12}$ requires that the elastic energy difference between the two locations be $\geq \sim 150$ mJ at 300 K. With $k_s = 20$ N/m, this condition is satisfied at transition-state separations $d_t \geq 0.12$ nm. Accordingly, with modest constraints on the chemistry of the reacting species, suppression of unwanted reactions by factors of better than 10^{-12} should be routinely achievable. With $k_s = 30$ N/m and $d_t \geq 0.15$ nm, this model yields $P_{\text{err}} \leq 10^{-27}$. (Different ratios of accessibility of differing transition points at different times complicates the description, but leaves the essential conclusions intact.)

8.3.3.7. *Precise control of highly reactive reagents*

It might seem that the most useful reagents would be those (e.g., strong free radicals, carbenes) that can react with many other structures. In solution-phase chemistry, however, unwanted reactions are the chief limit to synthesis, and reagents are prized less for their reactivity than for their selectivity. The ideal reagent in solution-phase synthesis is inert in all but a few circumstances, and it need not react swiftly when it reacts at all.

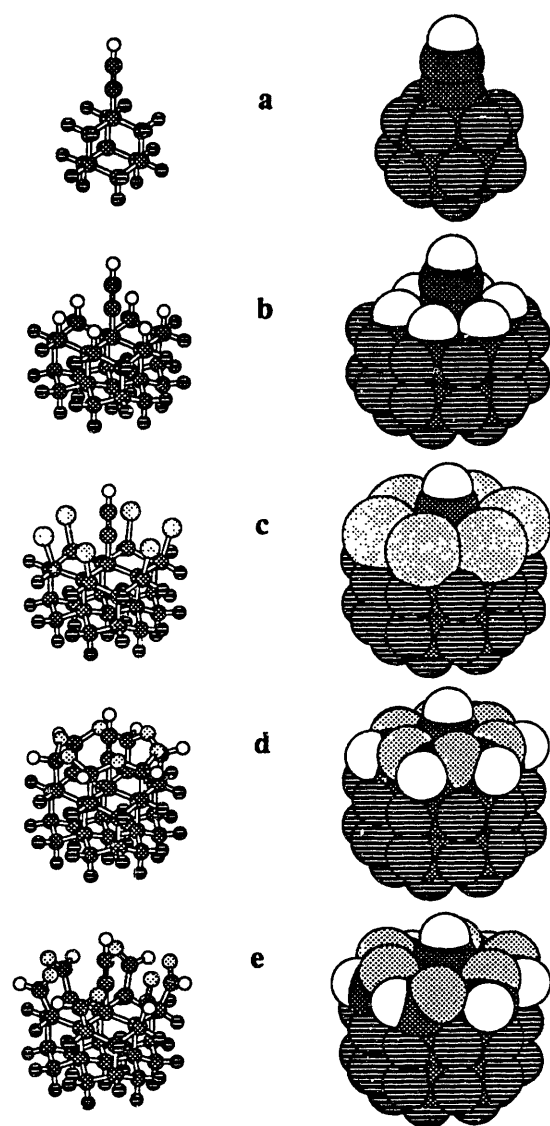


Figure 8.2. Models of a reagent moiety of low intrinsic stiffness with varying degrees of support from a surrounding “solvent-like” structure (MM2/C3D potential). Each model includes an alkyne moiety representing (for example) an alkynyl radical of the sort that might be used as a hydrogen abstraction tool (Sec. 8.5.4.3). In (a), the moiety is supported by an adamantyl group; the bending stiffness of the moiety at its most remote carbon atom, taking the hydrogen atoms at the attached to the six-membered ring opposite the moiety as fixed, is ~ 6 N/m. In (b), a more crowded environment increases the stiffness to ~ 11 N/m (relative to those hydrogen atoms not either bonded to or in contact with the alkyne moiety); (c) replaces H atoms with Cl, increasing the stiffness to ~ 20 N/m; greater stiffness could be achieved by adding structure outside the Cl ring to press its atoms inward. Structure (d) surrounds the alkyne with a ring of oxygen-linked carbon atoms, yielding a calculated stiffness of ~ 65 N/m. Structure (e), however, represents a possible failure mode of (d) in which a rearrangement has cleaved six C–O bonds, yielding six aldehyde groups. Although this process is quite exoergic, the transition state may be effectively inaccessible at room temperature.

In mechanosynthesis, however, *selectivity* based on the local steric and electronic properties of the reagents themselves can be replaced by nearly perfect *specificity* based on positional control of reagent moieties by a surrounding mechanical system. Accordingly, highly reactive reagents gain utility, including reagents that are highly unstable to bimolecular reactions among molecules of the same type (e.g., benzyne, reactive dienes, and other fratricidal molecules). The use of reagents with increased reactivity will increase typical reaction frequencies, adding to (or, more properly, multiplying together with) effective-concentration, eutaxic-environment, and piezochemical effects.

8.3.4. Preview: molecular manufacturing and associated constraints

The capabilities of mechanosynthetic processes can be gauged by comparing the limitations discussed in Section 8.3.2 with the strengths discussed in Section 8.3.3, taking the capabilities of diffusive synthesis as a baseline. This comparison, summarized in Table 8.1, suggests that mechanosynthetic operations are (overall) more capable than diffusive operations. These capabilities must, however, be judged in light of their proposed applications. This requires a preview of basic aspects of proposed molecular manufacturing systems.

8.3.4.1. Molecular manufacturing approaches

Mechanochemical devices are of interest in this volume chiefly as components of mechanosynthetic systems capable of building large, complex structures, including systems of molecular machinery. The requisite positioning and manipulation of reagent moieties can be achieved in any of several ways, but these can be roughly divided into *molecular mill* approaches and *molecular manipulator* approaches. The former use simple, repetitive motions; the latter use flexible, programmable positioning. Mill-style devices are well-suited to the task of preparing reagent moieties for manipulator-style devices, but can also produce final products directly. In all cases considered, mechanical transport mechanisms replace diffusion.

8.3.4.2. Reagent preparation vs. application

Product synthesis will use reagent moieties that have been prepared from other reagent moieties, starting ultimately with simple feedstock molecules. The preparative steps can typically take place in an environment significantly different from that of the final synthetic steps.

Table 8.1. Comparison of solution-phase synthesis and mechanosynthesis.

	Solution-phase synthesis	Mechanosynthesis
Parallelism	Natural	Requires many devices
Reagent structure	Unconstrained	Bound to “handle”
Electrostatic envir.	Control dielectric constant	Control dielectric constant, field
Electrochemistry	Useful	Useful
Photochemistry	Useful	Poor loc. w/o exciton mediation
Temperatures	Can be high	Must be modest
Max. effective conc.	$\sim 100 \text{ nm}^{-1}$	$> 10^9 \text{ nm}^{-1}$
Available pressures	Up to $\sim 2 \text{ GPa}$	$> 500 \text{ GPa}$
Control of forces	Mag. of uniform pressure	Mag., location, direction, torque
Positional control	None	All three degrees of freedom
Orientational control	None	All three degrees of freedom
Reagent requirements	Selective reactivity	Rapid reactivity
Reaction site selectivity	Steric, elect. influences	Direct choice of location
Max. synth. complexity	$\sim 100\text{--}1000 \text{ steps}$	$> 10^{10} \text{ steps}$

In *reagent preparation*, the entire surrounding environment can be tailored to further the desired transformation. In this respect, the environment can resemble that of an enzymatic active site, but with the option of exploiting a wider range of structures, more active reagents, piezochemical processes, and so forth. The freedom to tailor the entire reaction environment during reagent preparation is a consequence of the small size and consequent steric exposure of the structures being manipulated.

In the final synthetic step, however, one reacting surface must be a feasible intermediate stage in the construction of the product, and a prepared reagent must be applied to that surface.* Construction strategies can be chosen to facilitate the sequence of synthetic reactions, but the freedom to tailor the entire environment solely to facilitate a single reaction is not available. In these *reagent application* steps, highly reactive moieties will be of increased utility and moieties compatible with supporting structures of low steric bulk will be desirable. These remarks apply both to manipulator-based systems and to the reagent-application stages of a mill-style system engaged in direct synthesis of complex products.

8.3.4.3. Reaction cycle times

One measure of the productivity of a manufacturing system is the time t_{prod} required for it to make a quantity of product equaling its own mass. Mill-style systems are anticipated to contain $\sim 10^6$ atoms per processing unit, with each unit responsible for converting a stream of input molecules into one portion of each member of a stream of product structures. Manipulator-style systems are anticipated to contain $\sim 10^8$ atoms per manipulator, with each unit responsible for a series of constructive transformations on a product structure. Processing units and manipulators each perform mechanosynthetic operations on products with some frequency f_{synth} and transfer a mean number of atoms per operation, n_{synth} . For a given operation, the number transferred may be positive or negative, as sites on the product structure are altered to prepare them for further steps, but a typical value will be $n_{\text{synth}} \approx 1$. Accordingly, for $f_{\text{synth}} = 10^3$ Hz, $t_{\text{prod}} = 10^3$ s for a mill-based system and 10^5 s for a manipulator-based system; for $f_{\text{synth}} = 10^6$ Hz, the corresponding values of t_{prod} are 1 and 10^2 s respectively. Values of $f_{\text{synth}} \geq 10^3$ Hz will be acceptable for many practical applications.

* Note, however, that a convergent synthesis can reduce the asymmetry of the reagent application process.

In the example calculations of the following sections will assume that the time available for a reaction $t_{\text{trans}} = 10^{-7}$ s. This is compatible with $t_{\text{act}} \approx 10^{-6}$ s, or with $f_{\text{synth}} \approx 10^6$ Hz. These times are all long compared to the characteristic time scale of molecular vibrations, $\sim 10^{-13}$ s.

8.3.4.4. Limiting misreactions

In molecular manufacturing, two basic classes of error are (1) those that damage one product structure (fabrication errors), and (2) those that damage the manufacturing mechanism (destructive errors). Overall reliability can be increased in each instance by dividing a system into smaller modules and replacing those that fail. Where fabrication errors occur, the module can be discarded before being used in further assembly. Where destructive errors occur in a module in a functioning system, a backup module can assume its function. Simplicity of design typically favors the use of relatively large modules.

From a systems-engineering perspective, 10^8 -atom modules can often be considered large. Fabrication errors at a rate of $\leq 10^{-10}$ per reagent application operation, resulting in a fabrication failure rate of $\leq 10^{-2}$, can thus typically be considered low. If we demand that each module in a manufacturing system process $\geq 10^2$ times its own mass before failing (and assume $n_{\text{synth}} \approx 1$), then a destructive-error rate of $\leq 10^{-10}$ per reagent application operation is again acceptable. The following calculations will take as a requirement that both classes of error occur at rates $\leq 10^{-12}$ per reagent application *and* reagent processing application. Inasmuch as the ratio of reagent-preparation operations to reagent-application operations is expected to be ≤ 100 (values on the order of 10 seem likely), this is a relatively stringent condition.

8.3.4.5. Conditions for limiting failed reactions

Section 8.3.4.4 has discussed conditions for keeping the probability of misreactions $\leq 10^{-12}$; a related but distinct problem is to ensure the the desired reaction occurs, with a probability of failure $\leq 10^{-12}$. As noted in Section 8.3.3.6, a reaction with $\Delta F \leq -115$ mJ at 300 K (at the time of kinetic decoupling, in terms of the approximation discussed in Section 6.3.3) can proceed to completion with an equilibrium probability of remaining in the starting state $< 10^{-12}$. For a reaction characterized by a rate constant k_{react} (s^{-1}), the probability of a failure to react falls to $< 10^{-13}$ when the available reaction time $t_{\text{trans}} \geq 30/k_{\text{react}}$ s. Mechanochemical reactions with relatively rigid, well-aligned reagent moie-

ties will commonly have frequency factors of 10^{12} s^{-1} or more. A one-dimensional model based on classical transition state theory (Sec. 6.2.2), together with the bounds just described, then yields a bound on the allowable barrier height for the reaction

$$\Delta V^\ddagger \leq kT \ln \left(\frac{t_{\text{react}}}{30 \times 10^{-12}} \right) \quad (8.11)$$

At 300 K, with $t_{\text{react}} = 10^{-7}$, $\Delta V^\ddagger \leq 34 \text{ mJ}$ is acceptable.

How does this compare to typical solution-phase reactions? In the laboratory, characteristic reaction times (the reciprocals of the rates) vary widely, from $< 10^{-9}$ to $> 10^6 \text{ s}$, but a not-atypical value for a practical reaction is 10^3 s at a reactant concentration of $\sim 1 \text{ nm}^{-3}$. Relative to this, achieving a reliable reaction in $t_{\text{trans}} = 10^{-7} \text{ s}$ requires a speedup of $\sim 3 \times 10^{11}$. Increases in effective concentration resulting from mechanical positioning can easily provide a speedup of $> 3 \times 10^4$ (Sec. 8.3.3.1). Achieving the remaining factor of 10^{-7} requires that the energy barrier be lowered by $\sim 70 \text{ mJ}$. Earlier sections have shown that electrostatic effects in eutaxic environments can exceed 100 mJ , and that (crudely estimated) piezochemical effects can exceed 500 mJ ; shifts from less-active to more-active reagents can likewise have large effects. Achieving adequate reaction speeds does not appear to be a severe constraint.

Intersystem crossing from the singlet state to a low-lying triplet can cause unwanted behavior, including failed reactions and misreactions. Singlet transition-state geometries often resemble triplet equilibrium geometries (Salem and Rowland 1972). To ensure reliability, it will be sufficient to ensure either (1) that the singlet-triplet gap $\Delta \mathcal{V}_{s,t}$ always exceeds 115 mJ ($\approx kT \ln[10^{12}]$ at 300 K), or (2) that as the $\Delta \mathcal{V}_{s,t}$ increases, it exceeds 115 mJ at some time t_0 (prior to the time t_1 at which the geometry is no long suitable for correct bond formation), and that the integral of the intersystem crossing rate k_{isc} meets the condition that

$$\int_{t_0}^{t_1} k_{\text{isc}}(t) dt \geq \ln(10^{12}) \quad (8.12)$$

The conditions just described assume that reliability must be assured in a single trial. An alternative approach to ensuring reliable reaction, *optional repetition*, is available in systems in which the outcome of a measurement can determine whether an operation is repeated. For example, in a reaction process, the reaction rate and time might ensure equilibrium, but with $\Delta \mathcal{F} = 0$ (rather than -115 mJ). The reaction will then have a 0.5 proba-

bility of success in any single trial. If a series of trials can be terminated whenever success is achieved, then the probability of failure after 40 trials is $< 10^{-12}$, and the mean number of trials required to achieve success is 2. A similar process with an exoergicity of 25 mJ would (at 300 K) have a 0.9976 probability of success in a given trial, requiring a sequence of 5 optional repetitions to achieve $P_{\text{err}} \leq 10^{-12}$; the mean number of trials ≈ 1.002 .

8.3.4.6. Requirements for reagent stability

To meet the reliability objectives stated in Section 8.3.4.4, reagent instability must not cause errors at a rate greater than that of misreactions. If the mean time between reactions is $\leq 100t_{\text{act}} = 10^{-4}$ s, and the frequency factor for the instability is $\leq 10^{13}$ Hz, then keeping the instability-induced error rate per reaction $\leq 10^{-12}$ requires a barrier height ≥ 200 mJ. Since all the failure mechanisms discussed in this and the previous sections are exponentially dependent on energy parameters, exceeding the specified objectives will yield large improvements in reliability.

8.3.5. Summary of the comparison

Solution-phase chemistry has enabled the synthesis of small molecules with an extraordinary range of structures (Sec. 8.2), but has not yet succeeded in constructing larger structures while maintaining eutaxic control. The specific comparisons made in Sections 8.3.2–8.3.4 support some general conclusions regarding the relative capabilities of mechanosynthesis:

8.3.5.1. Versatility of reactions

Relative to diffusive synthesis, mechanosynthesis imposes several significant constraints on the kinds of reagents that can be effectively employed. It requires that reagent moieties be bound, which can reduce reactivity and impose steric constraints. It requires that reactions be fast at room temperature, limiting the magnitude of acceptable activation energies. It requires that reagent moieties have substantial stability against unimolecular decomposition reactions, precluding the use of some reagents that are acceptable in the diffusive synthesis of small molecules.

Offsetting these limitations, however, are several advantages. Most fundamentally, *mechanochemical processes permit the control of more degrees of freedom than do comparable solution-phase processes*; these degrees of freedom include molecular positions, orientations, force, and torques. As a consequence, highly-reactive moieties can be

guided with great specificity, enabling the exploitation of reagents that are too indiscriminate for widespread use in solution-phase synthesis. Bound reagent moieties are subject to mechanical manipulation, enabling piezochemical effects to speed reactions and overcome substantial steric barriers through localized compression and deformation of molecular structures. Finally, entirely new modes of reaction become available when reagent moieties can be subjected to forces of bond-breaking magnitude. Overall, these gains in versatility appear to exceed the losses, and hence the range of local structural features that can be constructed by mechanosynthesis should equal or exceed the range feasible with diffusive synthesis.

8.3.5.2. *Specificity of reactions*

In diffusive synthesis, most reactions are associated with substantial rates of misreaction, and the probability of a misreaction during any given step tends to increase with the size of the product structure. Experience suggests that the cumulative probability of error becomes intolerable for product structures of more than a few hundred to a few thousand atoms.

In mechanosynthesis, reliable exclusion of misreactions can be achieved given (1) a sufficient distance between alternative transition states and (2) a sufficient mechanical stiffness resisting relative displacements of the reagent moieties. Distances on the order of a bond length combined with stiffnesses comparable to those of bond bending yield $P_{\text{err}} < 10^{-12}$. Error rates (per step) are independent of the size and complexity of the product structure, given that the product is either stiff or well-supported.

8.3.5.3. *Synthetic capabilities*

Within the constraints required to achieve reliable, specific reactions, the versatility of the set of chemical transformations that will eventually become available in mechanosynthesis can be expected to equal or exceed the versatility of the set of transformations available in solution-phase synthesis. This versatility is sufficient to suggest that most kinetically-stable substructures will prove susceptible to construction (the challenging class of diamond-like structures is considered in more detail in Section 8.6). Transformations that (1) satisfy these reliability constraints and (2) yield kinetically-stable substructures can then be composed into long sequences that maintain eutaxic control and yield product structures of 10^{10} or more atoms.

8.4. Reactive species

8.4.1. Overview

This section examines several classes of reagents from a mechanosynthetic perspective. *Numerous classes are omitted, and those included are discussed only briefly.* As noted in Section 8.2, chemistry is a vast subject; introductory textbooks on organic chemistry commonly exceed 1000 pages.

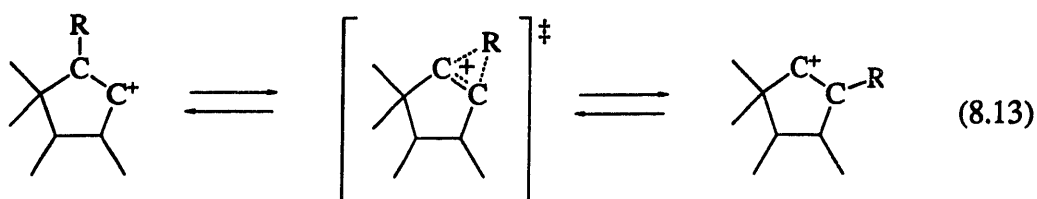
Polycyclic, broadly diamondoid structures are the products of of greatest interest in the present context, hence this discussion will focus on the formation of carbon-carbon bonds; much of what is said is applicable to analogous nitrogen- and oxygen-containing compounds. As Section 8.3.3.7 indicates, highly reactive species are of particular interest. Many of the following species would be regarded as reaction intermediates (rather than reagents) in solution-phase chemistry.

8.4.2. Ionic species

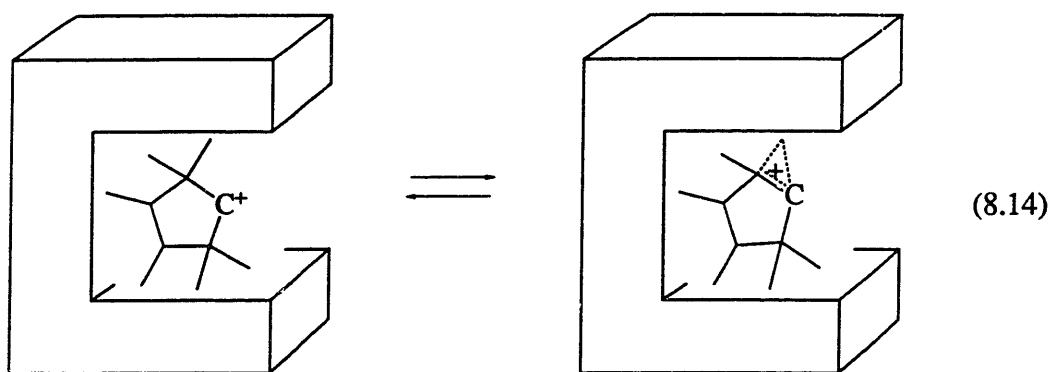
Ionic species play a major role in solution-phase chemistry but are rare in the gas phase, where their greater electrostatic energy decreases their stability by ≥ 500 mJ, favoring neutralization by charge transfer. In mill-style reagent preparation, where reactions can occur in an electrostatically-tailored environment, ionic species can be as stable as those in solution and desired transformations can be driven by local electric fields. Accordingly, the utility of ionic species will, if anything, be enhanced. In manipulator-based reagent application, however, it may frequently be desirable to expose reagent moieties on a tip moving through open space toward a product structure that is not tailored for favorable electrostatics. The electrostatic energy of ionic species will then be high (making solution precedents inapplicable), and the utility of ionic species may be relatively limited.

8.4.2.1. Rearrangements and neutralization

Ionic species vary in their susceptibility to unimolecular rearrangement. Carbocations, for example, are prone to 1,2-shifts with low (or zero (Hehre, Radom et al. 1986)) barrier energies:



A reagent moiety in which this process can occur will likely fail to meet the stability criterion for use in molecular manufacturing processes. If, however, each of the R-groups is part of an extended rigid system, this rearrangement will be mechanically infeasible, as in Eq. (8.14). Rearrangement could also be prevented by steric constraints from a surrounding matrix, or (since the rearrangement involves charge migration) by local electrostatics.



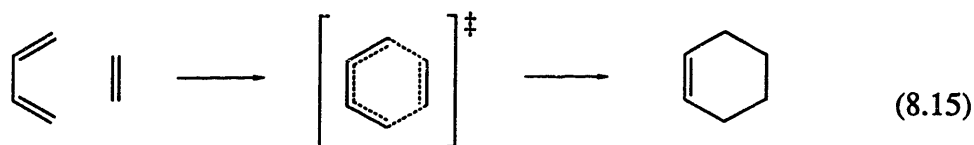
Carbanions, having filled orbitals, are somewhat less prone to rearrangement (Bates and Ogle 1983) and will more readily meet stability criteria. Again, structural, steric, and electrostatic characteristics can in many instances be used to suppress unwanted rearrangements.

Charge neutralization provides another, non-local failure mechanism. To avoid this will require a design discipline that takes account of the ionization energies and electron affinities of all sites within a reasonable tunneling distance (several nanometers) of the ionic site, ensuring that charge neutralization is energetically unfavorable by an ample margin (e.g., ≥ 115 mJ).

8.4.3. Unsaturated hydrocarbons

In diffusive chemistry, unsaturated hydrocarbons (alkenes and alkynes) find extensive use in the construction of carbon frameworks. Their reactions characteristically redistribute electrons from relatively high-energy π -bonds to relatively low-energy σ -

bonds, thereby increasing the number of covalent linkages in the system while reducing the bond order of existing linkages. This process has a strongly negative ΔV^\ddagger and ΔV_{react} , since the increase in interatomic separation resulting from a reduction in bond order is far outweighed by the decrease resulting from the conversion of a nonbonded to a bonded interaction. Accordingly, these reactions are subject to strong piezochemical effects. Typical examples are Diels–Alder reactions such as (8.14),



which have ΔV^\ddagger in the range of ~ 0.05 to 0.07 nm^3 .

Unsaturated hydrocarbons undergo useful reactions with ionic species, and their reactions with other reagents will be touched on below. Sections 8.5.5–8.5.9 describe certain classes of reactions in somewhat more detail.

From the perspective of solution-phase chemistry, mechanosynthesis will have a greater freedom to exploit reactions involving strained (and therefore more reactive) alkenes and alkynes. The cyclic and polycyclic frameworks necessary to enforce strain will be the norm, hence use of strained species can be routine; their use will be desirable and practical for reasons discussed in Sections 8.3.3.7 and 8.3.4.3. Planar alkenes have minimal energy, but molecules as highly pyramidalized as cubene **8.6** (Eaton and Maggini 1988) and as highly twisted as adamantene **8.7** (Carey and Sundberg 1983a) have been synthesized.

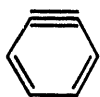


8.6

8.7

The reduced bonding overlap in these species (zero, for **8.7**) makes the energetic penalty for unpairing of the π -electrons small enough to permit them to engage in diradical-like reactions under mechanochemical conditions.

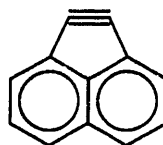
Alkynes are of lowest energy when linear, but such highly-reactive species as benzyne, **8.8**, cyclopentyne **8.9**, and acenaphthyne **8.10**, have been synthesized (Levin 1985).



8.8



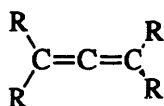
8.9



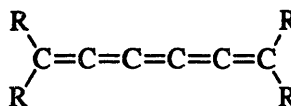
8.10

Structures 8.8–8.10 have short lifetimes in solution owing to bimolecular reactions, but all will have mechanically-anchored analogues that are kinetically stable in eutaxic environments while displaying high reactivity in synthetic applications. Again, one π -bond is sufficiently weak to permit diradical-like reactivity under suitable conditions (Levin 1985).

Also of high energy are allenes 8.11, cumulenes 8.12, and polyynes 8.13, (Patai 1980).



8.11



8.12



8.13

Cumulenes and polyynes are of particular interest in building diamondoid structures consisting predominantly of carbon (Sec. 8.6); they bring no unnecessary atoms into the reaction.

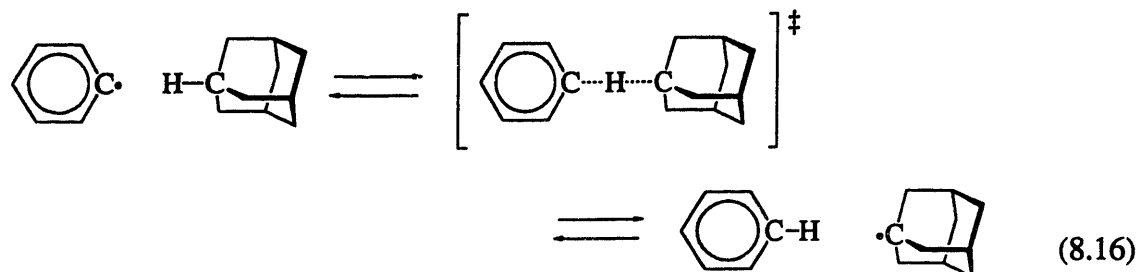
Unsaturated hydrocarbons are prone to a variety of rearrangements, subject to constraints of geometry, bond energy, and orbital symmetry. Those shown here are stable, however, and (as with carbocations) mechanical constraints from a surrounding structure can inhibit many rearrangements of otherwise-unstable structural moieties.

8.4.4. Carbon radicals

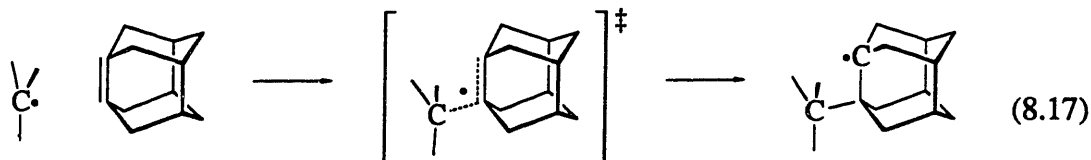
Free radicals result when a covalent bond is broken in a manner that leaves one of the bonding electrons with each fragment. Radicals thus have an unpaired electron spin and (in the approximation that all other electrons remain perfectly paired) a half-occupied orbital. Radicals can be stabilized by delocalization, for example in π -systems, but are typically highly reactive.

8.4.4.1. Reactions

Among the characteristic reactions of radicals are *abstraction*, in which the radical encounters a molecule and removes an atom (e.g., hydrogen), leaving a radical site behind,



addition to an unsaturated hydrocarbon (here, a reactive pyramidalized species), generating an adjacent radical site on the target structure,



and *radical coupling*, the inverse of bond cleavage.



Radical addition and coupling have significant values of ΔV^\ddagger (for addition, $\sim 0.025 \text{ nm}^3$ (Jenner 1985)) and ΔV_{react} . Abstraction reactions typically have a smaller ΔV^\ddagger , and no significant ΔV_{react} . Their susceptibility to piezochemical acceleration is analyzed in Section 8.5.4.

8.4.4.2. Radical coupling and intersystem crossing

Electron spin complicates radical coupling and related reactions. Bond formation demands that opposed spins be paired (a singlet state), but two radicals may instead have aligned spins (a triplet state), placing them on a repulsive PES. Bond formation then requires an electronic transition (triplet \rightarrow singlet intersystem crossing). As the radicals approach, the gap between the triplet and singlet state energies grows, but this *decreases* the rate of intersystem crossing. In delocalized systems, bond formation can occur with-

out intersystem crossing, at the energetic cost of placing some other portion of the system into a triplet state. If intersystem crossing is required during the transformation time, however, then a condition like Eq. (8.12) must be met, but with t_1 representing the time by which bond formation must have occurred (if, that is, the system is to operate correctly). The condition that $\Delta\mathcal{V}_{s,t} \geq 115 \text{ mJ}$ imposes a significant constraint because k_{isc} varies inversely with the electronic energy difference $\Delta\mathcal{V}_{isc}$, which (in the absence of mechanical relaxation) would equal the difference in equilibrium energies $\Delta\mathcal{V}_{s,p}$ and will frequently be of a similar magnitude.

Values of k_{isc} for radical pairs in close proximity vary widely, and intramolecular radical pairs (diradicals) provide a model (Salem and Rowland 1972). For 1,3 and 1,4 diradicals, k_{isc} has been estimated to be comparable to k_{isc} for the $S_1 \rightarrow T_1$ intersystem crossing in aromatic molecules (Salem and Rowland 1972). In the latter processes, k_{isc} is commonly in the range of 10^6 – 10^8 s^{-1} (Cowan and Drisko 1976; Salem and Rowland 1972). Experiments with 1,3 diradicals in cyclic hydrocarbon structures found more favorable values of k_{isc} , ranging from $\sim 10^7$ to $> 10^{10} \text{ s}^{-1}$ (Adam, Hössel et al. 1987); the differences were attributed to conformational effects on orbital orientation, confirming rules proposed in (Salem and Rowland 1972).

The presence of high- Z atoms relaxes the spin restrictions on intersystem crossing by increasing spin-orbit coupling (Cowan and Drisko 1976; Salem and Rowland 1972). In an aromatic-molecule model, k_{isc} for the $S_1 \rightarrow T_1$ transition in naphthalene is increased by a factor of ~ 50 by changing the solvent from ethanol to propyl iodide, and bonded heavy atoms have a larger effect: k_{isc} increases from $\sim 6 \times 10^6 \text{ s}^{-1}$ in 1-fluoronaphthalene to $> 6 \times 10^9 \text{ s}^{-1}$ in 1-iodonaphthalene (Cowan and Drisko 1976). The energy-gap condition is met at ordinary temperatures in this system: $\Delta\mathcal{V}_{isc}$ for 1-iodonaphthalene is 209 mJ (Wayne 1988).

From these examples, it is reasonable to expect that, the absence of special adverse circumstances, the inclusion of high- Z atoms bonded in close proximity to radical sites can be used to ensure values of $k_{isc} \geq 10^9$ in radical coupling processes where $\Delta\mathcal{V}_{s,t} \geq 115 \text{ mJ}$. This is consistent with $P_{err} < 10^{-12}$ and $\tau_{trans} < 10^{-7} \text{ s}$. Failure to achieve intersystem crossing rates of this magnitude would increase the required value of τ_{trans} for a particular operation, but would have only a modest effect on processing rates in a system as a whole, and no effect on the set of feasible transformations.

8.4.4.3. Types of radicals

Carbon radicals can broadly be divided into π -radicals (e.g., 8.14) and σ -radicals (e.g., 8.15–8.17), depending on the hybridization of the radical orbital. Of these, σ -radicals are significantly higher in energy and hence more reactive; examples include radicals at pyramidalized sp^3 carbon (e.g., the 1-adamantyl radical) 8.15, aryl radicals 8.16, and the alkynyl radical 8.17.



8.14



8.15



8.16



8.17

The latter forms the strongest bonds to hydrogen and has excellent steric properties. *Ab initio* calculations on the ethynyl radical ($\text{HCC}\cdot$), however, predict a low-energy electronic transition, $A\ ^2\Pi \leftarrow X\ ^2\Sigma$, with an energy of only ~ 40 mJ (Fogarasi and Boggs 1983). If alkynyl radicals have a similar state at a similar energy, they will have a significant probability of being found in the wrong electronic state. Interconversion, however, requires no intersystem crossing and has no symmetry forbiddenness; it should be fast compared to typical values of t_{trans} . Figure 8.3 illustrates an exposed alkyne stiffened by nonbonded contacts.

8.4.4.4. Radical rearrangement

Radicals are far less prone to rearrangement than are carbocations. Intramolecular abstraction and addition reactions are common, where they are mechanically feasible, but

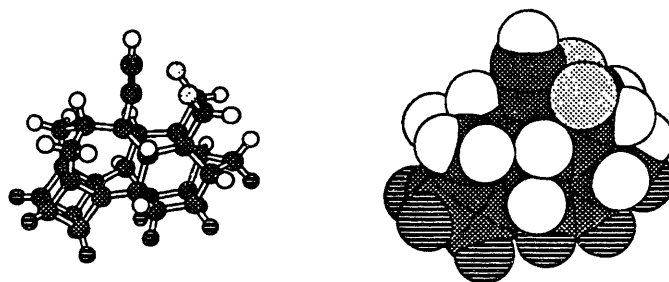


Figure 8.3. A structure with a sterically-exposed alkynyl carbon (here in a model alkyne group) having MM2/C3D stiffnesses of 4.5 and 21 N/m in orthogonal bending directions.

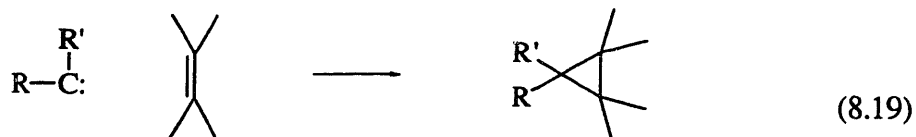
shifts analogous to that shown in Eq. (8.13) are almost unknown unless the migrating group is capable of substantial electron delocalization (e.g., an aryl group).

8.4.5. Carbenes

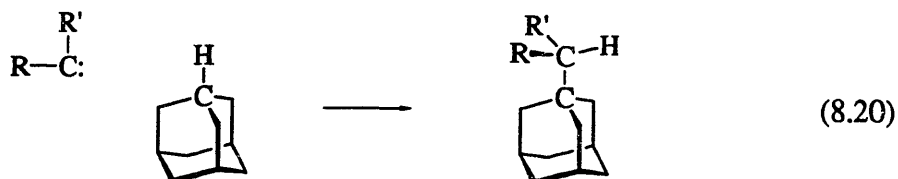
Carbenes are divalent carbon species, formally the result of breaking two covalent bonds. The two nonbonding electrons in a carbene can be in either a singlet or a triplet state; the unpaired electrons in the latter species behave much like those in radicals. Some carbenes are ground-state singlets in which the singlet-triplet energy gap $\Delta\mathcal{V}'_{s,t} \geq 115$ mJ, making the probability of occupancy of the triplet state $\leq 10^{-12}$ at 300 K. Singlet carbenes can react directly to form singlet ground-state molecules; to achieve analogous results with triplet carbenes requires intersystem crossing.

8.4.5.1. Carbene reactions

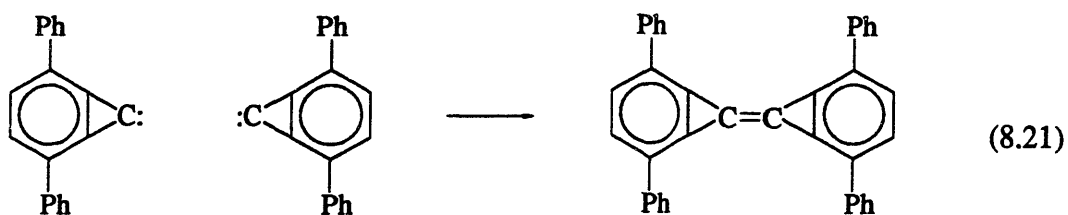
Carbenes can undergo addition to double bonds, yielding cyclopropanes



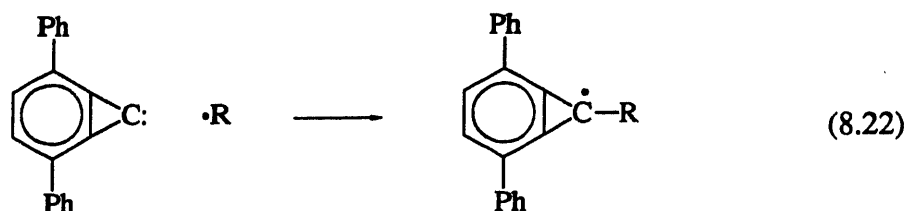
insertion into C–H bonds,



and coupling (Neidlein, Poignée et al. 1986).



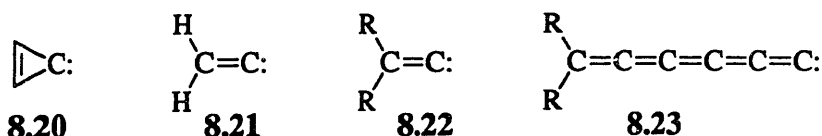
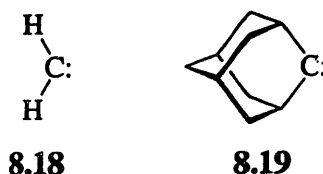
These reactions typically proceed with energy barriers of ≤ 20 mJ; many have a barrier of zero (Eisenthal, Moss et al. 1984; Moss 1989). Coupling of a carbene and a radical



should likewise proceed with little or no barrier.

8.4.5.2. Singlet and triplet carbenes

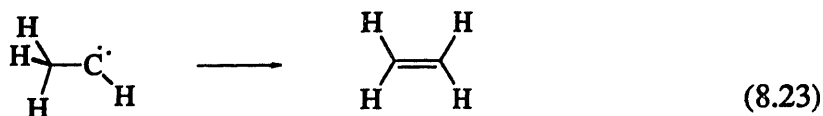
The prototypical carbene is methylene, **8.18**, a ground-state triplet with $\Delta\mathcal{V}_{s,t} = -63$ mJ (Schaefer 1986).



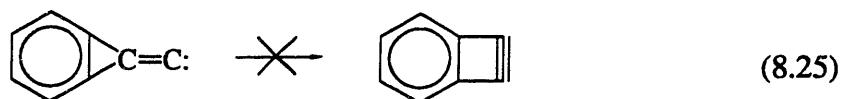
Adamantylidene **8.19** is thought to have a triplet ground state (Moss and Chang 1981). Carbenes with better steric properties (i.e., with supporting structures occupying a smaller solid angle) tend to be ground-state singlets: reducing the carbene bond angle stabilizes the singlet state, and the limiting case of a double bond does so very effectively. In cyclopropenylidene **8.20** the predicted value of $\Delta\mathcal{V}_{s,t}$ is ~ 490 mJ (Lee, Bunge et al. 1985); in vinylidene **8.21**, a prototype for alkylidenecarbenes **8.22**, $\Delta\mathcal{V}_{s,t} \approx 320$ mJ (Davis, Goddard et al. 1977); and in cumulenylidenecarbenes **8.23**, $\Delta\mathcal{V}_{s,t} > 400$ mJ (based on studies of odd-numbered carbon chains (Weltner and van Zee 1989)). These gaps are consistent with reliable singlet behavior in a molecular manufacturing context. Since singlet states of carbenes are appreciably more polar than triplets, singlet states can be significantly stabilized by a suitable electrostatic environment. For diphenylcarbene, experiment indicates that the shift from a nonpolar solvent to the highly polar acetonitrile increases $\Delta\mathcal{V}_{s,t}$ by ~ 10 mJ (Eisenthal, Moss et al. 1984). A preorganized environment (Sec. 8.3.3.2) having a polarization greater than that induced in a solvent by the singlet dipole (and doing so without imposing an entropic cost) would increase the stabilization. Substituents including N, O, F, and Cl also tend to stabilize the singlet state.

8.4.5.3. Carbene rearrangements

Carbenes have a substantial tendency to rearrange; alkylcarbenes, for example, readily transform into alkenes:



Among unsaturated carbenes, alkadienylidenecarbenes and cumulenylidenecarbenes have no available local rearrangements at the carbene center, and although vinylidene itself readily transforms to ethyne (indeed, it may not be a potential energy minimum (Hehre, Radom et al. 1986)), the reaction (8.24) is not observed (Sasaki, Eguchi et al. 1983), and related cyclic species, as in (8.24), are presumably yet more stable against this process.

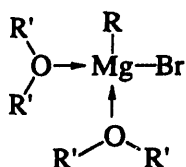


8.4.6. Organometallic reagents

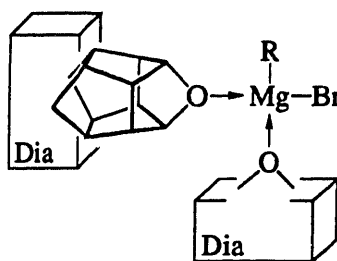
A variety of reagents with metal-carbon bonds are used in organic synthesis; this section briefly discusses only a few classes.

8.4.6.1. Grignard and organolithium reagents

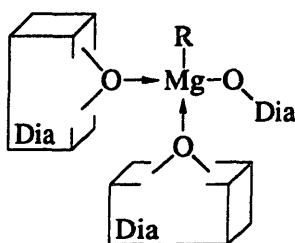
Grignard reagents (RMgX) and organolithium reagents (RLi) find extensive use, providing weakly-bonded carbon atoms with a high electron density (Carey and Sundberg 1983b). As is common with organometallic species, their metal orbitals can accept electron pairs from coordinated molecules; this provides options for improving stability and mechanical manipulability. For example, in an ether solution a typical Grignard reagent structure is 8.24; organometallic species such as 8.25–8.27 are typical of species that might be used in mechanosynthesis.



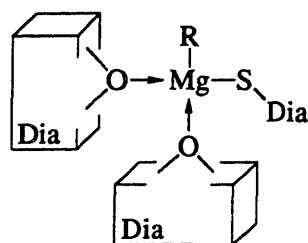
8.24



8.25



8.26



8.27

8.4.6.2. Transition metal complexes

Complexes containing a *d*-block transition metal atom exhibit versatile chemistry; such complexes are prominent in catalysis, including reactions that make and break carbon-carbon bonds. The presence of accessible *d* orbitals in addition to the *p* orbitals available in first-row elements changes chemical interactions in several useful ways: the orbital-symmetry constraints of reactions among first-row elements are relaxed, and bonded structures can have six or more ligands (rather than four); further, the relatively long bonds (typical M–C lengths are ~0.19–0.24 nm) reduce steric congestion, thereby facilitating multi-component interactions. (Longer bonds result in coordination shells with areas ~1.6–2.6 times larger than those of first-row atoms). Many transition metal complexes readily change their coordination number and oxidation state in the course of chemical reactions. Electronic differences among transition metals are large; complexation further increases their diversity. Multi-metal-atom clusters can approximate the reactivity of metal surfaces, which also find widespread use in catalysis.

Transition metals in bulky complexes will be useful in reagent preparation and small-molecule processing, rather than in sterically-constrained reagent-application operations. In enfolded sites, ligand arrangements can be determined by mechanical constraints in the surrounding structure and placed under piezochemical control. Further discussion of the mechanochemical utility of these species is deferred to Section 8.5.10.

8.5. Forcible mechanochemical processes

8.5.1. Overview

Section 8.3.3 delineated some fundamental characteristics of mechanochemical processes, giving special attention to the use of mechanical force. Section 8.4 described a variety of reactions and reactive species, weighing their utility in a mechanochemical context. The present section examines a selected set of forcible mechanochemical processes in more detail. It starts by expanding on the discussion of piezochemistry in Section 8.3.3.3, introducing the issue of thermodynamic reversibility. Tensile bond cleavage and hydrogen abstraction are then presented as model reactions and examined in quantitative detail. Selected other reactions (involving alkene, alkyne, radical, carbene, and transition-metal species) are considered, building on results from the cleavage and abstraction models.

8.5.2. General considerations

8.5.2.1. Force and activation energy

Forces in piezochemical processes alter the reaction PES, reducing the activation energy; in some instances, they can eliminate energy barriers entirely, thereby merging initially distinct states. Section 8.3.4.5 calculates that barrier reductions of ~ 70 mJ will suffice to convert reactions that take 10^3 s in solution into reactions that complete reliably in 10^{-7} s. How much force is required to have such an effect?

Initial motions along a reaction coordinate typically resemble either the stretching of bonds or the compression of nonbonded contacts. As these motions continue, the resisting forces increase, but (usually) not so rapidly or so far as they would in simple bond cleavage or in the compression of an unreactive molecular substance. The potential energy curve instead levels off, passes through a transition state, and falls into another well. Since the energy stored in a given degree of freedom by a given force is proportional to compliance, the energy stored by a force applied through a simple bonded or nonbonded interaction will usually be lower than that stored in a reactive system.

As shown by Figure 3.7, a compressive load of $\sim 2\text{--}3$ nN will store 70 mJ in a nonbonded interaction in the MM2 model; the ~ 30 nN per-bond compressive load in a diamond anvil cell (Sec. 8.3.3.3) is an order of magnitude larger. The energy stored in bonds

is more variable, but for a C–C bond in the Morse model (Table 3.7), a tensile force of ~ 5 nN stores 70 mJ.

It should be noted that pressure in piezochemistry plays a different role from pressure in gas-phase reactions. In the latter, so-called “pressure effects” on kinetics and equilibria have no direct relationship to the applied force per unit area, being mediated entirely by changes in molecular number density and resulting changes in collision frequencies. In the gas phase, the PES describing an elementary reaction process is independent of pressure, since each collision occurs (locally) in vacuum, free of applied forces. Likewise, so-called pressure effects in liquid and solid-surface environments exposed to reactive gases usually result more from changes in molecular number density than from piezochemical modifications to the PES of the reaction.

8.5.2.2. Applied forces and energy dissipation

When actuation times are relatively long ($\sim 10^{-6}$ s), acoustic radiation from time-varying forces (Sec. 7.2) will be minimal, as will be free-energy losses resulting from potential-well compression (Sec. 7.5), given reasonable values for critical stiffnesses. Likewise, with small displacements (~ 1 nm) and low speeds ($\sim 10^{-3}$ m/s), phonon-scattering losses (Sec. 7.3) will be small. In an elementary reaction process, the most significant potential source of dissipation will be transitions among time-dependent potential wells (Sec. 7.6).

Although the issue is distinct from the basic qualitative question of mechanosynthesis (i.e., what can be synthesized?), minimizing energy losses is of practical interest. Losses can broadly be divided into three classes: (1) those that are many times kT , resulting from the merging of an occupied high-energy well with an unoccupied low-energy well; (2) those on the order of kT , resulting from the merging of wells of similar energy; and (3) those of negligible magnitude, resulting from the merging of a low-energy, occupied well with a high-energy, unoccupied well. The simplest way to achieve high reaction reliability is to follow route (1), dissipating ≥ 115 mJ per operation. During forcible mechanochemical processes, however, it will in many instances be possible first to follow route (2) or (3) to a state in which the wells are merged (or rapidly equilibrating over a low barrier), then to use piezochemical effects to transform the PES to a type (1) surface before separation. This yields a process with reliability characteristic of (1), but with energy dissipation characteristic of (2) or (3). A system capable of altering relative well depths by ≥ 125 mJ in mid-transformation can combine error rates $< 10^{-12}$ with an

energy dissipation $< 0.1kT$ at 300 K. Opportunities for this sort of control will be discussed in several of the following sections.

8.5.3. Tensile bond cleavage

Cleavage of a bond by tensile stress is perhaps the simplest mechanochemical process, providing an instance of the conversion of mechanical energy to chemical energy and illustrating the relationship between stiffness and thermodynamic reversibility. Further, tensile bond cleavage plays a role in several of the mechanosynthetic processes described in later sections.

As Figure 6.11 suggests, a typical C–C bond has a relatively large strength. As Table 3.7 shows, k_s for such a bond is lower than that for bonds to several other first-row elements, but higher than that for bonds to second-row elements. In many practical applications, the bond to be cleaved will be of lower strength and stiffness than a typical C–C bond. The latter will be considered in some detail, however, and can serve as a basis for comparison to other bond cleavage processes.

8.5.3.1. Load and strength

The 300 K bond-lifetime curves in Figure 6.11 indicate the tensile loads required to effect rapid bond cleavage. To achieve a level of reliability characterized by P_{err} requires a rate meeting the condition of Eq. (8.12); for a C–C bond in this model, achieving $P_{\text{err}} \leq 10^{-12}$ within $\tau_{\text{trans}} = 10^{-7}$ s requires a tensile load of ~ 4.2 nN. The Morse potential underestimates bond tensile strengths, but the problem of achieving sufficient tensile loads for rapid bond cleavage essentially pits the strength of one bond against that of others, hence errors in estimated strengths approximately cancel.

So long as a carbon atom occupies a site with tetrahedral symmetry, straining one bond to the theoretical zero-K, zero-tunneling breaking point necessarily does the same to the rest. To concentrate a larger load on one bond requires that the angle θ_{bond} be increased (Figure 8.4), thereby increasing the alignment of the back bonds with the axis of stress and reducing their loads. Figure 8.5 illustrates a structure that has a geometry of this sort when at equilibrium without load. With load, however, even an initially tetrahedral geometry will distort in the desired fashion. Increase of θ_{bond} from 109.5° to 115° reduces back-bond tensile stresses to 0.79 of their undistorted-geometry values (neglecting the favorable contributions made by angle-bending forces in typical structures). Breaking of a back bond under these conditions would of necessity be a thermally-

activated process, and the energy barrier for breaking more than one bond at a time would be prohibitive. Moreover, in the structures considered here, breaking of a single bond is strongly resisted by angle-bending restoring forces from the remaining bonds (a ~ 0.1 nm, bond-breaking deformation is associated with an angle-strain energy of ~ 300 mJ); as has been discussed, this solid-cage effect invalidates the model used in Section 6.4.4.1 and strongly stabilizes structures. These stress and energy differences are more than adequate to ensure a $> 10^{12}$ difference in rates of bond cleavage.

8.5.3.2. Stiffness requirements for low-dissipation cleavage

Transitions between time-dependent potential wells can cause energy dissipation, and the occurrence of distinct wells is associated with regions of negative stiffness in the potential energy surface. The potential energy for a pair of atoms undergoing bond cleavage can be described as the sum of (1) the bond energy and (2) the elastic deformation energy of the structures in which the atoms are embedded:

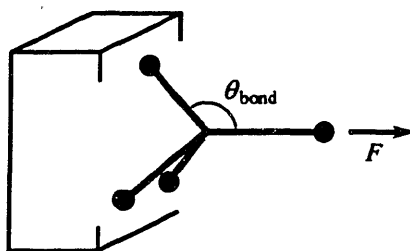


Figure 8.4. Bond angle in a distorted tetrahedral geometry.

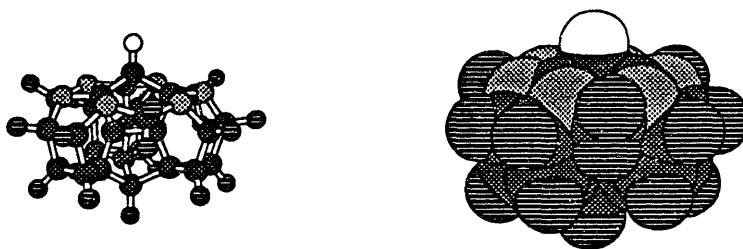


Figure 8.5. A structure having a surface carbon atom with a significantly non-tetrahedral geometry ($\theta_{\text{bond}} = 116.6^\circ$). In this structure, $k_{sz} = 225$ N/m (MM2/C3D) for vertical displacement of the central surface carbon atom with respect to the lattice-terminating hydrogens below (shown in ruled shading); with an approximate correction for compliance of a surrounding diamondoid structure (Sec. 8.5.3.4, Fig. 8.9), $k_{sz} \approx 190$ N/m.

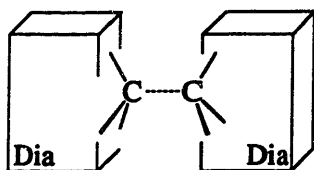
$$\mathcal{V}_{\text{cleave}}(\Delta\ell, \Delta d) = \mathcal{V}_{\text{bond}}(\Delta\ell) + \frac{1}{2}k_{s,\text{struct}}(\Delta d - \Delta\ell)^2 \quad (8.26)$$

where Figure 8.6 and its caption describe Δd and $\Delta\ell$ and the function $\mathcal{V}_{\text{bond}}(\Delta\ell)$ is the bond potential energy. The elastic deformation energy (neglecting modes orthogonal to the reaction coordinate) is a function of the displacement between the bonded atoms and their equilibrium positions with respect to the supporting structure, and is associated with positive stiffness $k_{s,\text{struct}}$. The bond energy, however, is associated with a negative stiffness in the bond-breaking separation range. The Morse potential predicts an extreme value of $-0.125k_s$, where k_s is the stretching stiffness of the bond at its equilibrium length; the Lippincott potential predicts negative stiffnesses of greater magnitude (-115 rather than -55 N/m, for a standard C–C bond), and hence is more conservative in the present context. It is adopted in the following analysis.

Figure 8.7 illustrates potential energy curves as a function of $\Delta\ell$ for various values of Δd for one set of model parameters. As can be seen, a steady increase in Δd causes the evolution of the system from a single well, to a pair of wells, to a single well again; larger values of $k_{s,\text{struct}}$ would reduce and then eliminate the barriers that appear.

At finite temperatures and modest speeds, transitions over low barriers can occur without causing substantial dissipation. The transiently-formed wells will remain in near-equilibrium so long as the barrier between them is low enough that the mean interval between transitions is short compared to the time required for significant changes in relative well depth to occur. A transition rate of 10^9 s^{-1} will ensure low dissipation (small

Structural diagram:



Mechanical model:

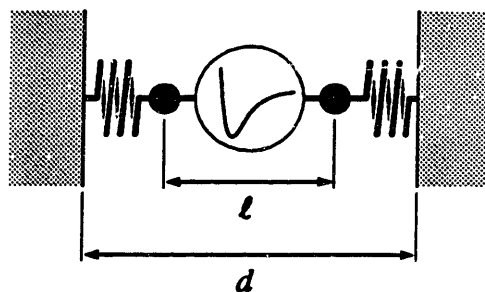


Figure 8.6. Diagrams illustrating tensile bond cleavage and the associated coordinates. The lengths d_0 and ℓ_0 associated with unstrained springs and an unstrained bond determine the values of the coordinates $\Delta d = d - d_0$ and $\Delta\ell = \ell - \ell_0$.

compared to kT) when the characteristic time for the evolution of the wells is $\sim 10^{-7}$ s. Assuming a frequency factor of 10^{13} s^{-1} , this is achieved for barrier heights $\leq 38 \text{ mJ}$ at 300 K.

For estimating energy dissipation, the height of the barrier when the two wells are of equal depth ($\Delta\mathcal{V}_{\pm}^{\ddagger}$) is a conservative measure for the process as a whole (assuming substantial values of $k_{\text{s,struct}}$ to limit the entropic differences between the two wells). Figure 8.8 plots $\Delta\mathcal{V}_{\pm}^{\ddagger}$ as a function of $k_{\text{s,struct}}$ for several bond types. For $k_{\text{s,struct}} > 90 \text{ N/m}$, and characteristic times $\geq 10^{-7}$ s, dissipation will be small compared to kT for all the single bonds shown; for standard C–C bonds, $k_{\text{s,struct}} > 60 \text{ N/m}$ is sufficient.

8.5.3.3. Spin, dissipation, and reversibility

In the absence of intersystem crossing, bond cleavage yields a singlet diradical. In a well-separated diradical, however, the singlet-triplet energy gap typically approaches zero, and thermal fluctuations soon populate the triplet state. If bond cleavage is fast compared to intersystem crossing, this equilibration process results in $\Delta\mathcal{F} = -\ln(2)kT$ (corresponding to the loss of one bit of information). Conversely, if intersystem crossing is fast,

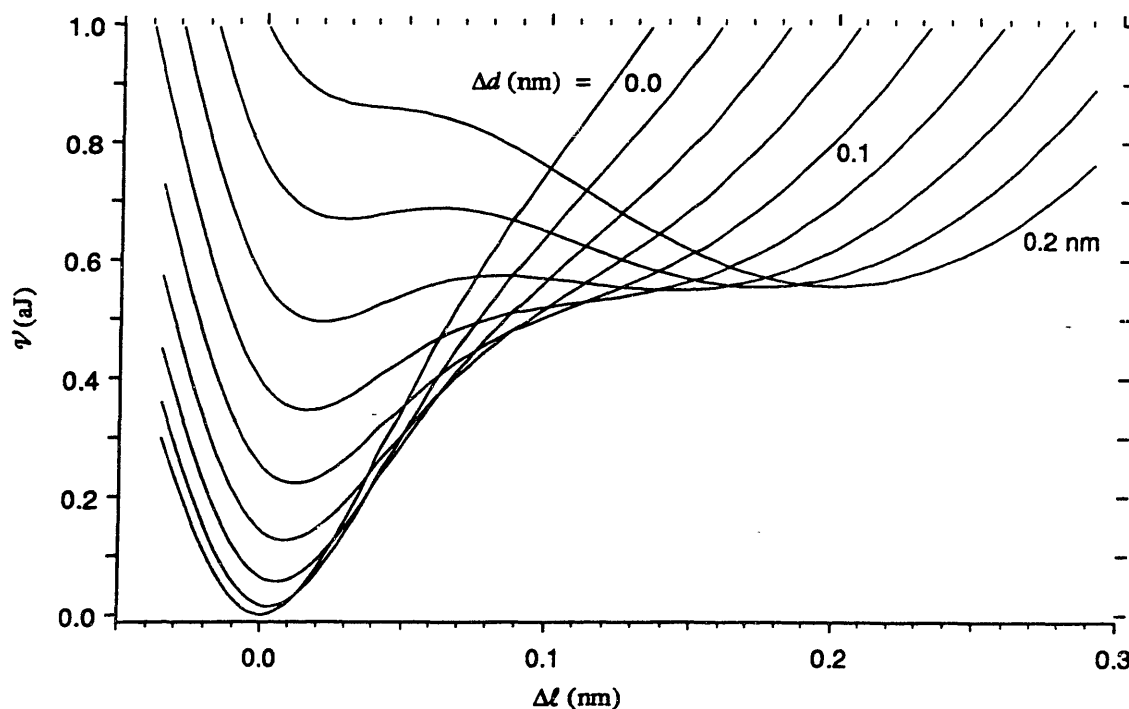


Figure 8.7. Potential energy as a function of CC distance, for several values of support separation. Stiffness of support = 50 N/m.

the thermal population of the (repulsive) triplet state results in a reduction of the mean-force bond potential energy during cleavage, and no significant dissipation. Note that slow intersystem crossing can cause large energy dissipation in mechanically-forced radical coupling, even when the reverse process has a dissipation $< kT$.

8.5.3.4. Atomic stiffness estimation

In the linear, continuum approximation, the z -axis deformation of a surface at a radius r from a z -axis point load is $\propto 1/r$ (Timoshenko and Goodier 1951). Accordingly, most of the compliance associated with displacement of an atom on a surface results from the compliance of the portion of the structure within a few bond radii.

A carbon atom on a hydrogenated diamond (111) surface can be taken as a model for sp^3 carbons on the surfaces of diamondoid structures. The MM2 model value for the z -axis stiffness k_{sz} of such a carbon atom on a semi-infinite lattice can be accurately approximated by measuring the stiffness in a series of approximately-hemispherical, diamond-like clusters of increasing radius (Figure 8.9), holding the lattice-terminating hydrogen atoms fixed. In the continuum model, the compliance of the region outside a

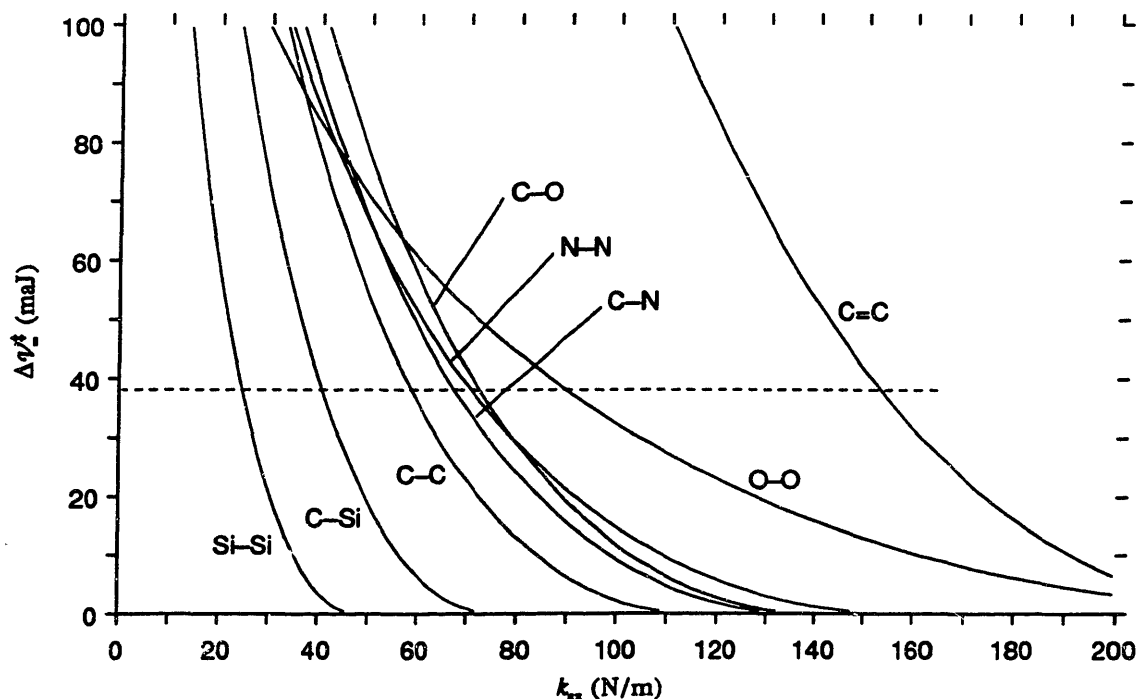


Figure 8.3. Barrier heights vs. stiffness, for various bonds placed under a tensile load which equalizes the well depths (bond parameters from Table 3.7).

hemisphere is $\propto 1/r$; treating the number of carbon atoms as a measure of r^3 and fitting the four cluster-stiffness values with this model yields $k_{sz} = 153 \pm 2$ N/m.

The above holds for small displacements, but at the larger displacements associated with peak forces in the bond-cleavage process, bond angle and lengths are significantly distorted; this affects stiffness. Examination of the energy as a function of z-axis displacements shows that the decrease in bond stretching stiffness resulting from tension is, under loads in the range of interest, more than offset by the increase in k_{sz} resulting from changes in bond angle. Under tensile loads of 3 to 7 nN, the stiffness is increased by a factor of ~ 1.05 . The MM2 model is known to have low bond-bending stiffnesses; increasing these stiffness values by a factor of 1.49 (to approximate MM3 results) increases k_{sz} by a factor of 1.14, resulting in an overall estimate of $k_{sz} \approx 183$ N/m.

Since the compliances of the two bonded atoms are additive, $k_{s,struct} = k_{sz}/2 \approx 90$ N/m. This significantly exceeds the required stiffness for low-dissipation cleavage of all the single bonds in Figure 8.8, save (surprisingly) for O–O. For the C–C bond, available stiffness exceeds the requirement by a factor of 1.5; accordingly, low-dissipation bond cleav-

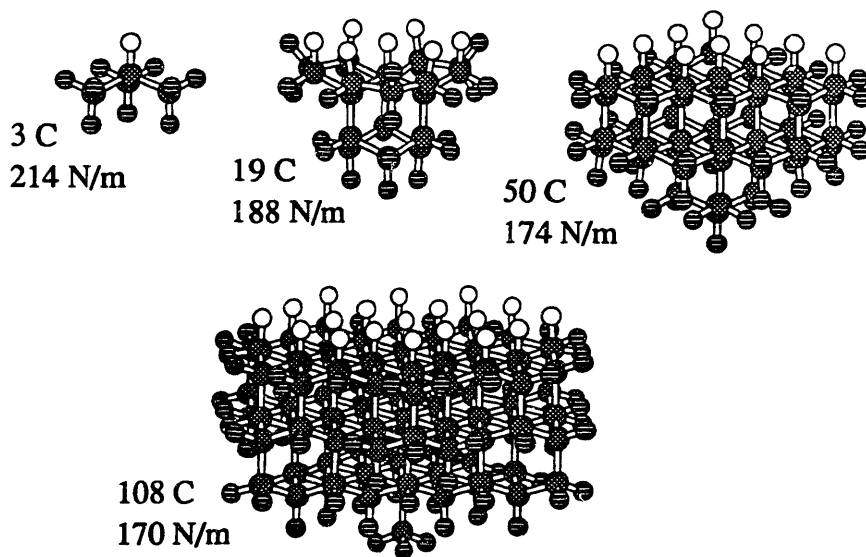
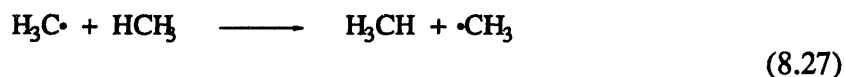


Figure 8.9. A series of structures modeling the diamond (111) surface, with associated values of k_{sz} (MM2/C3D) for vertical displacement of the central surface carbon atom with respect to the lattice-terminating hydrogens below (shown in ruled shading). Note that this carbon atom has poor steric exposure, but that a better-exposed carbon atom need not sacrifice stiffness (e.g., Fig. 8.5).

age should be a feasible process in a broad range of circumstances.

8.5.4. Abstraction

Abstraction reactions can prepare radicals for use either as tools or as activated workpiece sites. Although a variety of species are subject to abstraction, the present discussion will focus on hydrogen. Exoergic or energetically-neutral hydrogen abstraction reactions typically have activation energies ≤ 100 mJ (Bérces and Márta 1988). The abstraction reaction



can serve as a model for reactions involving more complex hydrocarbons, including diamondoid moieties. This process has a substantial energy barrier, ~ 100 mJ (Wildman 1986), making it representative of relatively difficult abstraction reactions.

8.5.4.1. Abstraction in the extended LEPS model

The piezochemistry of abstraction reactions can be modeled using the extended LEPS potential, Eq. (3.24). This three-body potential automatically fits the bond energies, lengths, and vibrational frequencies of each of the three possible pairwise associations of atoms. In a symmetrical process such as (8.27), the two independent Sato parameters can be used to fit the barrier height and the CC separation, $r(\text{CC})$, at the linear, three-body transition state (calculated to be 0.2669 nm for reaction (8.27) using high-order *ab initio* methods (Wildman 1986)). Fitting a LEPS function to these values using parameters based on standard bond lengths and energies for methane and ethane (Kerr 1990) and bond stretching stiffnesses from MM2, yields Sato parameters of 0.132 for the two CH interactions and -0.061 for the CC interaction. (The pure LEPS potential predicts deflection of the hydrogen atom from the axis at small CC separations, but a modest H-C-H angle-bending stiffness suffices to stabilize a linear C-H-C geometry.)

With this function in hand, it is straightforward to evaluate the energy barrier $\Delta\mathcal{V}^\ddagger$ as a function of the compressive load F_{compr} as shown in Fig. 3.10. In this model, $\Delta\mathcal{V}^\ddagger = 0$ at $F_{\text{compr}} \approx 3.6$ nN and $r(\text{CC}) \approx 0.242$ nm; this load is well within the achievable range. At ordinary temperatures, however, there is little practical reason to drive $\Delta\mathcal{V}^\ddagger$ to zero (indeed, with tunneling, there is little reason to do so even at cryogenic temperatures). The barrier is reduced to $2kT_{300}$ at a load of 2 nN, with $r(\text{CC}) \approx 0.251$ nm at the transition state. At a load of 1 nN, $\Delta\mathcal{V}^\ddagger \approx 29$ mJ and $r(\text{CC}) \approx 0.258$ nm. Assuming a frequency

factor of 10^{13} s^{-1} , this barrier would be consistent with a failed-reaction probability $\leq 10^{-12}$ in a mechanochemical system with a transformation time of $\sim 3 \times 10^{-9} \text{ s}$ (if, that is, the reaction were also sufficiently exoergic). These conditions are consistent with achieving high reaction reliability via the optional-repetition mechanism.

8.5.4.2. Exoergic abstraction reactions

A reliable single-step reaction must meet exoergicity requirements (Sec. 8.3.4.5) at the time of kinetic decoupling. Each of the following model reactions is consistent with $P_{\text{err}} < 10^{-12}$ without cycles of repetition and measurement:

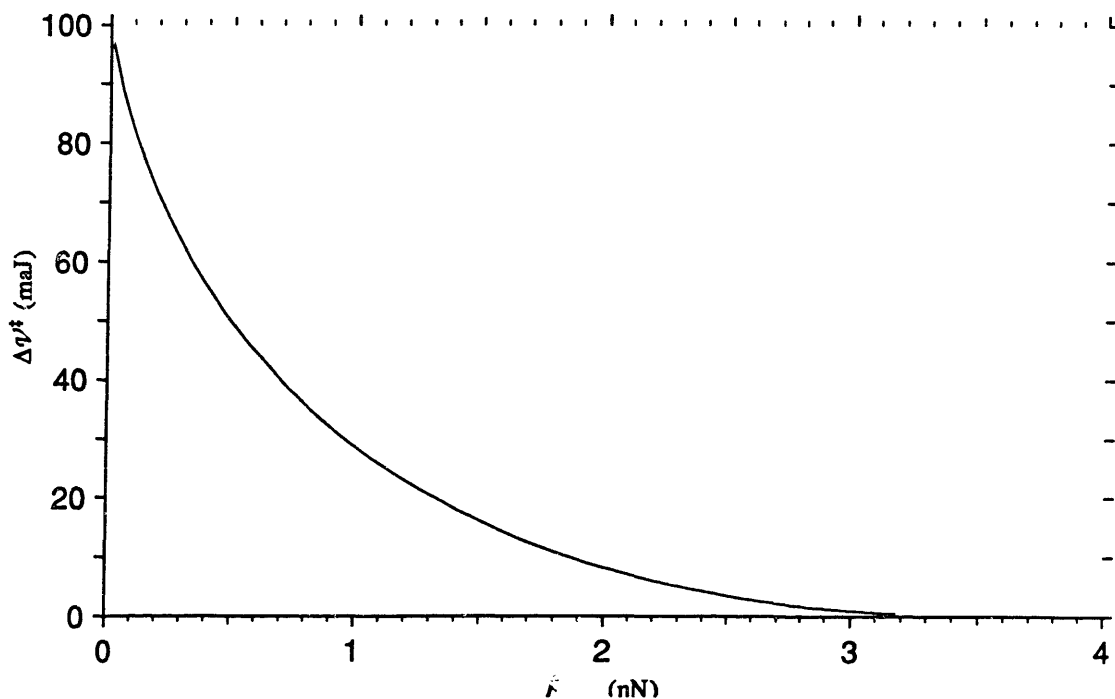
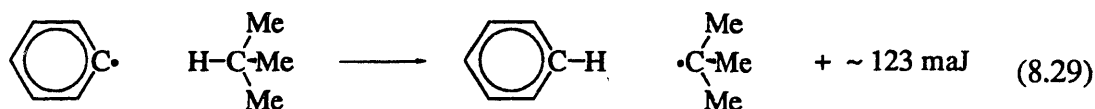
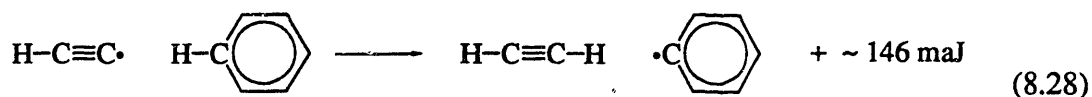


Figure 8.10. Barrier height for abstraction of hydrogen from methane by methyl, plotted as a function of the compressive load applied to the carbon atoms (linear geome-

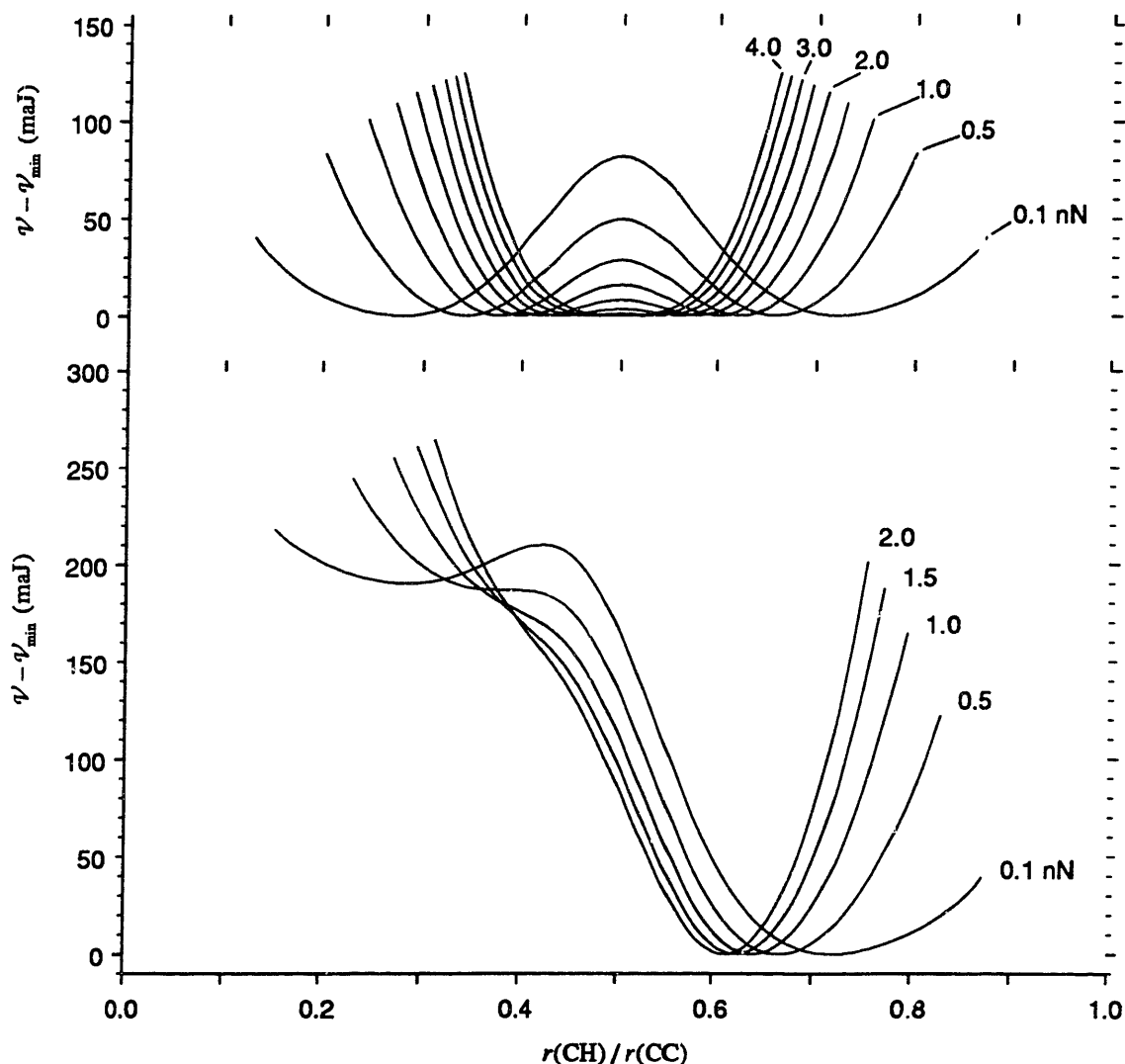
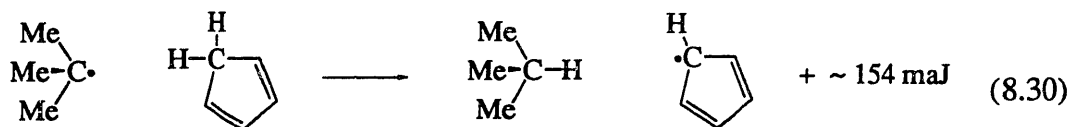


Figure 8.11. Energy surfaces for abstraction reactions under various compressive loads, plotted as a function of a reaction coordinate, the ratio of a CH distance to the (variable, optimized) CC distance. The curves in the upper panel are for the abstraction of hydrogen from methane by methyl, based on the extended LEPS model described in the text. The curves in the lower panel are representative of low-barrier exoergic processes, but do not fit a particular reaction (they result from an extended LEPS model with bond lengths and energies appropriate to the abstraction of hydrogen from methane by ethynyl, and with all Sato parameters arbitrarily = 0.15).



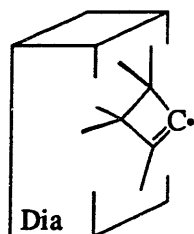
(The above energies are derived from differences in CH bond strengths (Kerr 1990).)

Figure 8.11 compares potential energy curves under compressive loads for the methyl-methane abstraction reaction (discussed previously) to a qualitatively-correct set of curves for a strongly exoergic reaction; with greater exoergicity, the barrier tends to disappear under a substantially smaller load.

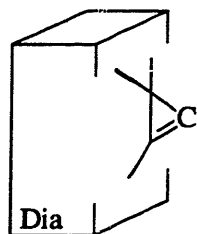
8.5.4.3. Hydrogen abstraction tools

The large C–H bond strength of alkynes (~ 915 kJ) will enable alkynyl radicals to abstract hydrogen atoms from most exposed non-alkyne sites with a single-step $P_{\text{err}} < 10^{-12}$. As shown in Figure 8.2, alkyne moieties can be buttressed by nonbonded contacts to increase their stiffness, thereby minimizing the probability of reacting with sites more than a bond length from the target. In a typical situation, it will suffice to supply supporting structure that resists displacements in a particular direction and permits larger excursions in directions that lack nearby reactive sites. Figure 8.3 illustrates an alkyne moiety that achieves greater steric exposure through selective buttressing.

Moieties with C–H bond energies between those of alkynes and aryls will suffice to abstract hydrogen atoms from most sp^3 carbon sites with good reliability. Strained alkynyl radicals such as **8.28** or **8.29** should bind hydrogen with the necessary energy (≥ 790 kJ, vs. ~ 737 kJ for ethenyl).



8.28



8.29

These structures can exhibit good stiffness without buttressing from nonbonded contacts. Their steric properties are attractive, and their supporting structures strongly suppress bond-cleavage instabilities that would otherwise be promoted by the radical site.

8.5.4.4. Hydrogen donation tools

A variety of structures capable of resonant stabilization of a radical permit easy abstraction of a hydrogen; these include cyclopentadiene, shown in (8.30), and a variety of other unsaturated systems. Some non-carbon atoms bind hydrogen weakly, including tin and a variety of other metals. Moieties like these, with C–H bond energies less than ~ 530 mJ, will be able to donate hydrogen atoms reliably to carbon radicals where the product is an ordinary sp^3 structure. Section 8.6 will accordingly assume that the abstraction and donation of hydrogen atoms can be performed at will on diamondoid structures.

8.5.5. Alkene and alkyne radical additions

Addition of radicals to alkenes and alkynes is typically exoergic by ~ 160 – 190 mJ; this is sufficient to ensure reaction reliability. Because radical addition converts a (long) nonbonded interaction into a (short) bond, compressive loads can couple directly to the reaction coordinate, reducing the activation energy and driving the reaction in the forward direction. Activation energies for alkyl additions to unsaturated species are typically 30–60 mJ (Kerr 1973). Since this barrier is lower than in the methyl-methane hydrogen abstraction reaction, and since piezochemical effects are expected to be larger, modest loads (~ 1 nN) can be expected to reduce the energy barrier sufficiently to permit fast, reliable reactions with $t_{\text{trans}} = 10^{-7}$ s. A semi-empirical study (MNDO method) of the addition of the phenyl radical to ethene yielded a substantial overestimate of the barrier height (~ 98 mJ) and a maximal repulsive force along the reaction coordinate of ~ 1.5 nN (Arnaud and Subra 1982).

In these reactions, the radical created at the adjacent carbon destabilizes the newly formed C–C bond, permitting the addition reaction to be reversed. The energy of the destabilized bond (~ 160 – 190 mJ), together with the barrier on the path between bonding and dissociation (~ 30 – 60 mJ, above), will often make such bonds acceptable with respect to the stability requirements for reagents and reaction intermediates (barriers ≥ 200 mJ). Stability problems can be remedied either by satisfying the radical (e.g., by hydrogen donation) or by mechanically stabilizing the newly-added moiety using other bonds or interactions.

The energy dissipation associated with radical addition (and its inverse) depends on an interplay of stiffness and reaction PES like that examined in Section 8.5.3 for bond cleavage. Obtaining the requisite PES data would make an interesting *ab initio* study.

8.5.6. Alkene torsion

Rotation of one of the methylene groups of ethene by 90° breaks the π -bond, yielding a diradical. Analogous structures and transformations can be used to modulate the strength of adjacent σ -bonds. The transition (a) \rightarrow (b) in Figure 8.12 represents an R–C bond cleavage yielding a twisted alkene; (c) \rightarrow (d) represents an R–C cleavage yielding a planar alkene. (Figure 8.13 illustrates a specific structure with good steric properties.) The differences in alkene energy around the illustrated cycle can be used to estimate the difference in R-group bond energies between (a) and (b). The (a) \rightarrow (c) transition involves torsion of a single bond that links centers of roughly twofold and threefold symmetry; the energy difference will typically be small in the absence of substantial steric interference between the end groups. The difference in R-group bond energies between (a) and (c) will thus approximate the energy difference between twisted and planar alkenes (b) and (d). The analogous energy difference is 0.453 aJ for ethene (Ichikawa, Ebisawa et al. 1985). Alternatively, one can assume that the R-group bond energy in a structure like (a) is unaffected by the presence of an adjacent *twisted* radical (i.e., would be unchanged if the radical site were hydrogenated), then use the bond weakening caused by an *unconstrained* radical to estimate the energy difference; from thermochemical data

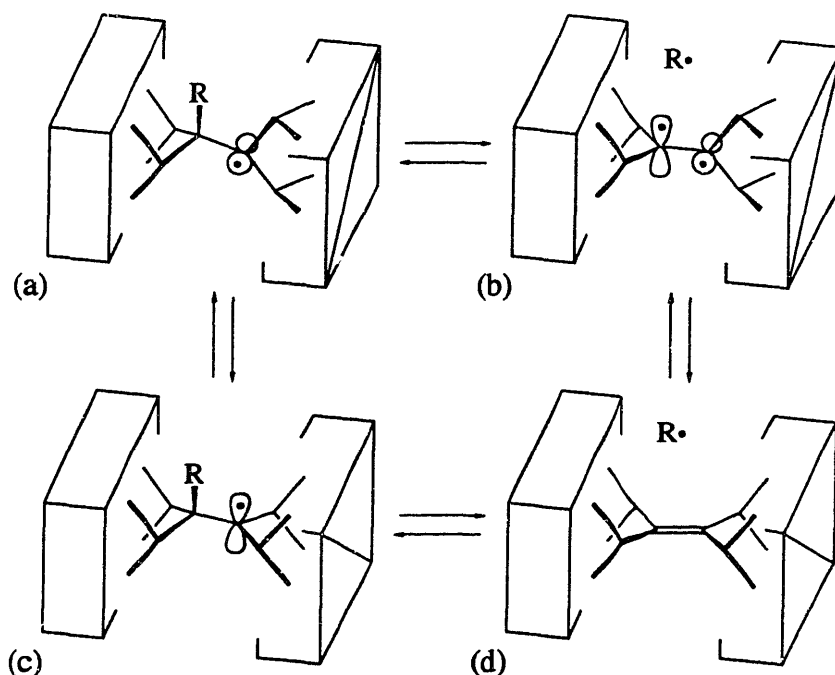


Figure 8.12. Modulation of bond strength by alkene torsion; see text for discussion.

(Kerr 1990), this difference is $\sim 0.410\text{--}0.440$ aJ for hydrocarbon R-groups. Thus, the bond energy of a typical R-group in (a) (~ 0.55 to 0.65 aJ) drops dramatically when the structure is twisted to the configuration of (c). Indeed, elimination of a more weakly bonded group at this site can become exoergic.

A typical reaction cycle for a mechanism of this kind could involve (1) abstraction of a relatively tightly-bound moiety by a radical site like that in (b), yielding a structure like (a), (2) torsion to a state like (c) weakening the new bond, and (3) abstraction of the moiety by another radical, delivering it to a more weakly-bound (i.e., high-energy) site and leaving a structure like (d). If steps (1) and (2) are exoergic by 115 mJ, then the net increase in the energy of the transferred moiety can be ≥ 180 mJ. Moreover, by modulation of the torsion angle during (rather than between) reaction steps, the transition states for steps (1) and (2) can be approached forcibly under conditions that would make the transitions ergoneutral, with separation under exoergic conditions. This meets the conditions stated in Section 8.5.2.2 for a low-dissipation process. (Note that in the latter process, the isolated twisted-alkene state (b) never occurs.) Alternatively, using the optional

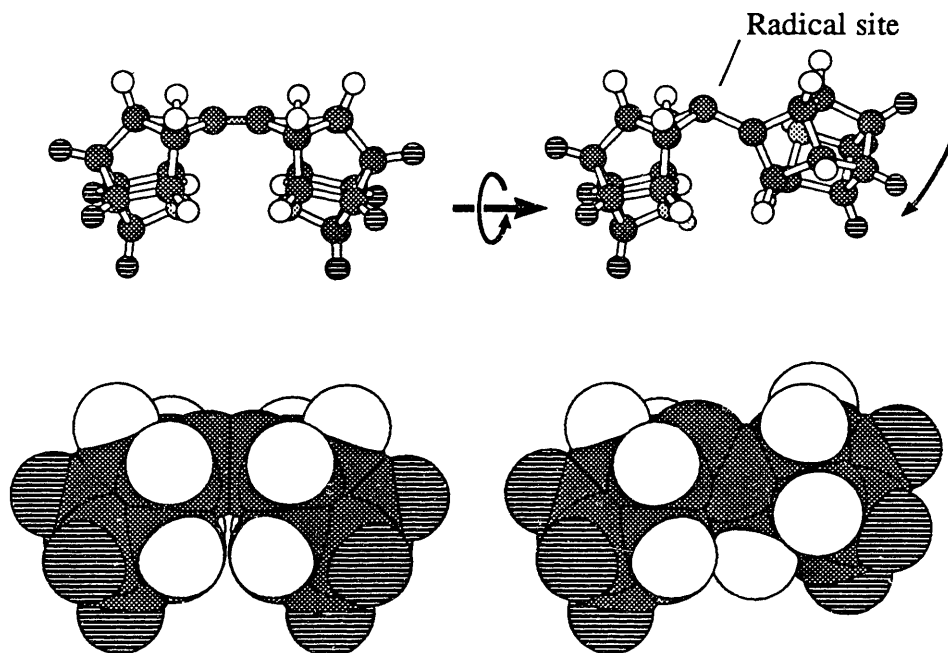


Figure 8.13. A structure suitable for imposing torsion on an alkene while maintaining good steric exposure at one of the carbon sites (pyramidalization of the radical site increases the change in energy).

repetition approach to remove exoergicity requirements, the net increase in the energy of the transferred atom can be > 400 mJ.

Note that abstraction of a moiety by a radical to yield an alkene resembles radical coupling: it requires spin pairing, raising questions of intersystem crossing rates. As discussed in Section 8.4.4.2, k_{isc} will typically be adequate in the presence of a suitably coupled high- Z atom. Bismuth ($Z = 83$), with its ability to form three (albeit weak) covalent bonds, is a candidate for inclusion at a nearby site in the supporting diamondoid structure.

Mechanochemical processes involving alkene torsion appear to have broad applications. For example, reactions of dienes (e.g., Diels-Alder and related reactions) can be accelerated by torsions that weaken the initial pair of double bonds. In the reverse direction, the sigma bonds resulting from the reaction can (for reactions yielding non-cyclic products) later be cleaved with the aid of radicals generated by torsion of the new double bond.

8.5.7. Radical displacements

A variety of mechanochemical processes are analogous to alkene torsion in that they modulate the strength of a σ -bond by altering the availability of a radical able to form a competing π -bond. For example, the addition of a radical to an alkyne can facilitate an abstraction reaction as shown in Figure 8.14.

A process of this sort could employ a weakly-bonding moiety (or one with modifiable bonding, as in Section 8.5.6) as the attacking radical R, generating a strong alkynyl radical by tensile bond cleavage (step (d) \rightarrow (e)). A mechanochemical cycle based on these steps can first use an alkynyl radical to abstract a tightly-bound atom from one location, then deliver the atom to a moderately-bound site R', and finally regenerate the original alkynyl radical. Given the strong weakening of the bond to hydrogen in (b), each of these steps should proceed rapidly and reliably under moderate mechanical loads. (Note that step (b) \rightarrow (c) requires spin pairing.) Analogous displacement operations can be used to regenerate alkene- and aryl-derived sigma radicals (e.g., 8.16, 8.28, 8.29).

8.5.8. Carbene additions and insertions

As noted in Section 8.4.5.1, the standard carbene addition and insertion reactions have low barriers. The changes in geometry resulting from these reactions show that mechanical loads can be directly coupled to the reaction coordinate, resulting in strong piezochemical effects; the increase in bond number points to the same conclusion. The

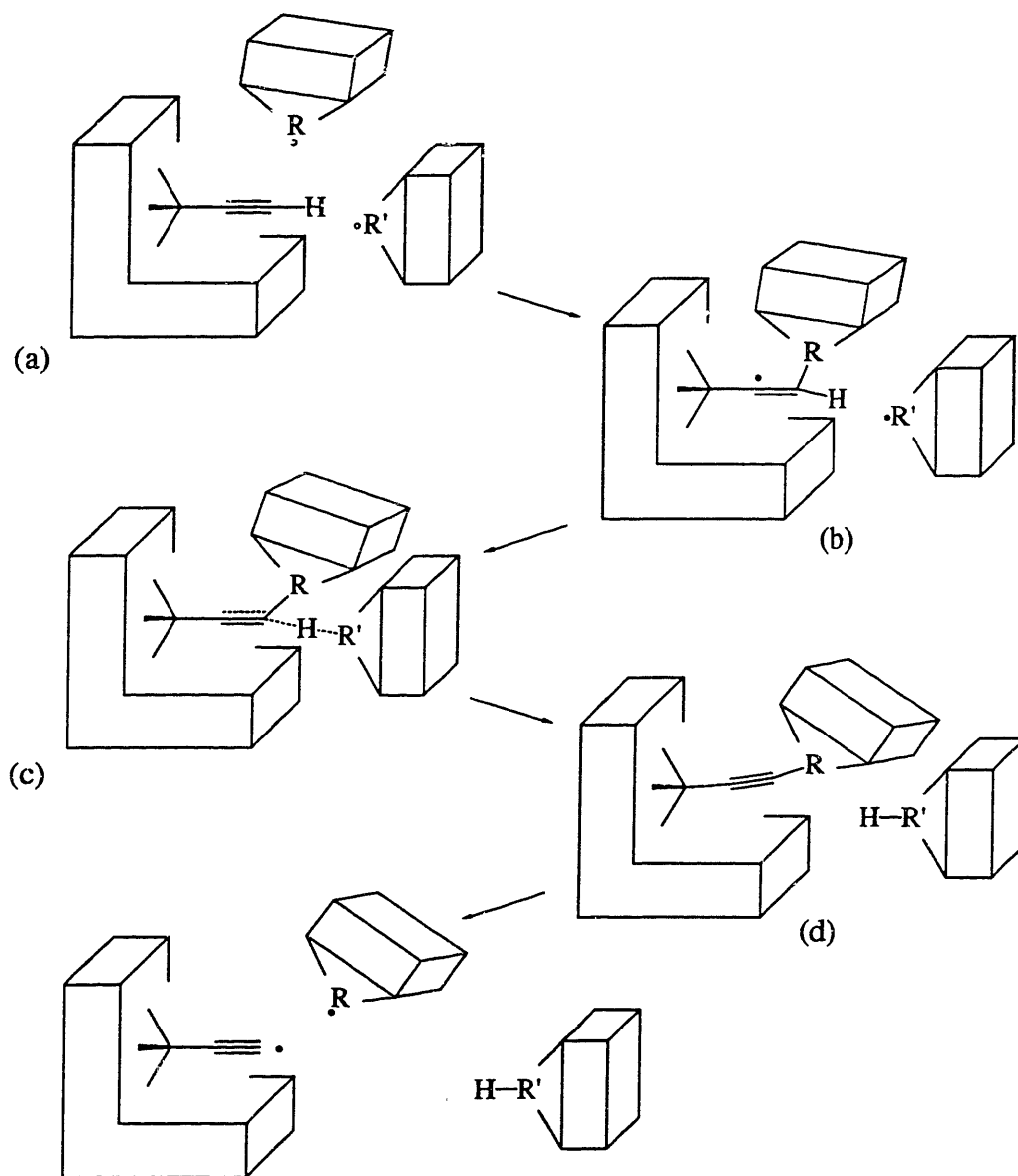


Figure 8.14. Removal of a tightly bound alkynyl hydrogen atom, facilitated by the addition of a radical. A supporting structure provides nonbonded contacts to one side of the alkyne; this enables the application of force to accelerate the radical addition step, (a) \rightarrow (b). See text for discussion.

insertion of carbenes into C–C bonds, although analogous to metal insertion reactions (8.35), has not been observed in solution-phase chemistry; the ubiquitous presence of alternative reaction pathways with less steric hindrance, lower energy barriers, and greater exoergicity is presumably responsible, since this reaction is exoergic and permitted by orbital symmetry. Section 8.6.4.3 presents a reaction step which assumes that positional control and mechanical forces can effect carbene insertion into a strained, sterically-exposed C–C bond.

Since bond-forming carbene reactions are highly exoergic, single-step reactions can be highly reliable (assuming, as always, that conditions and mechanical constraints are chosen to exclude access to transition states leading to unwanted products). Fast, reliable reactions involving triplet carbenes will commonly require fast intersystem crossing. The potential for low-dissipation carbene reactions is presently unclear.

8.5.9. Alkene and alkyne cycloadditions

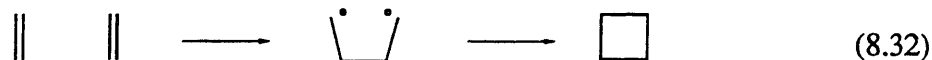
Cycloaddition reactions will likely find extensive applications in mechanosynthesis. The [4+2] Diels–Alder reactions (e.g., Eq. 8.15) form two bonds and a ring simultaneously, and as mentioned, have large volumes of activation at moderate pressures (~ 0.05 – 0.07 nm^3). These reactions will be more sensitive to piezochemical effects than are abstraction reactions. Energy barriers vary widely, from < 60 to $> 130 \text{ mJ}$. Reaction (8.15) proceeds at low rates at ordinary temperatures and pressures; piezochemical rate accelerations under modest loads should make them consistent with $t_{\text{trans}} = 10^{-7} \text{ s}$. Under these conditions, alkenes can be replaced with less-reactive alkynes, yielding less saturated products. Exoergicity ($\sim 280 \text{ mJ}$ for Eq. 8.15) is sufficient to yield stable intermediate products in a reliable process.

Cycloaddition reactions are subject to strong orbital-symmetry effects. The [2+2] cycloaddition reaction



is termed “thermally forbidden” (according to the Woodward–Hoffmann rules), because the highest-energy occupied orbital on one molecule fails to mesh properly with the lowest-energy unoccupied orbital on the other (the latter presents two lobes of opposite sign, divided by a node; the latter has no such node, and creating the required node is akin to breaking a bond). But, as with well-observed “forbidden” spectral lines, the ban is not

complete. Piezochemical techniques (including alkene torsion and direct compression) will suffice to force the formation of one of the new σ -bonds, even at the expense of breaking both π -bonds, thereby creating a pair of radicals that can then combine to form the second σ -bond:



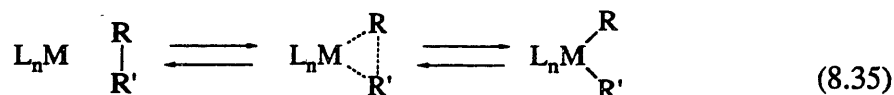
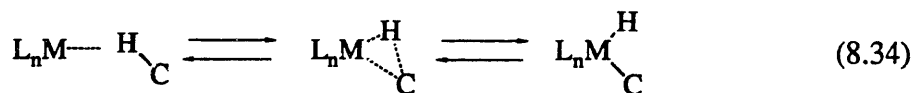
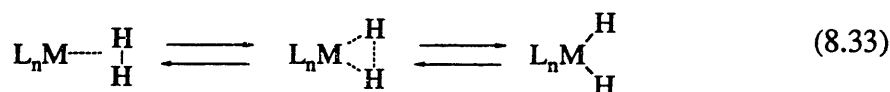
This reaction can also proceed through a polar, asymmetric intermediate, which is likewise subject to strong piezochemical effects. For the dimerization of ethene, the activation energy is ~ 305 mJ (Huisgen 1977).

8.5.10. Transition metal reactions

Section 8.4.6 surveyed some of the general advantages of transition metal compounds as intermediates in mechanosynthetic processes. The present section describes some reactions and mechanochemical issues in more detail.

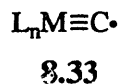
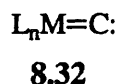
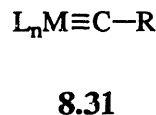
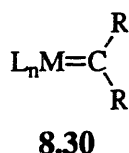
8.5.10.1. Reactions involving transition metals, carbon, and hydrogen

Transition metals participate in a wide variety of reactions, including many that make or break carbon-carbon and carbon-hydrogen bonds:



All illustrated states and transformations in each sequence have been observed, though not necessarily in a single complex (Crabtree 1987).

The above complexes have single bonds between metal and carbon, but double-bonded species ("metal carbenes," 8.30) and triple-bonded species ("metal carbynes," 8.31) are also known.



Species such as metal-carbene carbenes, **8.32**, and metal-carbyne radicals, **8.33**, can presumably exist (given stable ligands and a suitable eutaxic environment) and may be of considerable use in synthesis. (Note that high-Z transition metals will accelerate intersystem crossing in reactions in which they participate.)

Among the reactions of metal carbene complexes are the following:



(Dötz, Fischer et al. 1983; Hehre, Radom et al. 1986; Masters 1981). (Again, all illustrated states and transformations in each sequence have been observed, though not necessarily in a single complex). Transition metals can also serve as radical leaving groups in $\text{S}_{\text{H}}2$ reactions at sp^3 carbon atoms, transferring alkyl groups to radicals (Johnson 1983).

8.5.10.2. Ligands suitable for mechanochemistry

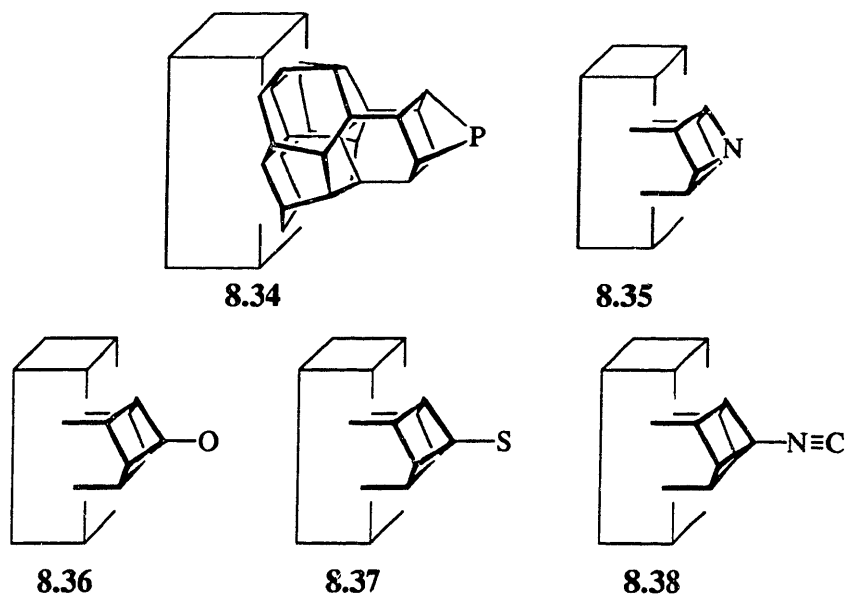
The reactivity of a transition metal atom is strongly affected by the electronic and steric properties of its ligands. These can modify the charge on the metal, the electron densities and energies of various orbitals, and the room available for a new ligand. A ligand can be displaced by a new ligand, or can react with it and dissociate to form a product.

In a molecular manufacturing context, reagent stability will be adequate if all components are bound with energies ≥ 230 mJ. Typical M–C bond strengths (Crabtree 1987) are ~ 210 – 450 mJ (vs. ~ 550 mJ for typical C–C bonds); the stronger bonds will be adequate by themselves, and weaker bonds will be acceptable if incorporated into a stabi-

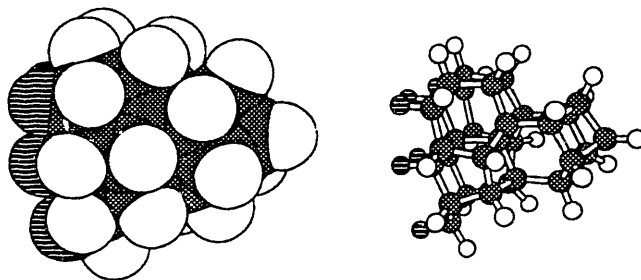
lizing cyclic structure. Stabilization by cyclic structures will often be necessary to prevent unwanted rearrangements in the ligand shell. Since M–H bonds are estimated to have strengths $\sim 100\text{--}200$ kJ greater than M–C bonds (Crabtree 1987), their strengths should be adequate in the absence of special destabilizing circumstances. Electronegative ligands such as F and Cl should also be relatively stable, particularly in structures that lack adjacent sites resembling anion solvation shells in polar solvents.

The stability and manipulability of ligand structures in a mechanochemical context will be greatest when they are bound to strong supporting structures. This is easily arranged for a wide variety of ligands having metal-bonded carbon, nitrogen, oxygen, phosphorus, or sulfur atoms. Of these, phosphorus (in the form of tertiary phosphines, PR_3) has been of particular importance in transition-metal chemistry. Carbon monoxide, another common ligand, lacks an attachment point for such a handle and may accordingly be of limited use. Isonitriles, however, are formally isoelectronic to CO at the coordinating carbon atom, exhibit broadly similar chemistry (Candlin, Taylor et al. 1968), and can be attached to rigid, extended R groups (e.g., **8.38**). (Many ligands not mentioned here also have useful properties.)

Ligand supporting structures can maintain substantial strength and stiffness while occupying a reasonably small solid angle. This will enable several ligands to be subjected to simultaneous, independent mechanochemical manipulation. The following structures provide one family of examples:



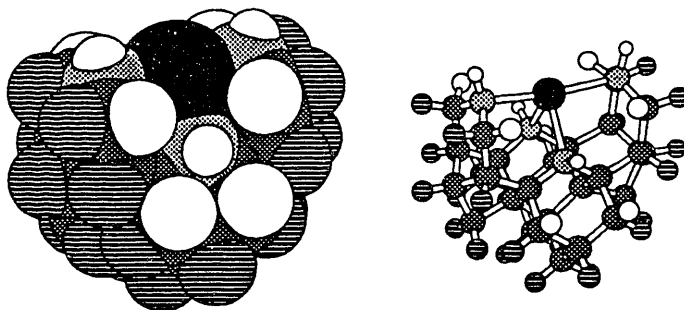
The MM2/C3D stiffness of the following diamondoid support structure (shown in two views):



8.39

is ~ 115 N/m for extension and ~ 20 N/m for bending, both measured for displacements of the carbon atom at the tip relative to the bounding hydrogen atoms at the base.

Alternatively, several ligand moieties can be integral parts of a diamondoid structure, as in the following two views of a bound metal atom with two of six octahedral coordination sites exposed:



8.40

The metal-nitrogen bond lengths in this structure are appropriate for octahedrally-coordinated cobalt.

Single ligands mounted on independently manipulable tips represent one extreme of mobility; multiple ligand moieties in a single rigid structure represent another. An intermediate class would incorporate several ligand moieties into a single structure, subjecting them to substantial relative motion by elastic deformation of that structure. This can assure large inter-ligand stiffnesses, facilitating low-dissipation processes.

8.5.10.3. Mechanically-driven processes

Low-stiffness, low-strength bonds are more readily subject to mechanochemical manipulation: the required forces are smaller, and (for low-dissipation processes) the

required stiffness of the surrounding structure is smaller. The stretching frequencies of M–H bonds (Crabtree 1987) imply stiffnesses in the 130–225 N/m range, ~ 0.25 – 0.50 the stiffness of a typical C–H or C–C bond; the stiffness of M–C bonds (which are longer and of lower energy) is likely to be still lower. Given that bonds as stiff as C–C can be cleaved in a low-dissipation process (Sec. 8.5.3), a wide range of mechanochemical processes involving transition metals can presumably be carried out in a positive-stiffness, zero-barrier manner (or, with similar effect, in a manner encountering only small regions of negative stiffness, and hence only low barriers).

In solution-phase chemistry, catalysts capable of inserting metals into alkane C–H bonds (8.34) have been unstable, either attacking their ligands or the solvent, and insertion into C–C bonds (8.35) has required a strained reagent (Crabtree 1987). With an expanded choice of ligands and the elimination of accessible solvent molecules, the first problem should be avoidable. Further, when loads of bond-breaking magnitude can be applied between a transition metal atom and a potential reagent, intrinsic strain is presumably no longer required in the reagent. Configurations like **8.40** seem well suited for insertion into a sterically-exposed bond. Since metal insertion in the above instances has the effect of replacing a strong, stiff bond with two weaker, more compliant bonds, it can facilitate further mechanochemical operations.

Transition metal complexes with large coordination numbers lend themselves to bond modulation based on manipulation of steric crowding. In octahedral complexes, for example, the metal atom can be anchored by (say) three ligands while two of the remaining ligands are rotated or displaced to modulate steric repulsion on the sixth. By analogy with processes observed in solution, the introduction of new ligands can be used to expel existing ligands by a combination of steric and electronic effects. Mechanochemical processes can forcibly introduce ligands almost regardless of their chemical affinity for the metal, driving the expulsion of other, relatively tightly bound ligands. Conversely, such processes can remove ligands (having suitable “handles”), even when they are themselves tightly bound. Again, the presence of multiple other ligands to anchor the metal atom facilitates such manipulations.

More subtly (and conventionally), the binding of a ligand in a square or octahedral complex can be strongly affected by the electronic properties of the ligand on the opposite side (the *trans*-effect); changes in this *trans* ligand can alter reaction rates by a factor of $\sim 10^4$, suggesting changes in transition state energy of ~ 40 kJ (Masters 1981). Accordingly, mechanical substitution or other alteration of ligands (e.g., double-bond tor-

sions) should be effective in modulating bonding at *trans* sites in mechanochemical reaction cycles.

In summary, although transition metals are of only moderate interest as components of nanomechanical products, they can be of substantial use as components of mechanochemical systems for building those products. Their comparatively soft interactions, relaxed electronic constraints, and numerous manipulable degrees of freedom suit them for the preparation and recycling of reagent moieties, and (where steric constraints can be met) for direct use as reagents in product synthesis. In light of the broad capabilities of other reagents under mechanochemical conditions, the use of transition metal reagents is unlikely to expand the range of structures that can be built; it is, however, likely to expand greatly that range of structures that can be built with low dissipation, in a nearly thermodynamically-reversible fashion.

8.6. Mechanosynthesis of diamondoid structures

Fundamental physical considerations (strength, stiffness, feature size) favor the widespread use of diamondoid structures in nanomechanical systems. "Diamondoid," as used in this volume, refers to a wide range of carbon-rich solids characterized by three-dimensional networks of covalent bonds or (in chemical terms) a wide range of polycyclic organic molecules consisting of fused, conformationally-rigid cage structures. This section considers the synthesis of such structures by mechanochemical means, based on reagents and processes of sorts described in the preceding section, and using diamond itself as an example of a target for synthesis.

8.6.1. *Why examine the synthesis of diamond?*

Diamond is an important product in its own right, but here serves chiefly as a test case in exploring the feasibility of more general synthesis capabilities. It is impractical at present to examine in detail the synthesis of numerous large-scale structures. Accordingly, it is important to choose a few appropriately-challenging models.

Diamond has several advantages in this regard, as can be seen by a series of comparisons. Synthetic challenges often center around the framework of a molecule, and diamond is pure framework. In general, higher valence and participation in more rings makes an atom more difficult to bond correctly. At one extreme is hydrogen placement on a surface; at the other is the formation of multiple rings through tetravalent atoms. (Divalent and trivalent atoms such as oxygen and nitrogen are intermediate cases.) Solid

silicon and germanium present the same topological challenges as diamond, but atoms lower in the periodic table will be more readily subject to mechanochemical manipulation owing to their larger sizes and lower bond strengths and stiffnesses. Thus, it seems that a structure built entirely of rings of sp^3 carbon atoms will tend to maximize the challenges of bond formation, and diamond is such a structure. Further, diamond has the highest atom and bond density of any well-characterized material at ordinary pressures, maximizing problems of steric congestion. Although diamond is a relatively low-energy structure, lacking significant strain or unusual bonds, existing achievements (Sec. 8.2.3) suggest that these features need not be barriers to synthesis, even in solution-phase processes.

Finally, diamond is a simple and regular example of a diamondoid structure. Accordingly, the description of a small synthetic cycle can suffice to describe the synthesis of an indefinitely large object; this avoids the dilemma of choosing between (1) syntheses too complex to describe in the available space, and (2) syntheses that might in some way be limited to small structures.

8.6.2. Why examine multiple synthesis strategies?

The identification of several distinct ways to synthesize a particular structure suggests that ways can be found to synthesize different but similar structures. Identifying multiple syntheses for diamond provides this sort of evidence regarding the synthesis of other diamondoids. Further, diamond itself has several sterically and electronically distinct surfaces on which construction can proceed, and sites on these surfaces can serve as models for the diverse local structures arising in the synthesis of less regular diamondoids. Accordingly, the following sections survey several quite different techniques, not to buttress the case for diamond synthesis (a process already known in the laboratory), but to explore the power of the mechanosynthesis for building broadly diamond-like structures.

8.6.3. Diamond surfaces

Corners and other exposed sites pose fewer steric problems than sites at steps in the middle of planar surfaces. Section 8.6.4 will consider a set of reaction cycles at such sites on low-index diamond surfaces; the present section introduces the surfaces themselves. (In the diagrams here and in the following section, structures are truncated without indicating bonds to missing atoms.)

8.6.3.1. The (111) surface

When prepared by standard grinding procedures, the closely-packed diamond (111) surface is hydrogen terminated (Pate 1986) (Fig. 8.15). When heated sufficiently to remove most of the hydrogen (1200–1300 K), this surface reconstructs into a structure with (2×1) symmetry (Hamza, Kubiak et al. 1988); the Pandey π -bonded chain model for this surface, Fig. 8.16(b), has received considerable support (Kubiak and Kolasinsky 1989; Vanderbilt and Louie 1985). One calculation (Vanderbilt and Louie 1985) yielded an exoergicity of ~ 50 mJ per surface atom for this reconstruction, driven by the conversion of radical electrons to bonding electrons (but offset by strain energy).

It is not presently known whether the bare diamond (111) surface, Fig. 8.16(a) would spontaneously transform to the (2×1) structure at room temperature. Calculations for the analogous transformation of the silicon (111) surface (Northrup and Cohen 1982) suggest an energy barrier of ≤ 5 mJ per surface atom. This energy barrier, however, cannot be equated with an energy barrier for the nucleation of a (2×1) domain on an unreconstructed (1×1) surface, since the latter process will require the simultaneous motion of a number of atoms and will impose an energetic cost associated with deformations at the domain boundary. Since hydrogenation is known to suppress the (2×1) reconstruction, and since the deposition and abstraction of hydrogen atoms from (111) surface sites will

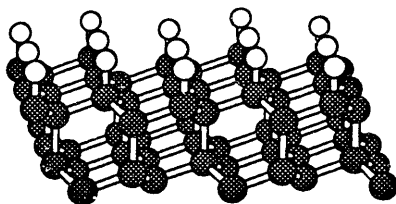


Figure 8.15. A hydrogen-terminated diamond (111) surface.

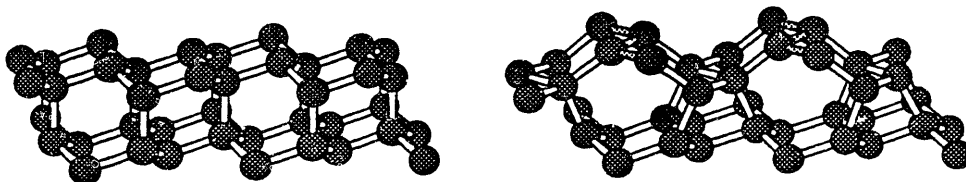


Figure 8.16. A bare, unreconstructed diamond (111) surface (left) and the Pandey (2×1) reconstruction (right)

be straightforward, an unreconstructed (111) surface can be maintained during construction.

8.6.3.2. *The (110) surface*

Termination of the bulk structure leads to a surface like that shown in Figure 8.17. The bonded chains along the surface resemble those formed in the (2×1) reconstruction of the (111) surface, but with the π -bonds subject to greater torsion and pyramidalization. Although this geometry generates no radicals, the π -bonds will be quite weak, and hence the p -orbital electrons can exhibit radical-like reactivity.

8.6.3.3. *The (100) surface*

Termination of the bulk structure would yield a surface like that shown in Figure 8.18(a), covered with carbene sites. Displacement of surface atoms to form π -bonded pairs yields the stable reconstruction shown in Figure 8.18(b) (Verwoerd 1981). (In silicon, a Jahn-Teller distortion causes further buckling of similar, but single-bonded, pairs.) The resulting surface alkene moieties are strongly pyramidalized and under substantial tension, increasing their reactivity.

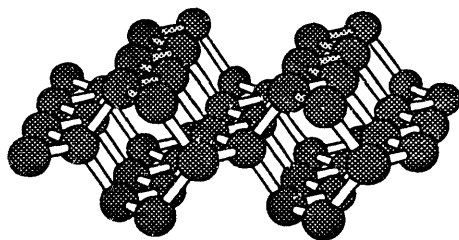


Figure 8.17. An unreconstructed diamond (110) surface.

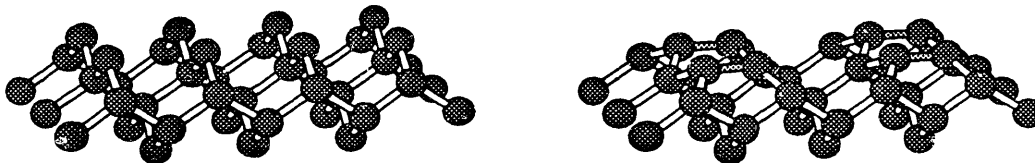


Figure 8.18. A diamond (100) surface, without (left) and with (right) reconstruction.

8.6.3.4. *Some chemical observations*

The diversity of surface structures possible on a single bulk structure, diamond, shows how a diamondoid structure (like most complex molecules) can be built up through intermediates having widely varying chemical properties. Further, the differing strained-alkene moieties found on (110) and (100) surfaces show that stable diamondoid intermediates can have highly reactive surfaces, facilitating synthetic operations. Finally, the (possible) instability of the bare, radical-dense, unreconstructed (111) surface shows how requirements for temporary, stabilizing additions (such as bond-terminating hydrogens) can arise.

8.6.4. *Stepwise synthesis processes*

The next section will describe synthesis strategies that exploit the regular structure of diamond by laying down reactive molecular strands. Synthesis based on the mechanical placement of small molecular fragments, in contrast, suggests how specific *irregular* structures might be synthesized, and thus provides a point of departure for considering the synthesis of general diamondoids.

8.6.4.1. *Existing models of diamond synthesis*

Models of synthesis *via* small molecular fragments have been developed to explain the low-pressure synthesis of diamond under non-equilibrium conditions in a high-temperature hydrocarbon gas, a process of increasing technological importance. Two models have been advanced and subjected to studies using semi-empirical quantum mechanics. Both propose mechanisms for the growth of diamond on (111) surfaces, one based on a cationic process involving methyl groups (Tsuda, Nakajima et al. 1986) and one based on the addition of ethyne (Huang, Frenklach et al. 1988). Of these, the latter appears better supported and more directly relevant to feasible mechanosynthetic processes.

Figure 8.19 (based on illustrations in (Huang, Frenklach et al. 1988)) shows the addition of two ethyne molecules to a hydrocarbon molecule which, with suitable positional constraints, was used to model a step on the diamond (111) surface. The overall set of calculations used MNDO and consumed ~ 35 hours of CPU time on a Cray XMP/48. The reaction mechanism originally proposed (Frenklach and Spear 1988) for the transformations $2 \rightarrow 4$ and $4 \rightarrow 6$ involved multiple steps, later shown to be concerted (steps 3 and 5). The initiation step involves abstraction of a hydrogen atom by atomic hydrogen

(1 \rightarrow 2, process not shown), and has a significant energy barrier. Steps 3 and 5 were calculated to proceed without energy barriers, suggesting that any energy barriers that actually occur are unlikely to be large.

In an analogous mechanosynthetic process, ethyne could be replaced by an alkyne moiety bonded to a structure serving as a handle. Step 5 liberates a free hydrogen atom, which is unacceptable in a eutaxic environment, but a related approach (Sec. 8.6.4.2) avoids this loss of control.

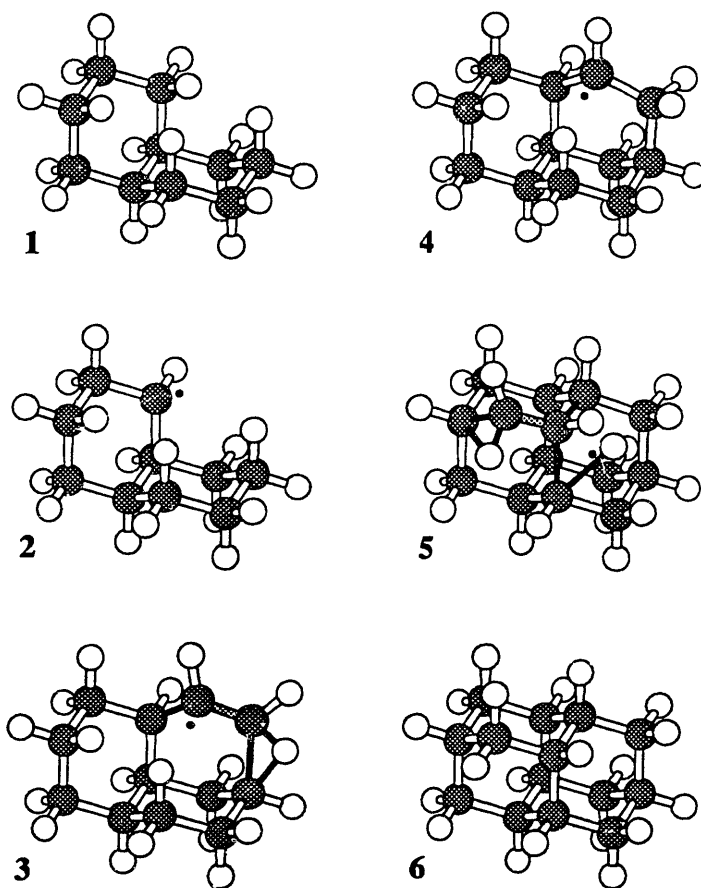
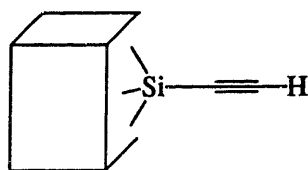


Figure 8.19. A sequence of reaction steps for the addition of ethyne to a compound serving as a model of a step on the hydrogen-terminated diamond (111) surface. Redrawn from (Huang, Frenklach et al. 1988) (see Sec. 8.6.4.1); bonds in the process of breakage and formation are shown in black.

8.6.4.2. Mechanochemistry on (111)

Step 1 of Figure 8.20 illustrates a kink site on a step on the hydrogenated diamond (111) surface; one bond results from a reconstruction, and two hydrogen atoms have been removed to prepare radical sites. In step 2 of the proposed synthetic cycle, an alkyne reagent moiety such as **8.41**



8.41

is applied to one of the radical sites, resulting in a transition structure analogous to step 3 of Figure 8.19. The chief differences are that insertion occurs into a strained C–C bond rather than an unstrained C–H bond, and that large mechanical forces can (optionally) be applied to drive the insertion process. In step 3, tensile bond cleavage occurs (use of Si or another non-first-row atom reduces the mechanical strength of this bond), and a bond forms to the remaining prepared radical. Deposition of a hydrogen at the newly-generated

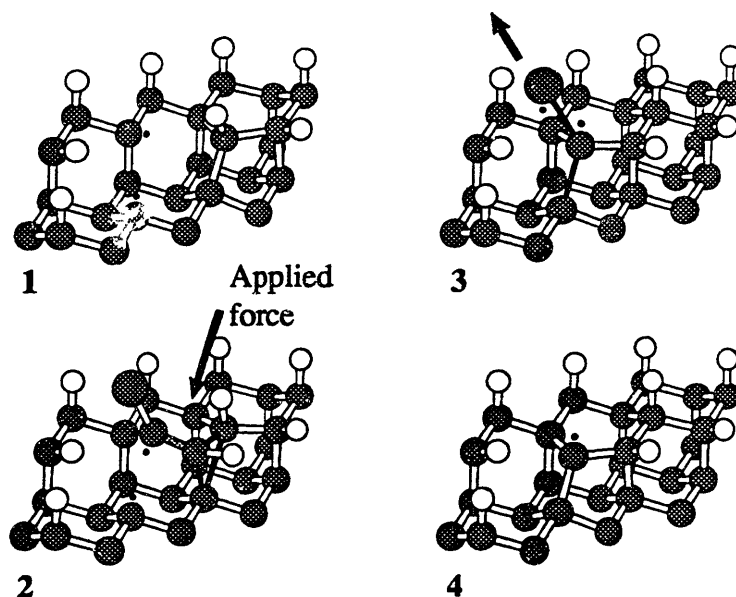


Figure 8.20. A sequence of reaction steps for the addition of a pair carbon atoms from an alkynyl moiety to a kink site in a step on the hydrogen-terminated diamond (111) surface (see Sec. 8.6.4.2).

radical site and abstraction of two hydrogens further along the step then completes the cycle (not shown), generating a kink site like that in step 1, but with the diamond lattice extended by two atoms. Save for accommodating boundary conditions at the edge of a finite (111) surface, this sequence of operations (like those in the following sections) suffices to build an indefinitely large volume of diamond lattice.

8.6.4.3. *Mechanosynthesis on (110)*

Figure 8.21 illustrates the extension of a π -bonded chain along a groove in the diamond (110) surface, using an alkylidenecarbene reagent moiety such as 8.42.

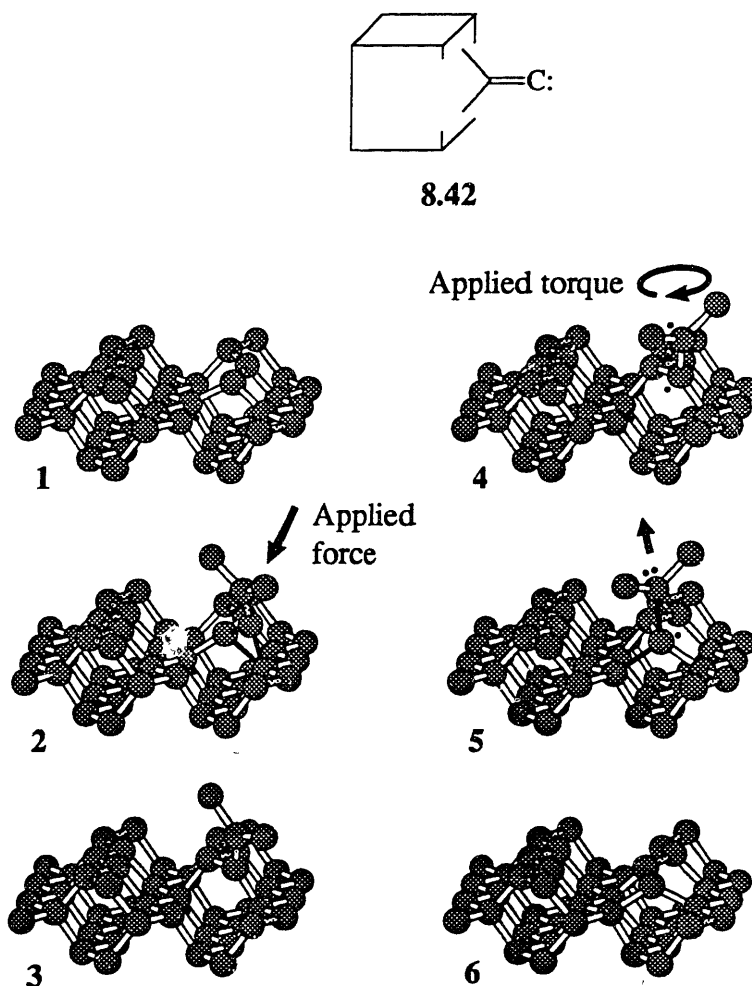


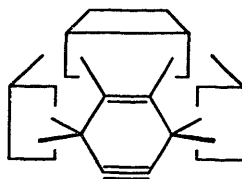
Figure 8.21. A sequence of reaction steps for the addition of a carbon atom from an alkylidenecarbene moiety to a π -bonded chain on the diamond (110) surface (see Sec. 8.6.4.3).

Step 1 illustrates the starting state; step 2 illustrates a transitional state in the insertion of the carbene into a strained C–C bond. This process takes advantage of mechanical constraints to prevent the addition of the carbene into the adjacent double bond to form a cyclopropane moiety. Instead, electron density is accepted from the terminal atom of π -bonded chain into the empty p -orbital of the carbene carbon, developing one of the desired bonds, while electron density is donated from the σ -orbital of the carbene carbon to form the other desired bond (shown in step 3). Substantial forces (≥ 4 nN; Sect 8.3.3.3) can be applied to drive this process, limited chiefly by mechanical instabilities.

Step 4 illustrates the application of a torsion to break a π -bond, thereby facilitating tensile bond cleavage (step 5). The final state, step 6, is (save for the extension of the lattice by one atom) much like that in step 1. A further cycle (restoring the state of step 1 exactly, save for the addition of two atoms) would be almost identical, except that the equivalent of step 5 would involve attack by a newly-forming radical on a weak π -bond, rather than its combination with an existing radical.

8.6.4.4. *Mechanosynthesis on (100)*

Figure 8.22 illustrates a synthetic cycle on diamond (100) in which reactions occur on a series of relatively independent rows of pairs of dimers. In step 1 of Figure 8.22, a strained cycloalkyne reagent moiety such as **8.43**



8.43

is applied. (The division of the supporting structure into blocks is intended to suggest opportunities for modulating bond strength by control of torsional deformations; alternative structures could serve the same role, providing weak bonds for later cleavage.) The reaction in step 1 (promoted by the nearly diradical character of the strained alkyne and by applied mechanical force) is formally a [4+2] cycloaddition process, provided that the resulting pair of radicals is regarded as forming a highly-elongated π -bond. Step 2 is then formally a (thermally-forbidden) [2+2] cycloaddition process, but the small energy difference between the bonding and antibonding orbitals in the “ π -bond” provides a low-lying orbital of the correct symmetry for bonding, hence the forbiddenness should be weak.

Moderate mechanical loads should suffice to overcome the associated energy barrier. Step 4 is an endoergic retro-Diels–Alder reaction yielding a high-energy, highly-pyramidalized alkene moiety (which is, however, less pyramidalized than cubene). The energy for this process is supplied by mechanical work.

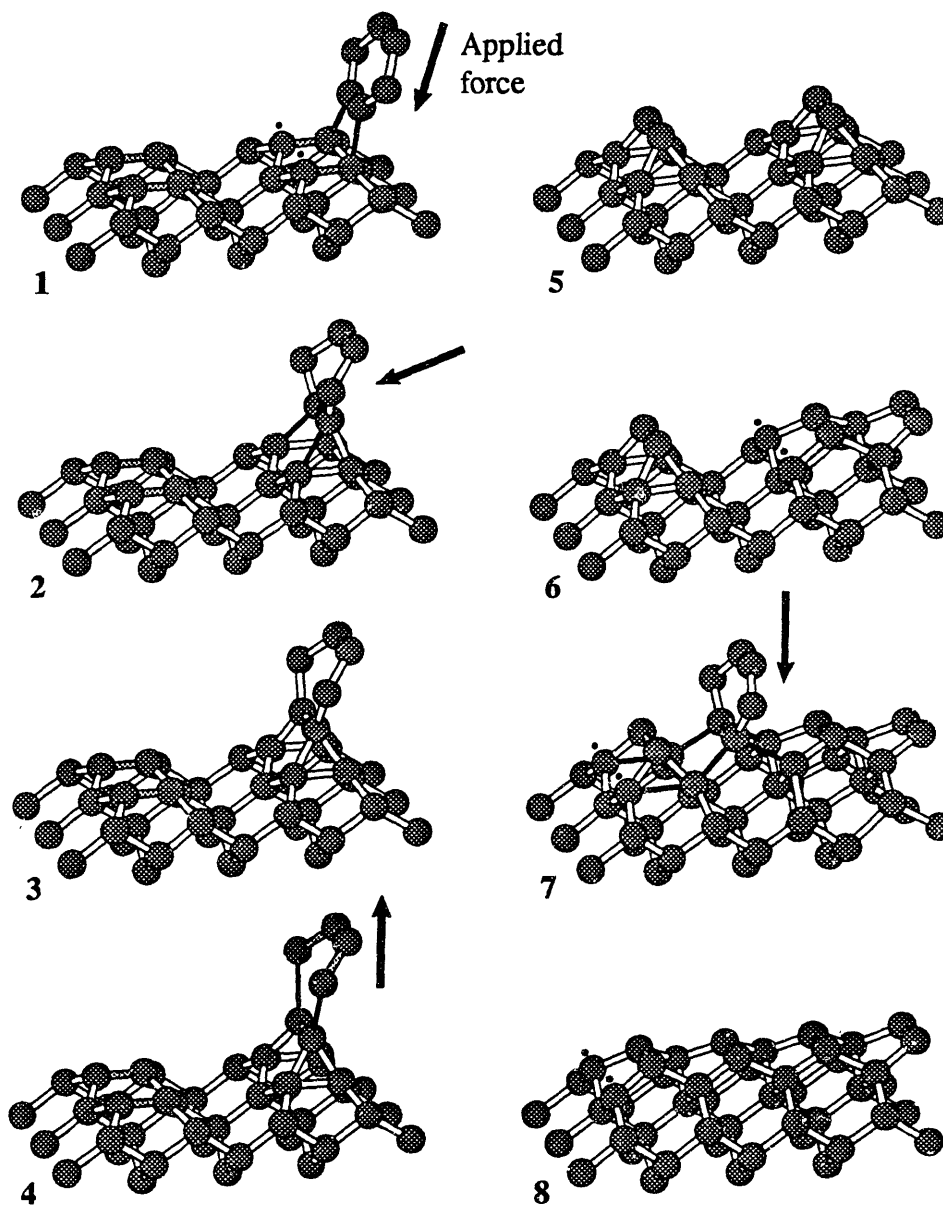


Figure 8.22. A sequence of reaction steps for the addition of pairs of carbon atoms from an strained alkyne moieties to a row of dimers on the diamond (100) surface (see Sec. 8.6.4.4).

Step 5 represents the state of the row after bridging dimers have been deposited at all sites. The resulting three-membered rings are analogous to epoxide structures in a model of the oxidized diamond (100) (2×1) half-monolayer surface (Badziag and Verwoerd 1987).

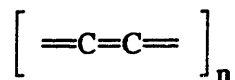
Steps 6–8 represent a cycle in which dimers are sequentially inserted along the row: In step 6, a dimer has already been added to the right, cleaving the strained rings and generating two radical sites adjacent to a cleft in the surface. Step 7 illustrates the bonds that undergo formation and cleavage as a result of the mechanical insertion of a strained alkyne into the cleft. The nature of the transition state will depend on the spin state of the radical pair and orbital symmetry considerations. The large exoergicity for this process can be seen from a comparison of the bonds lost (two π -bonds in the strained alkene and two strained sigma bonds in the three-membered rings) with the bonds formed (four almost strain-free sigma bonds). The geometry of the cleft will permit the application of large loads without mechanical instability, hence large energy barriers (in the unloaded state) would be acceptable. Afterward, the transition to step 8 (equivalent to step 6, but displaced) is achieved by a retro-Diels–Alder reaction like that in step 4.

8.6.5. Strand deposition processes

Stepwise synthetic processes like those suggested in Section 8.6.4 need not be applied in regular cycles to build up a regular structure, but could instead be orchestrated to build up diamondoid structures tailored for specific purposes. Where diamond itself is the target, synthesis can take direct advantage of structural regularities.

8.6.5.1. Cumulene strands

Cumulene strands, **8.44**, are high-energy, pure-carbon structures that represent promising precursors in the mechanosynthesis of diamond.



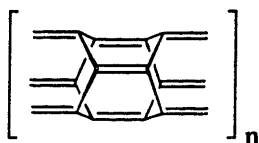
8.44

Figure 8.23 represents a pair of similar reaction processes on a dehydrogenated step on the diamond (111) surface and on a groove in the (110) surface: in each, a p -bonded chain is extended by the formation of additional bonds to a cumulene strand in an exoergic, largely self-aligning process. Substantial forces can be applied through non-bonded

interactions (which can also be used to constrain strand motions). The reaction on the hydrogenated (111) surface would require a series of step-wise reactions both to dehydrogenate atoms in the plane to be covered and to hydrogenate new atoms in the plane being constructed; the (110) reaction has no such requirement.

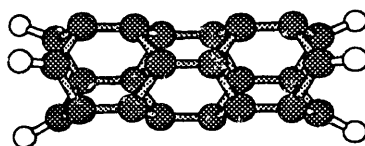
8.6.5.2. Hexagonal diamond from hexagonal strands

The unsaturated, hexagonal, columnar structure **8.45** can be regarded as a tightly-rolled tube of graphite; it could be made from a saturated structure by abstraction of all hydrogens. Like a cumulene strand, **8.44**, this is a pure-carbon structure; owing to pyramidalization and torsion of π systems, it is also relatively high in energy.



8.45

A semi-empirical quantum chemistry study using AM1 on the model structure **8.46** (aided by R. Merkle)



8.46

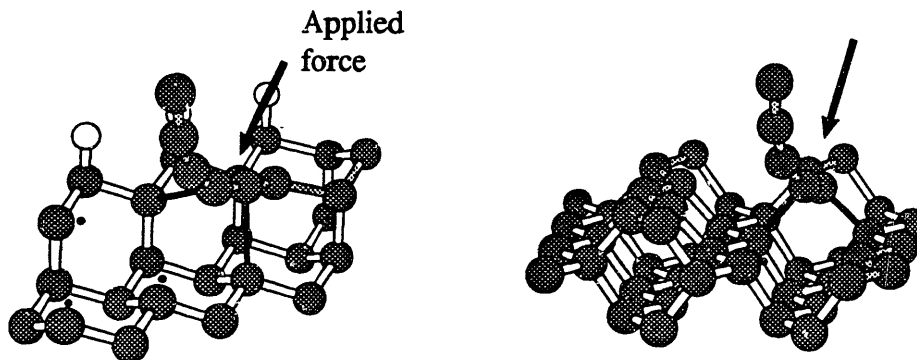


Figure 8.23. Reaction processes bonding cumulene strands to (left) a dehydrogenated step on a hydrogenated diamond (111) surface and (right) π groove on the diamond (110) surface (see Sec. 8.6.5.1).

yielded bond lengths of ~ 0.1384 nm for the three central, axially-aligned bonds, and ~ 1.510 nm for the twelve adjacent bonds. These values are close to standard values for pure double and pure single bonds (0.1337 and 0.1541 nm), hence structure 8.45, representing the hexagonal column as a network of strained alkenes, provides a good description.

Figure 8.24 represents a (100) surface of hexagonal diamond, bounded by similar strained alkenes. Figure 8.25 illustrates a reaction in which a hexagonal column bonds to a groove adjacent to a step on that surface; this can be regarded as proceeding by the attack of a strand radical on a surface alkene, generating a surface radical, which then attacks the next strand alkene, and so forth. Thus, each row of alkenes undergoes a process directly analogous to the free-radical chain polymerization that yields polyethylene, save for the greater reactivity of the participating alkenes and the presence of mechanical forces tending to force each radical addition.

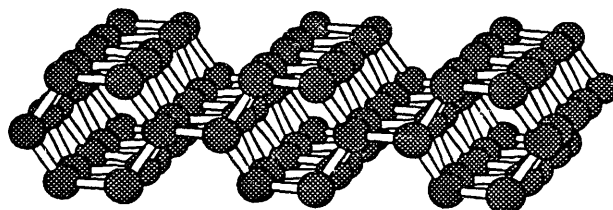


Figure 8.24. A (100) surface of hexagonal diamond.

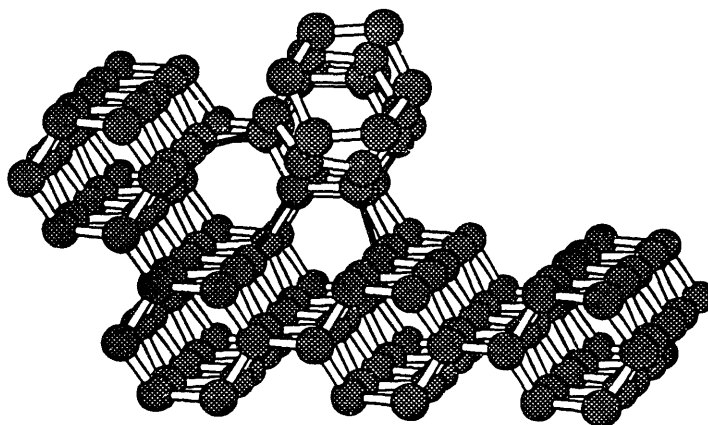


Figure 8.25. A reaction process bonding a tube formed of alkenes to a (100) surface of hexagonal diamond (see Sec. 8.6.5.2).

8.6.6. Cluster-based strategies

Syntheses based on cumulene and hexagonal-column strands suggest the feasibility of synthesizing diamondoid solids using reactive molecular fragments of intermediate size (e.g., 10 to 30 atoms). Fragments of this size can be strongly convex, relaxing steric constraints in their synthesis. Containing tens of atoms, they can embody significant structural complexity and deliver that pre-formed complexity to a workpiece. By incorporating unsaturated structures, radicals, carbenes, and the like, they can form dense arrays of bonds to a complementary surface.

This approach is a form of the familiar chemical strategy of convergent synthesis. Further, the contemplated size range of these fragments is familiar in organic synthesis today; their relatively high reactivity could be achieved (starting with more conventional structures) by a series of mechanochemically-guided abstraction reactions in the protection of a eutaxic environment.

8.6.7. Toward less diamond-like diamondoids

Members of the broad class of diamondoids can differ from diamond both in patterns of bonding and in elemental composition. Regarding the former, the differing bonding patterns created during intermediate stages of the syntheses proposed in this section, together with general experience in chemistry, suggests that non-diamond structures will be readily accessible. Deviations from the diamond pattern generally reduce the overall number density of atoms, thereby reducing steric congestion and (all else being equal) facilitating synthesis.

Regarding differences in elemental composition, it is significant that the classes of reagent species exploited in this section (unsaturated hydrocarbons, carbon radicals, and carbenes) have analogues among other chemical elements. For example, most other elements of structural interest (N, O, Si, P, S), all can form double (and sometimes triple) bonds (e.g., $C=N$, $C\equiv N$; $C=O$; $Si=Si$, $C=Si$ (Raabe and Michl 1989); $C=P$, $C\equiv P$ (Corbridge 1990); $C=S$). All can form bound radical sites. Silicon can form a carbene-like divalent species, silene (Raabe and Michl 1989). Of the monovalent elements of interest, fluorine, chlorine, and bromine can, like hydrogen, participate in abstraction reactions. The focus on hydrocarbon structures in this chapter has been driven more by limits on time and page space than by limits on chemistry.

The diversity of feasible syntheses for a challenging test-case, diamond, suggests that most reasonably-stable diamondoid structures will prove susceptible to mechanosynthesis. Part II of this volume will proceed on the assumption.

8.6.8. Mechanosynthesis of non-diamondoid structures

As we move away from diamondoids, while remaining within the class of reasonably-stable structures, synthesis appears to grow easier for reasons of the sort discussed in Section 8.6.1. Within the class of covalent structures, reduced rigidity might hamper mechanosynthesis, but the feasibility of using bound intermediates (using either covalent bonds or non-bonded interactions) can largely compensate. Further, the synthesis of flexible covalent structures is in the mainstream of existing chemical achievement.

Non-covalent solids present different challenges. In general, however, the ability to transfer atoms one at a time or in small clusters, and to perform piezochemical manipulation on a growing surface, will suffice to provide broad control. Accordingly, it seems reasonable to assume that most reasonably-stable structures, diamondoid or not, will prove susceptible to mechanosynthesis.

8.7. Conclusion

The achievements of solution-phase synthesis show that a remarkably wide range of molecular structures can be built, despite the absence of the standard basis for construction, that is, the ability to move and position parts to direct assembly. A comparison (Sec. 8.3.2) reveals several limitations of mechanosynthesis relative to solution-phase synthesis, but displays a more-than-compensating set of strengths. Chief among these strengths are (1) the ability to achieve reactions at specific sites while avoiding them elsewhere by exploiting direct positional control of reagent moieties, (2) as a consequence of this, the ability to employ highly reactive moieties, such as radicals and carbenes, in thoroughly selective manner, and (3) the ability to accelerate reactions by the application of localized forces of bond-breaking magnitude. Conditions for achieving reaction error rates on the order of 10^{-12} have been described, and it appears that these conditions do not excessively constrain the set of feasible synthetic operations. As a consequence, it will be possible to build precise structures having $> 10^8$ atoms, and to do so with high manufacturing yields.

A review of reactive species used in solution-phase chemistry suggests the attractiveness of unsaturated hydrocarbons, carbon radicals, carbenes, and transition metal com-

plexes in a mechanosynthetic context (Sec. 8.4). A consideration of the utility of these species when used in conjunction with positional control and mechanical force shows the feasibility of a wide range of useful transformations, some of which can be performed with energy dissipation $< kT$ (Sec. 8.5).

Finally, the application of mechanosynthetic processes to the production of diamondoids has been considered, using syntheses of diamond as an example. The diversity of feasible reactions (and the availability of analogous reactions involving elements other than C and H) indicates the feasibility of constructing a wide range of diamondoid (and other) structures. Accordingly, Part II of this volume will proceed on the assumption that a mature mechanosynthetic technology will suffice for the manufacture of most reasonably-stable structures.

Part II

Components and systems

Chapter 9

Nanoscale structural components

1.1. Overview

Nanomechanical designs can be specified in complete detail, permitting them to be modeled (with varying accuracy) molecular mechanics methods, and so forth. In pursuing the design and analysis of nanoscale systems, however, it will be of value to develop generalizations regarding structural components that can be applied at a less-than-atomic level of resolution. The two fundamental questions addressed in this chapter are the feasible *shapes* of nanoscale components, and the extent to which they keep their shapes, that is, their *stiffness*. In preparation, Section 9.2 discusses the context in which nanomechanical components are to be used.

To maximize strength and stiffness, it will typically be desirable to make nanoscale structural components from diamond and broadly diamondoid covalent solids. In the systems analyzed in the following chapters, stiffness requirements are usually more constraining than are strength requirements, and stiffness (unlike strength) can be directly estimated by computational experiments using molecular mechanics models. Examples of such calculations appear in Section 9.3, along with guidelines for estimating the stiffness of diamondoid components.

Nanomechanical engineering diverges from macroscale engineering in that the number of possible component shapes is constrained by atomic sizes and feasible bonding patterns. In sufficiently small components, the set of possible shapes can be exhaustively described. Accordingly, for each type of mechanism, there will be some minimum size below which components of the required shapes cannot be constructed (e.g, it is fruitless

to attempt to build a gear having 50 teeth in a volume that can contain only 10 atoms; the limiting size for a 50-tooth gear will be substantially larger). Near this lower limit of size, a design must be specified in complete molecular detail if one is to have any confidence in its validity.

Somewhat above this lower limit, however, complete molecular specification becomes unnecessary for the same reason that it is unnecessary in macroscale engineering: any one of many arrangements of atoms will serve the purpose. On this larger size scale, a specific molecular structure must be chosen before undertaking production via mechanosynthesis, but this choice can be made *after* performing a reasonably accurate analysis describing size, shape, strength, stiffness, dynamics, energy dissipation, and performance. Section 9.4 discusses how the number of available diamondoid structures varies with volume and structural constraints, drawing conclusions regarding the scale at which molecular detail can be omitted and a modified continuum model approach can be adopted. (Designs that follow this approach will often be relatively conservative, thus setting only lower bounds on the performance of nanomechanical systems.)

Looking ahead to the analysis of bearings and gears in Chapter 10, Section 9.5 examines a variety of structures having high-order rotational symmetry; this can be regarded as a special case of the question of feasible shapes. This discussion raises the issue of the maximum acceptable working strain for bonds in a diamondoid structure.

9.2. Nanomechanical components in a structural context

Nanomechanical systems of the sort proposed here differ from systems of biological molecular machinery in their basic architecture: the nanomechanical systems will be supported and constrained by stiff housings, while biological systems permit relatively free motion of most components with respect to one another.* Most of the sections that follow (in this and later chapters) will focus on moving parts, yet a large portion of the mass of a typical system will take the form of a stiff housing. Gears, bearings, springs, screws, sliding rods, motors—all should be pictured as anchored to or embedded in an extended diamondoid structure that is tailored to support those components in functional positions and orientations with respect to one another (Chapter 10 describes conditions under which nonbonded interfaces permit relatively free sliding motions).

The analogy to macroscopic engineering practice is clear; machines and other systems are typically supported by an extended structure, whether this is termed a housing, chassis, casing, airframe, engine block, or framework. This structure need not be mono-

lithic. Macroscopic housings can be made from pieces held together by fasteners, adhesives, or welding; nanomechanical housings can likewise be made from pieces held together by fasteners or adhesive interfaces.

The volume spanned by a housing will typically have a modest volume-fraction occupied or traversed by moving parts, and the rest of the volume (allowing for small clearances) will be available for solid structure with a stiffness that can be as large as that of diamond. Accordingly, the overall stiffness of a housing structure can typically be quite large, if need be. Note that the positional uncertainty of one region of a homogeneous three-dimensional structure with respect to another (as measured by the standard deviation of their separation) increases only as the logarithm of the separation, and is almost constant once the regions are separated by many region-diameters. Since the stiffness of the housing can typically be far greater than that of the moving parts, it is often reasonable to treat it as rigid on a large scale, making corrections (when necessary) for local compliance.

9.3. Surface effects on stiffness in nanoscale components

In the design and analysis of nanoscale components, it is often convenient to regard them as small regions of a bulk material, estimating their stiffness in the standard fashion, based on modulus, size, and shape. Each of these concepts, however, has some ambiguity in nanoscale systems. Size and shape are ill-defined, owing to the smoothly-graded properties of molecular surfaces; in the absence of well-defined cross-sectional areas, modulus likewise becomes ill-defined. Further, the elastic properties of nanoscale components will in general be anisotropic and inhomogeneous (which can be useful, when subject to intelligent design control). Nonetheless, a model based on defined surfaces and isotropic, homogeneous elastic properties can provide useful design-phase estimates of stiffness. These estimates can be used to find design conditions that will keep elastic deformation within acceptable bounds, although they may not provide an accurate description of the residual displacements.

* D. Tribble has emphasized the importance of such supporting structures for separating degrees of freedom in a system and thereby simplifying design and analysis.

9.3.1. Assigning sizes

Ordinarily, one regards the region occupied by a component as the region from which other components are excluded. This concept of component size is useful in designing mechanisms, describing how parts constrain one another's motion. The nature of nonbonded interactions, however, ensures that this region is imperfectly defined, being dependent on the normal force applied between the components and on the shapes and chemical natures of the impinging surfaces. A reasonable standard choice is the region occupied by the component atoms of the structure, assigning each a summable 0.1 nN radius; this has been the basis of the space-filling molecular representations throughout these chapters.

A glance at Figure 9.1. and Figure 9.2 immediately suggests reasons for correcting this size estimate when estimating stiffness. Component stiffness chiefly results from the polycyclic framework structure, while the exclusion region extends beyond this framework, both, as a result of the relatively large separation between nonbonded atoms (both figures), and as a result of any monovalent surface atoms present (Figure 9.1). Where stiffness must be maximized while component size is minimized, good design practice will avoid monovalent surface atoms, resulting in structures more like that in Figure 9.2. The required size correction is then approximately half the difference between the nonbonded and bonded separations; using the 0.1 nN nonbonded radius, this is $\delta_{\text{surf}} \approx 0.07$ nm for both first and second row atoms. A larger size correction ($\delta_{\text{surf}} = 0.1$ nm is used in Chapter 11) yields a more conservative estimate of stiffness. (For H-terminated diamond, a comparable correction is $\delta_{\text{surf}} \approx 0.14$ nm.)

A further correction can arise from differences between the stiffness contributions of the outermost atoms in the structural framework and those in the interior; with fewer constraints on their motion, they may be subject to relaxation processes that reduce the stiffness they contribute. This effect can be explored by examining the scaling of stiffness with cross-sectional area in a molecular mechanics model such as MM2.

9.3.2. Computational experiments to estimate rod modulus

Molecular mechanics predictions for the linear modulus of a rod, E_{ℓ} (which has the dimensions of force), can be measured by computing the energy of a series of rods placed under differing strains. For an ideal measure of area S_{rod} , the product $S_{\text{rod}}E = E_{\ell}$, for some value of Young's modulus E .

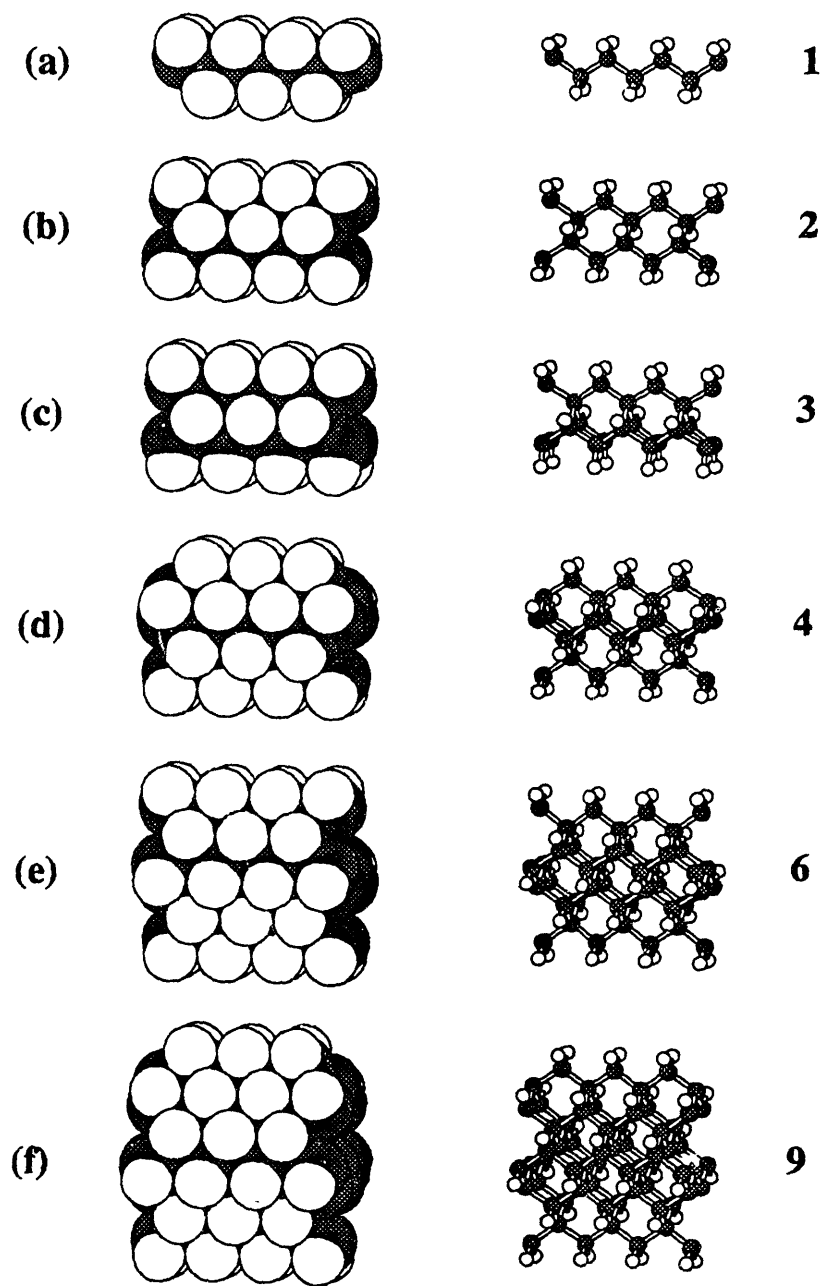


Figure 9.1. A series of rod structures consisting of a hydrogen-terminated region of the diamond lattice. Rod (a) is a polyethylene chain; (b)–(f) can be described as additional chains linked side-by-side (the digits to the right indicate the number of such chains in the associated rod.) The structure of rod (c) departs from the diamond pattern, adding bonds perpendicular to the rod axis to replace nonbonded H|H contacts.

Figure 9.1 shows a series of segments from rods of diamond and diamond-like structure, with increasing cross-sectional area. For rods of defined structure, a reasonable measure of this area is the minimum number of framework bonds n_{bonds} crossing a surface that divides the rod. In Figure 9.1(a), a strand of polyethylene, $n_{\text{bonds}} = 1$; for two joined strands (Fig. 9.1(b)), $n_{\text{bonds}} = 2$, and so forth, through $n_{\text{bonds}} = 9$. Calculating the effective area of these rods from the bond-count and the bond density for the equivalent orientation in diamond ($\sim 1.925 \times 10^{19} \text{ m}^{-2}$) avoids the issues of surface definition just discussed; for this family of rods

$$S_{\text{rod}} \approx 5.2 \times 10^{-20} n_{\text{bonds}} \text{ (m}^2\text{)} \quad (9.1)$$

and, in the bulk-material limit,

$$E_t \approx 5.45 \times 10^{-8} n_{\text{bonds}} \text{ (N)} \quad (9.2)$$

If relaxation were to reduce the stiffness contributions of surface atoms, then thinner rods would have a lower value of E_t .

Figure 9.3 plots the results of computational experiments on the structures in Figure 9.1, using the MM2/C3D+ model. As can be seen, thinner rods have a slightly *greater* value of E than thicker rods (presumably owing to contributions from nonbonded interactions involving hydrogen), but rods of all sizes have a value substantially lower than that of bulk diamond. Since this trend continues to structures with substantial stiffness contributions from interior atoms (e.g., the $n_{\text{bonds}} = 18$ case indicated in Fig. 9.3), it appears that this modulus deficit results from a defect in the MM2 model. (Since this defect underestimates stiffness, while the usual design goal is to maximize stiffness,

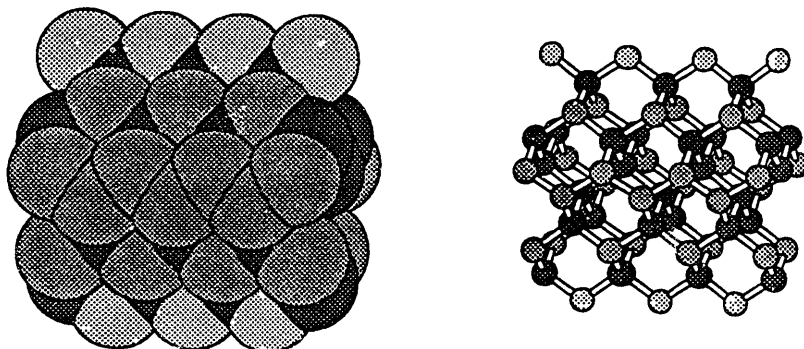


Figure 9.2. A segment of rod resembling Fig. 9.1(f), but with N and O termination in place of CH and CH₂. See text for discussion.

taking MM2 at face value will typically result in false-negative assessments of designs, rather than the more dangerous false-positives.)

Figure 9.2 illustrates a structure like Figure 9.1(f), but with N and O termination, rather than CH and CH₂l. (The stability of chains of *sp*³ nitrogen atoms in an analogous context is discussed in Chapter 10.) For a given value of n_{bonds} , structures of this kind are substantially more compact; as shown in Figure 9.3, the MM2 model also predicts substantially greater stiffness. This increase is a form of surface effect, but one which it is

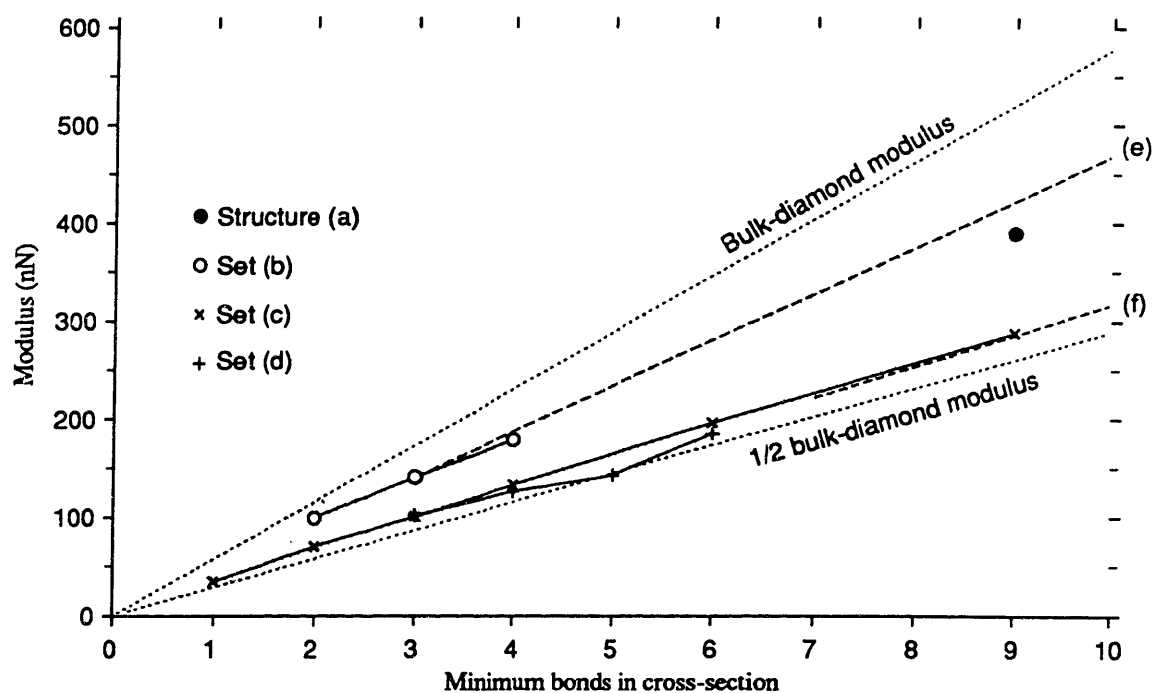


Figure 9.3. Results of computational experiments measuring rod moduli in the MM2/C3D+ approximation. Structure (a) is the N- and O-terminated rod in Fig. 9.2; set (b) consists of the structures in Fig. 9.4; set (c), Fig. 9.1; set (d), Fig 9.5. The dashed line labeled (e) passes through the origin and a point corresponding to a column of hexagonal diamond structure (with a compact cross-section) having $n_{\text{bonds}} = 16$; as can be seen, it almost exactly corresponds to an extrapolation from a similar column with $n_{\text{bonds}} = 3$. The dashed line (f) is analogous to (e), but for a structure like family (c) with $n_{\text{bonds}} = 18$. The upper dashed line bounding the shaded region corresponds to the modulus of diamond rods oriented along the axis, extrapolating from the properties of bulk diamond; the lower dashed line corresponds to a similar extrapolation with the modulus halved.

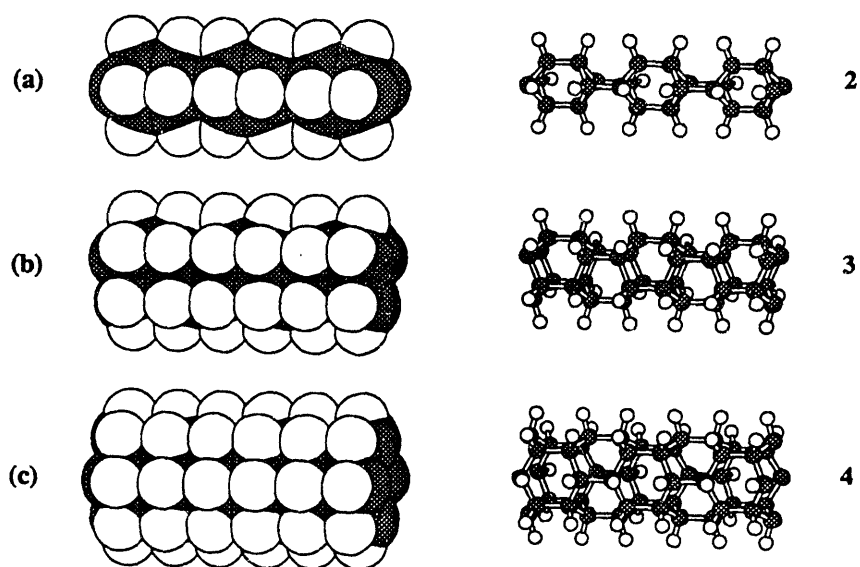


Figure 9.4. A series of rods forming a family of structures suffering increasing strain per bond with increasing diameter.

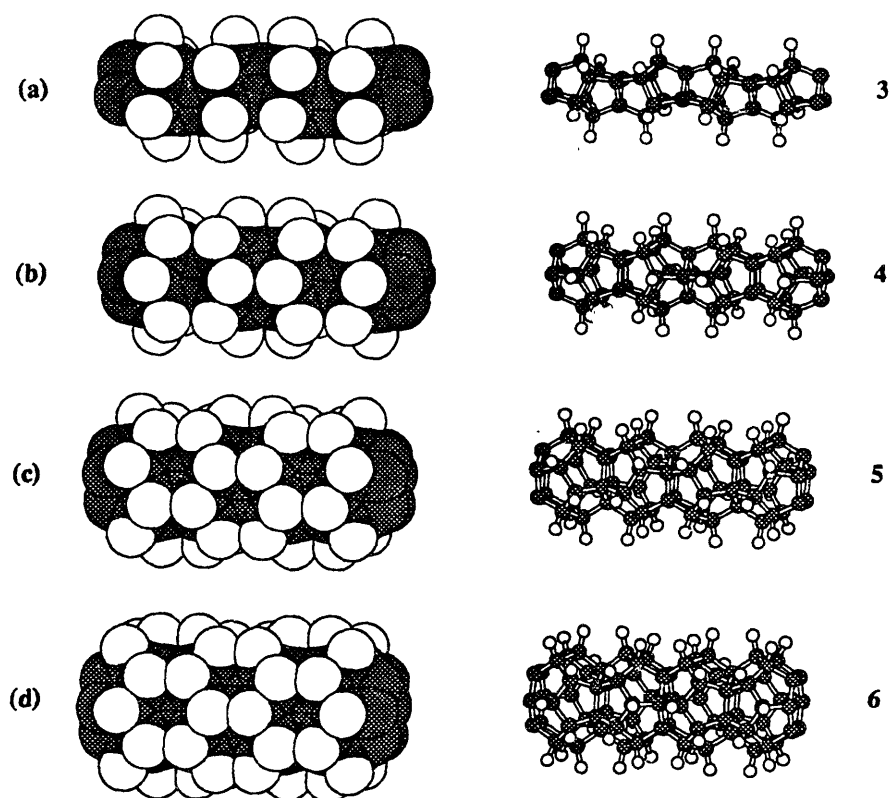


Figure 9.5. A second series of rods forming a family of structures suffering increasing strain per bond with increasing diameter.

conservative to neglect. Figure 9.4 and 9.5 illustrate other rods with moduli plotted in Figure 9.3. The hexagonal column ($n_{\text{bonds}} = 3$) of Figure 9.4(b) is the smallest structural unit that might plausibly exhibit the elastic properties of hexagonal diamond along the corresponding axis; it falls almost precisely on a line drawn through the result for a compact column of hexagonal diamond with $n_{\text{bonds}} = 16$. Surface relaxation effects on modulus appear negligible in this structure (within the MM2 model).

9.4. Control of shape in nanoscale components

In macroscale design, one may doubt that an object of a particular shape can be made from a particular material with a particular technology, but one need never doubt that an object of this shape can (in principle) exist, within macroscopically-negligible tolerances. In a nanoscale volume that can hold only three atoms, in contrast, the choice of structures (and hence shapes) is highly constrained. At what scale do structural possibilities become so numerous that essentially arbitrary shapes become available?

9.4.1. Estimates of the number of diamondoid structures

The number of distinct diamondoid structures, n_{struct} , that can fit in a given volume of reasonable shape will be an exponential function of the volume. This can be seen by considering a hypothetical atom-by-atom construction process: at each surface site in a partially-completed structure, there will be some number of distinct options (with a geometric-mean value n_{opt} which need not be an integer) for how to proceed, for what kind of atom to bond in what manner. The number of such steps is proportional to the number of atoms in the final structure (which will, for a given range of atomic number densities $\sim n$, be proportional to volume), and each choice leads to a different structure (by definition of “distinct choice,” above). For a suitable definition of n_{opt} , the number in question is

$$n_{\text{struct}} = n_{\text{opt}}^{Vn} \quad (9.3)$$

The value of n_{opt} is not obvious, even when the atoms are restricted to a single type. It will depend on the nature of the range of permissible local structures, constrained by (for example) the magnitudes and patterns of bond strain permitted in the interior of a stable solid. Knowledge of these constraints would be necessary, yet would not enable a simple computation.

The entropy of fusion of a crystal is a measure of the increase in the available volume in configuration space (Sect. 4.3.3) that occurs with a transition from a single, regular structure to an ensemble of states which includes many different structures. *If* the potential wells in the liquid state were as well-defined as those in the solid state, and *if* each of those potential wells were equally populated and corresponded to a stable amorphous structure (and vice-versa), then the entropy of fusion would be a direct measure of the increase in number of wells, and would thus provide a direct measure of the number of available structures. In practice, low-stiffness wells and the thermal population of unstable transition structures tend to make the entropy of fusion an overestimate; the negligible thermal population of high-strain states that would be kinetically stable as solids tends to make it an underestimate. The absence of a choice of atom types at various sites (in elemental crystals) introduces a strong bias toward underestimation, relative to the systems of interest here.

Measuring the entropy of fusion of carbon (melting point ≈ 3820 K) presents technical difficulties; the values for silicon and germanium, whose crystals share the diamond structure, are 4.6×10^{-23} and 4.7×10^{-23} J/K-atom respectively. These values are both about $3.4k$; if this increase in entropy were solely the result of the increased number of available potential wells per atom, that number would be $n_{\text{opt}} \approx 29$. An estimate of the number of distinct states in argon at liquid density (Stillinger and Weber 1984) yields an estimate equivalent to $n_{\text{opt}} \approx 3$ wells per atom.

In the set of broadly diamondoid structures built up of C, Si, N, P, O, S, H, F, and Cl (with a bias toward the earlier, higher-valence members of this series), n_{opt} will be increased as a result of the choice of atom type at each step; a factor of 5 is a not unreasonable estimate of this increase. Taking the argon estimate as a base, this suggests that a conservative estimate for reasonable diamondoid structures is $n_{\text{opt}} \approx 15$.

A region with a structural volume of 1 nm^3 , $n_{\text{opt}} = 15$, and $n = 100 \text{ nm}^{-3}$ will have $n_{\text{struct}} \approx 10^{118}$. A similar cubical region with an excluded volume of 1 nm^3 and $\delta_{\text{surf}} = 0.1 \text{ nm}$ will have a structural volume of $\sim 0.51 \text{ nm}^{-3}$, and $n_{\text{struct}} \approx 10^{60}$.

9.4.2. *The elimination of structures by constraints*

These enormous numbers imply that objects can be constructed with almost any specified shape, provide that the specifications do not specify complex contours on a sub-nanometer scale, and that the specifications permit adequate tolerances in the placement of surface atoms, relative to the ideal contours of the shape. Demanding that many sur-

face atoms simultaneously meet tight tolerances, however, can easily result in an over-constrained problem having no physically-realizable solution.

For example, one might ask that each of the ~ 100 atoms on the surface of a 1 nm object be within a distance ε of an ideal surface contour. For a randomly-chosen structure with $n = 100 \text{ nm}^{-3}$, the distance of surface atoms from some ideal contour will range over a distance of $\sim 0.2 \text{ nm}$. If we demand that each surface atom be within $\varepsilon = 0.01 \text{ nm}$ of this contour, then on the order of $(0.01/0.2)^{100} \approx 10^{-130}$ of a set of randomly-chosen structures will meet this condition. Given $n_{\text{struct}} \approx 10^{60}$, the probability that some structure satisfies this condition is on the order of 10^{-70} . If, however, one asks that 40 atoms on particular surfaces be accurately placed within $\varepsilon = 0.02 \text{ nm}$, then on the order of $(0.02/0.2)^{40} \approx 10^{-40}$ of a set of randomly-chosen structures will meet this condition, and the probability that a suitable structure does *not* exist is on the order of 10^{-20} . (Brute-force search procedures would be inadequate to find acceptable structures, but a variety of options exist for developing more sophisticated search procedures to solve problems of this general kind; doing so presents an interesting and important challenge.)

Tight constraints on object shape are expected to arise chiefly in the working interfaces of moving parts, for example, where surfaces must slide smoothly. Chapter 10 develops a variety of models for such interfaces; some are based on the choice of regular structures of high symmetry (to which the statistical arguments of this section do not apply); others describe the constraints on irregular structures imposed by the requirement that they slide smoothly over a regular surface. The major remaining class of working interface is used for binding molecules; this is considered in the next section.

9.4.3. Structural constraints from molecular binding requirements

In the acquisition and processing of molecules from solution (to name one application), it is useful to construct surfaces with selective binding sites for the desired molecules. The general principles of selective binding are familiar from molecular biology. The favorable conditions can chiefly be described as forms of detailed surface complementarity: matching surface shapes, to provide strong van der Waals attraction with little overlap repulsion; patterns of charge and dipole orientation that result in electrostatic attraction; the ability to form hydrogen bonds at appropriate positions and orientations, and so forth (in aqueous solutions, so-called hydrophobic forces often play a large role). These requirements for complementarity will impose constraints on surface structure, again reducing the size of the set of acceptable structures by many orders of magnitude.

Molecular biology provides an example in which structures capable of strong, selective binding are chosen from a random set of structures of calculable size. Mammalian immune systems can produce highly specific antibodies to novel molecules, including those unknown in nature. These antibodies are developed by a process of variation and selection within a space of possibilities defined chiefly by differing sequences of the 20 genetically-encoded amino acids within the hypervariable domains (variations at other sites are relatively rare). These domains contain ~ 38 amino acid residues, hence most antibodies are selected from a set of structures containing $\sim 20^{38} \approx 3 \times 10^{49}$ members. Equation (9.3) with $n_{\text{opt}} = 15$ suggests that a similar diversity can be found in the set of structures having < 50 atoms, although other constraints may require larger binding sites structures. (Note that the desirability of rigid structures for most machine components does not preclude the use of flexible structures elsewhere, e.g., to permit a flap on a binding site to fold over a ligand.)

9.5. Nanoscale components of high rotational symmetry

In rotating machinery, it will frequently be desirable to use components having high-order rotational symmetry. Chapter 10 discusses applications of such components in bearings and gears. (Note that perturbations from a less-symmetrical environment will break perfect symmetry, but perfection of this sort will seldom be of practical importance.) Rotationally-symmetric components divide roughly into two classes: *strained-shell* structures made (at least conceptually) by bending a straight slab of regular structure into a hoop, and *special-case* structures that are more intrinsically cylindrical. Members of an intermediate class of *curved-shell* structures resemble strained-shell structures, but with regular arrays of dislocations (and similar structures) that reduce the strain. These terms are intended as rough guides, not as precise categorizations. The first two are discussed in the following sections.

9.5.1. Strained-shell structures

Figure 9.6 illustrates a relatively small example of a strained shell with more than one layer of cyclic structure. The thickness-to-radius ratio (t/r) of strained shells is limited by the net hoop stress and the permissible bond tensile strain at the outer surface. For diamondoid structures with negligible net hoop stress and a benign chemical environment, permissible bond strains can be ≥ 0.035 nm ($\sim 1.23\ell_0 \approx 0.187$ nm); below this strain (given MM2 bond stiffnesses and a Morse potential) a C–C bond has positive stiff-

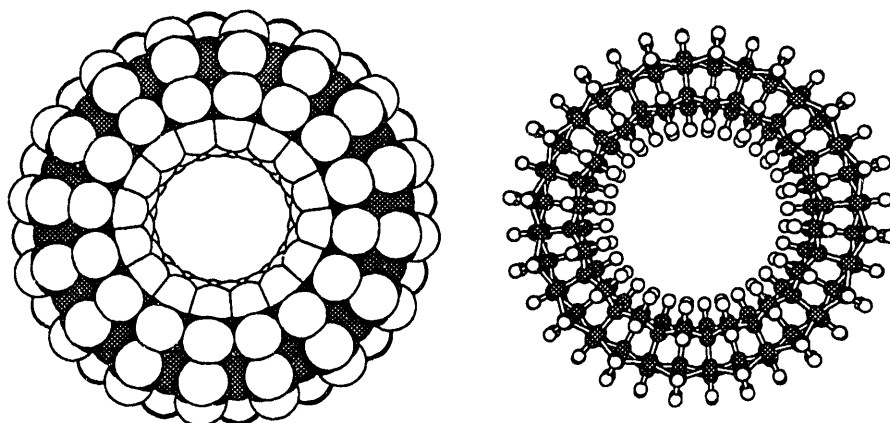


Figure 9.6. A strained shell structure with a relatively thin wall and small diameter.

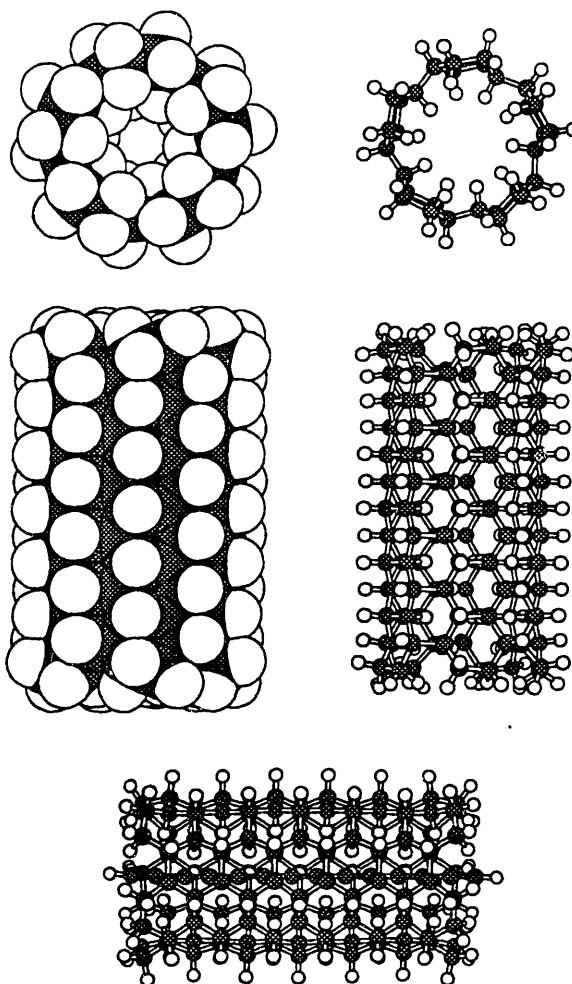


Figure 9.7. A strained shell structure with a very thin wall and small diameter; the structure below adds a nonbonded polyynic chain along the axis.

ness, tending to stabilize its length. In the presence of a strain gradient perpendicular to the local direction of strain, as in a strained shell, a bond at or somewhat beyond this limit will be stabilized by restoring forces resulting from the less-strained layers closer to the shell axis (the layers are coupled by shear stiffness).

If the material of the shell had linear elastic properties, the above calculations would indicate that $t/r \leq \sim 0.46$ is acceptable. The worst-case nonlinearity would prohibit bond compression entirely, resulting in a limit of $t/r \leq \sim 0.23$. Both of these estimates neglect the favorable effects of angle-bending relaxation. The appropriate measure of thickness for these calculations is smaller than that associated with the structural volume discussed

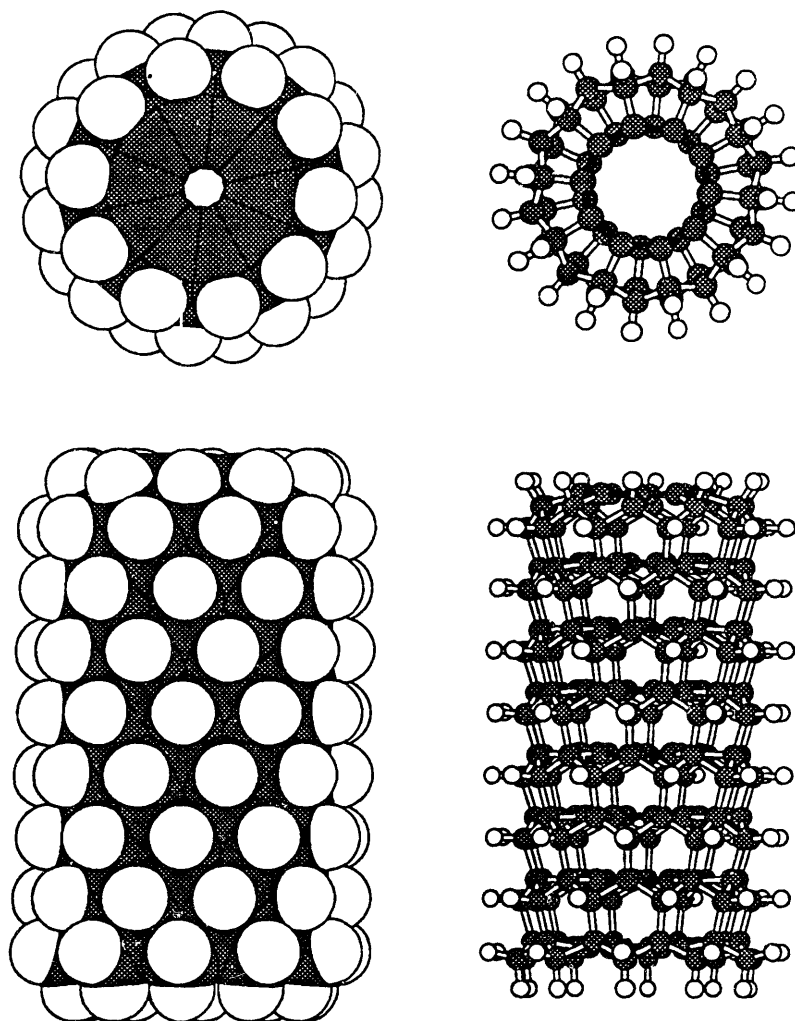


Figure 9.8. A cylindrical structure with sp^3 surface structure and an sp^2 inner structure (the ten-membered conjugated rings will exhibit aromatic stabilization).

in Section 9.3.1; atomic centers at the edge of the framework structure provide a conservative marker for the surface.

Figure 9.7 illustrates an example of a strained-shell structure with relatively little strain, consisting of a single layer of cyclic structure. It gains stiffness from the compressed nonbonded contacts in its interior; more nonbonded contacts can be provided by placing a polyyne chain along its axis (below).

9.5.2. Special-case structures

The rods of Figures 9.4 and 9.5 provide examples of small-radius structures having varying degrees of rotational symmetry. The cylindrical structure of Figure 9.8 provides an example of intermediate size; the outer ring of the bearing in Figure 1.1 provides a still larger example. The set of diamondoid structures of high rotational symmetry, small length, and diameters of ≤ 1 nm is clearly large, but may be small enough to permit the development of an exhaustive catalogue.

9.6. Conclusions

Nanoscale structural components will typically serve either as housings or as moving parts. Bulk-material stiffnesses for diamond (and diamond-like solids) can be extrapolated to components of subnanometer dimensions, provided that the modulus is applied to a cross-sectional area that is descriptive of the structural framework, omitting the surface regions of the occupied volume that result from the relatively long range of overlap forces, and those that result from monovalent surface atoms. At a ~ 1 nm size scale, the number of stable diamondoid structures becomes enormous, and essentially any shape can be built, provided that tight (< 0.02 nm) tolerances are not applied too widely. A discussion of the compatibility of irregular structures with smooth sliding motion is deferred to the next chapter. Components of high rotational symmetry can be constructed in several ways; Chapter 10 describes the use of such structures in building bearings and gears.

Chapter 10

Mobile nanomechanical components

10.1. Overview

A mechanical technology with broad capabilities must include mobile components. If these are to resemble the components familiar in macroscale technology, they will require interfaces that permit sliding and rolling motions, thereby enabling the construction of gears, bearings, and the like. Given the feasibility of strong stiff nanoscale structures having a wide variety of shapes (Chapter 9) and of nanoscale mobile components like those discussed in this chapter, the feasibility of a wide range of nanomechanical systems is immediately obvious. Chapter 11 will discuss one important and illustrative class of mechanical systems, computers, along with means for power supply and control.

The existence of a wide variety of stable molecular liquids—resistant to both polymerization and decomposition—shows that interfaces between molecules can be both stable and mobile at ordinary temperatures. Familiar examples include hydrocarbons, fluorocarbons, ethers, and amines; structures with analogous surface moieties will appear in the examples of this chapter. A yet wider range of surface moieties will be stable when incorporated into rigid polycyclic structures operating in well-ordered environments.

Two mutually-inert surfaces in contact can be characterized by their potential energy of interaction. Models like those described in Section 3.5 can give an estimate of the potential energy as a function of separation for two relatively smooth surfaces, but for use in gears and bearings, the potential energy function associated with sliding motions is of central importance. The two cases of greatest interest are those in which the potential energy function is nearly flat, permitting smooth sliding, and those in which it has large corrugations, entirely blocking sliding. The former interfaces can be used in bearings; the

latter can be used in gears.

The following section explore mobile interfaces and their applications in gears, bearings, and related devices. Section 10.2 begins by characterizing properties of the spatial fourier transforms of interatomic nonbonded potentials, which prove to be useful in analyzing the properties of sliding interface bearings. Section 10.3 considers the problem of sliding motion between irregular covalent objects and regular covalent surfaces, developing a Monte Carlo model that predicts the expected fraction of irregular structures that will exhibit smooth sliding motion with respect to such surfaces. Section 10.4 develops the theory of symmetrical sleeve bearings, presenting results from analytical models of idealized bearings, and characterizing two specific designs using molecular mechanics methods. Section 10.4 generalizes from these results to a variety of other systems incorporating sliding-interface bearings, including nut-and-screw systems, rods sliding in sleeves, and constant-force springs. Section 10.6 briefly describes bearings that exploit single atoms as axles. Section 10.7 moves from sliding interfaces to non-sliding interfaces, examining analytical models of gears, and using these as a basis for examining the properties of roller bearings and systems resembling chain drives. Finally, Section 10.9 briefly surveys devices that use surfaces intermediate between freely-sliding and non-sliding: dampers, detents, and clutches.

10.2. Spatial fourier transforms of nonbonded potentials

In the design of nanomechanical systems, smooth sliding motions are most directly achieved when one (or both) surfaces at an interface have a periodic or nearly-periodic structure of high spatial frequency, that is, when the surface has a series of features that repeat at regular intervals, with each found in essentially the same local environment. In one-dimensional sliding motion (by convention, along the x axis), only periodicity along the x axis is significant; the associated spatial frequency is κ rad/m, or f_x cycles/m ($= \kappa/2\pi$).

The variations in the potential $\mathcal{V}(x)$ associated with sliding of a component over a surface can in the standard molecular mechanics approximations be decomposed into a sum of the pairwise nonbonded potentials between the atoms in the object and those in the surface, together with terms representing variations in the internal strain energy of the object and the surface. For stiff components under small interfacial loads, the soft, non-bonded interactions will dominate the variations in $\mathcal{V}(x)$, and structural relaxation will result in only small deviations from straight-line motion of the interfacial atoms. Under

these conditions, variations in the total potential $\mathcal{V}(x)$ are accurately approximated by a sum of the pairwise nonbonded potentials between atoms in straight-line relative motion. Figure 10.1 plots a set of such interaction potentials for pairs sp^2 carbon atoms moving on paths with differing closest-approach distances d , based the MM2 exp-6 potential, Eq. (3.9).

10.2.1. Barrier heights and sums of sinusoids

Barrier heights $\Delta\mathcal{V}_{\text{barrier}}$ for sliding of components over periodic surfaces can be described in terms of a sum of contributions associated with integral multiples of the spatial frequency f_x . This sum can be divided into contributions each resulting from an atom in the sliding object interacting with a row of evenly-spaced atoms in the periodic surface. The energy of an atom with respect to a row consists of a sum

$$\mathcal{V}(x) = \sum_{i=-\infty}^{\infty} \mathcal{V}_{\text{vdw}} \left(\sqrt{d^2 + \left(x + \frac{i}{f_x}\right)^2} \right) \quad (10.1)$$

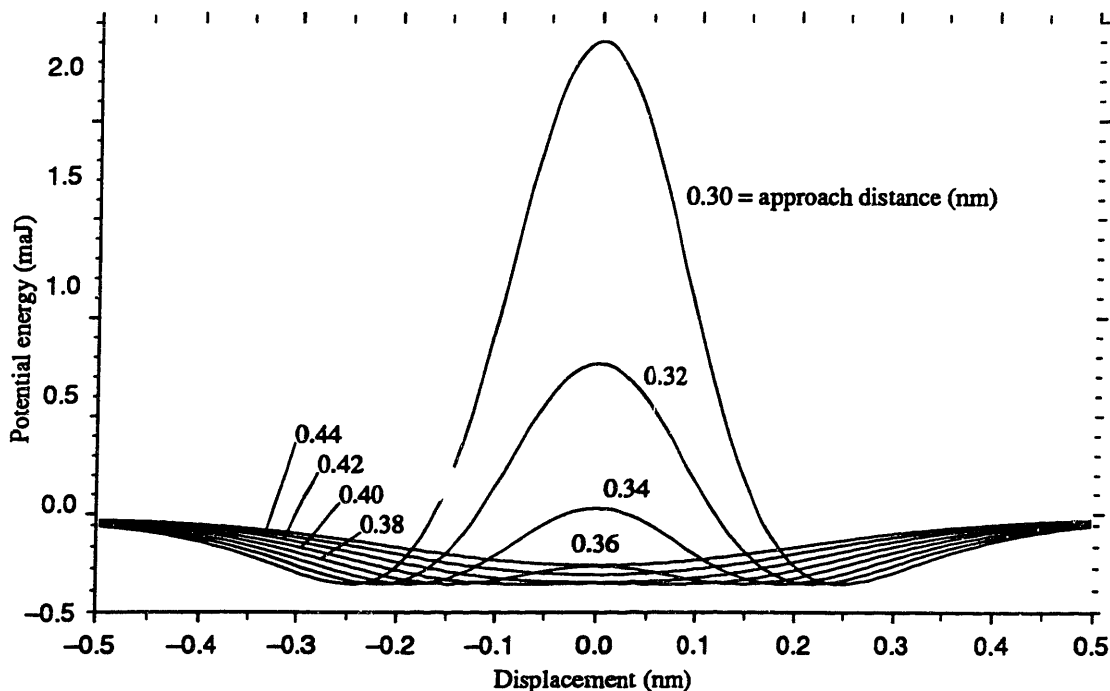


Figure 10.1. Nonbonded potentials, Eq. (3.9), as a function of x -axis displacement for two sp^2 carbon atoms moving in straight lines with various closest-approach distances d .

where $\mathcal{V}_{\text{vdw}}(r)$ is the nonbonded energy as a function of radius for the two atom types, Eq. (3.9). Graphically, such a sum can be represented as an evenly-spaced set of sample points, and the difference between two sums separated by one-half cycle can be represented by an evenly-spaced set with weights alternating between +1 and -1 (Fig. 10.2); in the typical cases of interest, the maximum value of this difference equals the barrier height for sliding of a single atom with respect to a single row.

Figures 10.3 and 10.4 plot amplitude spectral densities $|H(f_x)|$ derived from the spatial fourier transforms of set of interaction potentials like those in Figure 10.1. At large spatial frequencies f_x , the associated amplitudes drop steadily on a logarithmic scale as f_x increases. Exceptions to the "typical cases of interest" in the previous paragraph can arise when $f_x = f_{x,0}$, but these exceptional cases are associated with unusually low barrier heights. The designs discussed here do not exploit this phenomenon, as it depends on a delicate cancellation rather than on robust properties of intermolecular potentials.

The sampling shown in Figure 10.2 can be described as a sequence of positive and negative delta functions. The fourier transform for such a sequence yields the relationship

$$\Delta\mathcal{V}_{\text{barrier}} \propto |H(f_x) - H(3f_x) + H(5f_x) - H(7f_x) \dots| \quad (10.2)$$

From the results plotted in Figures 10.3 and 10.4, it can be seen that with reasonable val-

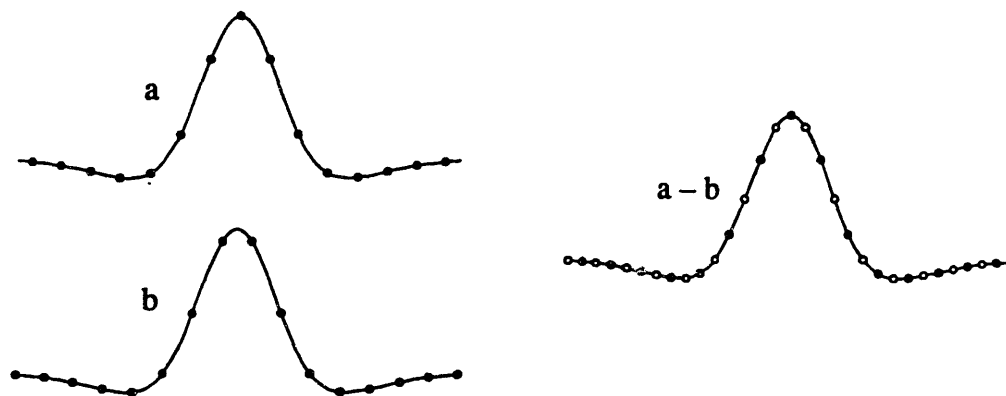


Figure 10.2. A nonbonded potential (as in Fig. 10.1) sampled at regular intervals (a), and sampled with a shift in phase (b). The difference in the sum of the sample energies as the phase shifts is the energy of a sum of samples with positive and negative weights (a - b); this observation can be used to relate spatial fourier transforms of pairwise potentials to energy barriers in sliding motion (see text).

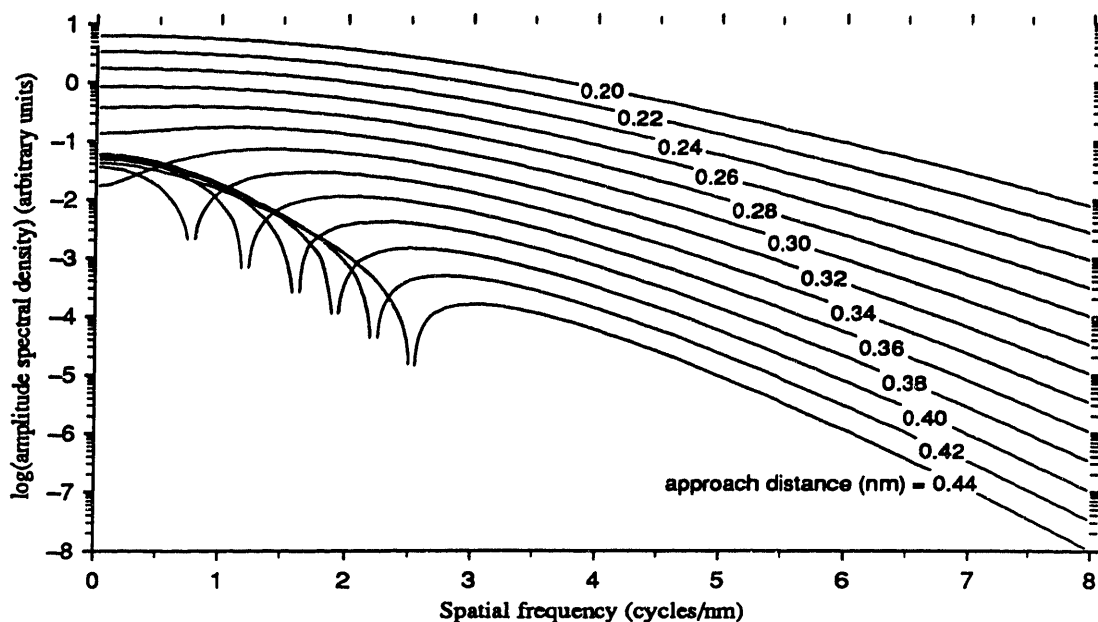


Figure 10.3. Amplitude spectral densities derived from spatial fourier transforms of nonbonded ClC (sp^2) potentials like those in Fig. 10.1, for a range of closest-approach distances d . For relatively large values of d , the fourier transform changes sign, resulting in a zero at some spatial frequency $f_{x,0}$.

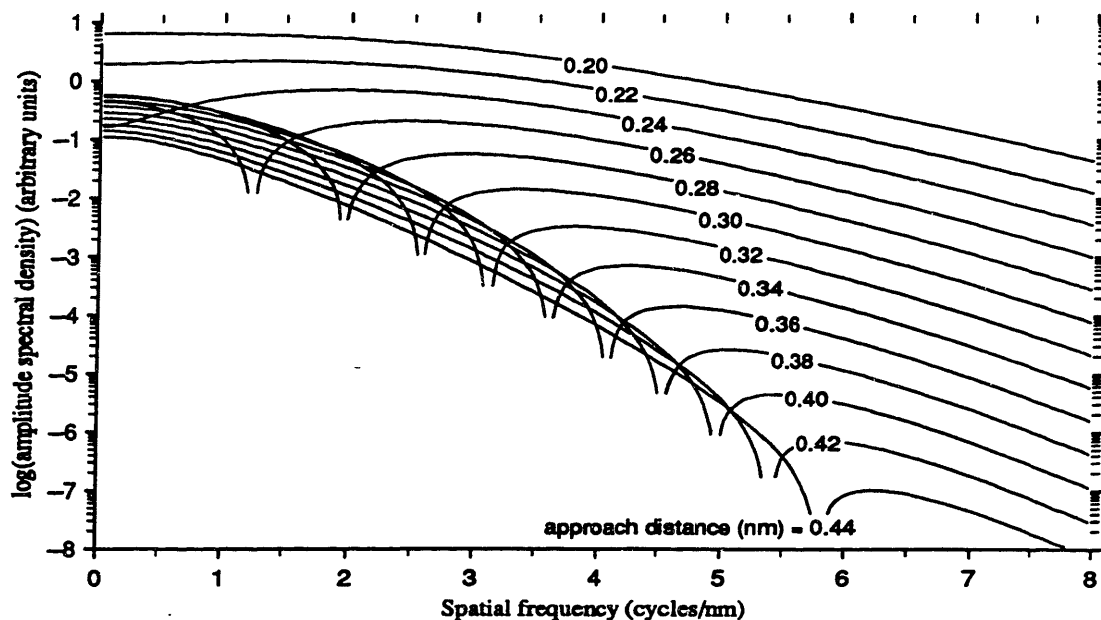


Figure 10.4. Amplitude spectral densities as in Fig. 10.3, but for nonbonded H1H potentials. Note that these graphs would be identical (within a constant energy factor) if all distances were measured in units scaled to the equilibrium nonbonded separation.

ues of d (e.g., ≥ 0.2 nm) and with a moderately high spatial frequency (e.g., $f_x \geq 3.0$ cycles/nm), the first term in Eq. (10.2) will dominate the rest by multiple orders of magnitude. For the diamond (111) surface in the high- f_x direction, $f_x \approx 4.0$ cycles/nm; for graphite, $f_x \approx 4.1$ cycles/nm. As a consequence, the potential of a component sliding over such a surface can be accurately approximated as a sum of sinusoidal contributions from interactions between the atoms of the component and each of the component rows of the surface, all of spatial frequency f_x , but of varying amplitudes and phases.

10.3. Sliding of irregular objects over regular surfaces

Chapter 9 estimated the (enormous) number of nanometer-scale diamondoid structures, observing that constraints on surface structure can drastically reduce size of this set. In particular, requiring that surface structures be regular imposes a requirement that interior structures be regular (to a depth dependent on the tolerance for residual irregularities); this reduces the set of possible structures to a minute fraction of that available in the absence of this constraint. Regular structures can make excellent bearings (as is demonstrated by the analysis and examples in Section 10.4) and this is their chief value in the present context. This section examines the bearing performance that can be achieved in the far larger set of irregular structures.

10.3.1. *Motivation: a random-walk model of barrier amplitudes*

Consider an atom sliding over a regular surface with spatial frequency f_x at a height h (note that $h \leq$ the minimum value of d with respect to any of the rows of the surface). As shown in Section 10.2, the potential energy of such an atom can be represented as a sum of sinusoids, each characterized by some amplitude $\Delta V/2$, phase ϕ . For a surface of high spatial frequency, and for values of h corresponding to modest loads, the sum will be dominated by a single sinusoid with a spatial frequency f_x .

Consider an irregular object sliding over a surface having N such atoms in contact. In the above approximation, the interaction energy of each will be dominated by a sinusoid of the same f_x , but with differing values of ΔV and ϕ . In the vector representation (see Fig. 10.5), the summing of these sinusoids can be visualized as a walk over a plane. For a set of irregular objects with randomly-distributed values of ϕ and bounded values of ΔV , the resulting random walk has familiar properties: in the limit of many steps, the probability density for the end points is gaussian, and the mean value of the radius is

$$\overline{\Delta\mathcal{V}}_{\text{sum}} = \overline{\Delta\mathcal{V}}_{\text{barrier}} \propto \sqrt{N} \quad (10.3)$$

hence the area over which the end points are scattered varies as N . For a set of irregular structures in which there are n_{opt} choices for the properties of the N atoms, the number of possible structures increases as n_{opt}^N , and thus the density of end-points in the plane (Fig. 10.5) near the origin varies as

$$\rho_{\text{area}} \propto \frac{n_{\text{opt}}^N}{N} \quad (10.4)$$

and the mean distance from the origin to the closest point (i.e., the value of the smallest barrier for any member of the family of structures) is

$$\overline{\Delta\mathcal{V}}_{\text{barrier}} \propto \sqrt{\frac{N}{n_{\text{opt}}}} \quad (10.5)$$

Thus, although the expected barrier height for any given irregular structure increases with increasing N , the minimum expected barrier height for a family of structures decreases.

10.3.2. A Monte Carlo analysis of barrier heights

The scaling arguments of Section 10.3.1 do not translate directly into an accurate statistical model; such a model would require a treatment taking account of the nature of the available choices of interacting atoms and their interactions with the regular surface, which in turn affects the value(s) of n_{opt} and the distribution of values of $\Delta\mathcal{V}$. Further, as values of $\Delta\mathcal{V}_{\text{barrier}}$ become small, sinusoidal terms of higher spatial frequency become

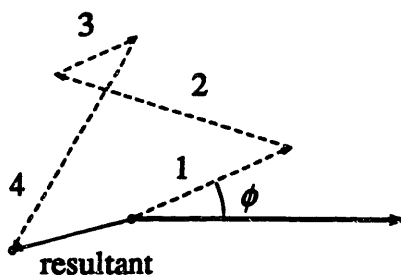


Figure 10.5. Sinusoidal energy terms (defined by magnitude and phase) represented as vectors, illustrating the magnitude and phase of the sum as the result of a random walk over the plane.

important, and the end-points must be treated as being scattered in a space of higher dimension. The complexity of these interacting physical effects and design choices suggests the use of a Monte Carlo model to estimate the distribution of $\Delta\mathcal{V}_{\text{barrier}}$ for model systems of interest.

10.3.2.1. Approximations and assumptions

To reduce the computational burden while retaining the essential physical and statistical features of the problem, a set of approximations was adopted:

- Use of the MM2 exp-6 potential to represent non-bonded interaction energies. The exp-6 potential is realistic over a wide range of separations, and the neglected electrostatic terms would have little effect on $\Delta\mathcal{V}_{\text{barrier}}$.

- Use of a straight-line translation model, which effectively treats structures as infinitely stiff. For x -axis sliding, neglect of y and z axis relaxation is conservative in this context, while neglect of x axis relaxation is the reverse; at modest loads, the neglected effects are minor.

- Lack of rotational relaxation of the sliding object. An irregular object pressed against a surface will tend to rotate to distribute load over several contact points; neglect of this will artificially increase the disparity in contact loads, which tends to increase barrier heights.

With these approximations, the interaction energy of an object is just the sum of the energies of a set of atoms, each interacting with the surface independently. Atomic interaction energies can be precomputed and then combined to represent different structures.

Families of structures can be modeled using a further set of assumptions and approximations:

- Each family is characterized by a framework structure having a set of N sp^3 surface sites within a certain range of distances (h_{\min} to h_{\max}) from the regular surface.

- To generate the members of a family, each surface site can either be occupied by an N atom with a lone pair, or by a C atom bonded to H, F, or Cl, or can be deleted (locally modifying the framework); each site thus has 5 states, giving each structural family 5^N possible members.

- To model irregular structures, sites are assumed to be randomly distributed with

respect to the unit cells of the regular surface, and bond orientations are assumed to be randomly distributed within a cone directed toward the regular surface, with a half-angle of $109.47 - 90^\circ$.

- Site positions are treated as fixed, with lone pair, H, F, and Cl positions determined by the site coordinates and bond orientation, together with standard bond lengths.

- To ensure that each structure examined will be in firm contact with the regular surface, site-states are characterized by their z -axis stiffnesses relative to the regular surface. Members of a structural family containing a site with a stiffness less than some threshold $k_{s,\text{thresh}}$ are discarded, as are those with less than three non-deleted sites. These exclusions ensure that the z -axis stiffness for each structure is $\geq 3k_{s,\text{thresh}}$, and typically reduce the number of retained members of a structural family to far fewer than 5^N .

10.3.2.2. Computational procedure

Based on the above, the computation proceeds by generating a set of sites characterized by locations and bond orientations (200 sites were used), constructing a set of site-states for each site, filtering this set for stiffness $\geq k_{s,\text{thresh}}$, and computing the energy of each remaining site-state at a series of displacements (16) spanning one cycle of motion over the regular surface. A structural family is defined by randomly choosing N sites from the above set, and a list of the acceptable members of that family is generated by forming all possible combinations consisting of one state from each site and discarding

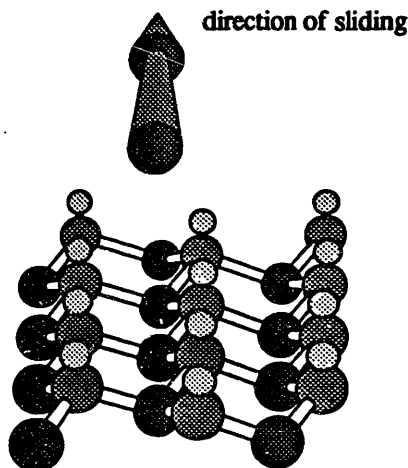


Figure 10.6. Structure of the nitrogen-substituted diamond (111) surface, showing the direction of sliding assumed in the calculations summarized in Figures 10.7 and 10.8 (this direction is taken as the x axis).

those with more than $N - 3$ deletion states. The energy of each structure as a function of displacement is computed by summing the corresponding precomputed values for each of its constituent site-states, and the barrier height for sliding motion of the structure is taken as the difference between the maximum and minimum energies in the resulting sum. Finally, $\Delta\mathcal{V}_{\text{barrier}}$ for the structural family is taken as the minimum of the energy differences found for any member of that family.

10.3.2.3. Results

The results of a set of calculations based on the above model for sliding motion of irregular structures over a strip of nitrogen-substituted diamond (111) surface (as illustrated in Fig. 10.6), are summarized in Figure 10.7, taking the initial sampling bounds for site generation as $h_{\text{min}} = 0.2$ nm and $h_{\text{max}} = 0.5$ nm. The statistical distribution of values

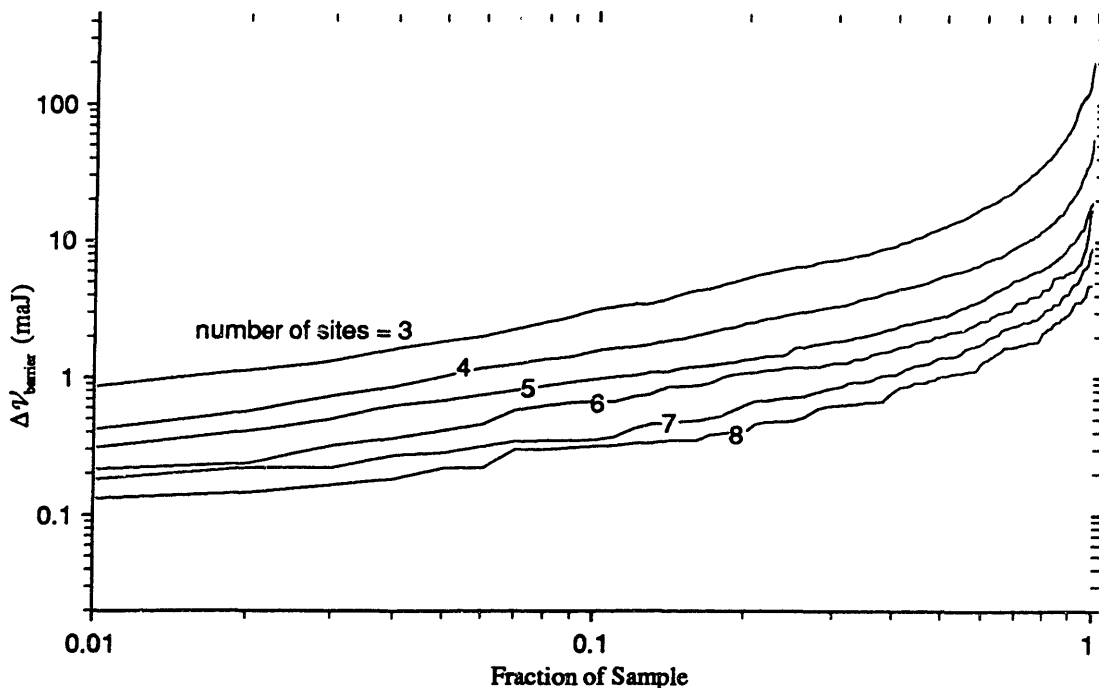


Figure 10.7. Cumulative distributions resulting from a Monte Carlo study of barrier heights encountered by irregular structures sliding over a regular surface, based on the model described in Sec. 10.3.2. Each curve is the result of 1000 trials, where each trial selects the best member of a particular family of structures (see description in text); note that sampling statistics result in substantial scatter, particularly toward the left tail of the distributions. The surface modeled is nitrogen-substituted diamond (111), Fig. 10.6.

of $\Delta\mathcal{V}_{\text{barrier}}$ is presented as a series of cumulative distributions for samples of randomly-generated structural families with $N = 3, 4, 5, 6, 7$, and 8 . As can be seen, for $N \geq 5$, about 10% of randomly-selected structural families have a member yielding $\Delta\mathcal{V}_{\text{barrier}} \leq 1.0 \text{ mJ}$ ($< 0.25kT$ at 300 K). Barriers this low will be surmounted on most encounters, and hence will for most purposes fail to act as barriers. Further, if one assumes that the potential is characterized by a sinusoid with the period of the lattice ($\sim 0.28 \text{ nm}$), then the peak negative stiffness during sliding along the x axis is $\sim -0.25 \text{ N/m}$.

Figure 10.8 presents the results of a similar calculation with $h_{\text{min}} = 0.19 \text{ nm}$, $h_{\text{max}} = 0.49$, and $k_{s,\text{thresh}} = 10 \text{ N/m}$. Values of $\Delta\mathcal{V}_{\text{barrier}}$ are higher owing to the combined effects of higher energies and a shift toward higher spatial frequencies in the interatomic potentials (Figs. 10.3 and 10.4).

10.3.3. Implications for constraints on structure

Assume that a nanomechanical component has been designed to meet some set of functional constraints along one surface (the *constrained surface*), and that some other surface of the component (the *sliding surface*) is required to slide smoothly along a regu-

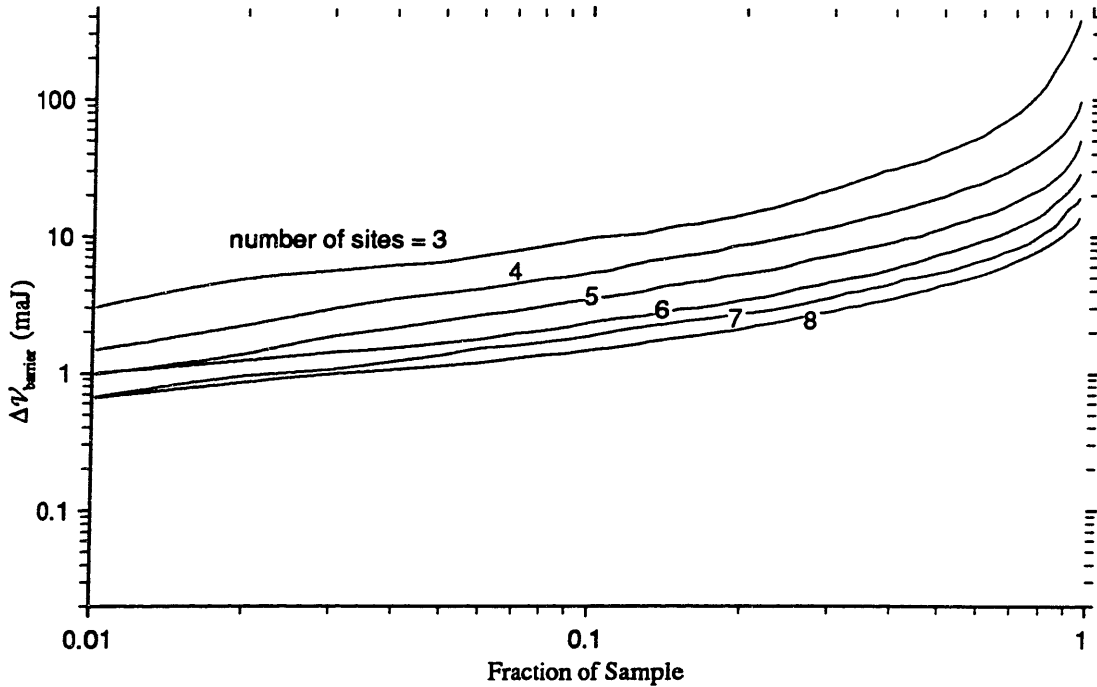


Figure 10.8. Cumulative distributions like those in Fig. 10.7, but with minimum values of z -axis stiffness 10 times higher (see text). The author thanks L. Zubkoff for providing the computational resources used to acquire the data plotted in Figures 10.7 and 10.8.

lar surface. One would like the design of the sliding surface to be relatively unconstrained. One can define a set of *compatible framework structures* that consists of diamondoid frameworks that satisfy the constraints of the constrained surface and extend for some indefinite distance toward (and past) the desired location of the sliding surface. If the two surfaces are separated by a distance on the order of 1 nm, then the set of compatible framework structures (now considering only variations in the region that falls short of the desired sliding surface) will be a large combinatorial number, typically $> 10^{10}$. Candidate families of sliding surface structures can be generated (in a design sense, not a fabrication sense) by truncating these compatible framework structures so as to generate a set of sp^3 sites falling between h_{\min} and h_{\max} ; members of each family can then be generated by modifying these sites in the manner suggested by the model of Section 10.3.2.1. (It is here assumed that these surface-site modifications usually do not make a compatible framework structure incompatible.)

The results of the above model suggest that (for modest loads applied to regular surfaces of high spatial frequency) a substantial fraction of these truncated compatible framework structures will permit low values of $\Delta\mathcal{V}_{\text{barrier}}$ provided that the number of potential sites between $h_{\min} = 0.2$ nm and $h_{\max} = 0.5$ nm is > 5 . If all atoms in the slab could serve as sp^3 sites, then for structures with an atomic number density $n_a = 100 \text{ nm}^{-3}$, a cross-sectional bearing area $S_{\text{bear}} = 0.25 \text{ nm}^2$ would suffice. If the area-density of sites equaled that of atoms in the (111) surface of diamond, then a somewhat larger area, $S_{\text{bear}} = 0.28 \text{ nm}^2$, would be necessary. For design work, values of $S_{\text{bear}} \geq 0.5 \text{ nm}^2$ should prove ample. Larger areas (or use of structures selected from a larger set of compatible frameworks) will permit both greater z-axis stiffness and lower $\Delta\mathcal{V}_{\text{barrier}}$.

In summary, it is safe to assume that any component with a surface in contact with a strip of regular, high-spatial-frequency structure can be made to slide smoothly with respect to that strip, provided that loads are modest, that the contact surface is of a reasonable shape (with a length of at least $2\pi/\kappa$ in the x direction) with $S_{\text{bear}} \geq 0.5 \text{ nm}^2$, and that the contact surface is not too tightly constrained in structure. The latter condition can typically be satisfied a few atomic layers from a surface that is tightly constrained. This conclusion generalizes to irregular structures that slide along grooves and ridges, and to irregular structures sliding along curved surfaces in rotary bearings.

10.3.4. Energy dissipation models

10.3.4.1. Phonon scattering

Interaction between a small sliding contact and ambient phonons can be modeled as scattering from a moving harmonic oscillator. With surface interaction stiffnesses of several N/m per atom, a total stiffness on the order of 30 N/m will not be atypical. This corresponds to the example in Section 7.3.4, which yielded an estimated energy dissipation of $\sim 3 \times 10^{-16}$ W at a sliding speed of 1 m/s and $\sim 3 \times 10^{-20}$ W at 1 cm/s.

10.3.4.2. Acoustic radiation

A small sliding contact will exert a time-varying force on the surface with an amplitude roughly proportional to amplitude of the energy variation ($= \Delta\mathcal{V}_{\text{barrier}}/2$). From Eq. (3.19), the force amplitude can be estimated at

$$F_{\text{max}} \approx 1.7 \times 10^{10} \Delta\mathcal{V}_{\text{barrier}} \quad (10.6)$$

(F_{max} in N, $\Delta\mathcal{V}_{\text{barrier}}$ in J). Assuming sinusoidal variations in force, Eq. (7.8) and (10.6) yield

$$P_{\text{rad}} \approx 2.9 \times 10^{20} \Delta\mathcal{V}_{\text{barrier}}^2 v^2 \kappa^2 \sqrt{\rho} \frac{1}{8\pi M^{3/2}} \quad (10.7)$$

With $\Delta\mathcal{V}_{\text{barrier}} = 1$ mJ, a surface with a stiffness and density like those of diamond, and $\kappa \approx 2.2 \times 10^{10} \text{ m}^{-1}$, Eq. (10.7) yields a radiated power of $\sim 3 \times 10^{-19}$ W at 1 m/s, and $\sim 3 \times 10^{-23}$ W at 1 cm/s. These losses are small compared to those resulting from phonon scattering; in geometries in which the energy variations reflect variations in pressure (for example, sliding in a tube or a slot with balanced forces on either side), then Eq. (7.21) applies and losses will be still lower.

10.3.4.3. Thermoelastic damping

Eq. (7.49) can be applied to estimate losses resulting from thermoelastic damping. Using diamond material parameters for the surface (save for K_T , taken as 10 W/m·K), and treating the alternating forces in the system as being applied to square nanometer areas and cubic nanometer volumes, the estimated energy loss per cycle (at a sliding speed of 1 m/s) is $\sim 3 \times 10^{-30}$ J, or $\sim 1 \times 10^{-20}$ W. In addition, the total (non-alternating) force is time-varying from the perspective of a site on the surface. Assuming that the

force is adequate to ensure a nonbonded stiffness of ~ 30 N/m (~ 1 nN) and that the length scale ℓ of the loaded region is ~ 1 nm, the pressure is ~ 1 GPa; at a sliding speed $v = 1$ m/s the estimated energy loss per cycle ($t_{\text{cycle}} \approx \ell/v$) is $\sim 10^{-27}$ J, or $\sim 10^{-18}$ W. Thermoelastic damping losses thus are small compared to phonon scattering losses.

10.3.5. Static friction

Where $\Delta\mathcal{V}_{\text{barrier}} \ll kT$, static friction is effectively zero. At low temperatures, however, and in the absence of tunneling, the static friction of a sinusoidal potential can be identified with the maximum value of $d\mathcal{V}/dx$, where x measures the displacement of the surface. Where the sinusoid has a period d_a (as will be the case whenever amplitudes are significant),

$$F_{\text{frict}} = \left(\frac{d\mathcal{V}}{dx} \right)_{\text{max}} = 4\pi \frac{\Delta\mathcal{V}_{\text{barrier}}}{d_a} \quad (10.8)$$

For a typical value of d_a (~ 0.25 nm), $F_{\text{frict}}/\Delta\mathcal{V}_{\text{barrier}} \approx 0.05$ nN/maJ.

10.3.6. Coupled sites

An extended object (such as a rod) can contact a periodic surface at regions spread over a considerable distance; this raises the issue of the interaction between the stiffness of the elastic coupling within the object and the negative stiffness associated with the sinusoidal potentials of the regions taken individually. In particular, one would like to ensure that structures can be found in which these negative stiffnesses are small compared to the positive stiffness of the coupling between adjacent regions, and that this holds true on all length scales.

Consider a contact region for which families of compatible structures yield N vectors (in the space considered in Sec. 10.3.1) falling within a disk of radius $D\mathcal{V}_1$. Within the approximations used here, the distribution of these vectors will (for small values of $D\mathcal{V}_1$) be essentially random. Now consider a second such contact region, with an independent set of families of structures. Given that the density of vectors is approximately constant within a region of radius $1.5D\mathcal{V}_1$, each vector \mathbf{V} in the first region will have (on the average) $N/4$ vectors in the second region that are within a radius $D\mathcal{V}_1/2$ of $-\mathbf{V}$. Choosing a pair of structures for which this holds would, in the rigid coupling approximation, yield a system with $D\mathcal{V}_{\text{barrier}} \leq D\mathcal{V}_1/2$; the negative stiffness associated with the interaction between the two sites is bounded by

$$|k_s| \leq 3\Delta\mathcal{V}_1 \left(\frac{\pi}{d_s} \right)^2 \quad (10.9)$$

For $d_a = 0.25$ nm and $D\mathcal{V}_1 = 1$ mJ, $|k_s| \leq 0.5$ N/m. So long as the stiffness coupling the two regions is large compared to this small value, the two regions can be treated as rigidly coupled for purposes of the present analysis. Only extraordinarily compliant structures will fail to meet this criterion for points separated by ten nanometers or less.

The mean number of candidate structures with this property is $N^2/4$, if we require

$$\frac{N^2}{4} \geq N \quad (10.10)$$

then the same argument can be applied between *pairs* of regions of the sort just described, with $D\mathcal{V}_2 = D\mathcal{V}_1/2$ playing the role of $D\mathcal{V}_1$. By induction, sets of regions can be constructed on all length scales, with $D\mathcal{V}_{\text{barrier}}$ varying inversely with size. Note that, in a rod, stiffness varies as ℓ^{-1} , but for a system with these scaling properties, the magnitude of the negative stiffness between regions likewise varies as ℓ^{-1} . A threshold value for the above argument to proceed (in the absence of statistical fluctuations) is $N \geq 4$; larger values ensure reliability in the face of statistical fluctuations, and can ensure that the magnitude of the negative stiffness between regions varies as ℓ^{-b} , $b > 1$.

10.4. Symmetrical sleeve bearings

The results of the previous section indicate that irregular sleeve bearings with small energy barriers will be feasible provided (1) that either the shaft or the sleeve has a rotational symmetry such that a rotation corresponding to a small tangential displacement is a symmetry operation and (2) that the other component is sufficiently weakly constrained in its structure that a design can be selected from a large set of possible structures. Sleeve bearings, however, lend themselves to analysis and design that exploits symmetry in both components. This section extends a preliminary study (Drexler 1987) that indicated the promise of this class of structures. The resulting analyses can in several instances be extended directly to sliding-interface bearings with non-cylindrical geometries.

10.4.1. Models of symmetric sleeve bearings

For calculations involving bearing stiffness, interfacial stiffness, and dynamic friction, a sleeve bearing can often be approximated as a cylindrical interface with a certain

stiffness per unit area k_a for displacements perpendicular to the surface, and a distinct stiffness per unit area $k_{a,para}$ (which can be low) for displacements of the surface parallel to the axis of the bearing. These approximations will be applied in Section 10.4.5.

Where static friction is concerned, sleeve bearing models must take account of atomic detail. Both the outer surface of a shaft and the inner surface of a sleeve can be decomposed into rings of atoms, each having the rotational symmetry of the corresponding component. In the no-relaxation approximation, the potential energy of the system as a function of the angular displacement of the shaft with respect to the cylinder can be treated as a sum of the pairwise interaction of each inner ring with each outer ring. These pairwise potentials will be well approximated by sinusoids of a single frequency, which will in the worst case add in phase, and in the best case will substantially cancel. A single ring-ring interaction thus captures the essential characteristics of a shaft-sleeve interaction, save for the omission of (potentially favorable) cancellations. For concreteness, coplanar rings are used as a model in the following section, with parameters illustrated in Figure 10.9.

10.4.2. Spatial frequencies and symmetry operations

Consider an inner ring with n -fold rotational symmetry and an outer ring of m -fold symmetry. If $n = m$, then the inner ring must be displaced by an angle $\theta_{\text{symm}} = 2\pi/n$ to

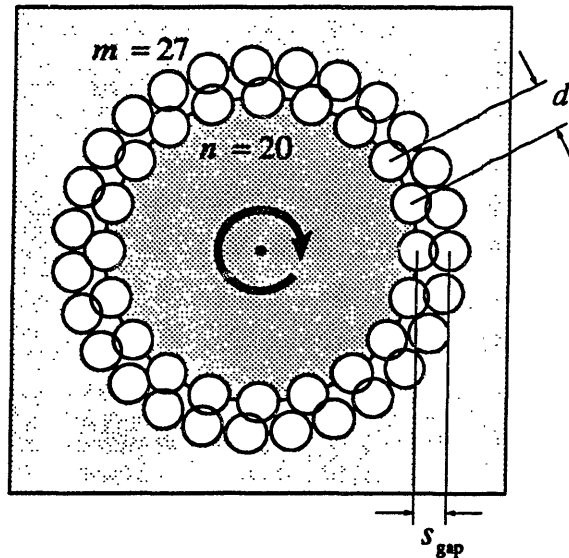


Figure 10.9. Coplanar ring model for a symmetrical sleeve bearing. The radii r_{inner} and r_{outer} have the obvious definitions.

restore the initial geometry and potential energy, and the spatial frequency associated with this symmetry operation is approximately that of the interatomic spacing in the inner ring, d_a ($r_{\text{inner}} = d_a n / 2\pi$). In a better approximation, it can be taken as the interatomic spacing of the inner ring projected to the mean radius

$$r_{\text{eff}} = \frac{r_{\text{inner}} + r_{\text{outer}}}{2} = r_{\text{inner}} + \frac{s_{\text{gap}}}{2} \quad (10.11)$$

yielding an effective spatial frequency for rotational displacements of the the inner ring

$$f_{\theta} = \frac{1}{r_{\text{eff}} \theta_{\text{symm}}} = \left[d_a \left(1 + \frac{s_{\text{gap}}}{2r_{\text{inner}}} \right) \right]^{-1}, \quad n = m \quad (10.12)$$

If $n \neq m$, then a smaller relative displacement can yield a geometry that is identical to one resulting a *combined* rotation of the two rings. Since a combined rotation leaves the energy unchanged, this smaller rotation is a symmetry operation for the potential energy function (note that this still holds in the presence of relaxation). The required angle is related to the least common multiple of n and m ,

$$\theta_{\text{symm}} = \frac{2\pi}{\text{lcm}(n, m)} \quad (10.13)$$

yielding an effective spatial frequency

$$f_{\theta} = \left[\frac{nd_a}{\text{lcm}(n, m)} \left(1 + \frac{s_{\text{gap}}}{2r_{\text{inner}}} \right) \right]^{-1} \quad (10.14)$$

10.4.3. Properties of unloaded bearings

Figure 10.10 presents a logarithmic plot of calculated barrier heights for coplanar-ring bearing models for $1 \leq n, m \leq 15$ in the concentric (unloaded) case. Figure 10.11 shows the closely corresponding pattern of spatial frequencies calculated from Eq. (10.14); this correspondence results from the smooth fall-off of amplitude spectral densities $|H(f_x)|$ with increasing f_x shown in Section 10.2; calculations using curved trajectories and f_{θ} would differ little. Figure 10.12 shows the effect of variations in s_{gap} for two examples with fixed n and m . (Note that H interactions, small values of s_{gap} and relatively large values of d_a have been used in these calculations because each choice is sig-

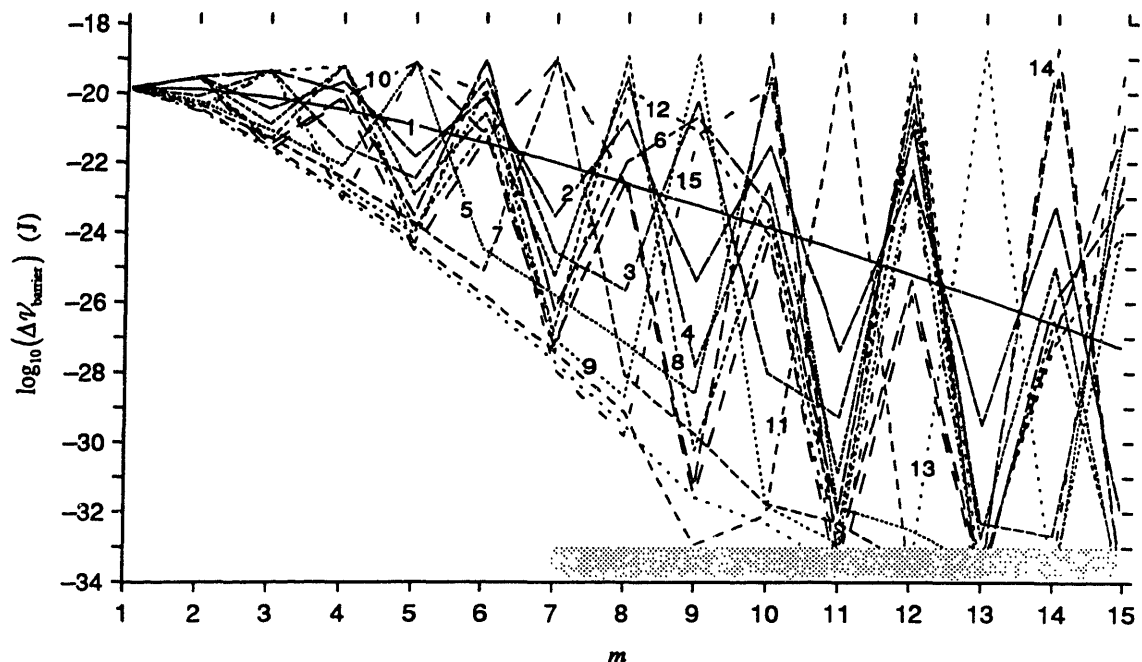


Figure 10.10. Barrier heights for rotation in the coplanar-ring model, based on the MM2 exp-6 potential, Eq. (3.9), for the HHH interaction, using parameters from Table 3.1. All rings were constructed with $d_a = 0.3$ nm and $s_{\text{gap}} = 0.2$ nm.

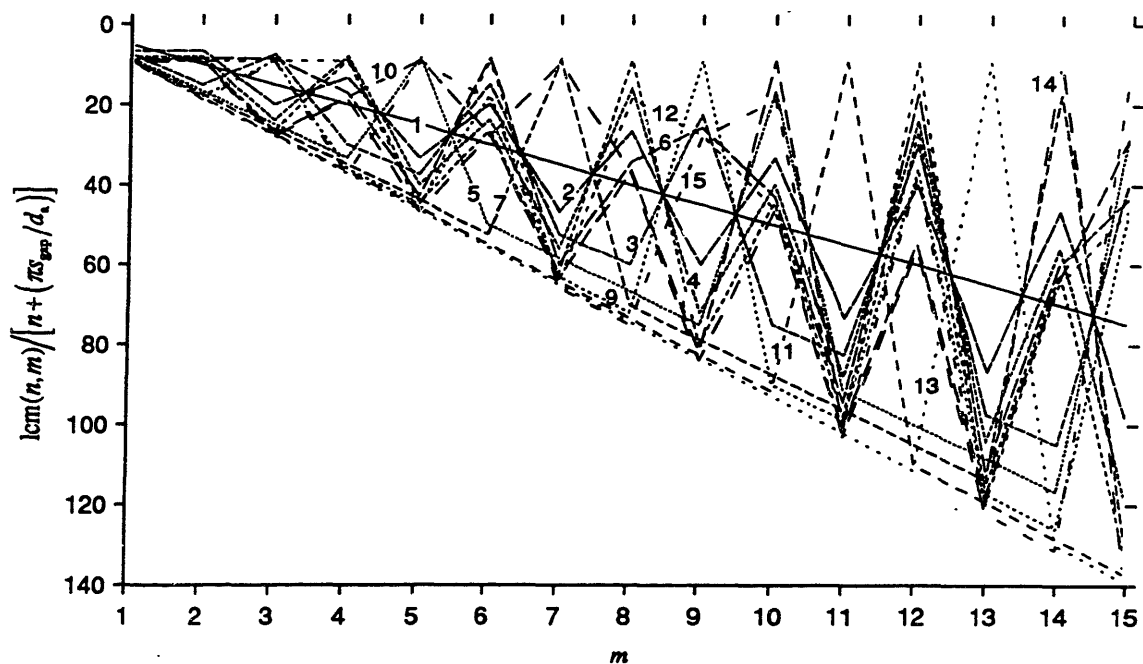


Figure 10.11 Spatial frequencies (in units of $1/d_a$) for the ring systems of Fig. 10.10, based on Eq. (10.14).

nificantly adverse; larger atoms, smaller gaps, and smaller interatomic spacings all are feasible and reduce $\Delta\mathcal{V}_{\text{barrier}}$.)

Unloaded bearings having $\text{lcm}(n,m) \gg n$ and d_a reasonably small will typically have minute values of $\Delta\mathcal{V}_{\text{barrier}}$; the best values of n and m are relatively prime. Small loads, however, destroy the symmetries that are required for this result to hold.

Anisotropies in the potential for displacements perpendicular to the axis (ideally characterized by a single stiffness k_s) give an indication of how rapidly barrier heights increase with load. Where k_s is nearly isotropic, small loads perpendicular to the axis will store nearly equal energies, independent of the angle of rotation; where k_s varies greatly, so will differences in stored energy. Where $n - 2 \leq m \leq n + 2$, problems of commensurability and anisotropic stiffness tend to be severe.

For relatively isotropic ring systems with the parameters used in Figure 10.10, $k_s \approx 10n^2$. Sleeves having multiple rings will have correspondingly greater stiffnesses.

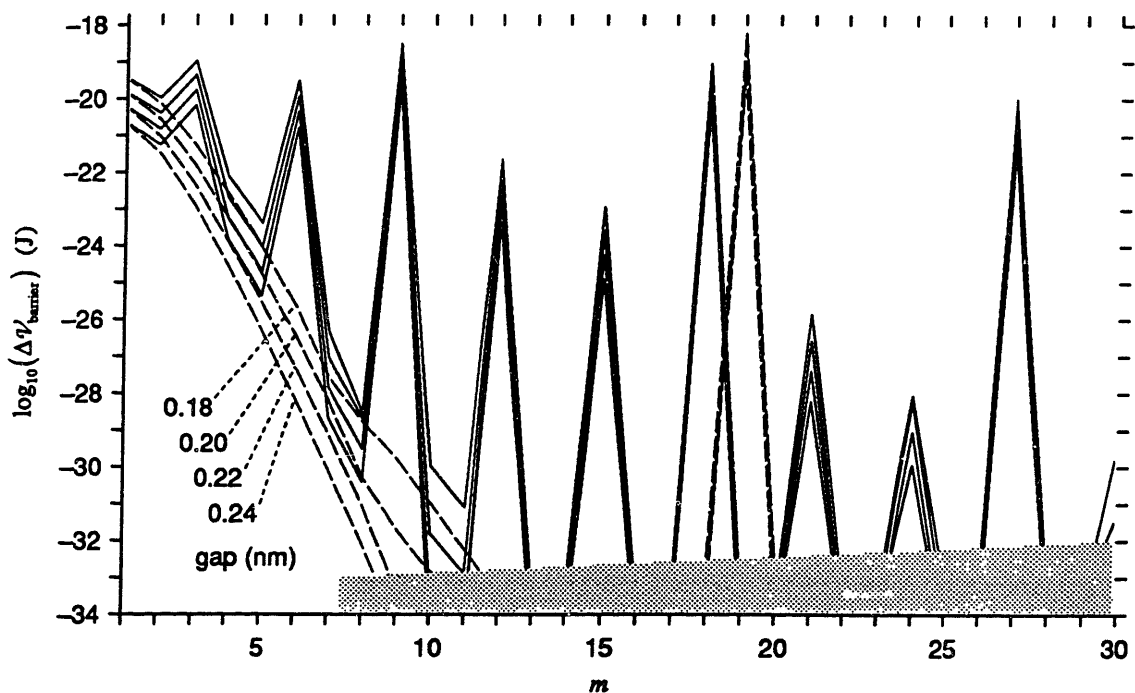


Figure 10.12. Barrier heights as in Fig. 10.10, for sleeve bearings with $n = 9$ and 19 , $m = 1$ to 30 , and values of s_{gap} varying from 0.18 to 0.24 nm. (Lines for $n = 9$ are solid, for 19 dashed.)

10.4.4. Properties of loaded bearings

The effects of load perpendicular to the bearing axis can be modeled by displacing the center of the inner ring by a distance Δx and examining the potential as a function of angular displacement of the inner ring about its own axis. Figures 10.13 and 10.14 show the results of such an investigation for a series of ring systems with $n = 9$ and 19 respectively. As can be seen, even small displacements destroy the delicate cancellations required for extremely low $\Delta\mathcal{V}_{\text{barrier}}$, but for suitably chosen systems with n and $m \geq 25$, values of $\Delta\mathcal{V}_{\text{barrier}}$ can be negligible compared to kT at 300 K, even at substantial displacements. This will continue to hold for bearings having multiple interacting rings, so long as those number are modest or (as can often be arranged) contribute sinusoidal potentials that add out of phase and approximately cancel.

For $n = 9$, $m = 14$, and the parameters of Figure 10.10, a displacement of $\Delta x = 0.01$ nm corresponds to a mean restoring force of ~ 0.96 nN and an energy barrier of ~ 1 mJ. The restoring force, however, fluctuates by ~ 0.1 nN, and since k_s is ~ 100 , the associated fluctuations in stored energy from this source $\sim (\Delta F^2/2k_s)$ would be ~ 0.05 mJ for shafts subject to no other source of stiffness. For $n = 19$, $m = 27$, the corresponding force and stiffness are ~ 2 nN, 210 N/m, with fluctuations in force < 0.0001 nN and fluctuations in stored energy from this source $\sim 10^{-29}$ J. These quantities have been neglected in the calculations described above.

10.4.5. Bearing stiffness in the transverse-continuum approximation

Section 3.5.2 developed a model of surfaces that averages overlap repulsion and van der Waals attraction over displacements transverse to the interface. This model can be used to estimate the energy, force, and stiffness per unit area as a function of separation for a pair of nonpolar, nonreactive, diamondoid surfaces, so long as those surfaces are smooth, regular, and out of register with one another, and provides a good approximation for a broad class of sleeve bearings.

A simple sleeve bearing can be characterized by a cylindrical interface of radius r_{eff} , length ℓ , and interfacial stiffness per unit area k_a . As indicated by Figure 3.11, k_a can range from large positive values to moderate negative values, depending on the separation s_{gap} (these models assume $r_{\text{eff}} \gg s_{\text{gap}}$). Where $k_a \approx 0$, the tensile stress across the interface can be ~ 1 GPa, but this only a few percent of the tensile strength of diamond.

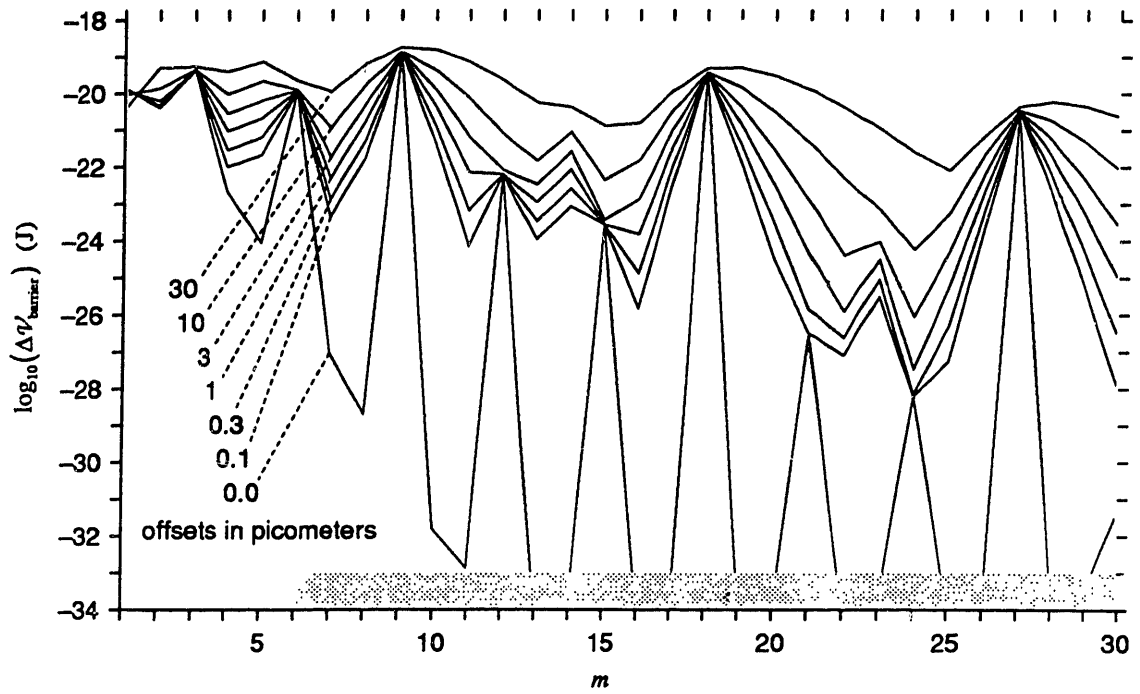


Figure 10.13. Barrier heights as in Figure 10.10, for sleeve bearings with $n = 9$, $s_{\text{gap}} = 0.2$ nm, and varying values of transverse offset.

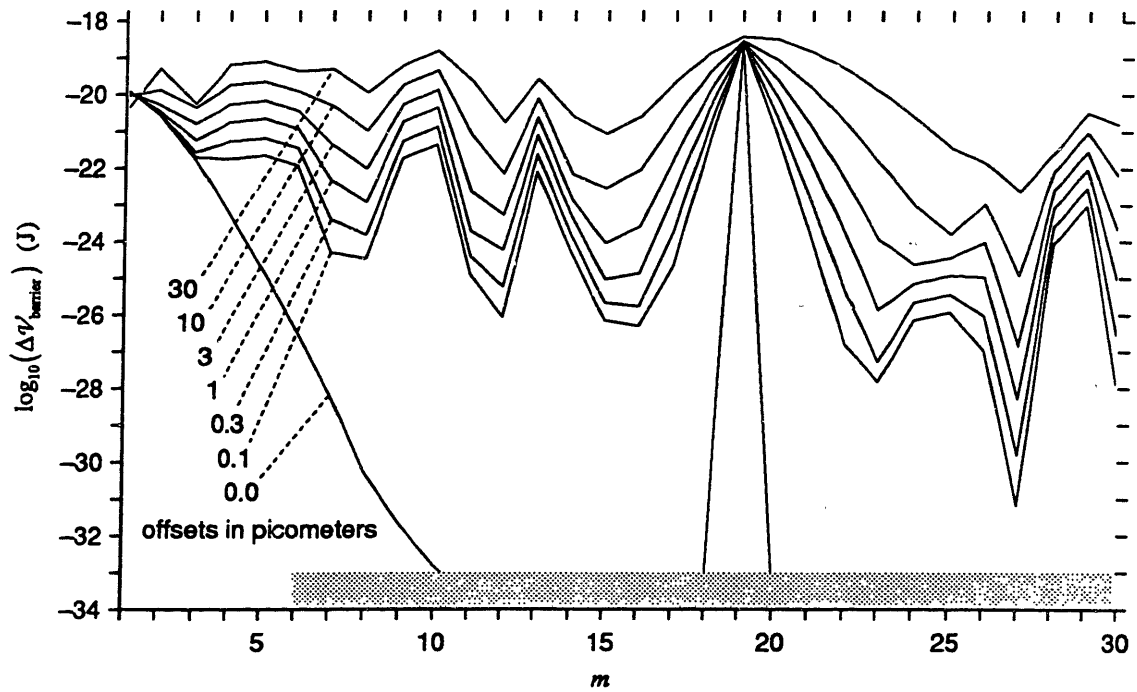


Figure 10.14. Barrier heights as in Figure 10.13, for sleeve bearings with $n = 19$.

So long as $k_a > 0$, an unloaded shaft will be centered in the sleeve; small displacements perpendicular to the bearing axis will be characterized by a positive stiffness $k_{s,bear}$ (larger displacements result in larger stiffnesses, owing to the nonlinearity of steric repulsions). The contribution of a patch of interface to $k_{s,bear}$ varies with the angle θ between the normal and the direction of displacement, yielding an expression for $k_{s,bear}$ in terms of the interfacial area and k_a :

$$k_{s,bear} = \int_0^{2\pi} k_a \ell r_{eff} \sin^2(\theta) d\theta = \pi k_a \ell r_{eff} \quad (10.15)$$

10.4.6. Mechanisms of drag

Drag in sleeve bearings has several sources. These include acoustic radiation), shear-reflection drag, band-stiffness scattering, band-flutter scattering, and thermoelastic damping. Aside from acoustic radiation, all of the following models estimate dissipation at the cylindrical interface between the shaft and the sleeve using relationships developed for the limiting case of a flat interface between indefinitely extended solids. Sample calculations are presented for bearings of large stiffness (1000 N/m) and moderate size ($r_{eff} = \ell = 2$ nm).

10.4.6.1. Acoustic radiation

For symmetrical sleeve bearings of high spatial frequency, oscillating forces will be zero and oscillating torques and pressures will be negligible. Significant radiation can be expected only from loaded bearings. For the latter, oscillating torques will be on the order of $\pi \Delta \mathcal{V}_{barrier} r_{eff} / d_a$ N·m, hence the radiated torsional acoustic power, Eq. (7.15), will be on the order of

$$P_{rad} \approx \left(\frac{\pi \Delta \mathcal{V}_{barrier} r_{eff}}{d_a} \right)^2 \left(\frac{2\pi v_{inter}}{d_a} \right)^4 \frac{\rho^{3/2}}{48\pi G^{5/2}} \approx (\Delta \mathcal{V}_{barrier} r_{eff})^2 v_{inter}^4 \frac{102\rho^{3/2}}{d_a^6 G^{5/2}} \quad (10.16)$$

With diamond-like material properties, $\Delta \mathcal{V}_{barrier} = 1$ mJ, $r_{eff} = 1$ nm, the interfacial sliding speed $v_{inter} = 1$ m/s, and $d_a = 0.25$ nm, the power resulting from torsional acoustic radiation $P_{rad} \approx 10^{-25}$ W.

Oscillating forces in loaded bearings will depend on the nature of the load, the frequency of the oscillation, and the mass of the rotor. From Eq. (7.8).

$$P_{\text{rad}} = F_{\text{max}}^2 \left(\frac{2\pi v_{\text{inter}}}{d_{\text{a}}} \right)^2 \sqrt{\rho} \frac{1}{8\pi M^{3/2}} \quad (10.17)$$

In most instances, $F_{\text{max}} \ll 0.1 \text{ nN}$, and hence (with assumptions like those above) $P_{\text{rad}} \ll 10^{-17} \text{ W}$ at $v_{\text{inter}} = 1 \text{ m/s}$; for the example with $n = 19$, $m = 27$, for example, $F_{\text{max}} < 0.0001 \text{ nN}$, and $P_{\text{rad}} \approx 10^{-23} \text{ W}$. With good designs, losses from acoustic radiation will decrease as bearing size increases.

The fluctuating forces in a loaded bearing can excite transverse vibrations of the bearing with respect to the sleeve. To avoid this, it is necessary to avoid more than transient operation near the condition

$$\frac{\omega_{\text{bearing}} r_{\text{inner}}}{d_{\text{a}}} = \frac{v}{d_{\text{a}}} = \frac{1}{2\pi} \sqrt{\frac{k_{\text{s,trans}}}{m_{\text{rotor}}}} \quad (10.18)$$

For rotors with dimensions of a few nanometers and transverse stiffnesses $\sim 1000 \text{ N/m}$, the critical speed $v \approx 100 \text{ m/s}$, and the critical frequency $\omega_{\text{bearing}} > 10^{12} \text{ rad/s}$.

10.4.6.2. Shear-reflection drag

Several of the following energy dissipation mechanisms depend on the thermally-averaged phonon transmission coefficient of the bearing interface, T_{trans} . Using the approximations of Eq. (7.41) and (7.42,) and assuming materials with the modulus, atom number density, and Debye temperature of diamond, the transmission coefficient at 300 K is

$$T_{\text{trans}} \approx \frac{z}{1+3z}, \text{ where } z = 2.4 \times 10^{-37} k_{\text{a}}^{1.7}, \text{ and} \quad (10.19)$$

$$T_{\text{trans}} \approx 2.4 \times 10^{-37} k_{\text{a}}^{1.7}, \text{ if } k_{\text{a}} \leq 300 \text{ N/m} \cdot \text{nm}^2. \quad (10.20)$$

As discussed previously, these approximations should overestimate the transmission coefficient for interfaces that are curved or that have mismatched acoustic speeds, and hence should prove conservative in the present context.

Applying the approximations of Section 7.3.6 and again assuming the material properties of diamond at 300 K yields an estimate of the drag power as a function of the sliding speed v , area S , and transmission coefficient T_{trans} of the interface:

$$P_{\text{drag}} < 1.6 \times 10^{-33} k_a^{1.7} v^2 S \quad (10.21)$$

or, applying Eq. (10.20) with the same restriction,

$$P_{\text{drag}} < 1.4 \times 10^{-33} \frac{k_{s,\text{bear}}^{1.7}}{(\ell r_{\text{eff}})^{0.7}} v^2 \quad (10.22)$$

For a bearing with $r_{\text{eff}} = \ell = 2 \text{ nm}$ and $k_{s,\text{bear}} = 10^3 \text{ N/m}$, $P_{\text{drag}} < 3 \times 10^{-16} \text{ W}$ at $v = 1 \text{ m/s}$, and $< 3 \times 10^{-20} \text{ W}$ at 1 cm/s .

Sleeve bearings will frequently be required to support loads along the shaft axis and exhibit stiffness in resisting axial displacements; this can be accomplished by ensuring that the shaft and sleeve have interlocking circumferential corrugations. An interface of this sort will exhibit a stiffness per unit area for transverse displacements (in the axial direction) of $k_{a,\text{trans}}$. This coupling mode permits phonons of a different polarization to cross the interface, providing an independent mechanism for energy dissipation characterized by expressions like those above, but with $k_{a,\text{trans}}$ substituted for k_a (and in a more thorough analysis, a different modulus, etc.). Where the axial stiffness equals $k_{s,\text{bear}}$, the increment in energy dissipation is $\sim 0.5^{1.7} P_{\text{drag}} \approx 0.3 P_{\text{drag}}$.

10.4.6.3. Band-stiffness scattering

Following the procedure in the previous section, expressions analogous to (10.21) and (10.22) for power dissipation from band-stiffness scattering, Eq. (7.35), are

$$P_{\text{drag}} < 2.7 \times 10^{-33} k_a^{1.7} R^2 \frac{\Delta k_a}{k_a} v^2 S \quad (10.23)$$

and

$$P_{\text{drag}} < 2.4 \times 10^{-33} \frac{k_{s,\text{bear}}^{1.7}}{(\ell r_{\text{eff}})^{0.7}} R^2 \frac{\Delta k_a}{k_a} v^2 \quad (10.24)$$

In the coplanar ring model, the parameter R equals $|m/(m-n)|$; if the interatomic spacings in the inner and outer rings are equal, then $R \approx 10$ when $s_{\text{gap}} \approx 0.2 \text{ nm}$ and $r_{\text{inner}} \leq 2 \text{ nm}$. Regardless of bearing radius, differences in surface structure or strain on opposite sides of the interface can be used to ensure that $R \leq 10$.

The parameter $\Delta k_a/k_a$ can be estimated from variations in the stiffness of nonbonded interactions between rows of equally-spaced atoms as a function of their offset from

alignment. Like many such differential quantities, it is strongly dependent on the spatial frequencies involved. For second-row atoms (taking carbon as a model), $\Delta k_a/k_a \approx 0.3$ to 0.4 (at a stiffness-per-atom of 1 and 10 N/m respectively) where $d_a = 0.25$ nm, and ~ 0.001 to 0.003 where $d_a = 0.125$ nm. The latter value of d_a cannot be physically achieved in coplanar rings, but correctly models the interaction of a ring with two other equidistant rings of $d_a = 0.25$ nm and a rotational offset of 0.125 nm.

With the parameters used in the previous section, and $R = 10$ and $\Delta k_a/k_a = 0.4$, $P_{\text{drag}} < 2 \times 10^{-14}$ W at $v = 1$ m/s; with $\Delta k_a/k_a = 0.003$, $P_{\text{drag}} < 1.4 \times 10^{-16}$ W.

10.4.6.4. Band-flutter scattering

Again following the procedure in Section 10.4.6.2, the expression for power dissipation from band-flutter scattering Eq. (7.37) becomes

$$P_{\text{drag}} < 1.1 \times 10^{-31} \frac{k_{s,\text{bear}}^{1.7}}{(\ell r_{\text{eff}})^{0.7}} R^2 \left(\frac{A}{d_a} \right)^2 v^2 \quad (10.25)$$

The amplitude of the interfacial deformation, A , can be roughly estimated from Δk_a , R , d_a , and a characteristic elastic modulus M . From Eq. (3.20) and the associated discussion, it can be seen that Δp in the interface $\leq 3 \times 10^{-11} \Delta k_a$. A pressure distribution varying sinusoidally across a surface with a spatial frequency κ produces displacements that decrease with depth on a length scale $w = 1/\kappa \approx R d_a / 2\pi$, or ~ 0.4 nm for systems with $d_a = 0.25$ nm and parameters like those described in Section 10.4.6.2. The amplitude $A \approx w \Delta p / M$; for $k_a = 8 \times 10^{19}$ (N/m \cdot m²) and $\Delta k_a/k_a = 0.4$, and other parameters as before, $P_{\text{drag}} < 5 \times 10^{-18}$ W at $v = 1$ m/s.

10.4.6.5. Thermoelastic damping

Using diamond material parameters and the effective thickness w from the previous section together with $\tau_{\text{therm}} = 10^{-12}$ s yields the following specialization of Eq. (7.49):

$$P_{\text{drag}} \approx 4.3 \times 10^{-27} 2\pi r_{\text{eff}} \ell w \left(\frac{\Delta p v}{d_a} \right)^2 \quad (10.26)$$

Applying the assumptions of Section 7.4.1 with the parameters assumed in Section 10.4.6.3 and the estimate of ℓ and Δp from the previous section yields $P_{\text{drag}} \approx 6 \times 10^{-16}$ W ($\Delta k_a/k_a = 0.4$) and $\sim 8 \times 10^{-20}$ W ($\Delta k_a/k_a = 0.003$), for $v = 1$ m/s.

10.4.6.6. Summary

For well-designed bearings on a nanometer scale at 300 K, acoustic radiation losses will typically be negligible compared with losses resulting from phonon interactions. The latter loss mechanisms all scale as v^2 . Combining results from the sample calculations on a bearing with $k_{s,bear} = 1000$ N/m and $r_{eff} = \ell = 2$ nm yields an estimated total value of $P_{drag} < 2 \times 10^{-14} v^2$ W ($d_a = 0.25$ nm), dominated by band-stiffness scattering, or $< 5 \times 10^{-16} v^2$ W ($d_a = 0.25$ nm), with a substantial contribution from shear-reflection drag. (Note that each drag model contains different approximations, often intended to provide a gross upper bound on the magnitude of the drag; constants written to two significant figures in the expressions of this section generally lack the accuracy this notation might imply.) Using the higher value and $v = 1$ m/s, such a bearing is estimated to dissipate $< 0.06 kT$ per rotation.

10.4.7. Sleeve bearings in molecular detail

Sleeve bearings can usefully be examined at two levels of molecular detail: interfacial structure and overall structure. The design of relatively large bearings can exploit strained cylindrical shells. Bearings of this sort can be viewed as forming families with specific unstrained structures and crystallographic orientations (relative to the interface and the bearing axis), and with specific surface terminations at the sliding interface. Within such a family, the bearing radius r_{eff} and the spacing between the surfaces s_{gap} are determined within broad limits by the choice of the inner and outer circumferences (in unstrained lattice units). For these bearings, specification of the interfacial structure is primary, since the overall structure is simple and repetitive.

For smaller bearings, in contrast, strained cylindrical shells become a poor model. The structure of the shaft then becomes a special case rather than a member of a parameterized family, and the overall structure must be considered as a unit. Examples drawn from both classes are given in the following sections.

10.4.7.1. Bearing interfaces for strained-shell structures

Figures 10.15 and 10.16 illustrate several pairs of terminated diamond surfaces, each forming an interface suitable for use in a symmetrical sleeve bearing. These interfaces have differing properties with respect to axial stiffness and drag.

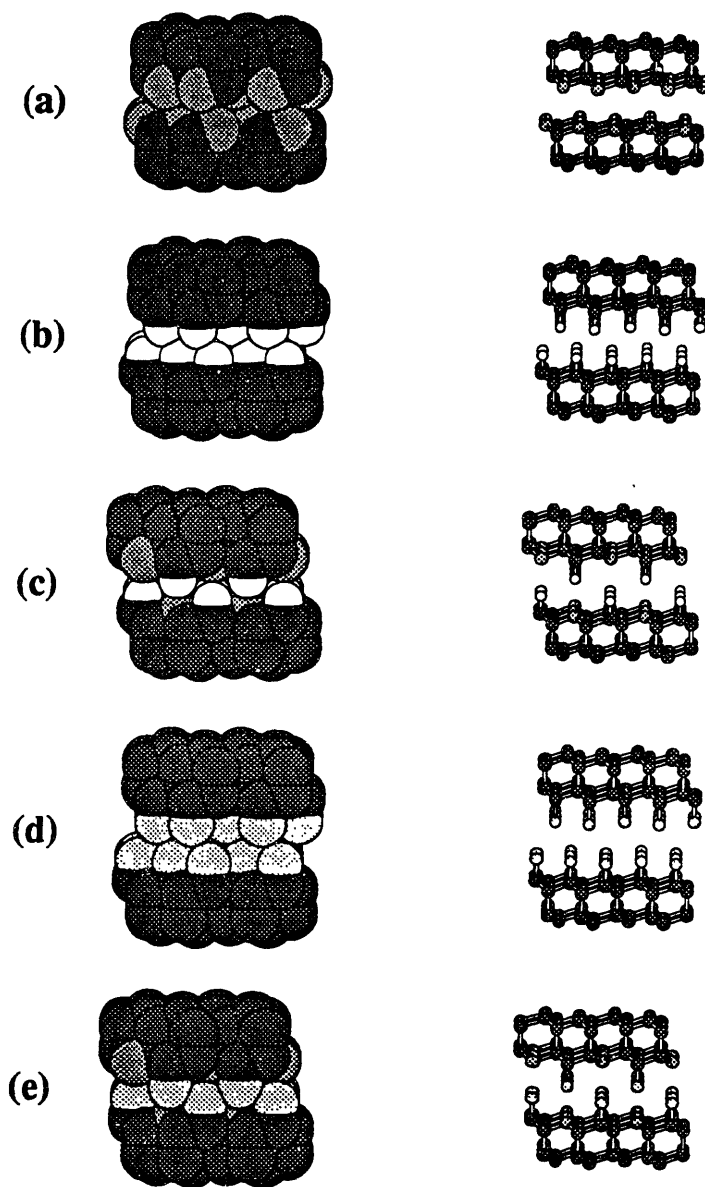


Figure 10.15. Several sliding interfaces based on diamond (111) surfaces, with nitrogen termination (a), hydrogen termination (b), alternating, interlocking rows of nitrogen and hydrogen termination (c), fluorine termination (d), and alternating, interlocking rows of nitrogen and fluorine termination (e). Note that pairs of surfaces from (a) (b) and (d) could be combined, as could a pair from (c) and (e).

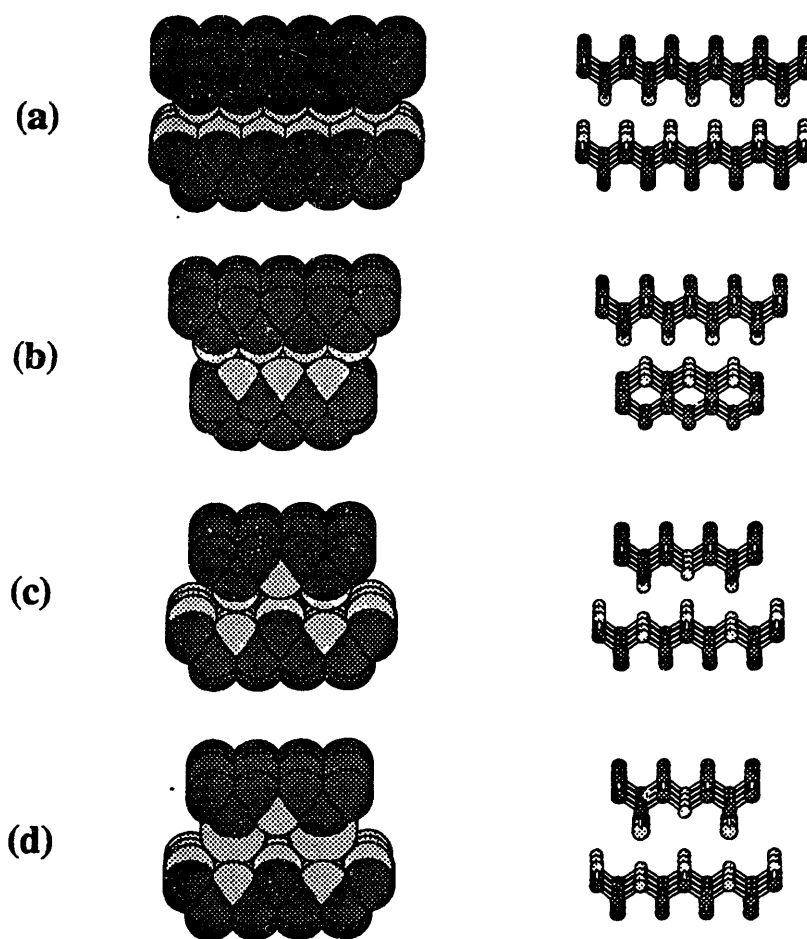


Figure 10.16. Several sliding interfaces based on diamond (100) surfaces, with oxygen termination and aligned rows (a), oxygen termination and crossed rows (b), oxygen termination with alternating deleted rows, providing an interlocking surface (c), and surfaces like (c), but with the exposed terminating rows on one surface consisting of sulfur rather than oxygen.

Each of the interfaces shown will exhibit substantial axial stiffness at suitably small values of s_{gap} . Along the axis, the surface atomic rows on opposite sides of each interface have identical spacings, hence the sinusoidal potentials for sliding in this direction add in phase over the entire interface. Interfaces 10.15(c), 10.15(e), 10.16(c) and 10.16(d) include interlocking grooves, increasing the energy barrier for axial sliding and (in most instances) the stiffness as well. In these structures, the axial stiffness can equal or exceed the transverse-displacement stiffness $k_{s,\text{bear}}$. An interface with mismatched spacings in both dimensions (e.g., combining 10.15(b) and 10.16(a)) will permit both rotation and axial sliding.

Interfaces 10.15(c), 10.15(e), and 10.16(a–d) exhibit values of $d_a \approx 0.25$ nm, corresponding to the high-drag cases in Sections 10.4.6.3 and 10.4.6.4. Interfaces 10.15(b) and 10.15(d), however, will (in the absence of significant axial loads) exhibit $d_{a,\text{eff}} \approx 0.125$ nm: an atom on one surface interacts equally with two staggered rows on the other, halving the effective spacing for most purposes. This value of $d_{a,\text{eff}}$ corresponds to the low-drag cases in Sections 10.4.6.3 and 10.4.6.4. The nitrogen-terminated (111) surface, Fig. 10.15(a), has $d_{a,\text{eff}} \approx 0.125$ nm where first-layer interactions are concerned, but the second atomic layer will introduce non-negligible interactions with $d_a \approx 0.25$ nm.

10.4.7.2. Interfacial stability

Each of the interfaces in Figures 10.15 and 10.16 will be stable enough for practical use under the baseline conditions assumed in this volume (no extreme temperatures, no UV exposure, no extraneous reactive molecules, no extreme mechanical loads). The low-valence atoms used to terminate each surface form strong bonds to carbon and (usually) weaker bonds to one another. A reaction between one surface and the other would typically form a bond between two low-valence atoms at the expense of cleaving two bonds to carbon. Since this would be a strongly endoergic process, the energy barrier will be large (> 500 mJ) and the rate of occurrence negligible ($< 10^{-39} \text{ s}^{-1}$). These remarks apply with equal force to a wide variety of sliding and rolling interfaces with similar termination by low-valence atoms.

Graphitic interfaces are also attractive, but the potential reactivity of their unsaturated tetravalent atoms demands attention. Experiments show that graphite transforms to a transparent, non-diamond phase at room temperature under pressures of ~ 18 GPa (Utsumi and Yagi 1991); this pressure corresponds to a compressive load of ~ 0.5 nN per atom (with an associated stiffness on the order of 20 N/m). The transformation is nucle-

ated at specific sites in a crystal, and can be observed to spread over a period of ~ 1 hour. Presumably small areas of graphitic bearing interface can be made that lack suitable nucleation sites, and so interfacial pressures approaching 18 GPa should be consistent with chemical stability.

10.4.7.3. A specific strained-shell structure

Figure 10.17 illustrates a strained-shell structure containing 2808 atoms, with a shaft of 34-fold rotational symmetry and a sleeve of 46-fold rotational symmetry; $\text{lcm}(n,m)/n = 23$. The dimensions of the interface are somewhat ill-defined, but approximate values are $r_{\text{eff}} \approx 3.5$ nm and $\ell \approx 1.0$ nm; the external radius at the illustrated termination surface is ~ 4.8 nm. This bearing was designed (perhaps over-designed) for large axial stiffness, achieved by combining interlocking ridges with large nonbonded contact forces (Fig. 10.18 shows a section through the interface). The molecular mechanics model (see Fig. 10.17 caption) predicts that the maximally-strained C–C bonds of the outer surface of the sleeve have lengths of ~ 0.166 nm. The closest nonbonded contacts across the interface are ~ 0.284 (for SiO) and ~ 0.270 (for SiN); the corresponding forces are ~ 0.56 and ~ 1.1 nN; the corresponding stiffnesses are ~ 20 and ~ 40 N/m. The compliance of the interface associated with nonbonded forces is $\sim 10^{-21}$ m²·m/N (from MM2/C3D+), comparable to the shear compliance of a ~ 0.5 nm thickness of diamond. Bond angle bending will add significant compliance to the interface, but the total compliance will not exceed the shear compliance of a sheet of diamond a few nanometers thick. For the bearing as a whole, the axial stiffness should exceed 2000 N/m.

Energy minimization of this structure yielded no rotation of the shaft with respect to the sleeve, regardless of angular displacement. Examination of energy differences for minimized structures as a function of angular displacement indicates values of $\Delta\mathcal{V}_{\text{barrier}} < 0.03$ mJ. Owing to high interfacial stiffnesses, however, the estimated drag in a bearing of this sort will be several times larger than that calculated the examples in Section 10.4.6.

10.4.7.4. Small sleeve bearing structures

In small sleeve bearings, the pursuit of high-order rotational symmetry dramatically limits the set of acceptable structures, although this set remains large enough to make enumeration challenging. A segment of any of the roughly cylindrical structures shown in Chapter 9. could serve as a shaft, given a suitable sleeve. Thin strained shells, backed up

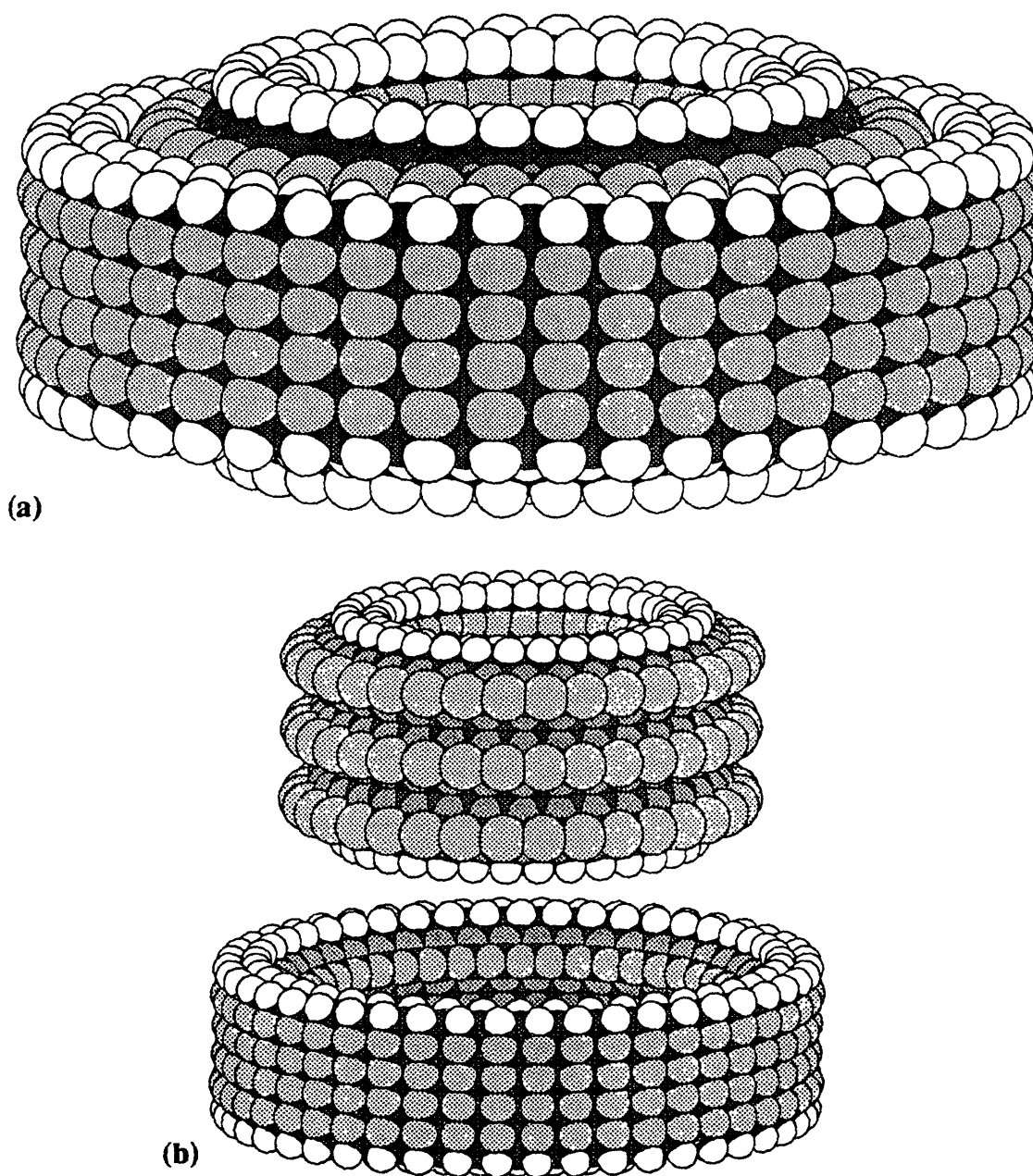


Figure 10.17. A 2808-atom strained-shell sleeve bearing with an interlocking-groove interface derived by modifying the diamond (100) surface; (a) shows the shaft within the sleeve, (b) shows and exploded view. This design was developed with the aid of an automated structure-generation package written by R. Merkle, then was energy-minimized and analyzed using the Dreiding potential energy function provided by the Polygraf[®] molecular modeling system (Molecular Simulations, Inc., Pasadena).

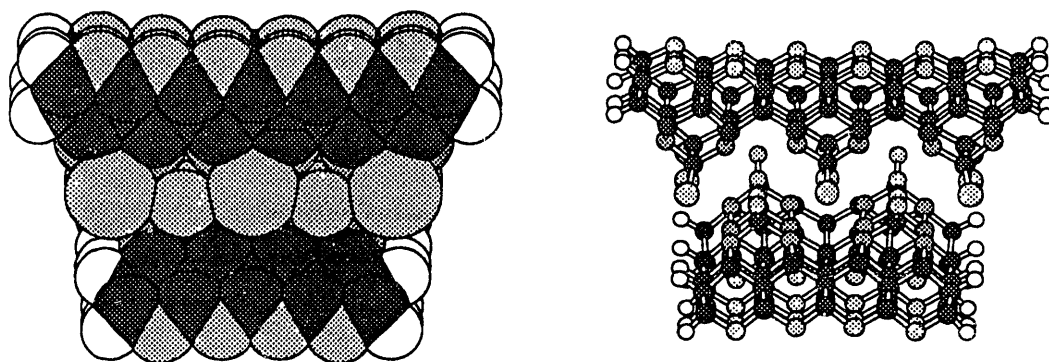


Figure 10.18. A section through the interface of the bearing shown in Fig. 10.17. The view roughly parallels the planar diagrams shown in Fig. 10.16, differing in the presence of curvature and in the use of a different (100)-based surface modification for the interface structure. The use of sulfur bridges on the outer shaft surface rather than oxygen both reduces strain (via longer bonds) and increases interfacial stiffness (via larger steric radii).

by layers with regular dislocation-like structures, could serve the latter function, as could less-regular, special-case structures.

10.4.7.5. A specific small sleeve bearing structure

An example combining a special shaft with a special sleeve is shown in Figure 1.1. This structure makes use of chains of sp^3 nitrogen atoms to form ridges (having high stiffness and spatial frequency) on both the shaft and the sleeve. These features require attention because N–N bonds are known to be weak and *isolated* chains of this sort are apparently unstable (a search of the chemical literature failed to identify a well-characterized example). These chains, however, are not isolated. Each N atom is also bonded to carbon and subject to the familiar constraints of a diamondoid structure: momentary thermal cleavage of an N–N bond will be resisted (and usually reversed) by elastic restoring forces from the surrounding structure. The most accessible failure mode in this system is apparently the transformation of a nitrogen chain in the shaft into a series of π -bonded dimers (Fig. 10.19), but estimates of bond energies and strain energies suggest that this transformation is only moderately exoergic (~ 70 mJ?) on a per-bond-cleaved basis. Moreover, the formation of a single π -bonded dimer by cleavage of two adjacent sigma bonds will be strongly endoergic (sacrificing two σ bonds to create, ini-

tially, one strongly twisted π -bond), and the simultaneous cleavage of six sigma bonds should be energetically prohibitive. Accordingly, these structures appear sufficiently stable for use (although none of the following design and analysis depends on this).

The stiffness of the bearing interface can be computed from the change in non-bonded interaction energies as a function of relative displacement of the shaft and sleeve, in the absence of structural relaxation. Computer modeling yields an axial stiffness of ~ 360 N/m and a transverse stiffness (which is isotropic) of ~ 470 N/m. Rotational energy barriers (computed with structural relaxation) were found to be < 0.001 mJ. All computations used the MM2/C3D+ model extended with parameters to accommodate N–N–N bond angles. The parameter θ_0 for N–N–N angle bending was set at 114.5° to fit AM1 semiempirical computations on H_2NNHNH_2 (performed with assistance from R. Merkle); k_θ was set at 0.740 aJ/rad², equaling the MM2 value for N–N–C angle bending. Bearing stiffnesses (but not barriers) are sensitive to the choice of θ_0 . Supplemental torsional parameters are of less significance; values were chosen to match analogous MM2 values.

The total strain energy in this structure is large (~ 12 aJ), but only ~ 0.53 aJ of this is in the form of bond stretching, and this energy is well-distributed over many bonds. $\sim 71\%$ of the strain energy is in the form of bond bending, much of this owing to the presence of 22 cyclobutane rings within the structure. The closest nonbonded distance between shaft and sleeve is ~ 0.26 nm (N|N). For a relaxed model of a shaft popping into (or out of) a sleeve, with the nitrogen chains approximately coplanar, the total energy is

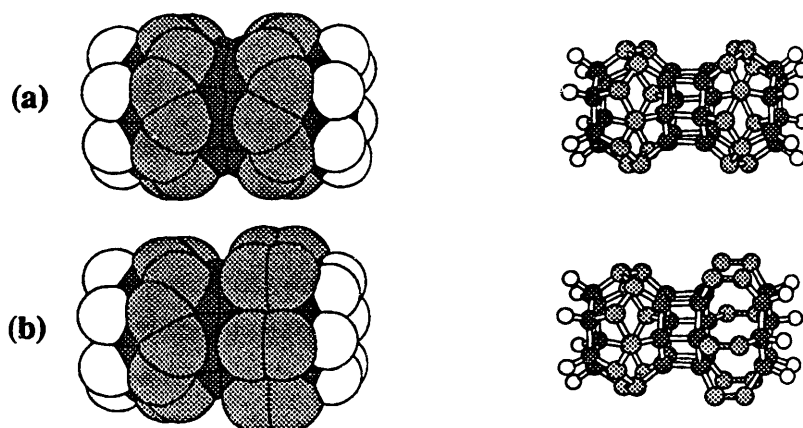


Figure 10.19. The shaft from the structure of Figure 1.1, undamaged (a) and after cleavage of six relatively weak N–N bonds (b); see text.

increased by ~ 1.7 aJ and the closest N|N distance is ~ 0.236 nm; bond lengths remain reasonable, the N–N bonds are under stabilizing, compressive loads, and the estimated peak forces in achieving this configuration are small compared to bond tensile strengths. Assembly of this bearing from separate components accordingly appears feasible.

10.4.8. Less symmetrical sleeve bearings

10.4.8.1. Asymmetries to compensate for load

Transverse loads on bearings shift the axis of the shaft with respect to that of the sleeve; the use of a symmetrical sleeve is then no longer motivated by the symmetry of the situation. Assume that the sleeve is fixed, and the shaft rotating under a constant transverse load with an orientation fixed in space (Fig. 10.20). The previous analysis in this chapter assumes that this asymmetric load is supported by the topmost atomic layers, chiefly through differences in steric repulsion caused by shaft displacement. Alternatively, however, the load on the shaft can (in many instances) be supported by differences in van der Waals attraction resulting from an asymmetric structure having

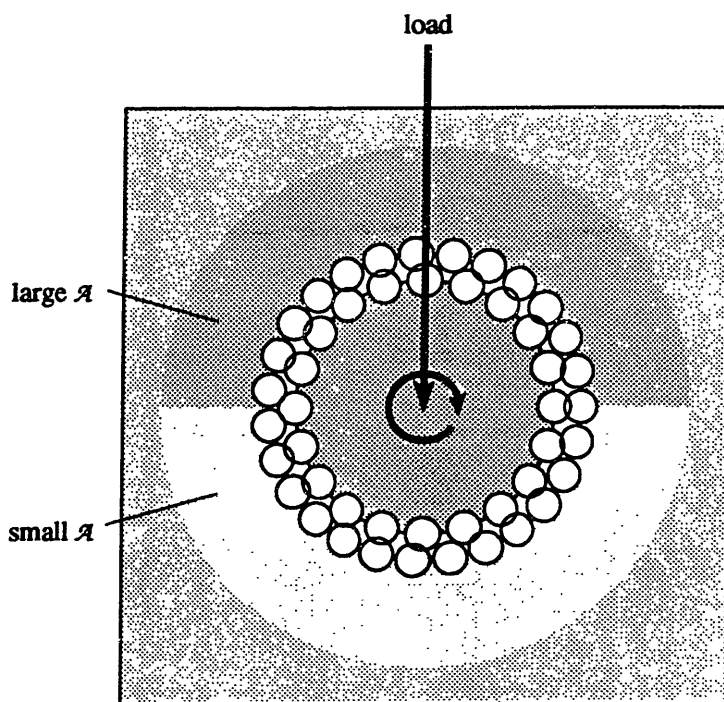


Figure 10.20. Schematic diagram of a loaded bearing, with compensating asymmetric van der Waals attractions.

regions of different Hamaker constant \mathcal{A} (e.g., differing atom number densities) immediately behind the surface layer. This approach can reduce $\Delta\mathcal{V}_{\text{barrier}}$, and can increase load-bearing capacity without increasing interfacial stiffness.

In small loaded bearings, $\Delta\mathcal{V}_{\text{barrier}}$ can be large. Where the shaft is of high symmetry and the sleeve has > 10 interfacial atoms, the calculations of Section 10.3 suggest that an asymmetric placement of sleeve atoms can be chosen to ensure that $\Delta\mathcal{V}_{\text{barrier}} < 1$ mJ. In many instances, a slightly perturbed version of a symmetrical structure will accomplish this.

10.4.8.2. Asymmetries to simplify construction

Bearings having interfaces with shallow grooves or relatively large s_{gap} can often be assembled from separate shafts and sleeves in the obvious manner, by pressing the shaft into the sleeve. Large loads can often be applied, and the final energy minimum (with respect to axial displacement) can be quite deep and stiff.

Structures with more strongly interlocking grooves cannot be assembled in this fashion. If they are to have full symmetry, the sleeve must typically be synthesized *in situ* around the shaft (e.g., building out from polymeric bands that become the ridges on the sleeve), or must have a final assembly step involving closure of an adhesive interface. The latter process will present little difficulty, provided that the constraint of perfect symmetry is dropped. A sleeve made from two C-shaped segments, with an atomic-scale discontinuity the two seams, will have lower symmetry but can nonetheless be designed to exhibit low $\Delta\mathcal{V}_{\text{barrier}}$.

10.5. Other sliding-interface bearings (and bearing systems)

The above results regarding irregular objects and symmetrical sleeve bearings shed light on wide range of other sliding-interface systems. Among these are nuts turning on screws, rods sliding in sleeves, and a class of constant-force springs. Energy dissipation analyses are omitted, but follow the principles discussed previously.

10.5.1. Nuts and screws

The thread structure of a nut-and-screw combination can formally be generated by dividing a grooved, strained-shell sleeve bearing parallel to the axis along one side, shifting one cut surface in the axial direction by an integral number of groove spacings, and reconnecting. The result is locally similar to the original bearing, save for the introduc-

tion of a helical pitch in the grooves and ridges, which accordingly must come to an end as some point (in a straight, finite structure).

What will be the static friction of such a structure, assuming that the helical atomic rows of the inner and outer surfaces have non-matching spacings? (Note that these spacings need not be commensurate; in good designs, the spacings will lack small common multiples.) Figure 10.5 approximated the potential of an atom with respect to a row as a single sinusoidal potential with some amplitude and phase, represented as a vector magnitude and angle in a plane. In this representation, the potential of a series of atoms in one surface of a uniform, non-matching interface takes the form shown in Figure 10.21: all vectors are of the same magnitude, and each is rotated with respect to the last by a fixed angle. Where that angle is not zero, the resultant vector always lies on a circle passing through the origin. Its magnitude oscillates between fixed bounds and periodically assumes a small value; $\Delta\mathcal{V}_{\text{barrier}}$ and the static friction accordingly do the same. The bound on the magnitude of the barrier is

$$\Delta\mathcal{V}_{\text{barrier,max}} = \frac{\Delta\mathcal{V}_{\text{barrier, single}}}{2} \sqrt{1 + \tan^2\left(\frac{\pi - \phi}{2}\right)} \quad (10.27)$$

where ϕ is the phase angle between succeeding atoms in the surface under consideration.

This result indicates that the static friction of a nut-and-screw system (under low loads) will depend chiefly on the end conditions. With the right choice of interface length, $\Delta\mathcal{V}_{\text{barrier}}$ will be low because the resultant vector will be of small magnitude. For other choices of interface length, $\Delta\mathcal{V}_{\text{barrier}}$ can be made low by the methods discussed in Section 10.3: tuned structural irregularities can be introduced at one end of a nut in such a way that the amplitude and phase of their contribution to the potential cancels the residual contribution from the regular portion of the nut. As shown in Figure 10.22, where the

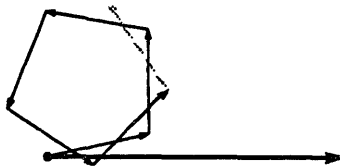


Figure 10.21. Interatomic potentials as in Fig. 10.5, for regularly-spaced atoms in one surface moving over regularly-spaced atoms in another.

overlap of the nut and screw is variable, minimizing the sinusoidal component of the potential can require tuned irregularities at two sites.

10.5.2. Rods in sleeves

The analysis of rod-in-sleeve systems is entirely analogous to that for nut-and-screw systems, save that the helical grooves take the degenerate form of straight lines. Again, end conditions determine the magnitude of the static friction, and again, choice of length of tuning of irregular structures can yield low values of $\Delta\mathcal{V}_{\text{barrier}}$. As noted previously, cylindrical interfaces can be designed to permit simultaneous axial sliding and rotation in any proportion.

10.5.3. Constant force springs

In the variable-overlap case described mentioned in Section 10.5.1, suppression of sinusoidal potentials does not leave a flat potential, because the energy of the system is in

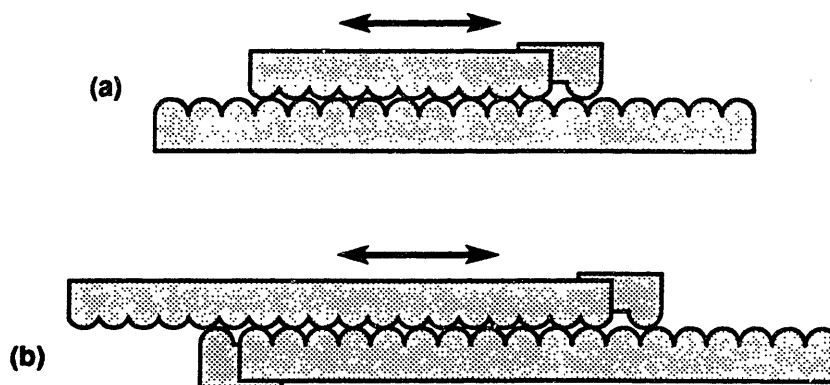


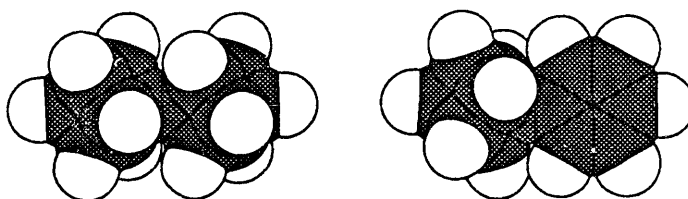
Figure 10.22. Linear representation of the sliding of two finite but otherwise regular surfaces over one another. In (a), the range of motion of one surface places it within the width of the other surface at all times; this corresponds (for example) to a nut turning on the middle of a long screw. In (b), the range of motion of the surfaces enables each to extend beyond one limit of the other at all times; this corresponds (for example) to a screw partially inserted into a deep threaded hole. In (a), the irregularities corresponding to both ends of the overlap region move together over a surface of a single spatial frequency; tuned irregularities at one end will suffice to keep $\Delta\mathcal{V}_{\text{barrier}}$ low. In (b), the irregularities move in opposite directions over surfaces of differing spatial frequency, and tuned irregularities are generally needed at both ends (that is, on both sides of the interface) if the sinusoidal component of the potential is to be minimized.

general a function of the overlap. Where local interactions are dominant, the potential energy will be proportional to the overlap, with a positive or negative constant of proportionality depending on the net energy of the surface-surface interaction. A sliding rod in a sleeve of this sort, tuned for smooth motion, will act as a constant-force spring over its available range of motion. Since the characteristic attractive interaction energies of surfaces are on the order of 100 mJ/nm^2 , the force with which a rod can be made to retract into a sleeve will be on the order of $0.3r \text{ nN}$, where r is in nanometers (this relationship breaks down for small r owing to the finite range of van der Waals attractions). Since repulsive interaction energies can be far larger, forces for a constant-force spring operating in this mode can be far larger. In the latter case, large additional energies can be stored in elastic deformation of the rod and sleeve.

10.6. Atomic-axle bearings

10.6.1. Bonded bearings

Sigma bonds permit rotation, in the absence of mechanical interference between the bonded moieties. Barriers for rotation vary. Two model systems of interest are cubylcubane **10.1** and phenylcubane **10.2**



10.1

10.2

with MM2/C3D+ values of $\Delta\mathcal{V}_{\text{barrier}}$ barriers of 11.5 and 0.3 mJ respectively; note that the latter has a structure with 3-fold rotational symmetry interacting with one with 2-fold symmetry. The stiffness for shearing displacements of sigma bearings of this sort will be about twice the transverse-displacement stiffness of a sigma bond, $\sim 60 \text{ N/m}$ for structures like cubylcubane. A rotor supported by two such bearings would have a transverse displacement stiffness of $\sim 120 \text{ N/m}$, and (with a proper choice of relative phases) values of $\Delta\mathcal{V}_{\text{barrier}} \ll 0.3 \text{ mJ}$. This class of bearing is extremely compact, requiring (by one set of accounting rules) no atoms and no volume.

Using a $\text{—C}\equiv\text{C—}$ unit in place of a sigma bond will drop $\Delta\mathcal{V}_{\text{barrier}}$ to near zero, at the expense of increasing the volume and greatly decreasing resistance to shearing

displacements.

10.6.2. Atomic-point bearings

A rotor with protruding atoms on opposite sites can be captured between a pair of surfaces with matching hollows. With a suitable choice of geometry for the interacting surfaces, values of $\Delta\mathcal{V}_{\text{barrier}}$ can be low. Under substantial compressive loads, stiffness can be moderately high. A macroscopic analogue of a bearing of this sort would have sliding interfaces, but the reduction of the contact region to a single atom (on one side of the interface), together with the placement of that atom on the axis of rotation, makes atomic-point bearings qualitatively different.

10.7. Gears, rollers, and belts

10.7.1. Spur gears

Spur gears find extensive use in machinery, chiefly for transmitting power between shafts of differing angular frequency. Spur gears achieve zero slip (so long as the teeth remain meshed), and ideally exhibit minimal energy barriers during rotation, minimal dynamic friction, and maximal stiffness resisting interfacial shear.

Nanomechanical gears will be able to exploit a variety of physical effects to implement “teeth.” These include complementary patterns of charge, of hydrogen bonds, or of dative bonds. The most straightforward effect to analyze, however, is steric repulsion between surfaces with complementary shapes (which is, after all, how macroscopic gears work).

In conventional gearing, teeth are carefully shaped (e.g., in involute curves) to permit rolling motion of one tooth across another. Because a tooth on one gear meshes between two teeth on the opposite gear, but can only roll across one of them, a clearance (backlash) must be provided. Accordingly, on reversal of torque, teeth lift from one rolling contact before making another. These complexities of geometry are both impossible and unnecessary in nanomechanical gears. Since sliding of one tooth over another is just another variation on a sliding-interface bearing, there is no need to construct involute surfaces. Since the steric interaction between atoms is soft, and since sliding is acceptable, no backlash is required (and would not be a well-defined quantity if it existed).

Small nanomechanical gears of this class will use single atoms or rows of atoms as steric teeth; any of the grooved interfaces in Figures 10.15 and 10.16 could serve this

role, in strained-shell structures. As with bearings, a variety of small special-case structures will also be feasible. And again, models based on rigid, circular arrays of atoms reduce computational costs while preserving the essential physics.

10.7.1.1. A relaxation-free model of meshing gear teeth

Figure 10.23 illustrates a model of steric gears as rigid, coplanar rings of atoms. The calculations described in this section, like those in the discussion of bearings, assume the MM2 exp-6 nonbonded potential Eq. (3.9) with H atoms separated by $d_a = 0.3$ nm. Again, the general nature of the results is insensitive to the choice of potential, atom type, and interatomic spacing.

Figure 10.24 plots barrier heights as a function of the number of teeth n (here, equal for both gears) and the separation s_{gap} . The lower family of curves shows barriers for corotation at a uniform angular velocity without slippage. The upper family of curves shows barriers for slipping of one ring with respect to another, based on a search for minimum-energy pathways at a range of rotational angles. As can be seen, with small values of s_{gap} (< 0.12 nm) and moderate numbers of teeth (> 20), energy barriers to slippage are large (> 500 mJ) and energy barriers to corotation are small (< 0.01 mJ).

Under interfacial shear loads (required, for example, to transmit power), the symmetry of the system is degraded, and corotational barriers are larger. The effects of load can be modeled by a constant angular offset between the rings relative to their minimum-energy, symmetrically-meshed geometry. Figure 10.25 shows corotational barrier heights for two values of n at several values of the offset (measured by displacement at the ring circumference). For n as small as 10 and offsets as large as 0.01 nm, the corotational barrier remains < 1 mJ; at small values of s_{gap} , this offset is adequate to transmit a shear

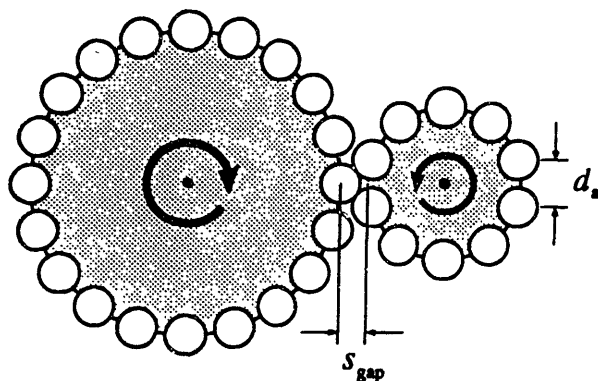


Figure 10.23. Coplanar ring model for a steric gear.

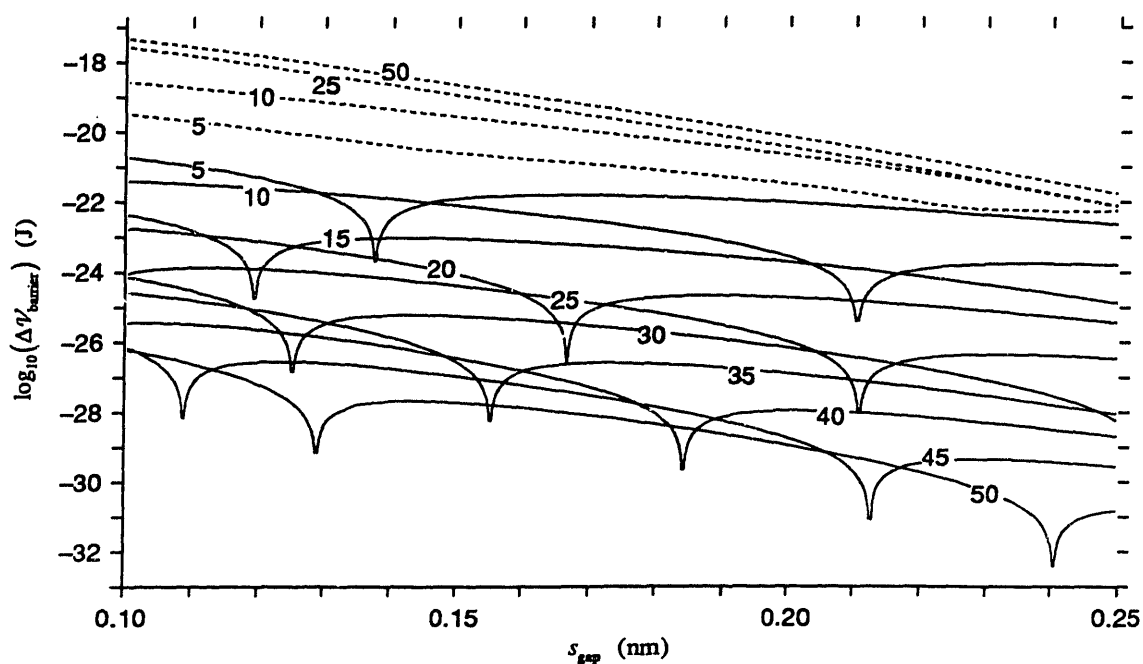


Figure 10.24. Barriers for gear corotation (solid lines) and slippage (dashed lines) in the rigid coplanar ring model (see text), for a range of inter-ring separations s_{gap} and a variety of tooth-numbers n .

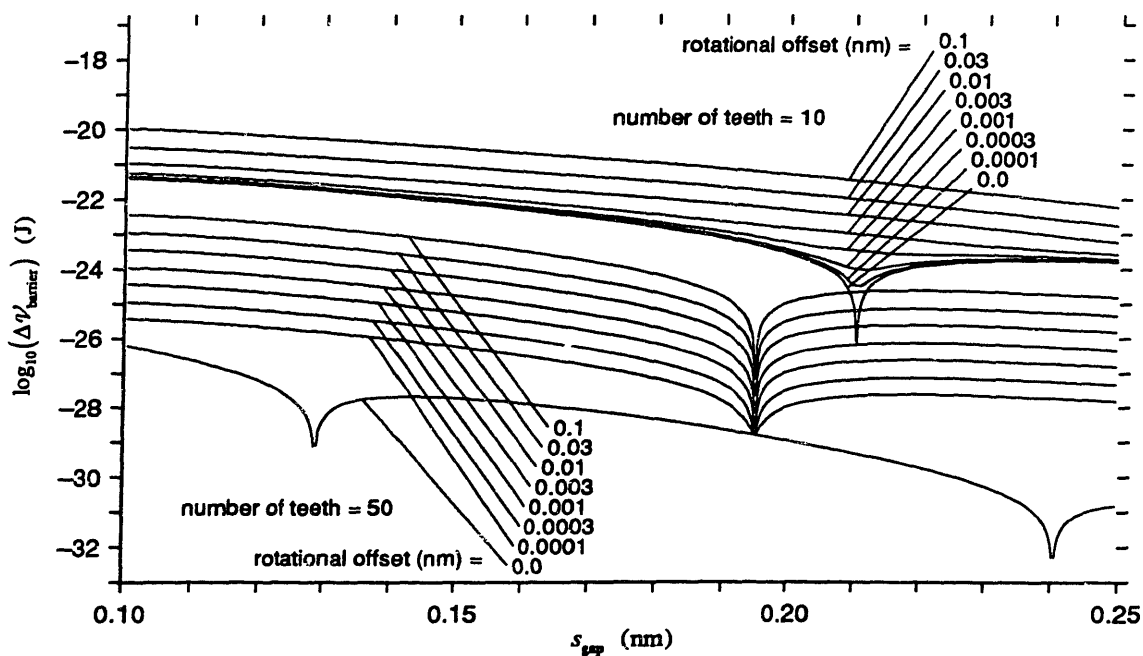


Figure 10.25. Barriers for gear corotation as in Fig. 10.24, for $n = 10$ and 50 , modeling various torsional loads with varying values of inter-ring rotational offset (see text).

force of ~ 1 nN per fully-meshed tooth. At a gear rim speed $v = 1$ m/s, this corresponds to a transmitted power of 1 nW.

A prominent feature of Figures 10.24 and 10.25 is the presence of sharp dips in the barrier height at locations that depend both on n and s_{gap} . Sharp dips in barrier height are a robust feature of a broad class of nanomechanical systems, for reasons illustrated by the diagrams in Figure 10.26. For a single tooth, the energy is low (and nearly constant) before and passing through the meshing zone, and is high (and nearly constant) as it passes directly between the two gears (as in the potential energy curves shown in Fig. 10.26). The potential energy of a gear (in the present model) is simply the sum of the nonbonded interaction energies of its teeth, occupying evenly-spaced points along a suitable potential curve. When a single tooth, or a gap between a pair of teeth, is precisely between the two gears, the slope of the potential energy function is zero by symmetry (see diagrams (a) and (b) in Fig. 10.26). The potential will be periodic (again by symmetry); in the sinusoidal approximation, the slope of the potential will be a maximum at a position (c), half-way between (a) and (b). The slope of the bearing potential energy function at position (c) can be represented as the sum of the slopes at the points in diagram (d), folding points from the left side to the right side, and giving them negative weights in the sum.

The sign of this slope will depend on the positions and weights of the points on the shoulder of the potential curve. Where this shoulder is well away from the center (as it

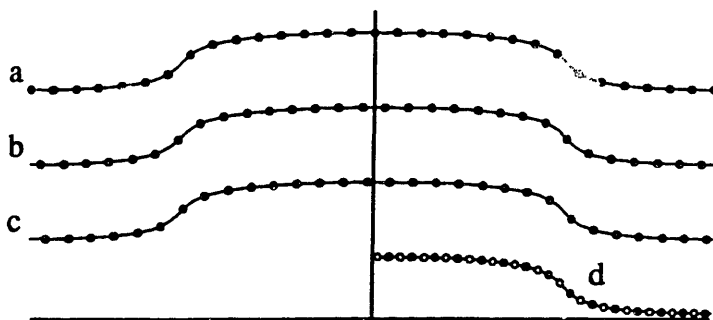


Figure 10.26. Potential energies of gear teeth as a function of rotational angles. Diagrams (a) and (b) represent symmetrical positions, and (c) an intermediate position. Diagram (d) represents (c) with the left-hand points folded over and reversed in sign, to illustrate certain properties of the slope of the potential function with respect to gear rotation angle (see text).

was not in Fig. 10.2), changes in the spacing of the points, or in the position of the shoulder (e.g., by changing n or s_{gap}), can readily change the sign of the slope. For a clear example, consider an inter-point spacing that places only one point in a region of high slope: by varying the parameters, points of either positive or negative weight can be placed in that position.

Finally, where state (c) is of zero slope, it becomes an extreme of the potential energy function, the spatial frequency of barrier-crossing is doubled, and the barrier heights are greatly reduced (much as barrier heights are reduced by higher spatial frequencies in bearings). In moving along a curve like those in Figures 10.24 and 10.25, the energy difference between states like (a) and (b) in Figure 10.26 changes sign. A result of this sort can be expected when the interatomic potential has a form like that in Figure 10.26, with a low curvature at the point of maximum energy.

10.7.1.2. Energy dissipation in gear contacts

The energy dissipation mechanisms in gears parallel those for sliding contacts on a surface (within the approximations already adopted). Acoustic radiation losses will be small for well-designed gears. Relative to the figures used in Section 10.3.4, typical stiffnesses and compressive loads are anticipated to be ~ 10 times larger (~ 300 N/m, ~ 10 nN), resulting in phonon scattering cross sections and thermoelastic effects ~ 100 times larger. The associated power levels (at a 1 m/s interfacial rolling speed) are $\sim 3 \times 10^{-14}$ and $\sim 10^{-16}$ W for phonon scattering and thermoelastic drag, respectively. For a gear system operating at a shear force of 1 nN, these losses amount to $\sim 0.003\%$ of the transmitted power.

10.7.1.3. Integrated bearings and gears

Gears of the sort discussed in this section will require substantial compressive loads at the interface to ensure good meshing and stiff contact between teeth. In macroscopic gearing this would be achieved (were it necessary) by means of loads transmitted through bearing surfaces on adjacent shafts. In nanomechanical gearing, however, the distinction between a gear interface and a bearing interface is chiefly a matter of relative interatomic spacings and curvatures of the opposing surfaces. The computational examples of bearings and gears in this chapter have used the same spacings and atom types on the convex surface of each interface, differing only in the nature of the opposed surfaces. Thus, a single surface can serve both roles, as in Figure 10.27, and the required compressive loads

can be transmitted to the housing structure by the gear surfaces themselves. In a system of this sort, symmetry constraints do not guarantee low values of $\Delta\mathcal{V}_{\text{barrier}}$, but tuning of the interface can readily be applied.

10.7.2. Helical gears

Gears can be constructed from concentric strained shells with opposing torsional shear deformations locked into each by bonding at the ends. These will have surface atomic rows subject to a gentle helical deformation. Matching gear surfaces of this sort can be made to mesh more smoothly than spur gears of the same radius; in effect, n is larger.

10.7.3. Rack-and-pinion gears and roller bearings

The principles of spur and helical gears are equally applicable when one of the toothed surfaces is flat (resulting in a rack-and-pinion gear system) or concave (as is found in planetary gearing). A gear without a shaft can serve as a roller bearing between two flat surfaces, or between two concentric cylindrical surfaces. The barrier to slippage in this instance need not be high enough to transmit large shear forces, but if it is sufficiently high to prevent thermally-activated transitions, then a series of roller bearings can be made to keep a uniform spacing around the raceway without requiring a cage.

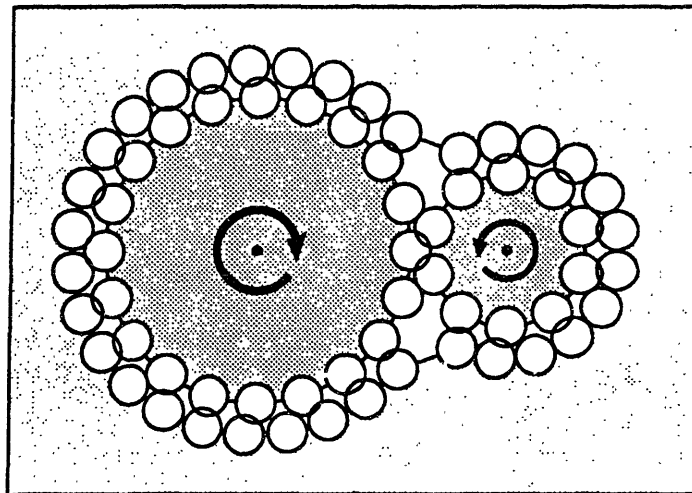


Figure 10.27. Model of two gears supported by bearing surfaces.

10.7.4. Bevel gears

Where the axes of two coplanar shafts intersect at an angle, power can be transmitted from one to another by means of bevel gears (essentially, rolling cones). In nanomechanical systems, the soft interactions of atomic teeth permit two non-coplanar rings of teeth to mesh essentially as well as coplanar rings; bevel gears with single-atom teeth thus present no special problem.

Larger nanomechanical bevel gears, however, cannot be directly patterned on macroscopic gears, owing to the impossibility of making atomic rows that converge smoothly (in spacing and steric radii) toward the tips of the cones. One alternative is to form a pair of bevel gears from conical surfaces with matching patterns of teeth and holes, without attempting to form rows like those in traditional bevel gears. The modified (111) surfaces in Figure 10.28, developed into cones, have this property; with a modest degree of strain ($\sim 6\%$), they can form cones having a 45° half-angle, suitable for shafts meeting at 90° . In small bevel gears based on this approach, the structure immediately beneath the surface would have to depart from that of a simple strained lattice, and the degree of achievable regularity is at present uncertain. Where all radii of curvature exceed 1 nm, however, irregularities can be buried more deeply and the structural choices for those irregularities

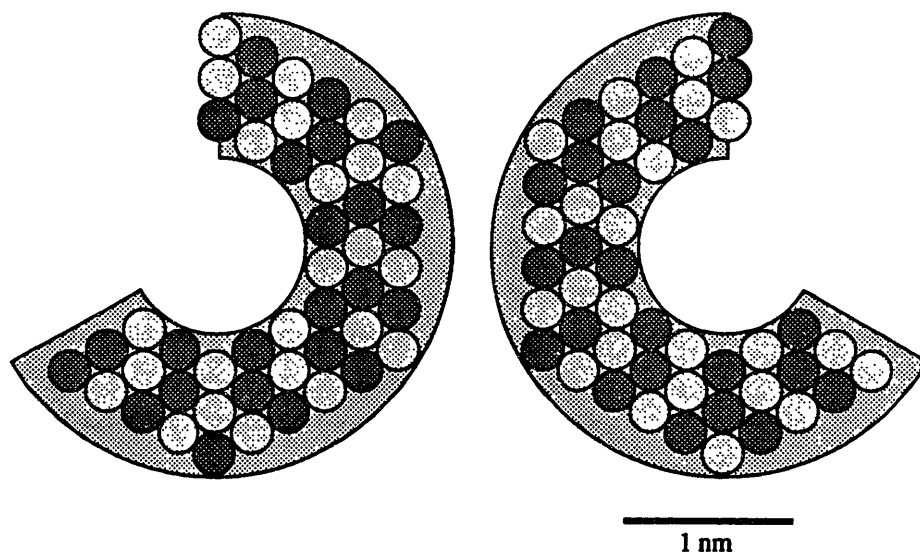


Figure 10.28. Surfaces for a matched pair of small bevel gears; light and dark circles represent protruding atoms (e.g., F) and hollows (e.g., N termination sites) on a modified (111) surface (Drexler 1987).

become ample to ensure solutions permitting low values of $\Delta\mathcal{V}_{\text{barrier}}$.

10.7.5. Worm gears

Worm gears combine characteristics of of nuts and screws with those of gears to yield large gear ratios in a small package. A simple implementation combines a driven screw tangent to (and meshing with) the rim of a helical gear, creating an interface which has short-term fluctuations in potential energy driven by the sliding of the screw with respect to the gear surface, with superimposed long-term fluctuations driven by the meshing and unmeshing of helical teeth with the thread of the screw.

Where static friction is concerned, the low-frequency components will be of relatively low importance, owing to the high gear ratio. Further, the smoothness properties of the low-frequency component will resemble those of spur (or more accurately, rack and pinion) gear system with comparable numbers of teeth; since these can be good even in the absence of mechanical advantage, their minimization will pose little problem. The high-frequency components of the fluctuations in potential energy can be made to cancel (to a good approximation) given suitable choices of gear geometry and compressive load, but the analysis is complicated by the superimposed rotation of the helical gear.

10.7.6. Belt-and-roller systems

In engineering practice, tension members stretched over rotors are commonly used to transmit either materials or power (or both). Examples include conveyor belts moving over rollers, drive belts moving over pulleys, and chain drives moving over sprockets.

In nanomechanical practice, surfaces commonly have the sort of periodicity associated with chains and sprockets; accordingly, it is natural to design devices in such a way that belts and rollers mesh. The meshing and unmeshing of a belt and roller then locally resembles the meshing and unmeshing of a rack and pinion, and values of $\Delta\mathcal{V}_{\text{barrier}}$ will have parallel behavior. A significant difference is that the larger region of contact (where the belt wraps around the roller) decreases the required contact pressure and areal stiffness for a given interfacial shear load, and this, in turn, will decrease thermoelastic damping and phonon scattering.

The meshing of bumps on rollers and belts can excite resonant transverse vibrational modes in the belts. For example, in a belt with a length of 20 nm and a ratio of tension to linear mass density equal to that of diamond under a tensile stress of 10 GPa, the transverse wave speed is $\approx 1.7 \times 10^3$ m/s, and the frequency of the lowest normal mode is

$\approx 1.3 \times 10^{11}$ rad/s. If bumps on the rollers have a spacing $d_a = 0.25$ nm, then the lowest resonant belt speed will be ≈ 5 m/s. Operation of equipment between (rather than below) resonant modes is common practice in macroscopic engineering, but requires attention to startup dynamics as resonant conditions are traversed; nanomechanical engineering will be no different in this regard.

10.8. Barriers in extended systems

In a typical nanomechanical subsystem, a series of components are mechanically coupled, moving as a nearly-rigid unit with respect to some motion coordinate q . Examples include rotating shafts with multiple bearings, sliding rods moving over multiple surfaces, and sets of shafts and rods linked by gearing. In each case, where the local negative stiffnesses associated with energy barriers in the components are small compared to the positive stiffnesses of the structures linking those components, the energy barriers for the system as a whole are not those resulting from the potentials of the components taken individually, but those resulting from the sum of the component potentials with respect to q . In systems of this sort, the overall barriers are bounded by the sum of the component barriers, but can be much lower where nearly-sinusoidal contributions from different components have the same period and are made to cancel.

More generally, a coupled subsystem of this sort can commonly be extended by adding a *tuning component* that undergoes simple linear or circular motion in an environment with no function save the modification of the overall subsystem potential. Using levers or gears, the ratio of physical displacements in the tuning component to those found elsewhere in the subsystem can be made ≥ 1 , and therefore the characteristic frequencies (with respect to q) of interactions between moving and stationary in the tuning component can be made greater than or equal to those in the subsystem as a whole. (Note that high spatial frequencies resulting from special symmetries are associated with low barriers, and will seldom motivate the introduction of a tuning component.)

10.8.1. Sliding of irregular objects over irregular surfaces

One class of tuning component could be used to smooth the potential associated with the sliding of an irregular object over an irregular surfaces, for example, a rod with irregular features sliding in a sleeve with irregular features. There is no obvious procedure for choosing object irregularities so as to result in a nearly-constant potential at all displacements with respect to a given irregular surface; in general, there will be some fluctuating

potential $\mathcal{U}(q)$. Given that $\mathcal{U}(q)$ does not result in excessive negative stiffnesses, however, a tuning component can be introduced (Fig. 10.29) in which one or more atoms on the moving part interacts with an irregular surface shaped so as to provide a compensating potential $\mathcal{V}(q) \approx -\mathcal{U}(q)$.

This design task will be feasible if each strip of the fixed surface of the tuning component interacts with only one moving atom. Each such moving atom can interact strongly with several stationary atoms at each point, and $\mathcal{V}(q)$ can be generated as a sum of the interactions of an indefinitely large number of moving atoms. Multiple strips of interacting atoms in a system of this sort suffice to provide (1) an indefinitely large energy range for $\mathcal{V}(q)$, (2) multiple, independent contributions to $\mathcal{V}(q)$ in each atomic-scale range of q , permitting fine control of its magnitude, and (3) control of $\mathcal{V}(q)$ with a spatial resolution comparable to that of the features in $\mathcal{U}(q)$. This provides sufficient freedom to design systems with $\mathcal{V}(q) \approx -\mathcal{U}(q)$.

10.9. Dampers, detents, and clutches

The focus in the earlier sections of this chapter has been on components that can move with smooth or flat potential energy functions (implying low static friction), and with low energy dissipation. Engineering practice on a macroscopic scale demonstrates the utility of components with properties quite different from these, which are briefly discussed here.

10.9.1. Dampers

Dissipation of energy (e.g., vibrational energy) is often desirable in a dynamical system, and can be accomplished in a variety of ways in nanoscale systems. Stiff, interlocking sliding interfaces with large values of R and $\Delta k_s/k_a$ will exhibit relatively large

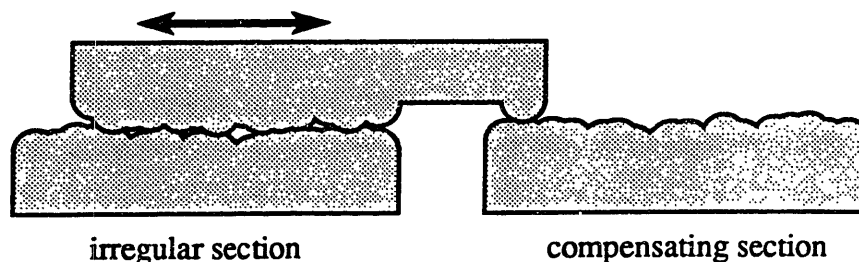


Figure 10.29. Linear representation of the sliding of two irregular surfaces over one another (see text).

viscous damping (although the models developed here should not be used to estimate this drag, as they are expected to be conservative when drag is to be minimized, not maximized). Alternatively, interfaces with short, flexible protrusions (e.g., $-\text{C}\equiv\text{CH}$) can be designed such that the protrusions hop from one potential well to another as they slide over a facing surface; this can provide substantial damping together with a threshold shear strength for the interface. A variety of other linear and non-linear damping devices can be developed based on the energy dissipation mechanisms described in Chapter 7.

10.9.2. Detents

Violations of the design principles for bearings lead to devices with relatively deep potential wells along some motion coordinate. These devices can serve as detents, snap fasteners, and the like. A class of devices of some interest would resemble a sleeve bearing with a moderately-large value of $\text{lcm}(n,m)$, but with flexible protrusions bearing the interfacial load. For a proper choice of (positive) bending stiffness and (negative) interaction stiffness between a protrusion and the opposite surface, the protrusions will approach flexural instability as they move from one potential well to the next, and the single-protrusion potential at that point can have arbitrarily-high spatial frequency components, relative to the overall rotational coordinate of the device. The result is a system with $\text{lcm}(n,m)$ equally-spaced, well-defined potential wells, where $\text{lcm}(n,m) \gg n$ or m .

10.9.3. Clutches

The depth of the potential wells encountered as one surface moves past another will depend on their relative positions and the magnitude of the interfacial load. Where two complementary, gear-like surfaces can be pressed together or separated under the control of one subsystem, while one of them moves in an orthogonal degree of freedom under the control of another subsystem, the result is a clutch. As in standard engineering practice, clutches can be used to couple and decouple mechanisms and power sources.

10.10. Conclusions

Both sliding and meshing interfaces can be constructed in a variety of nanoscale geometries, as can interfaces with intermediate properties and specially-tailored potential energy functions along a sliding coordinate. As a consequence, it will be feasible to construct nanomechanical rotary bearings, sliding shafts, screws and nuts, power screws, snaps, brakes, dampers, worm gears, constant-force springs, roller bearings, levers, cams,

toggles, cranks, clamps, hinges, harmonic drives, bevel gears, spur gears, planetary gears, detents, ratchets, escapements, indexing mechanisms, chains and sprockets, differential transmissions, Clemens couplings, flywheels, clutches, drive shafts, robotic positioning mechanisms, and suitably-adapted working models of the Jacquard loom, Babbage's Difference and Analytical Engines, and so forth.

It has not been established that all classes of sliding-interface bearing are feasible in the nanometer size range. For example, a smoothly-sliding ball-and-socket joint would require a potential energy function that is smooth with respect to three rotation degrees of freedom; neither the symmetry properties exploited in Sections 10.3 and 10.4, nor the tuning approaches discussed in Section 10.8 guarantee that this can be accomplished in a stiff, nanoscale device.

Together with the conclusions of Chapter 9 regarding the feasibility of building strong, rigid components of (almost) arbitrary shape on a nanometer scale, the conclusions of Sections 10.3 and 10.8 support what can be termed the *modified continuum model* for the design of nanomechanical systems. Working within this model, one assumes that components can be of any desired shape, so long as strong symmetry and precision requirements are not imposed, and so long as the minimum feature size is ≥ 1 nm. Further, one assumes that the mechanical properties of the components can approximate those of diamond, save for a surface correction to the effective component size (to allow for the difference between nonbonded and covalent radii), and degradation of stiffness by a factor no worse than 0.5, to account for less dense, regular, and stiff arrays of bonds. Finally, so long as one of the conditions for smoothly-sliding interfaces (as established in this chapter) holds, smooth sliding may be assumed.

Within the modified continuum model, as in the standard continuum model used in mechanical engineering, design work can proceed without reference to the positions of individual atoms. In nanomechanical engineering, of course, production of devices will require that this final step be taken, but—if one is willing to accept a performance penalty resulting from conservative assumptions regarding size, stiffness, and so forth—the writing of atomically-detailed specifications can in many instances be postponed for now.

Chapter 11

Nanomechanical computational systems

11.1. Overview

This chapter examines a representative set of components and subsystems for nanomechanical computers, chiefly within the bounds of the modified continuum model. The range of useful components and subsystems is, however, larger than that considered here. The following analysis describes systems capable of digital signal transmission, fan-out, and switching, together with registers for storing state and devices for interfacing with existing electronics; this will suffice to demonstrate the feasible scale, speed, and efficiency of nanomechanical technologies for computation. This discussion of logic rods in this chapter parallels that in (Drexler 1988), but applies a wider range of analytical tools to a different set of physical structures.

Within the modified continuum model, the design of nanomechanical systems largely parallels that of macromechanical systems. In neither case are structures specified in atomic detail, and in both, structural properties are described in terms of parameters such as strength, density, and modulus. The special characteristics addressed in the modified continuum model include surface corrections to strength, density, and modulus; constraints on feature size and shape; molecular PES based models for static friction; and phonon-interaction based models for dynamic friction. In addition, nanomechanical designs are commonly constrained by statistical mechanics and the resulting trade-offs involving structural stiffnesses, positional tolerances, and error rates.

11.2. Digital signal transmission with mechanical rods

11.2.1. *Electronic analogies*

In conventional microelectronics, digital signals are represented by voltages of conducting paths; for example, a high voltage within a particular range can be taken as a 1, and a low voltage within a particular range can be taken as a 0. Propagation of voltages through circuits requires the flow of current, with associated delays and energy losses.

Analogously, digital signals in nanomechanical computers can be represented by displacements of solid rods; for example, a large displacement within a particular range can be taken as a 1, and a small displacement within a particular range can be taken as 0. Propagation of displacements along rods requires motion, with associated delays and energy losses.

The parallels between existing microelectronics and proposed nanomechanical systems are not exact. The time constants in microelectronics are dominated by resistance, not inductance; the time constants in nanomechanical systems will be dominated by inertia (analogous to inductance), not by drag (analogous to resistance). Elastic deformation of rods plays a role resembling (but differing from) that of parasitic capacitance in conductors.

11.2.2. *Signal propagation speed*

The speed of signal propagation in rods is limited to the speed of sound, for diamond ~ 17 km/s ($\sim 6 \times 10^{-5}c$). To minimize energy dissipation resulting from the excitation of longitudinal vibrational modes in a rod, it suffices to make the characteristic motion times long compared to the acoustic transit time. This lowers the effective signal propagation speed, but delays can still be in the range familiar in microelectronic practice: at an effective propagation speed of only 1.7 km/s, for example, the delay over a 100 nm distance is ~ 60 ps, and over a 1 μ distance is ~ 0.6 ns. These distances are substantial relative to the size of typical nanomechanical logic systems. For propagation over longer distances, it will commonly be desirable to invest energy in an acoustic pulse, at the cost of either dissipating this energy or requiring accurate frequency control in the drive system to permit its recovery. (Further discussion of acoustic signal propagation is deferred to Section 11.5.4, after gates, drive systems, etc., have been introduced.)

11.3. Gates and logic rods

11.3.1. *Electronic analogies*

In conventional microelectronics, digital logic systems are built using transistors. In the case of CMOS logic, transistors can make the current-carrying ability of a path dependent on the voltage applied to a gate, either permitting current to flow at high voltage and blocking it at low, or blocking at high voltage and permitting at low, depending on the structure of the transistor.

Analogously, digital logic systems in nanomechanical computers can be built using *interlocks*. These can resemble CMOS logic, in that interlocks can make the mobility of a rod dependent on the displacement applied to a gate structure, either permitting motion when the gate is at a large displacement and blocking it at low displacements, or *vice versa*, depending on the structure of the interlock.

Again, the parallels are not exact. In particular, MOS gates have a large capacitance relative to a comparable length of simple conducting path, resulting in substantial propagation delays; interlock gates, in contrast, represent a more modest perturbation in the structure of a logic rod. Accordingly, fan-out has less effect on speed in the mechanical technology.

11.3.2. *Components and general kinematics*

Figure 11.1 schematically illustrates the components of a small logic rod system, some included only for compatibility with descriptions of rods with greater fan-out. In the following, *logic rod* will refer to a particular rod under consideration, and the otherwise-similar rods that interact with it will be termed *input* and *output* rods.

11.3.2.1. *Drivers and drive springs*

Rod logic systems of the sort described here are clocked, with a distinct clock signal for each level of gates in a combinational logic system (this approach is useful in minimizing energy dissipation). A rod is accordingly attached to a *driver*, a source of periodic, non-sinusoidal displacements; the implementation considered here achieves this motion using a follower sliding on a sinusoidally-oscillating cam surface. (The cam surface, in turn, would be part of a thick *drive rod*, part of a power distribution and clocking system ultimately driven by crankshafts coupled to a motor/flywheel system.) Typically, multiple rods will be attached to a single driver mechanism.

Driver displacements are coupled to a single rod via a *drive spring*. This can be implemented as a constant force spring that retains a fixed length until the force exceeds a threshold. If the rod is not blocked, Figure 11.1(b), displacements of the driver are transmitted through the drive spring to result in comparable displacements of the rod. If the rod is blocked, Figure 11.1(c), displacements of the driver chiefly result in stretching of the drive spring. The driver and drive spring thus form a *drive system* that periodically

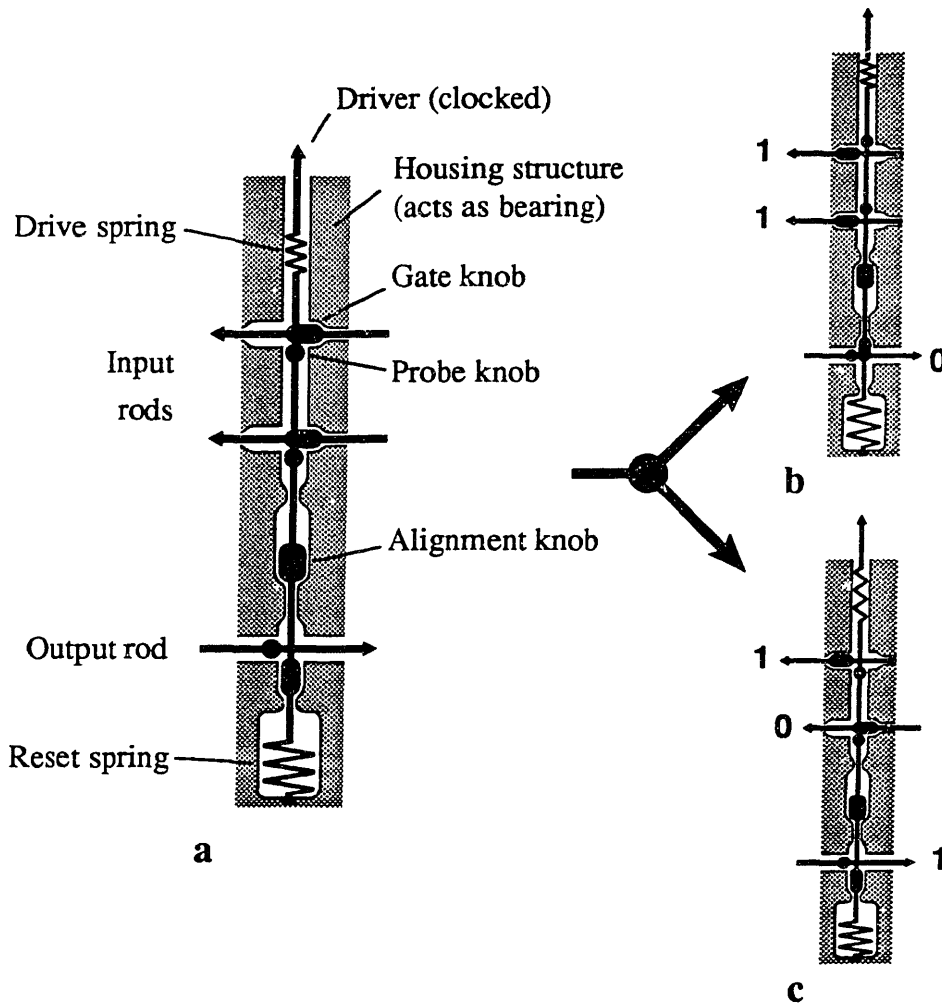


Figure 11.1. Schematic diagram of a gate knob, probe knob, and an interlock. Diagram (a) shows two rods, each bearing a knob. Diagram (b) shows the two rods and knobs in the correct geometry for the gate knob to block the motion of the probe knob toward the lower right; displacement of the gate knob rod along its axis would unblock the probe knob. Diagram (c) shows the rods surrounded by a housing structure that permits no large-amplitude motions of rods except longitudinal displacements.

tensions and de-tensions the rod, without forcing motion.

11.3.2.2. The housing structure

A stiff housing structure surrounds the moving parts of a rod logic system, constraining their motions (within small excursions) to simple linear displacements. The surface of this housing structure serves as a bearing for the sliding motions of rods, and the rod-housing system can be analyzed along the lines developed in Chapter 10.

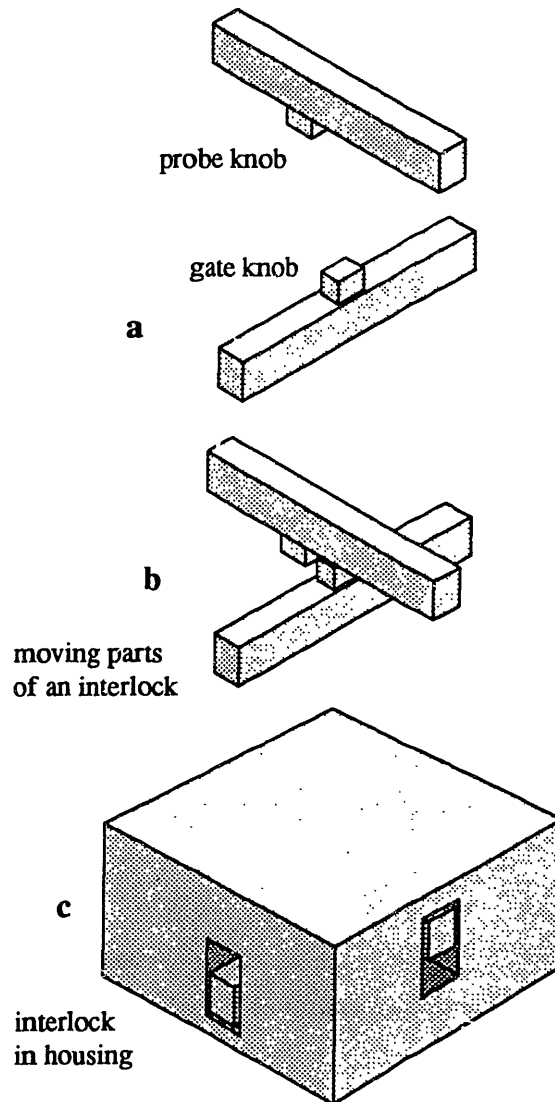


Figure 11.2. Components of an interlock: input rod with gate knob and output rod with probe knob, separated (a), in their working positions (b), and constrained by a housing structure (c).

11.3.2.3. Gate knobs, probe knobs, and interlocks

Each logic rod bears a series of protrusions, termed *knobs*. A *gate knob* and a *probe knob* in a suitable housing form an interlock, as shown in Figure 11.2. A gate knob on a rod in its 0 position can be positioned either to block or to fail to block its matching probe knob. Displacement of the gate-knob rod can thus either unblock or block a probe knob, depending on the position of the gate knob with respect to the rod. (A crossover will permit motion in either state if the gate knob is omitted.)

11.3.2.4. Input rods

The mobility of a logic rod can be determined by the state of an indefinitely large number of *input rods*, each bearing a gate knob and interacting with a probe knob on the logic rod. When none of the input rods blocks its corresponding probe knob, the logic rod becomes free to move when the drive system next applies tension. Figure 11.1 illustrates a NAND gate and hence has only two input rods, both of which must be displaced to unblock the vertical logic rod.

11.3.2.5. Alignment knobs

To define the displacements of the two distinct logic states requires an alignment mechanism. The *alignment knob* of a logic rod slides within a certain range, bounded by *alignment stops*, so that a net force in one direction results in one positional state, and a net force in the other direction results in the other positional state. If rods were rigid, the alignment knob could be located at any point; given finite rod compliance, a location immediately adjacent to the gate knob (output) segment of the rod reduces displacements resulting from thermal excitation, relative to more remote locations. Placing it between the probe knob and gate knob segments enables the gate knob segment to be isolated from fluctuations in tension resulting from the drive system, and thus ensures greater dimensional stability and better gate knob alignment relative to the output rods.

(A small advantage could be gained at the cost of a less regular structure by placing the alignment knob somewhat inside the gate knob segment. A greater advantage can be gained by placing a second alignment knob at the far end of the gate knob segment.)

11.3.2.6. Output rods

The displacement of a logic rod affects further steps in a computation by blocking or unblocking an indefinitely large number of *output rods*, each bearing a probe knob and

interacting with a gate knob on the logic rod. An interlock in the output segment of one rod is an interlock in the input segment of another. The state of the interlocks of an output segment switch when the input segment is unblocked, the drive system applies tension, and the seating knob shifts from its 0 to its 1 position. In the NAND gate of Figure 11.1, the single output switches from unblocked to blocked if the rod is mobile.

11.3.2.7. Reset springs

When the drive system de-tensions a logic rod, a restoring force must be provided to return the rod to its resting state. This can be provided by a constant-force spring (drawn in Figure 11.1 as a large, low-stiffness spring). With this choice, the tension (and hence the strain) in the gate knob segment remains fixed throughout the cycle, and so the alignment of the gate knobs and probe knobs remains uncompromised. Note that the tension in the probe-knob segment varies, but that the resulting fluctuations in length do not affect the reliability of the logic operations.

11.3.3. A modified continuum model

To explore how system parameters such as size, speed, error rate, and energy dissipation vary with device geometry and other parameters, a modified continuum model can be applied. For components of sufficient size, and for a suitable choice of material and interface parameters, it will provide a realistic description. (For smaller components, it will give a preliminary indication of the performance to be expected provided that a structure can be found having the appropriate geometry and properties.)

11.3.3.1. Geometric assumptions and parameters

For purposes of the present analysis, it will be assumed that probe knob segments and gate knob segments cross at right angles, and that rods and knobs can be approximated as rectangular solids (Fig. 11.3). Probe knobs will ordinarily have some regular spacing d_{knob} . The spacing of gate knobs will vary owing to differences in their logical function. Each interlock has two possible locations for its gate knob, and the spacing between locations of the same kind is here assumed to be the same as the spacing of probe knobs, d_{knob} . Accordingly, each interlock in a grid of intersecting probe segments and gate segments (as in a programmable logic array, Sec. 11.5) will occupy a square region with sides of length d_{knob} . Gate knobs and probe knobs will be assumed to have the same dimensions.

For some purposes, the positions of rod surfaces can be defined as hypothetical sliding-contact surfaces (with no gap and no overlap between objects). This definition is compatible with some choice of steric radius, which can (for uniformity) also be applied where surfaces do not make a sliding contact. For second-row atoms, typical radii of this kind will be ~ 0.14 to 0.17 nm. Rod dimensions can then be given, with widths w_{rod} and w_{knob} , and heights h_{rod} and h_{knob} defined as in Figure 11.3. The total height of the moving parts in an interlock is then $2h_{\text{rod}} + h_{\text{knob}}$.

The function of a rod logic system demands that the knob length ℓ_{rod} meet the condition

$$\ell_{\text{knob}} \leq d_{\text{knob}} - w_{\text{knob}} \quad (11.1)$$

a simple and attractive set of choices that meets this conditions is

$$w_{\text{knob}} = w_{\text{rod}} = \ell_{\text{knob}} = \frac{d_{\text{knob}}}{2} \quad (11.2)$$

With these choices, the thickness of the housing structure also equals w_{rod} , and the minimum feature dimensions are uniform throughout. For sample calculations in the following sections, it will be assumed that $w_{\text{rod}} = 1$ nm, and that $h_{\text{knob}} = 0.5$ nm.

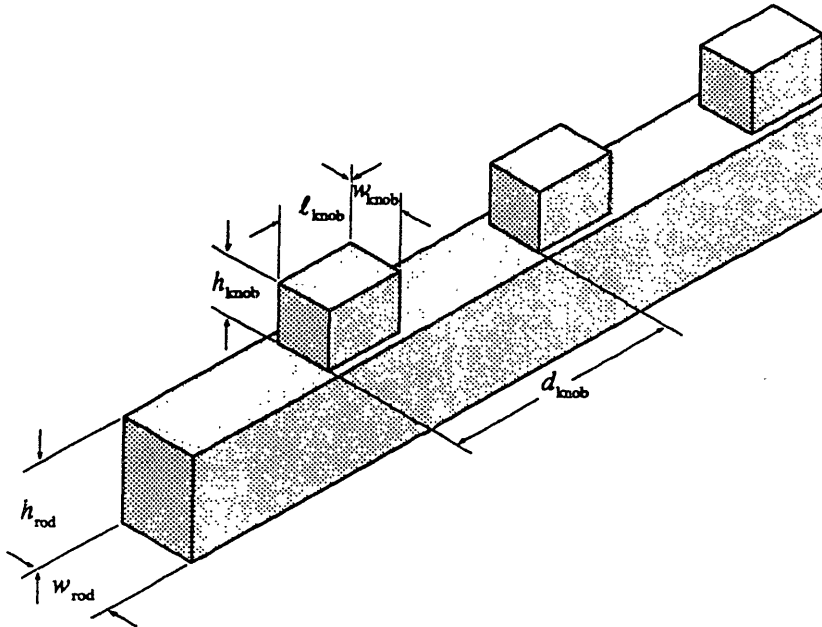


Figure 11.3. Definition of rod dimensions used in the analysis in the text.

The overall dimensions of a rod can be characterized by the number of input rods n_{in} and output rods n_{out} . The length of the input segment is then $\ell_{\text{in}} = d_{\text{knob}} n_{\text{in}}$; the length of the output segment is $\ell_{\text{out}} = d_{\text{knob}} n_{\text{out}}$; and (with a correction for the length of the alignment knob mechanism) the total length of the rod is

$$\ell_{\text{rod}} = d_{\text{knob}} (n_{\text{in}} + n_{\text{out}} + 1) \quad (11.3)$$

The above length will be use for estimating a variety of dynamical quantities; for estimating system dimensions, allowances must also be made for the drive system and reset spring, and their associated structures. The sample calculations will assume $n_{\text{in}} = n_{\text{out}} = 16$, implying $\ell_{\text{rod}} = 66 \text{ nm}$.

11.3.3.2. Interactions and applied forces

The dynamics and error rates of a logic rod can be described in terms of the potential energy function of the rod with respect to its environment (the housing and crossing rods) and the forces applied to it by the drive and reset springs. Regarding the former, from the description in Section 11.3.2, it can be seen that the environment of a specific *mobile* logic rod is the same in every cycle, with the possible exception of gate-knob free crossovers, which will permit mobility in either of two states. If the interaction energy of the crossing rod with the logic rod is made nearly equal in both states (e.g., with differences in van der Waals attraction compensated by differences in steric and electrostatic repulsion), then the potential energy function of the logic rod in its mobile state will be invariant, and hence can be tuned to near constancy within the permitted range of motion.

The force with which the alignment knob is pressed against its limit stops determines the magnitude of one contribution to the positional uncertainty and hence to the error rate. In the present design context, symmetry considerations suggest that the force applied to one stop in the tensioned state should equal the force applied to the other in the detensioned state, both of a magnitude F_{align} . Accordingly, the (constant) force applied by the reset spring equals F_{align} , and the peak force applied by the drive spring equals $2F_{\text{align}}$. The sample calculations will assume $F_{\text{align}} = 1 \text{ nN}$.

11.3.3.3. Stiffness and mass

In the modified continuum model, the stiffness of knobs and rods is estimated by combining moduli of elasticity with component dimensions modified by a surface correction. Assuming that rods are of diamondoid structure, with surface termination chiefly

using di- and trivalent atoms, rather than monovalent atoms, it is reasonable to compute stiffnesses and strengths on the basis of effective dimensions that discard a surface layer $\delta_{\text{surf}} = 0.1 \text{ nm}$ thick; this allows for the difference between steric and covalent radii, along with a further margin for the effects of surface relaxation. The effective cross sectional area of the rod is then

$$S_{\text{eff}} = (w_{\text{rod}} - 2\delta_{\text{surf}})(h_{\text{rod}} - 2\delta_{\text{surf}}) \quad (11.4)$$

or 0.64 nm^2 in the example case.

A conservative value of Young's modulus to use in computations is $E = 5 \times 10^{11} \text{ N/m}^2$; this is about half the value for diamond, and only moderately greater than the value for silicon carbide, silicon nitride, or alumina. In strong covalent solids, it is common for the shear modulus G to be $\sim 0.5E$, as it is in diamond.

For rods of a thickness range in which the modified continuum model is applicable, sliding in a housing suitable for logic rods, the bending stiffness and transverse constraint forces are large enough that the variance in length resulting from transverse vibrational modes of the rod (Sec. 5.6) can be neglected. The stretching stiffness of a segment of rod of length ℓ is then simply

$$k_s = \frac{S_{\text{eff}} E}{\ell} \quad (11.5)$$

(neglecting the stiffening effect of the knobs). For $\ell = \ell_{\text{rod}}$, $k_{s,\text{rod}} = 4.85 \text{ N/m}$.

In the modified continuum model, masses are estimated by combining a density with modified component dimensions. In general, a different value of δ_{surf} will be appropriate for mass calculations, but the use of $\delta_{\text{surf}} = 0.1 \text{ nm}$ will suffice for present purposes. With these assumptions,

$$m_{\text{rod}} \approx \rho \ell_{\text{rod}} (w_{\text{rod}} - 2\delta_{\text{surf}}) \left[(h_{\text{rod}} - 2\delta_{\text{surf}}) + h_{\text{knob}} \frac{(\ell_{\text{knob}} - 2\delta_{\text{surf}})}{d_{\text{knob}}} \right] \quad (11.6)$$

A density $\rho = 3500 \text{ kg/m}^3$ (comparable to that of diamond, somewhat higher than that of silicon carbide or silicon nitride) will yield conservative estimates of device performance. With the sample calculation parameters, $m_{\text{rod}} = 1.9 \times 10^{-22} \text{ kg}$.

11.3.4. Dynamics and energy dissipation in mobile rods

In the design regime of interest, drag forces are sufficiently small that they can be neglected in a first-order calculation of dynamics, then introduced later in estimating

energy dissipation. It will further be assumed that displacements resulting from elastic deformation of a mobile rod are small compared to those that would result from rigid-body displacement of the rod, and hence that accelerations and kinetic energies are well-approximated by a rigid-body analysis. Again, corrections are introduced later, both in estimating energy dissipation and in computing the requirements for the drive system. This section discusses mobile rods; Section 11.3.5 considers blocked rods.

11.3.4.1. Dynamics in the rigid-body approximation

The use of a cam surface in the drive mechanism permits flexible control of displacement $\Delta x(t)$ in the driver, and hence of the forces applied by the drive spring. In particular, these can be chosen such that a mobile rod (in the rigid-body approximation) executes a smooth motion that can be approximated by

$$\Delta x(t) = \begin{cases} 0, & t < 0 \\ \frac{d_{\text{kno}}}{4} \left[1 + \cos\left(\frac{\pi}{t_{\text{switch}}} t\right) \right], & 0 \leq t \leq t_{\text{switch}} \\ \frac{d_{\text{kno}}}{2}, & t > t_{\text{switch}} \end{cases} \quad (11.7)$$

which has the amplitude required to achieve switching (in a time t_{switch}), given the geometric parameters chosen in Section 11.3.3.1. (A more accurate model would include a finite rate of onset of acceleration.)

The rod mass

$$m_{\text{rod}} = \rho \ell_{\text{rod}} (w_{\text{rod}} - 2\delta_{\text{surf}}) \left(h_{\text{rod}} + \frac{h_{\text{kno}}}{2} - 2\delta_{\text{surf}} \right) \quad (11.8)$$

can be combined with the peak rigid-body acceleration to yield an estimate of the peak drive force for rod acceleration

$$F_{\text{accel}} \approx m_{\text{rod}} \left(\frac{\pi}{t_{\text{switch}}} \right)^2 \frac{d_{\text{kno}}}{4} \quad (11.9)$$

The peak speed in this model is

$$v_{\text{max}} \approx \frac{\pi d_{\text{kno}}}{4 t_{\text{switch}}} \quad (11.10)$$

Using the parameters of the example case, and adding the assumption that $t_{\text{switch}} = 0.1 \text{ ns}$, $m_{\text{rod}} = 1.94 \times 10^{-22} \text{ kg}$, $F_{\text{accel}} \approx 0.096 \text{ nN}$, and $v_{\text{max}} = 15.7 \text{ m/s}$. (The balance of this analysis assumes $F_{\text{accel}} \leq F_{\text{align}}$.)

11.3.4.2. An estimate of vibrational excitation

A detailed examination of the vibrational dynamics of a logic rod (e.g., taking account of the effects of alignment knob contacts and drive system force profiles on each vibrational mode) is beyond the scope of the present work. The chief interest in vibrational excitation in the present context is its role as a mechanism of energy dissipation. The energy of excitation can be estimated from a harmonic-oscillator model of the lowest vibrational mode of the rod (coupling of energy from the drive system to higher-frequency modes will generally be far lower).

A harmonic oscillator with mass m and stiffness k_s and natural frequency ω will if, suddenly subject to a constant longitudinal acceleration a , acquire a vibrational energy

$$\Delta \mathcal{E} = \frac{(ma)^2}{2k_s} = \frac{k_s a^2}{2\omega^4} \quad (11.11)$$

Substituting the peak acceleration derived from Eq. (11.7) yields

$$\Delta \mathcal{E} = \frac{k_s d_{\text{knob}}^2}{32} \left(\frac{\pi}{\omega t_{\text{switch}}} \right)^4 \quad (11.12)$$

For $\lambda \gg d_{\text{knob}}$, the speed of a longitudinal wave along a rod will be

$$v_s \approx \sqrt{\frac{E}{\rho(1 + h_{\text{knob}}/2h_{\text{rod}})}} \quad (11.13)$$

which includes an approximate correction for knob mass. (For the example system, $v_s \approx 11 \text{ km/s}$) The angular frequency of the fundamental mode of the rod (in which the far end is essentially free) is

$$\omega_0 \approx \frac{\pi v_s}{2\ell_{\text{rod}}} \quad (11.14)$$

To estimate the vibrational energy introduced into a rod, ω_0 can be substituted for ω in Eq. (11.11), and using a value of $\ell_{\text{rod}}/2$ for ℓ in Eq. (11.5), the resulting value of k_s can

(conservatively) be substituted for k_s in (11.12). Vibrational excitations induced by deceleration can add in phase with, or cancel, those induced by acceleration. Averaging over these cases introduces a factor of two in the total vibrational energy per displacement, and yields the estimate

$$\Delta E_{\text{vib}} \approx \frac{2S_{\text{eff}} d_{\text{knob}}^2 \rho^2 \ell_{\text{rod}}^3}{E t_{\text{switch}}^4} \left(1 + h_{\text{knob}}/2h_{\text{rod}}\right)^2 \quad (11.15)$$

As can be seen from this expression, energy dissipation via the excitation of irrecoverable vibrational energy in logic rods is strongly sensitive to design choices. In particular, choosing sufficiently small values of ℓ_{rod} or sufficiently large value of t_{switch} can reduce ΔE_{vib} to negligible values. For the example case, $\Delta E_{\text{vib}} \approx 5.6 \times 10^{-22}$ J.

A more detailed analysis would determine a force profile applied by the drive spring that results in the arrival of the alignment knob at the alignment stop with approximately zero mean relative velocity, followed by a tensioning phase in which the input is stretched by a load of $2F_{\text{align}}$. The smaller displacement (with the exemplar parameters, $\sim d_{\text{knob}}/8$) and higher characteristic frequency ($\sim 4\omega_0$) allow the latter phase to be fast, while inducing little vibrational excitation.

11.3.4.3. An estimate of sliding-interface drag

The sample calculations of Chapter 10 indicated that band stiffness scattering is the dominant energy dissipation mechanism in systems where $\Delta k_s/k_a = 0.4$ and $R = 10$. Since the sliding interface between a logic rod and a housing is analogous to the sliding interface in other bearing systems, Eq. (7.35), applied with these parameters, should yield a conservative estimate of the drag in the logic rod system.

The contact area can be taken as

$$S = \ell_{\text{rod}} (w_{\text{rod}} + 2h_{\text{rod}}) \quad (11.16)$$

and folding in the above values of $\Delta k_s/k_a$ and R yields the estimate

$$\Delta E_{\text{drag}} \approx 3.3 \times 10^{-32} (w_{\text{rod}} + 2h_{\text{rod}}) \frac{d_{\text{knob}}^2 \ell_{\text{rod}} k_a^{1.7}}{t_{\text{switch}}} \quad (11.17)$$

(note that the mean square speed of the rod introduces a factor of 1/2). Assuming an interfacial stiffness $k_a = 10 \text{ N/m}\cdot\text{nm}^2$ and the standard sample parameters assumed above, this yields an energy dissipation of 5.2×10^{-23} J per displacement. Since this energy loss is

small compared to the kinetic energy of motion ($\sim 2.3 \times 10^{-20}$ J), damping can be neglected in computing velocities, accelerations, and so forth.

11.3.4.4. Sliding-interface drag in the cam surface

If the maximum slope of the cam surface is $1/2$, then the sliding speed of the cam surface is

$$v_{\text{cam}} \approx 2v_{\text{max}} \left(1 + \frac{4F_{\text{align}} \ell_{\text{in}}}{d_{\text{knob}} S_{\text{eff}} E} \right) \quad (11.18)$$

where the expression in parentheses provides an upper-bound correction for the effect elastic deformation of the input segment. With the standard sample parameters, $v_{\text{cam}} \approx 38$ m/s. In the contemplated system context, multiple logic rods will share a drive-system cam. Reasonable values for the sliding interface in the cam are $\sim 4 \text{ nm}^2$ per logic rod, with an interfacial stiffness $k_a = 10 \text{ N/m} \cdot \text{nm}^2$. The energy dissipation per rod per 0.1 ns period is then $\sim 1.2 \times 10^{-23}$ J. Since the cam surface continues to move when the rod is stationary, this energy must be multiplied by the reciprocal of the fraction of time spent in rod motion, ~ 5 .

11.3.4.5. An estimate of thermoelastic losses

In the transition to the tensioned state, the tension in the probe segment increases by $2F_{\text{align}}$, with resulting thermoelastic effects. A bound on the energy dissipation in this process can be derived from Eq. (7.48)

$$\Delta \mathcal{E}_{\text{therm}} < 8.2 \times 10^{-5} \beta^2 \frac{(2F_{\text{align}})^2}{S_{\text{eff}}} \ell_{\text{in}} \quad (11.19)$$

in which a volumetric heat capacity $C_{\text{vol}} = 1.7 \times 10^6 \text{ J/K} \cdot \text{m}^3$ and temperature $T = 300 \text{ K}$ have been assumed. The value of the volumetric thermal coefficient of expansion β varies widely among materials, and is occasionally negative. The values found in Table 7.1 suggest that $\beta = 5 \times 10^{-6}$ will prove conservative for many diamondoid structures chosen with some attention to this parameter. With this assumption, the example parameters yield $\Delta \mathcal{E}_{\text{therm}} < 4.1 \times 10^{-22} \text{ J}$.

11.3.4.6. Other energy dissipation mechanisms in mobile logic rods

Treating the knobs as phonon-scattering centers of large mass leads to estimated scattering cross sections $\sim 10^{-23} \text{ m}^2$, and to comparatively negligible power dissipation.

Thermoelastic effects associated with the motion of knobs with respect to the housing are likewise small. Compression of the alignment knob against its limit stop will result in minimal nonisothermal losses, because the knob is in good thermal contact with the rest of the rod, which forms a large thermal inertia relative to the heat of compression.

11.3.4.7. The resetting process

A logic rod will be reset after its output rods have been reset and before its input rods are reset. Thus, the resetting process occurs in the same environment as the switching process. Further, the motion of the driver will approximate the time-reversal of the switching motion, since it is driven by the same follower tracking the same cam surface, but on its return stroke. Accordingly, the *overall* dynamics of resetting will approximate the time-reversal of the switching motion, in those systems for which vibrational excitation of the rod is small. Energy losses for a switching cycle, including resetting, will therefore be approximately twice those that result from the switching motion itself.

11.3.5. Dynamics and energy dissipation in blocked rods

If one or more probe knobs on a logic rod are blocked by an input gate knob, no large displacement of the rod will occur, and sliding interface losses will be minimal. As the driver displacement increases, the tension climbs to $\approx 2F_{\text{align}}$, the drive spring begins to extend, and the tension becomes constant. The lowest longitudinal modal frequencies of a blocked input segment will be $\geq \sim 4\omega_0$, reducing vibrational energy losses. The tensioned length will be $\leq \ell_{\text{in}}$, hence the thermoelastic losses will be comparable to or less than those for a mobile rod.

11.3.6. Fluctuations in stored energy

The flow of energy between the drive system and rod is large compared to the energy dissipated. Logic-state-dependent fluctuations in the energy stored in the rod (at a given point in the clock cycle) determine certain requirements for the drive system (Section 11.6).

The greatest difference in logic state, from a mechanical point of view, is the difference between a mobile rod and a rod that is blocked near the drive spring. In the former case, the energy stored by tensioning is dominated by the energy in the reset spring ($\Delta\mathcal{E}_{\text{reset}}$) and the strain energy of the input segment ($\Delta\mathcal{E}_{\text{input}}$). In the latter case, it is dominated by the energy stored in the drive spring ($\Delta\mathcal{E}_{\text{drive}}$). The energy difference is thus

$$\begin{aligned}
\Delta E_{\text{state}} &\approx \Delta E_{\text{drive}} - (\Delta E_{\text{reset}} + \Delta E_{\text{input}}) \\
&\approx 2F_{\text{align}} \left(\frac{d_{\text{knob}}}{2} + 2F_{\text{align}} \frac{\ell_{\text{in}}}{ES_{\text{eff}}} \right) - \left(F_{\text{align}} \frac{d_{\text{knob}}}{2} + \frac{(2F_{\text{align}})^2}{2} \frac{\ell_{\text{in}}}{ES_{\text{eff}}} \right) \\
&\approx F_{\text{align}} \left(\frac{d_{\text{knob}}}{2} + 2F_{\text{align}} \frac{\ell_{\text{in}}}{ES_{\text{eff}}} \right)
\end{aligned} \tag{11.20}$$

With the parameters of the exemplar system, $\Delta E_{\text{state}} \approx 1.2$ aJ.

11.3.7. Thermal excitation and error rates

Thermal excitation displaces gate knobs and can induce errors in switching events (given a low rate of energy dissipation, the magnitude of the non-thermalized vibrational energy will be relatively trivial). These errors can in principle occur in two ways: either a gate knob blocks a probe knob when it should let it pass, or a gate knob lets a probe knob pass when it should block it. Erroneous blockage, however, will be transient relative to the switching time because (given low damping) the maximum delay before the obstructing knob moves from the path is $\sim 2\ell_{\text{out}}/v_s$, or ~ 0.006 ns for the example system. Accordingly, the probe knob will begin to pass the gate knob early in the switching period, regardless of the initial state of displacement.

Erroneous probe knob passage, in contrast, will not undergo fast, spontaneous reversal, and causes a faulty output. The resulting error rate can be estimated from a model of positional uncertainty in gate knob position, in the approximation that the probe knob tests the position of the gate knob at a single moment in time, and passes if the displacement of the gate knob is sufficiently large.

11.3.7.1. A conservative bound on acceptable gate knob displacement

Gate knobs and probe knobs are regarded as blocks with facing surfaces of equal extent, w_{knob} . In the modified continuum model, corners are regarded as occupied by atoms, and hence are rounded. If the contact between gate knob and probe knob occurs solely through a glancing contact of corner atoms, the probe knob will exert a force tending to push the gate knob aside. For a wide range of geometries and stiffness and force parameter, this tendency can be resisted by gate knob restoring forces.

A simple, conservative model neglects these restoring forces, and assumes that any contact in which the line of travel center of the corner atom of the probe knob falls beyond the center of the corner atom of the gate knob as an error state. In terms of the effective steric radii for the atoms r_{eff} , the effective width of the knobs is $w_{\text{eff}} = \ell_{\text{eff}} = w_{\text{knob}} - 2r_{\text{eff}}$. The acceptable gate knob displacement is then $\Delta x_{\text{thresh}} = w_{\text{eff}}$ (assuming that the knobs are aligned at $\Delta x = 0$). For the example case, assuming $r_{\text{eff}} = 0.15$ nm, $\Delta x_{\text{thresh}} = w_{\text{eff}} = 0.7$ nm.

11.3.7.2. Probability density for gate knob displacement

Transverse elastic displacements of the probe knob with respect to the housing will be characterized by a stiffness $k_{s,p}$, and the parallel, longitudinal displacements of the gate knob with respect to the alignment knob will be characterized by a stiffness $k_{s,g}$. Since in the size range of interest (and at 300 K), quantum mechanical effects on positional uncertainty will be negligible (Sec. 5.3), the variance in relative positions resulting from these elastic displacements will be

$$\sigma_{\text{elastic}}^2 = kT \left(\frac{1}{k_{s,p}} + \frac{1}{k_{s,g}} \right) \quad (11.21)$$

and the associated PDF for the elastic contribution to the relative displacement is

$$f_{\Delta x_{\text{elastic}}}(\Delta x_{\text{elastic}}) = \frac{1}{\sqrt{2\pi}\sigma_{\text{elastic}}} \exp \left[-\frac{1}{2} \left(\frac{\Delta x_{\text{elastic}}}{\sigma_{\text{elastic}}} \right)^2 \right] \quad (11.22)$$

In the approximation of hard-surface interactions between the alignment knob and its limit stops, the PDF for the associated contribution to the relative displacement is the exponential

$$f_{\Delta x}(\Delta x) = \frac{F_{\text{align}}}{kT} \exp \left(-\frac{F_{\text{align}}}{kT} \Delta x \right) \quad (11.23)$$

The probability of a relative displacement greater than Δx_{thresh} is then

$$P_{\text{err-disp}} = \int_0^{\infty} f_{\Delta x_{\text{align}}}(\Delta x_{\text{align}}) \int_{\Delta x_{\text{thresh}} - \Delta x_{\text{align}}}^{\infty} f_{\Delta x_{\text{elastic}}}(\Delta x_{\text{elastic}}) d(\Delta x_{\text{elastic}}) d(\Delta x_{\text{align}}) \quad (11.24)$$

Applying the sample parameters, $k_{s,p} = 10$ N/m; assuming (conservatively) $k_{s,g} = 40$ N/m, $\sigma_{\text{elastic}} \approx 0.023$ nm. The condition for the above approximation then holds, and

$P_{\text{err-disp}} \approx 1.3 \times 10^{-67}$. Error rates of this order would be negligible in digital logic systems of conventional reliability.

11.3.8. Summary observations based on the sample calculations

The sample calculations in the preceding sections describe the characteristics and performance of a non-optimized logic rod design with 16 inputs and 16 outputs. One cycle of this rod represents one-half cycle for 32 interlocks, or a total of 16 complete switching events.

11.3.8.1. Volume per interlock

Allowing for a 0.5 nm thickness of material to partition one layer of interlocks from the next, the volume per interlock in a tightly-packed array is 12 nm^3 . Including the length of the alignment knob and adding 6 nm at one end for the drive system and 2 nm at the other for the reset spring (and adjusting the effective width of a logic rod assembly to account for similar additions to the length of the crossing rods) yields an estimated volume $\sim 16 \text{ nm}^3$ per interlock.

11.3.8.2. Energy per switching cycle

For a mobile rod with a switching time of 0.1 ns, the estimated energy loss in a switching cycle (including reset) is $\sim 2 \text{ mJ}$. This amounts to $\sim 0.013 \text{ mJ}$ per interlock per switching cycle, or $\sim 0.031 \text{ kT}$. (This low dissipation is possible because combinatorial rod logic systems “store” their inputs and intermediate results as rod displacements, rendering all of their computations logically reversible; for comparison, see (Bennett 1982; Feynman 1985; Fredkin and Toffoli 1982; Landauer 1961; Toffoli 1981).) Note that in the limit of slow motion all of the identified energy dissipation mechanisms in combinational rod logic systems approach zero.

11.3.8.3. Comparison to transistors

Like transistors, interlocks are switching devices with the essential properties of a “good computer device” (Keyes 1985): they support fan-out and fan-in, make reliable binary decisions, switch rapidly, pack densely, work with interconnections on their own scale, restore signals to a reference level at each step, and exhibit the equivalent of high gain (in the form of a large change in output rod displacement when the gate displacement passes a threshold). In this, they are superior to numerous failed device technologies (some of which are described by (Keyes 1989)), although they suffer from the disadvan-

tage of lacking an extant manufacturing technology.

In speed, interlocks are inferior to present experimental transistors, having switching times of ~ 100 ps rather than ~ 10 ps (but note that in 100 ps, the exemplar logic rod switches 16 outputs based on 16 inputs, including all relevant propagation delays). Transistors continue to improve, while the mechanical logic described here may well be within a modest factor of real physical limits. Accordingly, there is every reason to expect that the fastest devices in the future will exploit electronic rather than mechanical degrees of freedom; electronic devices will, of course, benefit from the precise control feasible with molecular manufacturing.

In density, interlocks are greatly superior to present transistors. The area occupied by an interlock ($\sim 4 \text{ nm}^2$) is orders of magnitude smaller than the area occupied by a modern transistor ($\sim 10^6 \text{ nm}^2$). Further, there is no barrier to stacking interlocks with a vertical spacing of $\sim 3 \text{ nm}$, while planar semiconductor processing, together with packaging constraints, results in vertical transistor spacings $> 10^6 \text{ nm}$. The overall improvement in volumetric packing density is $> 10^{11}$.

In energy dissipation, interlocks are greatly superior to present experimental transistors. The latter have a switching energy of $\sim 100 \text{ aJ}$, $> 10^6$ times the calculated switching energy of the example interlock. Relative to present CMOS devices, the advantage in energy dissipation is $> 10^9$. To a first approximation, this advantage in switching energy translates into a factor of $> 10^9$ increase in the quantity of computation that can be performed in a box of a certain size within given cooling limitations.

11.4. Registers

Logic rods suffice to implement combinational logic systems of arbitrary depth and complexity, but in the absence of other components, the resulting state is destroyed when the rods are reset. To implement computational systems capable of iterative operations requires some means for recording the output of one cycle of combinational logic such that they can serve as inputs for a later cycle. Registers serve this function.

11.4.1. Kinematics of an efficient class of register

Minimizing energy dissipation requires adherence to proper design principles. In particular, designs based on the use of forced transitions between potential wells can easily dissipate large amounts of energy; in naïve designs, the energy dissipated is equal to the height of the barrier partitioning the states, which must be many times kT for reliability.

More efficient designs avoid this problem. Figures 11.4 and 11.5 illustrate the structure and kinematics of an efficient class of register, which abstractly parallels previous proposals for the efficient writing and erasure of bits by manipulation of a particle in a time-varying system of potential wells (e.g., (Bennett 1982; Landauer 1961)).

The present description is motivated by the modified continuum model; accordingly, the rod widths in Figure 11.4 are taken to be ~ 1 nm. The four components are as follows:

The *input rod* is part of the output segment of a logic rod; it sets the state of the register.

The *state rod* is the component that stores the state of the register by virtue of its position.

The *spring rod* provides a clocked bias force, helping to set the state of the register without determining which state is achieved.

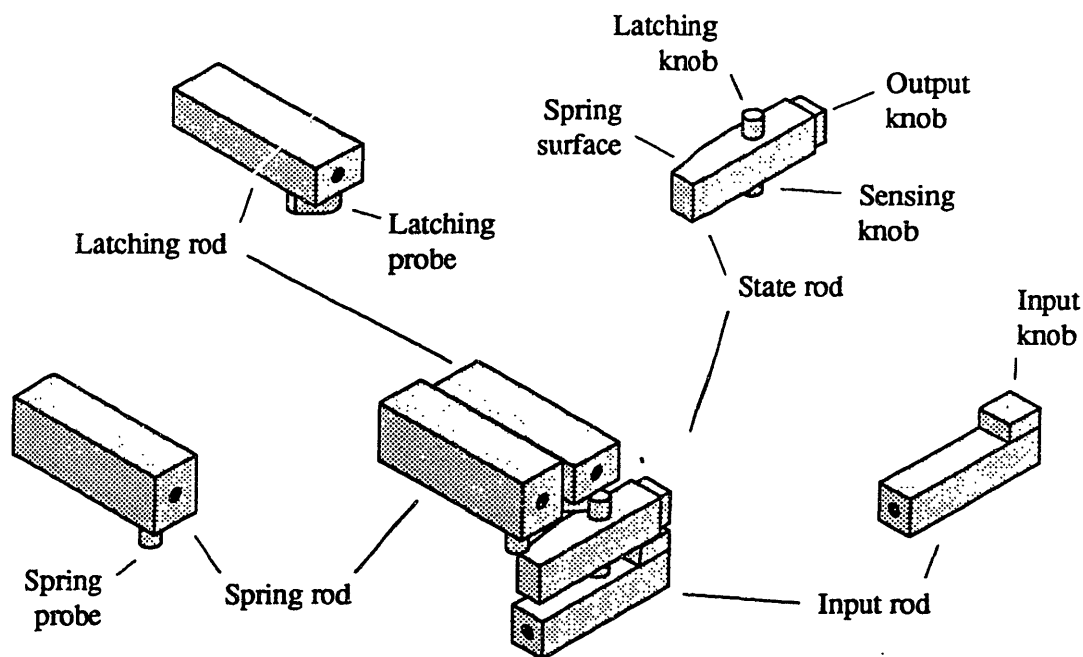


Figure 11.4. Moving parts for a thermodynamically-efficient class of register, with molecular detail suppressed in accord with the modified continuum approximation and housing structure suppressed for clarity. Rod widths are ~ 1 nm; sensing and latching knobs are assumed to be single protruding atoms (to minimize van der Waals interaction energies in the former case, and to minimize knob length in the latter). Rod ends marked with dots represent truncated surfaces of longer structures. See text for description.

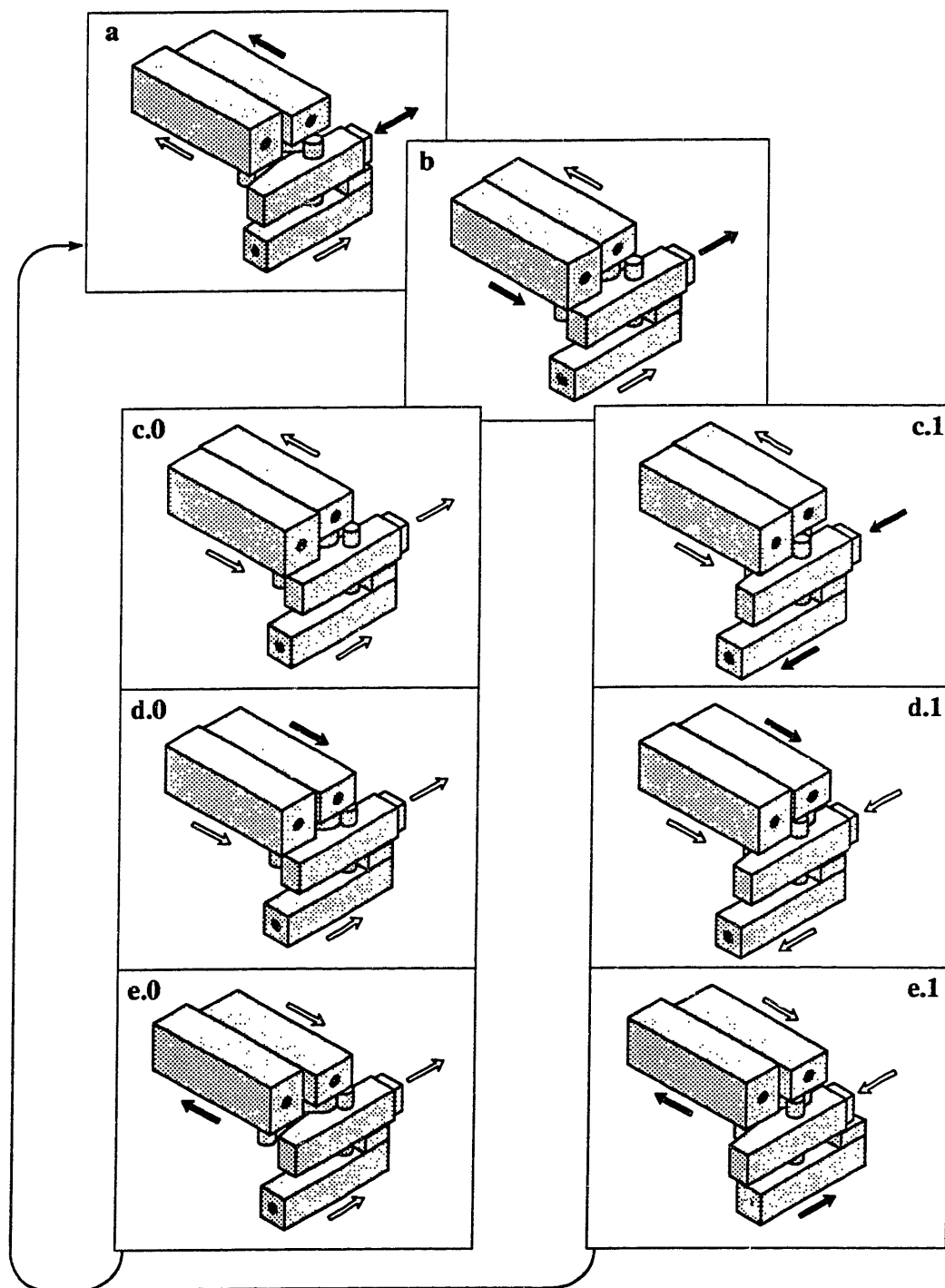


Figure 11.5. Steps in the write-erase cycle of a thermodynamically-efficient register, illustrated as in Fig. 11.4. A housing structure (omitted for clarity) constrains gross rod motions to axial displacements. See text for description.

The *latching rod* provides a clocked latching force, blocking movement of the state rod after the input rod is reset.

The interaction of these components is diagrammed in Figure 11.5 and described in the following sections.

11.4.1.1. Compression

In frame a, the state rod is constrained neither by the spring rod nor by the latching rod (both are retracted). It is accordingly free to move within a range that encompasses both blocking and non-blocking positions of the output knob (with respect to a probe knob on some logic rod to which the state of the register cell is an input). In the transition to frame b, the spring rod is advanced, a small, compliant knob on its surface presses against the spring surface of the state rod, and the state rod is forced to the right, into its 0 position.

11.4.1.2. Input

In the transition from b to c.1, the input rod moves to its 1 state, and the steric contact between the input knob and the sensing knob forces the state rod to the left, into its 1 position. In doing so, the input rod does $\sim 30kT \approx 120 \text{ mJ}$ of work against the force applied by the spring knob; assuming a displacement of 0.5 nm, the required force is $\sim 0.24 \text{ nN}$.

In the transition from b to c.0, the input rod remains in its 0 state, and no displacements occur. The state rod remains in its 0 position. (In either sequence, the state rod is now ready for reading by further logic rods.)

11.4.1.3. Latching

In the transition from c.0 to d.0, the latching rod advances, capturing the latching knob to the right and stabilizing the state rod in its 0 position. In the transition from c.1 to d.1, the latching probe captures the latching knob to the left, stabilizing the state rod in its 1 position.

11.4.1.4. Input reset

In the transition from d.0 to e.0, the spring rod retracts, leaving the state rod in place. In the transition from d.1 to e.1, both the spring rod and the input rod retract, again leaving the state rod in place. These rod retractions need not be synchronized. In either case, the resetting of the input rod (along with others controlled by the same clocking phase)

can initiate a chain of resetting operations that returns a pre-register combinational logic system to its initial state. Meanwhile, the latched state rod is serving as the input gate knob for a logic rod in the first stage of the post-register combinational logic system.

11.4.1.5. Expansion

After the post-register combinational logic system has completed its cycle and been reset, the state rod can be returned to the condition represented by frame a, by retracting the latching rod. This results in a re-expansion of the range of motion of the state rod and loss of the stored information.

11.4.2. Device size and packing

The state rod width can equal the logic rod width, and (for widths of ~ 1 nm), the separations can likewise be equal, making registers of this sort geometrically compatible with densely-packed arrays of logic rods with the parameters used in the exemplar system of Section 11.3. With a geometry like that shown in Figure 11.4, the length of the assembly is ~ 4 nm, and its height is ~ 5 nm, including an allowance for housing structure, yielding a total volume of ~ 40 nm³. Relative to a good design, this volume should prove conservative.

11.4.3. Energy dissipation estimates

Unlike logic rods in a purely combinational system, registers of this sort undergo a logically and therefore thermodynamically irreversible cycle. Their energy dissipation mechanisms can be divided into frictional mechanisms, which approach zero energy per cycle as speed approaches zero, and fundamental mechanisms, which impose a fixed loss per cycle, regardless of speed.

11.4.3.1. Vibrational, thermoelastic, and drag losses

Owing to small size and associated high modal frequencies, losses resulting from the excitation of vibrational modes by rod displacements will be negligible compared to that in logic rods. Thermoelastic losses depend on the square of the stress, which is substantial only in the spring rod, and then only if one spring rod serves many register cells. For a cross-sectional area ≈ 1 nm², a spring rod serving 16 register cells would have a mean-square tensile stress of ~ 2 GPa. This yields an estimate of thermoelastic losses per cell per cycle of $\sim 10^{-23}$ J.

The speed of sliding motions in this register design are uniformly equal to or less than those in the associated system of logic rods. Since the sliding interfaces will be similar, the energy losses per unit area per cycle will be similar. The sliding interface area of an interlock with the exemplar parameter is $\sim 12 \text{ nm}^2$; for the register system, it is roughly twice that. Accordingly, drag losses can be estimated at $\sim 10^{-23} \text{ J}$.

11.4.3.2. Nonisothermal compression and expansion losses

Losses from nonisothermal compression and expansion can be estimated using the model of Section 7.5.1. Given the stated dimensions and a density of 3500 kg/m^3 of (surface-corrected) volume, the mass of the state rod is $\sim 6 \times 10^{-24} \text{ kg}$, giving it an rms thermal velocity of $\sim 25 \text{ m/s}$ at 300 K .

During the compression phase, the extension of the spring rod reduces the range of motion of the state rod from $\sim 0.5 \text{ nm}$ to a small region (e.g., an effective width of $\sim 0.16 \text{ nm}$, assuming that the stiffness of the final potential well is $\sim 1 \text{ N/m}$). If the compression operation is assumed to take 0.1 ns , with an approximately linear motion over a distance of $0.5 - 0.16 = 0.34 \text{ nm}$, then the speed is 3.4 m/s . Assuming that the input knob and the spring probe are built for reasonably efficient energy exchange with state knob thermal motions, Eq. (7.60) with an accommodation coefficient $\alpha = 0.5$ should provide a reasonable estimate of the energy loss. With the above parameters, the result is $\sim 1.4 \text{ mJ}$.

During the expansion phase, the withdrawal of the latching rod allows expansion of the range of motion of the state rod from a small region to a region 0.25 nm long before free expansion occurs (assuming, for the moment, that barrier crossing is switched on instantaneously). Assuming that this also takes 0.1 ns , and assuming length and α as above, Eq. (7.60) yields an estimated energy loss of $\sim 0.14 \text{ mJ}$.

Nonisothermal compression and expansion results in the largest speed-dependent, non-fundamental losses identified here. Avenues for their minimization include reductions in speed (undesirable but direct and effective), reductions in compression and expansion lengths (for which there is only modest scope), reductions in state rod mass (which might be substantial in a good design), and improvements in the coupling of state rod motions to thermal modes (which will involve trade-offs between drag and nonisothermal compression losses).

11.4.3.3. Free expansion losses

As the latching probe withdraws, the barrier dividing the potential wells corresponding to 0 and 1 disappears; in a good design, these potential wells will have approximately equivalent free energies throughout the merging process. As discussed in Section 7.6.3, the merger of symmetrical potential wells leads to an increase in entropy of $\ln(2)k$, and hence to a loss of free energy of $\ln(2)kT$, or ~ 2.87 mJ at 300 K. This loss is irreducible.

11.4.4. Fluctuations in stored energy

The energy stored in a register cell during the input phase differs with the logic state: in writing a 1, the input rod does work against the spring probe, resulting in a net transfer of energy (on spring-probe withdrawal) from the input-rod drive system to the spring-rod drive system. This energy transfer will be on the order of 120 mJ, with the precise value depending on design details and the desired level of register reliability.

11.5. Combinational logic systems and finite-state machines

A standard way to implement combinational logic (i.e., logic that implements a function mapping one set of bit vectors into another) is through the use of programmed logic arrays (a.k.a. programmable logic arrays, programmable array logic, PALs and PLAs). PLAs can be combined with registers to build finite state machines and much of the control circuitry for a CPU (Mead and Conway 1980). They are used here as an example of how logic rods and registers can be combined into clocked computational systems.

11.5.1. Finite-state machine structure and kinematics

The PLA-based finite state machine illustrated in Figure 7.6.3 is nearly a device-for-device translation of a design exercise described in *Introduction to VLSI Systems* (Mead and Conway 1980), with the chief difference in the PLA being a reduction by a factor of two in device count of the AND-plane, resulting from the use of CMOS-like switching elements rather than NMOS. In addition, the feedback of state information from the output register to the input register is transmitted through a simple PLA structure, to provide the delay required to permit the rods in the main PLA to be reset.

Computation proceeds as follows: First, the vertical lines to the left (set 1 of PLA a) are tensioned; their mobility is determined by the bit vector latched into the register to the lower left (register ba). After a switching time, the horizontal lines (set 2 of a) are tensioned, with their mobility determined by the state of set 1. After another switching time,

the vertical lines to the right (set 3 of a) are tensioned, and move as determined by set 2. Finally, the state of set 3 is latched into the register to the lower right (register ab) by a tensioning of the latch line L2. A similar process then proceeds, starting with this register and propagating through the lower PLA (b) while the series of motions in the upper PLA is reversed.

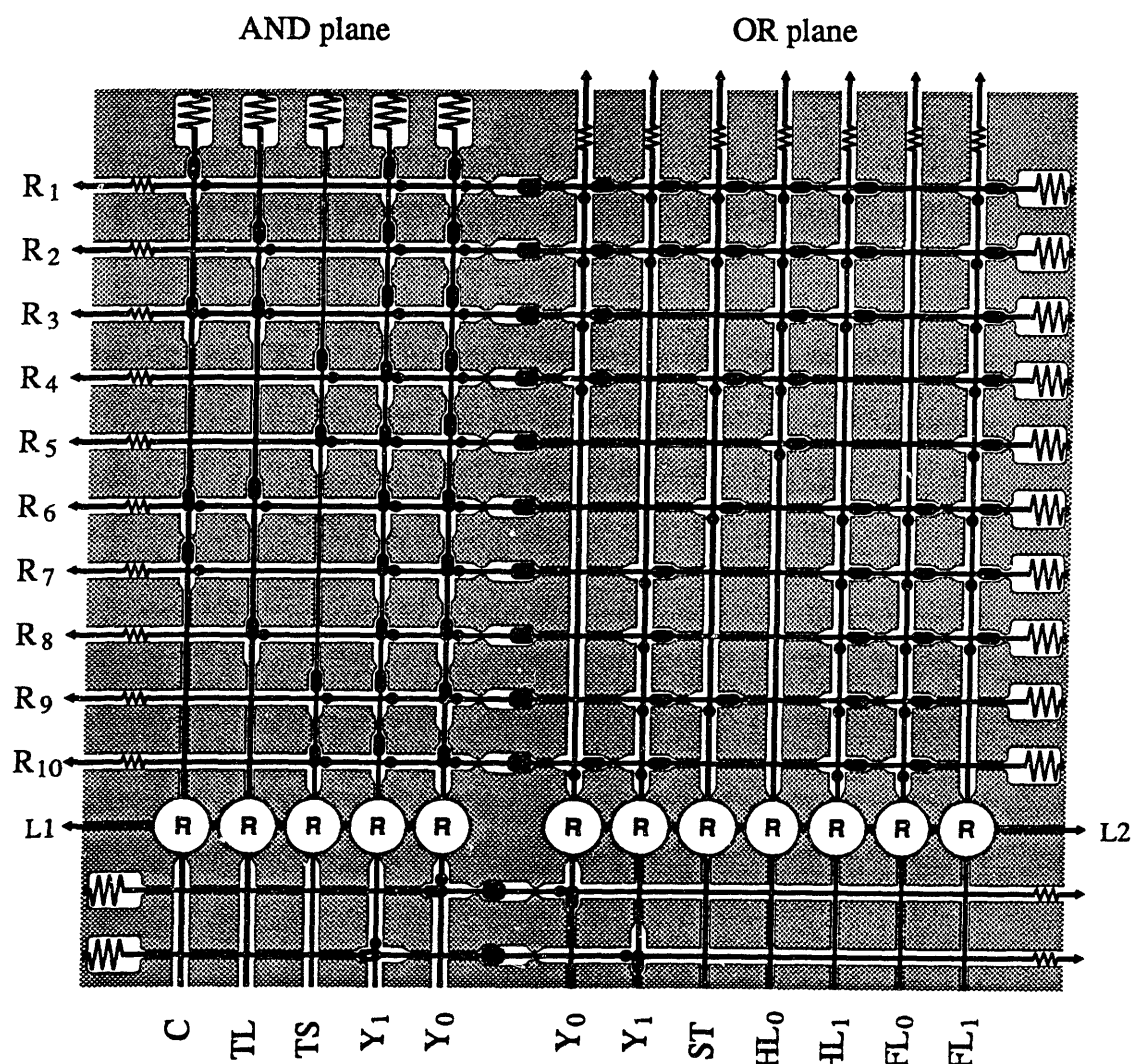


Figure 11.6. A PLA-based finite state machine translated from an NMOS design example (Mead and Conway 1980) to nanomechanical logic, in the same graphical notation as Fig. 11.1, but omitting the several drive systems owing to the constraints of a flat representation. Labeling of the lines corresponds to that in the source (which implements a traffic light controller with inputs from a timer and a car sensor).

11.5.2. Finite-state machine timing and alternatives

The upper panel of Figure 11.7 shows a timing diagram for a two-stage finite state machine. Note that the time for a cycle (which includes two register-to-register transfers with intervening combinational logic) is $10t_{\text{switch}}$, or 1 ns for the parameters used in the exemplar system. In this timing sequence, both the register compression processes (driven by the spring rod) and the latching processes (driven by the latching rod) take t_{switch} , as assumed in Section 11.4.3. Motion of the first set of rods in the post-register PLA, however, does not immediately follow the displacement of the last set of rods in the pre-register PLA (as it could, based on purely local constraints); it is instead delayed by $2t_{\text{switch}}$ in order to satisfy global timing constraints involving the requirement that a PLA and its input register be reset when the other PLA output is ready to be written. In the overall cycle, $4t_{\text{switch}}$ out of the cycle time of $10t_{\text{switch}}$ is consumed in these delays.

These delays can be eliminated by moving to a different architecture. The lower panel of Figure 11.7 shows a timing diagram for a four-stage finite state machine, in which a cycle involves four register-to-register transfers with intervening combinational logic. If PLA a and c are the same as a in the above example, and if b and d are the same as b in the above example, then (aside from an alternation in which register contains the most current bit vector) the system is equivalent to that above, with a full four-stage cycle corresponding to two cycles of the two-stage system. (With $a \neq c$ and $b \neq d$, the four-stage system is equivalent to a more complex system of two-stage register-to-register PLAs.)

The four-stage architecture performs register-to-register combinational logic operations at $5/3$ the frequency of the two-stage architecture, using twice the volume and device count to do so. The register compression process, moreover, can be allotted a time period of $3t_{\text{switch}}$ to complete; the energy dissipation model for this process predicts that this change will divide energy dissipation by a factor of 3. The estimates in Sections 11.3.4 and 11.4.3 suggest that register compression will be the largest (non-irreducible) energy loss mechanism. Accordingly, in systems where devices and volume are inexpensive, and where speed and low power dissipation are at a premium, it appears that a four-stage architecture will be superior to the two-stage architecture described above.

The time required for register-to-register transfer through a PLA (t_{PLA}) can be used to estimate the rate of instruction execution in a RISC machine (which for a typical mix of instructions is approximately equal to the clock rate). A clock cycle in a well-designed

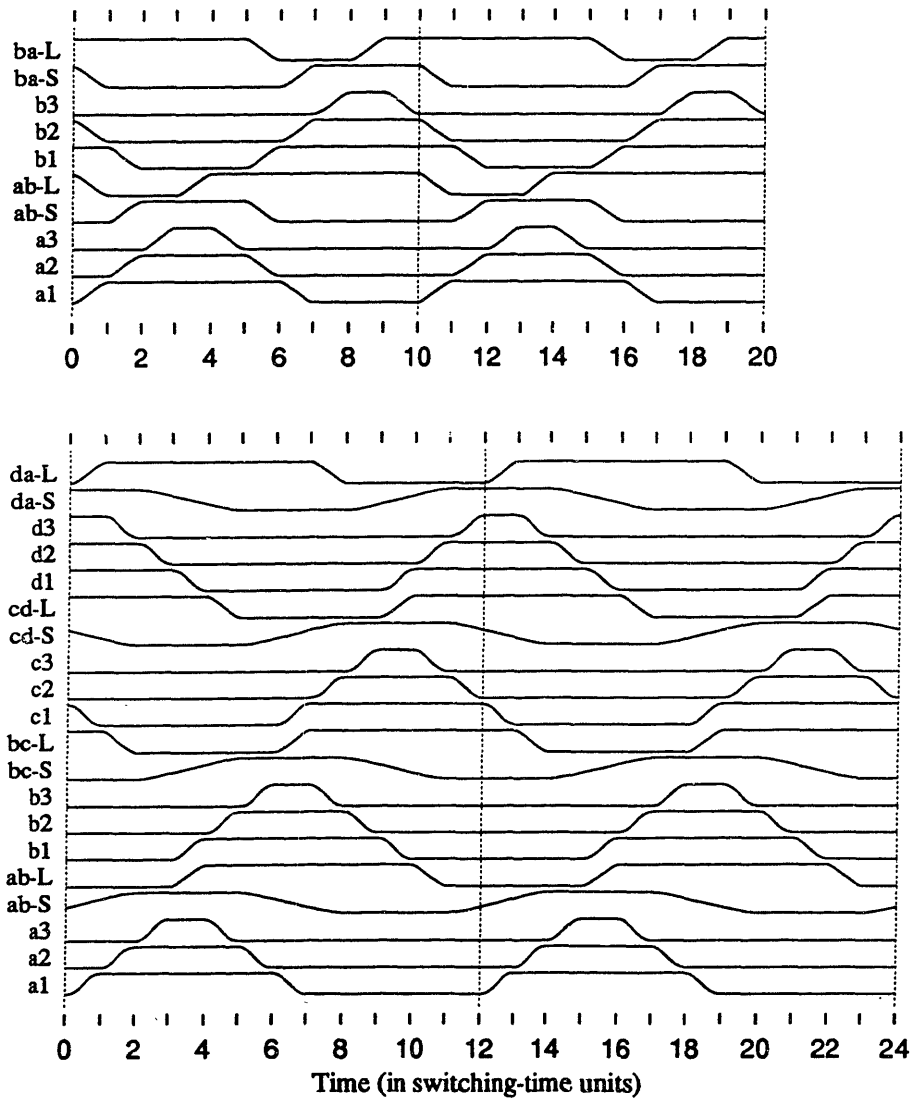


Figure 11.7. Timing diagrams for two PLA-based systems, with time measured in units of the logic rod switching delay t_{switch} , and an arbitrary scale for displacement (tensioned is high, de-tensioned is low). Upper panel: two PLAs (a and b), two registers (ab and ba). Lines a1, a2, a3 represent the motions of mobile members of the three sets of rods in PLA a; lines ab-S and ab-L graph the displacements of the spring rod and the latching rod (respectively); other labels are analogous. Lower panel: four PLAs and four registers, otherwise analogous to the above. (See text for discussion.)

RISC machine requires $2t_{\text{PLA}}$ to $3t_{\text{PLA}}$, with the latter being a conservative value for the present purposes (Knight 1991). Since the above calculations yield $4t_{\text{PLA}} = 1.2 \text{ ns}$, it appears that a conservatively-designed rod logic technology will enable the implementation of RISC architectures capable of $\sim 1000 \text{ MIPS}$.

11.5.3. Fan-in, fan-out, and geometric issues

The exemplar logic rod described in Section 11.3 has a fan-in and fan-out of 16, with a straight geometry. Real systems will seldom require a rod with exactly these properties. Some feasible variations are discussed in this section.

11.5.3.1. Changes in fan-in

Reductions in fan-in shorten the input segment and increase its stiffness. Accordingly, such system properties as feasible switching speed and total energy dissipation improve. The only disadvantage is that the overhead of the drive mechanism, alignment knob, and reset spring are spread over a smaller number of devices, increasing the volume and total energy dissipation per device.

Increases in fan-in are possible by simply lengthening the input segment, but at the cost of decreasing the feasible switching speed. Alternatively, multiple input segments can be mounted in parallel (e.g., yoked together at the drive end and at the alignment knob), resulting in a system with the same logical properties as a series connection, but with higher vibrational frequencies and hence a higher feasible switching speed. (Indeed, this strategy might be desirable for $n_{\text{in}} \leq 16$.)

11.5.3.2. Changes in fan-out

Reductions in fan-out parallel reductions in fan-in where stiffness, switching speed, energy dissipation, and volume are concerned. The resulting increases in stiffness also reduce error rates resulting from thermal noise; in optimized designs (where error rates are significant but acceptable), this shift will typically permit the use of either slimmer rods or lower values of F_{align} .

Increases in fan-out increase both switching speed and error rates, although the latter can be kept in check by proportional increases in S_{eff} (and hence stiffness). Again, yoking a set of parallel output segments together (e.g., at the alignment knob and the reset spring) results in a system with the same logical properties and higher frequencies and switching speeds. This strategy also reduces errors from thermal excitation, and reduces the required magnitude of F_{align} , on a per-rod basis. (And again, this strategy might be

desirable for $n_{\text{in}} \leq 16$.)

11.5.3.3. Flexible rod geometries

Although interlocks of the sort considered here are based on straight rods, a logic rod system can be made from multiple segments of this sort, linked by curved segments passing through curved housings. These segments can be made from structures that are one atom thick in the direction of curvature, such as polyyne chains or strips of graphite-like material. A radius of curvature ~ 1 nm will then be feasible. Potential energy functions for sliding of the curved segments can be made smooth using the design methods described in Chapter 10. Bent segments permit broad geometric flexibility in the design of rod logic systems, permitting (for example) PLAs stacked face to face to share access to a register.

11.5.4. Signal propagation with acoustic transmission lines

In the exemplar system, the state of an output gate is switched in 0.1 ns over a distance of 64 nm, yielding an effective signal propagation speed of only 640 m/s. In large logic systems, it will be desirable to transmit signals at the full acoustic speed, ~ 17 km/s in diamond. This can be accomplished using systems that launch acoustic pulses at one end of a rod, and probe the displacement of the rod at the far end.

For concreteness, consider a rod 5μ in length, with $S_{\text{eff}} = 3 \text{ nm}^2$ and $E = 10^{12} \text{ N/m}^2$. The one-way signal propagation time along this rod is ~ 0.3 ns. As with the exemplar logic rods, a drive system and drive spring apply forces to one end, and an input segment with one or more interlocks either blocks or fails to block motions of the rod as a whole. If the rod is in a mobile state, the force applied by the drive spring accelerates the rod. After 0.1 ns, a force of 1.8 nN ($\approx 2F_{\text{align}}$ in the exemplar system) will have displaced the end of the rod by 1 nm, and propagated a region of tensile stress 1.7μ along the rod. Ceasing to apply tension for a further 0.1 ns, then applying a reverse force for 0.1 ns restores the rod end to its initial position, while launching a wave which, as it passes, displaces the rod by 1 nm for 0.1 ns. At the far end, this displacement can be probed by a conventional interlock associated with a short, briefly-displaced logic rod that writes into a register and is immediately reset.

The positional uncertainty at the receiving end of the rod is adequately constrained by positional control at the sending end. The energy associated with a 1 nm stretching deformation of the rod is ~ 0.6 aJ, or $\sim 150 kT$. Non-thermal vibrational modes are a con-

cern here, but clocked damping can be introduced to remove energy from these modes between signal pulses, without dissipating energy from the signal pulses themselves.

The energy of a wave of this sort is ~ 3.6 aJ. The round-trip signal propagation time is ~ 0.6 ns, hence a pulse can be launched and recovered in a single clock period. In systems with stable clocking at the correct speed, the energy delivered by the reflected pulse can be recovered by a suitably-structured drive system. Since an outgoing pulse consisting of a rarification followed by a compression will be reflected from a free end as a compression following a rarification, the return pulse resembles the time-reversal of the outgoing pulse, and a simple reversal of drive shaft motion produces a sequence of cam displacements that will couple the energy of the return pulse into the drive system with reasonable efficiency. Energy dissipation will be larger than that in the exemplar rods, but the need to propagate signals over such a distance should be reasonably rare; a sphere of $5\text{ }\mu$ radius can contain $> 10^{10}$ interlocks.

11.6. Clocking and power distribution for CPU-scale systems

In electronic digital logic, clocking systems with two, four, or eight phases are not uncommon. What were termed two- and four-stage architectures in Section 11.5.2 might seem similar, but are in fact substantially different: each stage includes three levels of logic and a register with two control rods, requiring a total of twenty distinctly-clocked inputs in a four-stage system. Further, the clocking systems in electronic digital logic are used to modulate power distributed by a DC system, while the proposed rod logic systems combine power and clocking. The overall design of the clocking and power supply system is accordingly quite different.

To estimate clocking and power-distribution parameters appropriate for CPU-scale systems requires a model describing the size and content of a CPU. The following will assume a device containing 10^6 interlocks, 10^5 logic rods, and 10^4 register cells, which (together with interconnects, power supply mechanisms, wasted space, etc.) occupy a cubical volume 400 nm on a side. (Assuming a mean density of 2000 kg/m^3 implies a mass of 1.6×10^{-17} kg and hence a half-life against radiation damage in Earth ambient background radiation of ~ 100 years; see Fig. 6.13.)

11.6.1. Clocking based on oscillating drive rods

Consider the timing diagram for a two-stage system (shown in the upper panel of Figure 11.7). The tensioning and de-tensioning of each set of rods can be characterized

by the duration of the tensioned and de-tensioned intervals, and by the time at which the tensioned interval is half-completed; the latter time can be taken to define the phase of the clock for that set of rods. Inspection of this diagram shows four phases: one for the logic rods of PLA a and the spring rod of register ab, another for the latching rod of register ab, and another two phases for PLA b and register ba.

As illustrated in Figure 11.8, a sinusoidally-oscillating cam surface on a drive rod can generate clocked drive-system impulses of with intervals that depend on the position of the follower with respect to the mean position of the ramp on the cam surface. The

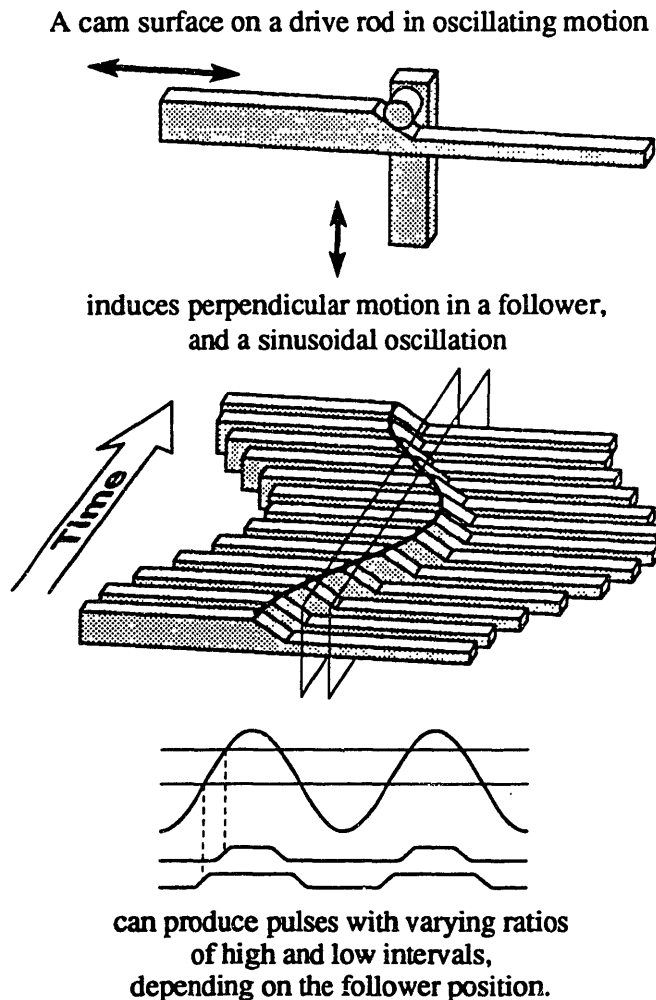


Figure 11.8. Diagrams and text illustrating the generation of clocked drive-system impulses of varying intervals from the sinusoidal motion of cam surface (only one ramp and one follower are shown).

drive system for a set of rods could use several followers and several ramps on the same drive rod, all positioned to move in synchrony. A single drive rod (or system of drive rods moving in phase) can serve multiple drive systems having different intervals of tensioning. Further, since a ramp can make a transition in either direction, sets of rods in opposite phases can likewise be driven by a set of drive rods having a single phase. Accordingly, the ten different patterns of clocked impulses required for the two-stage system in Figure 11.7 can be generated by drive rods having only two distinct phases; the four-stage system similarly requires four phases.

11.6.2. A CPU-scale drive system architecture

A drive system requires a source of motive power to compensate for energy losses, a mechanism for buffering energy to compensate for fluctuations in the energy stored in the logic system in different logic states, and a mechanism for coupling drive rod motions such that different rod systems have the correct relative phases. The source of motive power can be a DC electrostatic motor (Section 11.7). The mechanism for buffering energy can be a flywheel, rotating with a frequency equal to that of the system clock. The coupling mechanism can be a crankshaft, which can convert a single rotary motion into an indefinitely large number of linear, approximately-sinusoidal motions of differing relative phases. Note that subsidiary crankshaft mechanisms can be used at remote locations to couple drive subsystems of differing phase, permitting energy transfer between them and increasing the resistance to local desynchronization. These subsidiary crankshafts also provide one of several mechanisms for transmitting drive power around corners, permitting design-level flexibility in the location and orientation of drive rods within the logic system.

This mechanism for clock distribution has dynamics differing from that used on present integrated circuit chips. The latter have properties dominated by damping, leading to diffusive spread of a clock signal along the conducting paths. The rod logic drive system, in contrast, has dynamics dominated by inertia, leading to resonant behavior in which all parts of the system could, in the absence of load, share a single phase. Nonetheless, clock signal distribution will likely have to differ in larger-scale synchronous systems. Control of local phase using optical or electrical signals would provide one approach; as M. Miller observes, propagation of acoustic pulses at known speeds over known distances (e.g., with propagation times that are an integral number of clock periods) can also serve as a basis for large-scale synchronization.

11.6.3. Energy flows and clock skew

During a clock cycle, a drive rod executes an oscillation, forcing movement in a set of logic rods, spring rods, or latching rods. *If* the energy stored in these rods as a function of time were invariant from cycle to cycle, and *if* there were no energy dissipation, then the potential energy function of the drive rod as a function of position could be adjusted such that the total energy of the system was invariant, with the drive rod moving in an effectively harmonic potential energy field and executing sinusoidal oscillations with no external energy input or constraining force. In practice, energy must flow into the drive rod from an external source to compensate for energy losses, and variations in stored energy from cycle to cycle result in fluctuating forces and displacements in the drive system.

11.6.3.1. Energy fluctuations

In a clock cycle, fluctuations in stored energy are > 100 times greater than the energy dissipated, and hence dominate the cycle-to-cycle fluctuations in drive-system dynamics. Further, fluctuations in the energy stored in logic rods are ~ 10 times greater than comparable fluctuations in register cells. An energy-based analysis, focusing on logic rod states, can provide an estimate of the magnitude of the required drive system stiffness and inertia required to ensure that these fluctuations do not result in excessive disturbances in local clock phase.

An examination of the timing diagrams in Figure 11.7 shows that no more than $1/3$ of the 6 sets of logic rods in the two-stage system are tensioned simultaneously, and as are no more than $5/12$ of the 12 sets in the four-stage system. Accordingly, the estimated $1.2 \text{ aJ } \Delta \mathcal{E}_{\text{state}}$ derived for logic rods with the exemplar system parameters (Section 11.3), together with the assumed 10^5 rods for a CPU-scale system yields an estimated maximum variation in stored energy, relative to the mean (more accurately, mid-range) stored energy, of $\Delta \mathcal{E}_{\text{max}} = 1/2 \times 5/12 \times 10^5 \times 1.2 \text{ aJ} = 2.5 \times 10^{-14} \text{ J}$.

(Note, however, that it is possible to implement a logic system in which every logic rod set to a 1 state is mirrored by a rod set to a 0 state, thus cancelling the major contributions to $\Delta \mathcal{E}_{\text{max}}$; this approach would approximately double system volume and energy dissipation. A similar scheme is used to avoid fluctuations in Cray computers.)

11.6.3.2. Acceptable clock skew and required drive-system stiffness

In a rod logic system with the exemplar parameters, switching produces a 1 nm displacement in the gate knobs, and a comparable displacement in the probe knobs. An interlock switching from a non-blocking to a blocking state can produce an error if substantial probe knob displacements precede substantial gate knob displacements; significant added energy dissipation and an increased probability of error can occur if the motion periods overlap more than slightly.

A system designed such that a premature probe-knob displacement of 0.05 nm occurs in the worst-case logic state will operate reliably, and should experience little added energy dissipation under ordinary conditions. With the exemplar parameters, this displacement corresponds to a clock skew of ~ 0.014 ns. At the assumed drive-rod speed of ~ 38 m/s, (Sec. 11.3.4.4) this skew corresponds to a displacement of ~ 0.53 nm, relative to the optimal drive-rod position (a phase-angle error $\Delta\phi \approx 0.073$ rad).

For a system with the exemplar parameters, the peak speed of motion of a drive rod was estimated at 38 m/s (Sec. 11.3.4.4). If the entire peak stored energy $\Delta\mathcal{E}_{\max} = 2.5 \times 10^{-14}$ J were delivered in a single period $t_{\text{switch}} = 0.1$ ns, over a stroke length $0.1 \text{ ns} \times 38 \text{ m/s} = 3.8$ nm, the mean (fluctuating-component) force applied by the drive system during this period would be $\sim 6.6 \times 10^{-6}$ N.

To transmit this force with a displacement of ~ 0.53 nm requires a system with a stiffness of $\sim 1.2 \times 10^4$ N/m. If this is to be achieved using drive mechanism components with a typical length of 200 nm ($\sim 1/2$ the system diameter) and a modulus $E = 10^{12}$ N/m², their required total cross sectional area is ~ 2500 nm², and their volume is ~ 0.0078 that of the reference CPU-scale system. Accordingly, a drive system stiff enough to limit clock skew to acceptable values occupies need occupy only a small fraction of the total system volume.

11.6.4. Energy buffering

Within the constraints of a 400 nm CPU-scale package, one can allocate space for a cylindrical flywheel of radius $r = 195$ nm and height $h = 20$ nm. The kinetic of such a flywheel will be

$$\mathcal{E}_f = \frac{1}{2} I_f \omega_f^2 = \frac{\pi^2 r^4 \rho h}{t_{\text{clock}}^2} \quad (11.25)$$

With the above geometric parameters, $\rho = 3500 \text{ kg/m}^3$, and $t_{\text{clock}} = 1.2 \text{ ns}$, $\mathcal{E}_f \approx 6.9 \times 10^{-13} \text{ J}$. Since $\Delta\mathcal{E}_f/\mathcal{E}_f \approx 2\Delta\omega_f/\omega_f$ ($\Delta\mathcal{E}_f/\mathcal{E}_f \ll 1$), $\Delta\mathcal{E}_{\text{max}} = 2.5 \times 10^{-14} \text{ J}$ implies $\Delta\omega_f/\omega_f \approx 0.018$.

A flywheel remains below its lowest critical angular velocity so long as the condition

$$\sqrt{\frac{k_{s,t}}{m}} > \omega_f \quad (11.26)$$

is met (where $k_{s,t}$ is the stiffness of the bearing structure in resisting displacement of the shaft in a direction perpendicular to its axis, and m is the flywheel mass). For the parameters given above, this requires a bearing with $k_{s,t} > 230 \text{ N/m}$, a condition which is easily met. Note that the rim velocity, $\sim 1000 \text{ m/s}$, is well below the bursting speed for a diamond hoop, and that the presence of a significant vacuum gap avoids frictional drag.

11.6.5. Logic-system power dissipation

The total power dissipation for the model CPU-scale system can be estimated from the component energy dissipations and the clock rate. Assuming a 1.2 ns clock, 0.03 mJ per switching operation per interlock (Sec. 11.3.4), 4.4 mJ per register-cell storage cycle (Sec. 11.4.3), and operation of every device on every clock cycle yields an estimated power dissipation of 74 aJ per clock cycle, or $\sim 90 \text{ nW}$. The associated mechanical loads on the drive system are trivial, but these losses must be made up by a power supply system.

11.7. Power supply systems

Electrostatic DC motors provide one straightforward means of supplying power to nanomechanical systems. In one implementation (Fig. 11.9), electric charge is placed on the rim of a rotor as the rim passes within a dee electrode, and is then transported by rotation across a gap to the interior of the opposite dee electrode, where it is removed and replaced with a charge of the opposite sign. If a voltage of the proper sign is applied across the dees, the charges in transit between them experience a force that applies a torque to the rotor, delivering power (this resembles a Van de Graff generator operating in reverse). In a nanomechanical computer, the rotor of the electrostatic motor can conveniently serve as the flywheel; the above geometric parameters will accordingly be used for sample calculations.

11.7.1. Charge carriers and charge density

Feasible voltages and separations (and hence charge densities) are constrained by the appearance of leakage currents resulting from tunneling and field emission. For small voltages (≤ 1 V), an insulating gap of ~ 3 nm can limit tunneling current densities to magnitudes that are negligible in a nanomechanical context (e.g., $\leq 10^{-18}$ A/nm²); metals exposed to vacuum can sustain fields of ~ 2 V/nm, with a negligible current density from field emission (Farrall 1980).

The rim of a rotor can be made of insulating material with embedded conductive electrodes (Fig. 11.9) that serve as charge carriers. If the electrodes are separated by 3 nm and no large voltages are imposed between them, interelectrode charge transfer can be neglected. Neglecting the beneficial effects of insulator polarization, the surface charge density of the electrodes associated with a (modest) field of 0.2 V/nm is $\sim 0.0018 \text{ C/m}^2$.

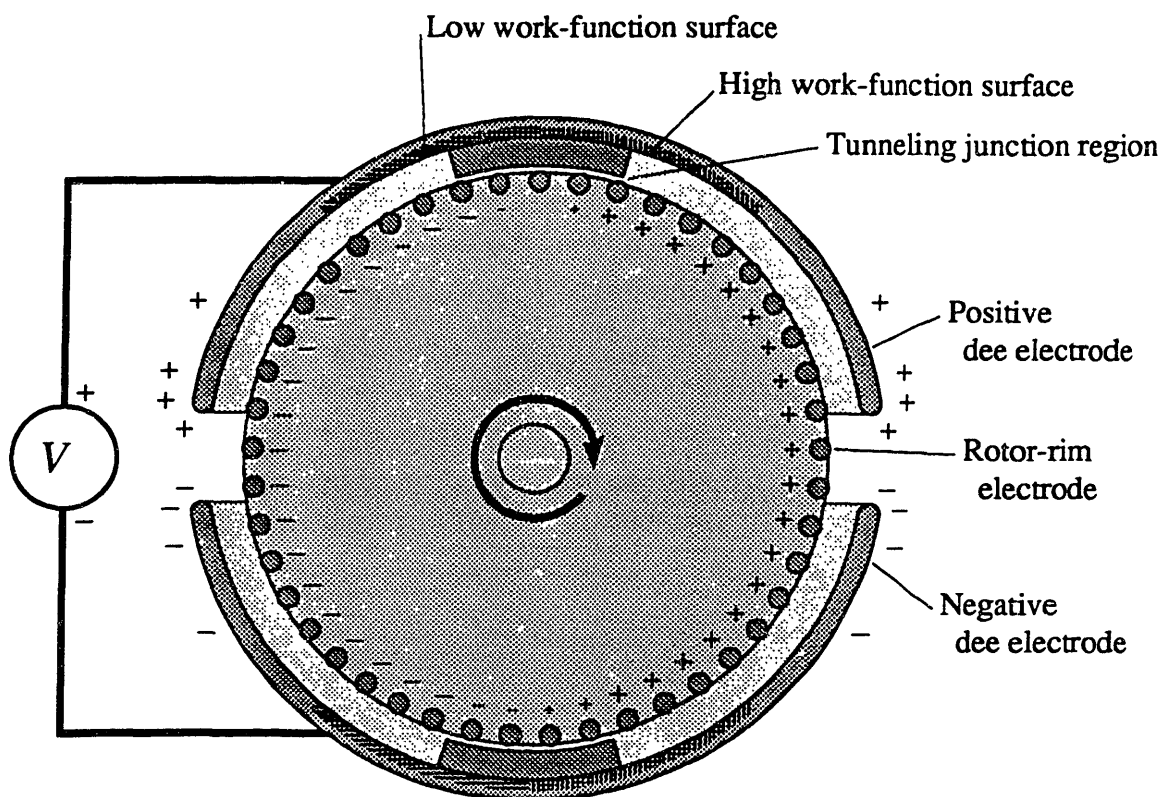


Figure 11.9. Schematic diagram of an electrostatic motor (not to scale). The dee electrodes and rotor-rim electrodes are conducting, the rotor structural material is insulating. Rotation and torque are in the direction shown (see text for description).

Assuming an electrode diameter of 3 nm and length of 20 nm, the charge per electrode is $\sim 3.3 \times 10^{-19}$ C, or ~ 2 electronic charge units; the charge per unit rotor circumference is $\sim 5.5 \times 10^{-11}$ C/m. A vacuum gap of several nanometers will suffice to insulate the electrodes from a nearby dee electrode surface.

11.7.2. Tunneling junction properties

Experiment shows that tunneling currents crossing a vacuum gap in an STM apparatus can exceed 300 nA at 20 mV (Gimzewski 1988), implying a resistance near zero voltage of $< 10^5 \Omega$; resistance falls with increasing field. STM theory assumes that the bulk of the current is transmitted through a single atom at the tip, hence devices with parallel surfaces of several square nanometers should exhibit resistances of $\sim 10^3 \Omega$.

A single electrode can be charged in 0.1 ns by a current of ~ 3 nA. If charge is transferred using a single tunneling contact with properties like those observed experimentally, this can be accomplished with a voltage differential of ~ 0.2 mV. Achieving a reversal of charge in the same time would require ~ 0.4 mV with these junction parameters.

11.7.3. Electrode charging mechanism

11.7.3.1. Work functions and contact potentials

The work function ϕ_w of a material depends strongly on its surface condition, and extreme values of the ϕ_w will presumably be achieved by careful tailoring of surface structure. The range of work functions observed on clean surfaces of elemental metals of reasonably high melting point places a lower bound on breadth of the range to be expected in a broader class of stable structures. Among the lower values is $\phi_w = 2.7$ eV (for samarium, melting point ≈ 1350 K); among the higher values is $\phi_w = 5.7$ eV (for platinum, melting point ≈ 2050 K) (Lide 1990).

Placing two metals of differing work function in electrical contact equalizes electron energies through the system, and results in a potential difference (the *contact potential*) between the region of space immediately outside one surface and the region immediately outside the other. This potential difference implies an electric field, which is generated by a corresponding surface charge density. This provides a convenient mechanism for charging the rotor-rim electrodes.

11.7.3.2. Charging via contact potentials

Consider a rotor-rim electrode in electrical contact with the interior of a dee electrode, in an environment where its surface and the surrounding dee electrode surface share a common work function $\phi_{w,\text{middle}} = 4.2 \text{ eV}$. The surface charge density will be zero. Now, while maintaining electrical contact through a tunneling junction, allow the rotor-rim electrode to move into a region in which the interior of the dee electrode is covered by a material with a work function $\phi_{w,\text{high}} = 5.7 \text{ eV}$. The magnitude of the contact potential will be 1.5 V; with a capacitance of $\sim 2.1 \times 10^{-19} \text{ F}$, the charge on the rotor-rim electrode will have a magnitude of 2 electronic units. Modeling the rim electrodes as well-separated cylinders parallel to a plane at a distance of 3 nm from their surfaces yields a capacitance per cylinder of $\sim 6.3 \times 10^{-19} \text{ F}$; modeling them as a continuous strip at the distance of their centers yields a capacitance per rim-electrode length of $\sim 2.4 \times 10^{-19} \text{ F}$. Accordingly, the assumed charge could be induced on the electrodes with a substantially smaller value of $\phi_{w,\text{high}} - \phi_{w,\text{middle}}$. (Since $\phi_{w,\text{high}} - \phi_{w,\text{middle}}$ can equal $\phi_{w,\text{middle}} - \phi_{w,\text{low}}$, charge of the opposite sign can be applied in a similar fashion.)

11.7.3.3. The rim electrode charge/discharge cycle

A rotor-rim electrode can be made to go through the following cycle in a revolution of the rotor: Starting in the middle of the positive dee, it is in electrical contact with the dee via a tunneling junction, and in an environment where the induced charge on it is zero. Moving clockwise (Fig. 11.9), the rim electrode enters a region in which the neighboring dee surface has a higher work function, inducing a charge of $+2e$. This can occur gradually, since the mean work function of the surrounding electrode can be made to vary gradually by changing the fraction of area devoted to surfaces of different kinds. Once this charge is induced, the tunneling contact (made to a narrow ridge on the inner surface of the dee) is gradually broken in surroundings that keep the equilibrium charge on the rim electrode constant. After traveling a further distance, the positively-charged rim electrode exits the positively-charged dee and enters the negatively-charged dee, dropping through a potential difference of ϕ_{motor} as it does so. Once beyond the fringing fields at the entrance to the negative dee, the electrostatic environment is the same as it was before exiting the positive dee, and the approach to the middle of the negative dee is a nearly-precise reversal of the sequence of environments and charge-transfers that occurred in the positive dee, ending with no net charge on the rim electrode. Continuing through the neg-

ative dee, exiting, entering the positive dee, and returning to the initial position proceeds in the same manner as the first half of the cycle, except that low surrounding work functions induce a negative charge on the rim electrode.

This cycle avoids making electrical contact between conductors at differing potentials; this is directly analogous to chemical cycles that avoid merging an occupied potential well with a deeper potential well. By only connecting conductors at equilibrium, current flow is postponed until the contact resistance is low, and the process can remain close to thermodynamic equilibrium at all times. (Note that reversing the direction of rotor motion would reverse the direction of current flow and of work, making the system function as a DC generator.)

11.7.4. Motor power and efficiency

11.7.4.1. Motor power

Neglecting resistive and frictional losses, DC electrostatic motors of this kind deliver a shaft power equal to the product of the current and the voltage. With a rim speed of ~ 1000 m/s, and a charge per unit length of $\sim 5.5 \times 10^{-11}$ C/m, the current (including contributions from both sides of the rotor) is ~ 55 nA; at $\phi_{\text{motor}} = 10$ V, the delivered power is ~ 1.1 μ W. This is substantially more power than is necessary to operate a CPU-scale nanomechanical logic system of the sort previously described. (If it were insufficient, there is ample room for improvement in the design.) The power density in this system is large compared to that of macroscopic motors: $\sim 10^{15}$ W/m³. For comparison, Earth intercepts $\sim 10^{17}$ W of solar radiation. (It is not suggested that cubic-meter-scale systems of this sort be constructed; cooling requirements for practical systems are considered in Section 11.8.)

11.7.4.2. Energy dissipation and efficiency

Energy dissipation results chiefly from electrical resistance in the conduction paths and contacts in the motor, and from friction in the bearing at the axle of the rotor. These are readily estimated.

Since electrostatic motors do not exploit current-generated fields, and accordingly, lack long, coiled conductors, they lack large internal resistive losses. (Bearing friction might be considered a resistive loss associated with mechanical charge transport by the rotor.) At a current density of 1 nA/nm², conventional aluminum conductors do not suffer from electromigration problems (Mead and Conway 1980); this criterion can be used

to choose conductor cross sections, 55 nm^2 for the design discussed here. Resistance in a pair of leads, each of a length equal to the motor diameter, assuming a resistivity equal to the bulk value for aluminum ($\sim 2.7 \times 10^{-8} \Omega\cdot\text{m}$) is $\sim 390 \Omega$, yielding an estimated resistive power dissipation from this mechanism of $\sim 1.2 \times 10^{-12} \text{ W}$. The current flow responsible for distributing charge along a rim electrode contributes $\sim 10^{-13} \text{ W}$, calculated on a similar basis.

(The use of bulk resistivity values neglects the resistance added by surface scattering of electrons in conductors in which the diameter is not large compared to the electron mean free path. Conductors bounded by lattice-matched, epitaxially-aligned insulators, however, presumably would not suffer from significant surface scattering. Moreover, there exist one-dimensional organic conductors of lower bulk resistivity—in their direction of maximum conductivity—than aluminum; these should not be subject to significant size effects.)

The resistance of a single-atom tunneling contact is already known from STM work to be reducible to $< 10^5 \Omega$. Given the size of the rim electrodes, providing each with a contact region having a total resistance of $< 10^3 \Omega$ will be straightforward. Since the length of the rim contained in a dee is $\sim 400 \text{ nm}$, ~ 50 rim electrodes can be in contact in a given dee at any given time, with a steady mean rate of charge transfer. Allowing for the sequence of two transfers required to move charge from dee to dee, the contact resistance of the system is estimated to be $< 40 \Omega$, with an associated power dissipation of $< 1.2 \times 10^{-13} \text{ W}$.

Applying the band-stiffness scattering model to the bearing at the axle of the motor with a stiffness of 500 N/m (which meets the rotor-stability criterion), $r_{\text{eff}} = 10 \text{ nm}$, $\ell = 20 \text{ nm}$, $v = 50 \text{ m/s}$, $R = 10$, and the pessimistic value $\Delta k_a/k_a = 0.4$, yields an estimated bearing drag power of $\sim 9 \times 10^{-13} \text{ W}$.

Summing the above estimates yields an estimated power dissipation of $\sim 2.3 \times 10^{-12} \text{ W}$. At a total motor power output of $\sim 1.1 \mu\text{W}$, the estimated fractional power dissipation is $\sim 2 \times 10^{-6}$.

11.7.5. Motor startup

In the absence of acoustic transmission lines, all of the computational mechanisms work continue to work at lower speeds, and dissipated less energy per operation as they do so. Accordingly, a motor torque that is adequate to operate a system at full speed is more than adequate to start it (note that static friction can be made effectively nil, with

barriers less than kT). In the presence of acoustic transmission lines (which are more dissipative just below their design frequency), startup may be more complex. Use of a clutch to permit flywheel spinup before engaging the computational engine is one possibility; locking of transmission lines until the clock speed is within tolerances is another; use of extra motor power is a third.

If the motor is initially disengaged from a load, startup can occur spontaneously. The mean thermal speed of the rotor is ~ 0.01 m/s, with little damping. Rotation in reverse is effectively prohibited when the dees are charged, and forward rotation will begin to carry charge across the gap, producing torque, in ~ 10 μ s.

Positive motor startup can be ensured if the rim electrodes in the gap between the dees have the proper charge, resulting in a strong torque. This can be accomplished by providing extensions from each dee that come close enough to provide a weak tunneling current to the rim electrodes in the gap (on the side where they have the same charge in normal operation). If this rate is low, it has little effect on the above analysis of an operating motor, yet would ensure charging of electrodes in the gaps of any motor that was stationary and uncharged.

11.7.6. Speed regulation

Precise speed regulation calls for a mechanism based on a frequency standard, such as a mechanical oscillator or an external AC source (use of an AC synchronous motor would avoid this issue, at least locally). For a single computer system operating asynchronously with respect to other systems in its environment, approximate speed regulation will suffice. A spring-loaded variant of the Watt governor can serve this purpose, opening and closing a tunneling contact instead of a steam valve. If designed for bistability, it will oscillate between having a closed contact and having an open contact, dissipating little energy in either state.

11.8. Cooling

In nanomechanical systems of sufficiently large scale, the $\sim 10^{12}$ W/m³ power dissipation density of ~ 1 GHz nanomechanical logic systems of the sort described in this chapter will exceed any possible means of cooling. On a sufficiently small scale, however, cooling poses no problem. The ΔT of an isolated 100 nW system of 400 nm dimensions will be < 0.01 K in a medium with a thermal conductivity of 10 W/m-K.

On an intermediate scale, convective cooling will be appropriate. Water cooling of silicon chips can remove $\sim 10^7 \text{ W/m}^2$, suggesting that power dissipation of $< 10^3 \text{ W}$ in a cubic centimeter volume can be sustained. This is sufficient to operate $\sim 10^{10}$ CPU-scale systems, with an aggregate clock rate of $\sim 10^{19}$ per second.

11.9. Interfacing to conventional microelectronics

I/O constraints are likely to place major bounds on the nature of the problems to which large arrays of CPUs can be efficiently applied. In systems that minimize these constraints, it is unlikely that interfacing of nanomechanical computational devices to conventional microelectronics will play any role. Nonetheless, demonstrating the feasibility of this provides a straightforward way of establishing that nanomechanical systems can be interfaced to systems that are already well-connected with human operators and other systems in the macroscopic world.

11.9.1. Input using electrostatic actuators

Transduction of digital signals from standard 5 V microelectronics can be performed by electrostatic actuators of the sort diagrammed in Figure 11.10. This approach uses a structure resembling a standard logic rod (including gate and alignment knobs and a reset spring), but with the input segment and drive system replaced by an electrostatic mechanism driven by an external source of digital signals. In this mechanism, one plate of a small capacitor is fixed and connected to the signal source; the other plate is kept at

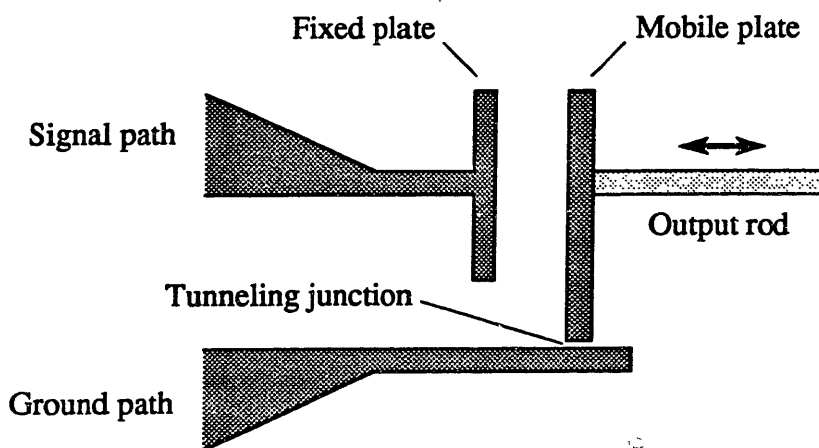


Figure 11.10. Schematic diagram of an electrostatic actuator for interfacing electrical inputs to rod logic outputs (moving parts and electrical components only); see text.

ground potential using a tunneling junction, while leaving it free to move (and so drive a rod) within a small range of displacements.

In a concrete example, an actuator with a stroke length of 1 nm (from a plate separation of 4 to 3 nm) and a potential ranging from 0 to -5 V can apply a force ranging zero to $2F_{\text{align}} = 2$ nN with a plate area of $\sim 290 \text{ nm}^2 \approx (17 \text{ nm})^2$. If the plate structure had a mass equaling that of a 1 nm thickness of diamond, and were free to move, crossing a 1 nm distance under this accelerating force would take ~ 0.03 ns, hence input data rates can be consistent with the rod logic speeds calculated in Section 11.5.

11.9.2. Output using modulated tunneling junctions

Transduction of digital signals from mechanical displacements in rod logic to electrical signals can be performed by mechanically-modulated tunneling junctions of the sort diagrammed in Figure 11.11. In this approach, the output gates of a logic rod are replaced by a mobile contact resembling the mobile plate in the electrostatic actuator. Here, however, the area is smaller, and the range of motion includes contact separations small enough to permit large tunneling currents; the tendency of the two surfaces to snap together under the influence of attractive forces can be resisted by small-area compressive contacts between relatively-inert protrusions, and by other means of tailoring the potential energy as a function of displacement.

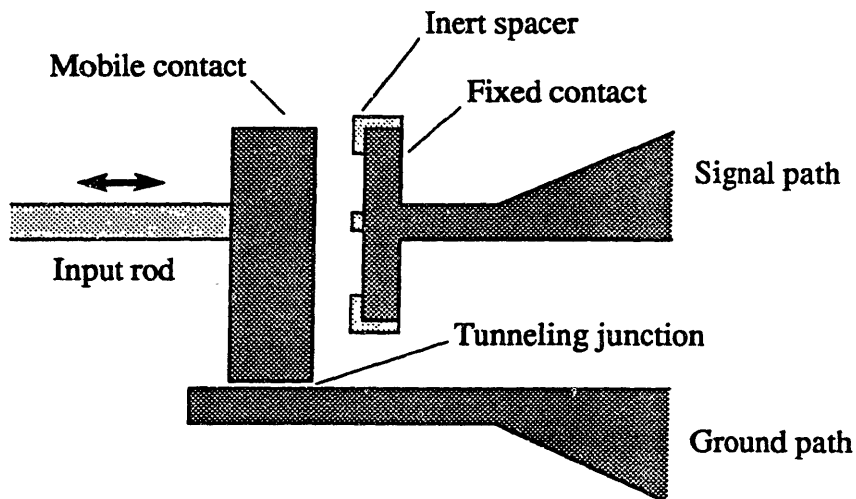


Figure 11.11. Schematic diagram of a modulated tunneling junction for interfacing rod-logic inputs to electrical outputs (moving parts and electrical components only); see text.

Section 11.7.2 concluded that tunneling junction resistances as low as $10^3 \Omega$ should be feasible with junction areas of a few square nanometers. Separation of the surfaces to a distance of a few nanometers will increase the resistance by many orders of magnitude and permit substantial voltages to be supported. At a junction potential of 1 mV, the current in a closed junction would be $1 \mu\text{A}$, and the power dissipation would be 1 nW. Transistors at 300 K can switch with $\sim 10^{-15}$ C of charge, requiring $\sim 10^{-9}$ s of current flow at this rate; accordingly, data can be output at clock rates like those calculated in Section 11.5. Sliding-contact switches (e.g., based on conductors terminated by graphitic planes) might well improve on these characteristics.

11.10. Conclusion

Nanomechanical computational systems can be constructed with logic systems based on sliding rods having switching times of ~ 0.1 ns, with energy dissipation $\ll kT$. Register cells can be constructed that approach the theoretical minimum energy dissipation of $\ln(2)kT$. Logic rods and registers can be joined to build register-to-register combinational logic systems that achieve four register-to-register transfers in ~ 1.2 ns; this performance (together with analogies to existing microelectronic practice) suggests that RISC machines implemented with this technology base can achieve clock speeds of ~ 1 GHz, executing instructions at ~ 1000 MIPS.

A CPU-scale system containing 10^6 transistor-like interlocks (constructed with the parameters used in previous example calculations) can fit within a volume 400 nm on a side. Compatible systems for clocking, power supply, cooling, and I/O have been described and analyzed. Omitting power dissipation resulting from interfaces to conventional microelectronics, the power consumption for a 1 GHz, CPU-scale system is estimated to be ~ 100 nW.

Chapter 12

Molecular manufacturing systems: a survey

12.1. Overview

This chapter presents a brief survey of the application of nanomechanical mechanisms and processes to the task of molecular manufacturing. Previous chapters have shown that nanometer-scale structural components can have a wide range of shapes, with good stiffness (Chapter 9); that nanometer-scale gears, bearings, chain drives, cams, and the like are feasible (Chapter 10); that mechanical and electromechanical components can be combined to build systems of power-driven machinery capable of executing complex, programmed patterns of motion (Chapter 11); and that mechanical manipulation of reactive moieties can be used to direct a broad range of synthetic operations with reliable positional control (Chapter 8). The present chapter considers how these and related capabilities can be combined to perform the acquisition, sorting, and processing of feedstock molecules, leading to the application of reagent moieties to workpieces to build up complex structures.

As outlined in Section 8.3.4.1, molecular manufacturing systems (and sub-systems) can be roughly divided into mill-style systems and manipulator-style systems. These will play roles analogous to familiar components of industrial and biological systems.

Molecular mills will perform repetitive operations with good efficiency, commonly working by means of simple linear and rotary motions. They will be useful in preparing standard reagent moieties for application to workpieces; where production volume is large enough to justify the existence of a complex mechanism specialized for each synthetic operation, they can also perform the final application step. Molecular mills can be viewed as analogous to special-purpose machines (stamping machines, screw-making

machines, transfer lines) in industrial production, or to enzymes in cellular metabolism.

Molecular manipulators will perform complex sequences of motions under programmable control; relative to mills, they will work with lower efficiency (at a given frequency of operation), within the constraint of a lower maximum feasible frequency. They will be useful in making smaller numbers of complex devices, such as the special-purpose molecular mills themselves. Molecular manipulators are analogous to robotic assembly systems in industrial production (indeed, many of the kinematic and design issues are identical), and to numerically-controlled machine tools. They are likewise analogous to ribosomes (which are also numerically-controlled devices), in that they are used to construct relatively complex devices from components prepared by simpler, repetitive operations.

12.2. Molecule acquisition and concentration

The previous chapters have assumed that operations, including mechanosynthetic processes, are conducted in a eutaxic environment. If molecular manufacturing systems are to expand the supply of eutaxic structures, then they must accept disordered feedstocks. A reasonable choice is a mixture of small molecules in solution: a mixture can provide a variety of material inputs, and (if it is kinetically stable yet far from thermodynamic equilibrium) can serve as a source of energy as well. To simplify the discussion, one can assume that the solution is of high purity, including isotopic refinement if necessary (e.g., for the construction of small, high-speed rotors in which mass distribution is critical).

In biological systems, molecules are acquired from a surrounding solution by binding sites, and feasible rates and selectivities of binding are well known from biological examples. The characteristic rate at which an exposed, empty binding site acquires a molecule, k_a , depends on the concentration of molecules of that type in the solution, on their rate of diffusion, and on properties of their interaction with the solvent and the binding site. Typical values (for proteins and small molecules in water) are $k_a \approx 10^8$ to $10^{11} \text{ nm}^3\text{s}^{-1}$ (Creighton 1984). Tens of different small molecular species can be simultaneously present in solution with concentrations $\geq 1 \text{ nm}^{-3}$; accordingly, characteristic times for molecule acquisition can be small compared to 10^{-6} s , which (as in Chapter 8) will be taken as the standard cycle time for processing operations.

Binding can be selective. In biological systems, a binding site exposed to a mixture of similar small molecules will commonly favor binding of a single type of molecule by

one or more (often many) orders of magnitude relative to the others.

A device of the sort illustrated in Figure 12.1 can perform the initial steps in molecule acquisition and concentration. A rotating mechanism can expose a series of binding sites to an external solution, transport them past a barrier, and expose them to an internal cavity. As this occurs, the mechanism (here, using a cam surface driving a set of rods) can modulate the properties of the sites such that they binds with high affinity when exposed to the external medium and moving inward, and with low affinity when exposed to the internal cavity and moving outward. This can effect transport of the molecules from a region of low concentration to one of high concentration, increasing their free energy at the expense of mechanical power. (Indeed, mechanisms of the sort illustrated can serve as pumps transporting molecules from a low-pressure region to a region with pressures in the GPa range.)

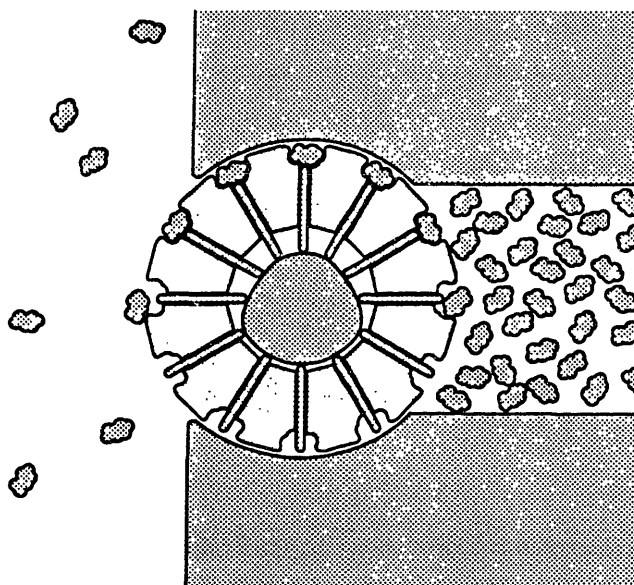


Figure 12.1. Diagram of molecule transport and concentration using a rotating mechanism with modulable binding sites. Transport occurs from left (low concentration) to right (high concentration), given clockwise rotation of the mechanism. An interior, non-rotating cam surface forces oscillating radial motions in the rods, affording a steady stream of high-affinity binding sites traveling inward, with low-affinity (indeed, mechanically obstructed) binding sites traveling outward. An analysis of such mechanisms in terms of time-dependent potential wells shows that this transport process can approach thermodynamic reversibility.

12.3. Molecule sorting

Several complexities can be avoided in designing a molecular manufacturing system if its material inputs can consist of streams of sites, each of which reliably contains a molecule (or moiety) of the correct type. Acquisition and concentration of molecules of a particular kind (with other mechanisms doing likewise for molecules of other kinds) is a step toward this objective. In addition, it must be ensured that (1) no incorrect molecules are present and (2) correct molecules are reliably present. The first task requires sorting molecules, that is, purification.

Numerous macroscopic purification processes work by passing streams of material through a succession of stages, each capable of increasing the concentration of desired materials (and decreasing the concentration of undesired materials) in accord with some separation factor α_{sep} , where

$$\alpha_{\text{sep}} \frac{[A_{\text{in}}]}{[B_{\text{in}}]} = \frac{[A_{\text{out}}]}{[B_{\text{out}}]}$$

and the bracketed quantities represent concentrations of substances A and B in the input and output streams. With an n -stage process of this sort,

$$\alpha_{\text{sep}}^n \frac{[A_{\text{in}}]}{[B_{\text{in}}]} = \frac{[A_{\text{out}}]}{[B_{\text{out}}]}$$

permitting the production of materials with arbitrarily low impurity concentrations, in the absence of a source of impurities in mid-process (semiconductors are routinely produced with impurity levels of 10^{-10} to 10^{-12}). A nanomechanical sorting process can do likewise. In particular, a series of active transport devices like the one illustrated in Figure 12.1 can move molecules selectively and increase their concentration in this manner. The use of selective binding as the basic discrimination mechanism will typically permit $\alpha \gg 1$, and hence the use of a modest number of stages.

12.4. Ensuring that sites are occupied

A molecule sorting process can, by an iterated, statistical process, produce a flow of material along a tube, consisting (with high reliability) of molecules of a single kind. To insert these molecules into a deterministic mechanosynthetic process requires a step in which the occupancy of specific sites is assured with high reliability. This can be

achieved in one of two ways.

First, failure to occupy a site can be made energetically prohibitive. This can be accomplished in any of several ways. Pressures of ~ 12 GPa suffice to make voids of molecular dimensions ($\sim 0.01 \text{ nm}^{-3}$) extremely rare (occurring with a probability $< 10^{-12}$); a vacant site will then have a prohibitive free energy, even in the (hypothetical) absence of a binding interaction. Alternatively, this stage can itself be an exoergic mechanochemical process, forming a bond. Binding energies for processes of this sort can again be large enough to make a vacant site energetically prohibitive. More generally, some combination of pressure, binding interactions, and chemical reaction can serve this function.

In a second approach, a site may have a large but inadequate probability of occupancy on a single exposure. Chapter 11 described mechanisms capable of following alternative sequences of actions based on the results of a mechanical probing operation. Analogous mechanisms can be used to expose a binding site to a liquid if and only if it is empty, and to repeatedly. A series of operations of this sort can decrease the probability of nonoccupancy by a fixed (and large) ratio per cycle, yielding an occupied site at the end of the process with high reliability.

These two approaches to achieving high reliability directly parallel those described in Section 8.3.4.5. Similar remarks apply regarding the intrinsic energy dissipation of these processes.

12.5. Molecule processing

In a molecular manufacturing system, several parallel sequences of mechanochemical processes will transform input molecules of various kinds into a set of reagent moieties ready for application to a workpiece. A typical input molecule might be ethyne; a typical reagent moiety might be a strained alkyne like **8.43**. The intermediate stages in this process might include two radical additions (Sec. 8.5.5 and 8.5) and hydrogen abstractions (Sections 8.5.4 and 8.5.7), followed by the regeneration of the abstraction tools in a process that ends with the elimination of molecular hydrogen from a transition metal (Eq. 8.33). Each of these steps (save the last) requires an encounter between two reactive groups, and each results in the transfer of some number of atoms between one site and other.

To perform molecule processing operations like these, exploiting the mechanochemical capabilities described in Sections 8.3 and 8.5, requires suitable devices. In a typical

stage, a mechanochemical device will apply a substantial force to drive a piezochemical process, while providing substantial stiffness to enable reliable positional control. In complex reactions (e.g., exploiting alkene torsion, Section 8.5.6), a device must also drive several coordinated, properly-sequenced motions. The former, simpler conditions can be met by a system like that illustrated in Figure 12.2, in which reactive moieties are brought together by belts (Sec. 10.7.6), and aligned and compressed by a pair of rollers supported by sleeve bearings (Sec. 10.4). The latter, more complex conditions can be met by elabo-

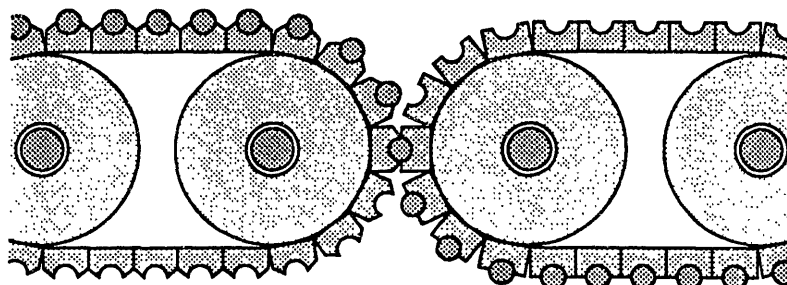


Figure 12.2. Schematic diagram of a roller-driven contact between belts in a mill-style molecule processing system. The stiffness of the roller-axis sleeve bearings, together with (optional) self-aligning contacts between the reagent-site structures, can ensure good alignment of the reagent moieties. Applied forces can be large, but this mechanism permits only brief contact and simple relative motions.

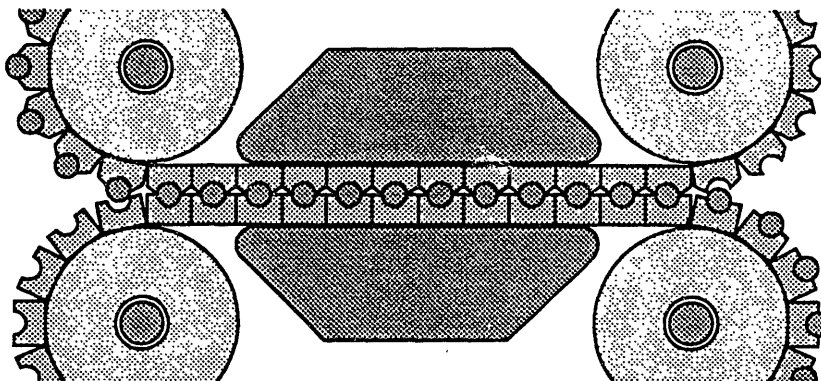


Figure 12.3. Schematic diagram of contact between belts driven by a pair of backing surfaces in a system otherwise like that of Figure 12.2. The prolonged period of contact while a pair of reagent sites moves between the backing surfaces (and auxiliary cam surfaces and rollers) can permit a complex pattern of motion to be imposed on mechanisms built into the reagent sites (e.g., the sequence of compression, torsion, and decompression in the low-dissipation alkene torsion cycle discussed in Sec. 8.5.6).

rating the structure of the reagent-bearing sites to include moving parts actuated by auxiliary rollers and cam surfaces (Chapter 10), and extending the encounter time (if need be) by pressing two belts together as they slide between a pair of stiffly-supported surfaces (Figure 12.3).

The approximate size of a mill-style molecule processing system can be readily estimated. An ample diameter for a typical reagent-bearing site (including surrounding structures that provide a favorable reaction environment and drive simple motions during the reactive encounter) is roughly that of a typical enzyme, ~ 4 nm; this can be taken as the site-to-site separation along the belt. Including devices for molecule acquisition, sorting, and processing (with recycling), a total of ~ 20 rotary mechanisms per kind of output-moiety seems adequate. If these mechanisms have an effective radius of 10 nm (including the thickness of the belt and reagent-bearing sites) and a thickness of 4 nm, the volume of material in moving parts for a single reagent-preparation line is $\sim 2.5 \times 10^4$ nm³. Allowing 100 atoms per cubic nanometer and doubling to allow for supporting structure yields an estimate of $\sim 5 \times 10^6$ atoms. Allowing a factor of ten in volume for packing inefficiencies yields an estimated volume of $\sim 5 \times 10^5$ nm³. The mass-based radiation damage model developed in Section 6.6 indicates that the mean lifetime of a single processing line in Earth ambient background radiation will be $\sim 10^{12}$ s. Assuming that the application of a typical reagent moiety adds ~ 1 atom to a workpiece (some moieties will subtract atoms; some will add several), a typical processing line prepares its own mass in deliverable atoms in ~ 5 s.

If the system has an output of 10^6 finished moieties per second, the belt speeds will be $\sim 3 \times 10^{-3}$ m/s. Using sleeve bearings of radius of 2 nm and length 4 nm, and stiffnesses $k_{s,bear} = 5 \times 10^3$ N/m (stiff enough to add only a trivial compliance in a mechanochemical context), P_{drag} per bearing is $\sim 2.8 \times 10^{-19}$ W, applying Eq. (10.24) with the conservative interfacial parameters of Section 10.4.6.3. With 20 rotary devices and an output of 10^6 s⁻¹, the energy dissipated per product moiety by this mechanism is $\sim 6 \times 10^{-23}$, $\ll kT$. Other energy dissipation mechanisms associated with belt and roller motion appear comparable or smaller. Overall energy dissipation will likely be dominated by phenomena associated with the mechanochemical processes themselves, which can in some instances approach thermodynamic reversibility (Section 8.5.2.2).

12.6. Reagent application

Mechanosynthesis of structures like those discussed in this volume requires the application of precisely-aligned reagent moieties to workpieces in a complex sequence. This can be achieved either by extending the molecular mill mechanism (a hard-automation approach), or by exploiting flexible molecular manipulation (a numerically-controlled machine tool approach).

12.6.1. *Reagent application by molecular mills*

The application of reagents by molecular mills differs little from the steps already described, save in the disparity between the size of the workpiece and the size of the reagent moiety being applied. As a consequence of this disparity, the spacing between reagent-bearing sites in the final stage may need to be larger, to match the increased spacing required by the bulk of the workpiece. (A single reagent-preparation line can accordingly serve several reagent-application mechanisms.)

Construction of complex objects (e.g., computer CPUs) by molecular mills will require a system with approximately as many reagent-application mechanisms as there are atoms in the product. To build a 100 nm scale product containing $\sim 10^8$ atoms would require a mill system tens of microns in diameter, a scale at which fault-tolerant design (Sec. 6.7.2) can become important.

Assuming that the most common mechanochemical operations (i.e., steps in reagent preparation) are responsible for most of the mass of the system, the above estimate of the time required for a system to process its own mass remain approximately correct: allowing for a further factor of two, ~ 10 s. How does this compare to macroscopic manufacturing systems? The magnitude of current economic growth rates suggests that the capital stock of modern industrial systems requires $> 10^7$ s to convert simple inputs into machinery equaling its own mass and complexity. This difference of $\sim 10^6$ in productivity is consistent with scaling laws for mechanical frequencies (Eqs. 2.7, 2.8), and the observation that manufacturing processes (save for photolithography and the like) tend to take an (approximately) constant number of machine cycles per component-feature produced, independent of scale.

12.6.2. *Reagent application by molecular manipulators*

It will frequently be desirable to use a system to build other systems of equal or greater complexity (if this were not done, an infinite regress would make molecular man-

ufacturing impossible). This can be achieved by applying reagent moieties using a multiple-degree-of-freedom positioning mechanism, controlled either by a local nanocomputer or through an interface to more remote macroscopic systems: a single mechanism of this kind can perform an arbitrarily complex sequence of operations.

The design of such positioning mechanisms is familiar in the engineering of industrial robots. With the availability of nanometer-scale digital logic systems, motors, gears, bearings, and so forth, similar designs are feasible here. A novel design constraint in this instance is the maintenance of positional accuracy in the presence of thermal excitation, but this can be minimized by use of either a stiff arm design, alignment of tools to workpieces by local nonbonded contacts immediately before reagent application, or operation at reduced temperatures.

Molecular manipulators will typically execute many internal motions (e.g, rotations of drive shafts and gears, displacements of logic rods) per reagent moiety applied. Accordingly, their peak operating frequencies will be reduced relative to mill-style systems, and their energy dissipation per operation (at a given frequency) will be greater. Preliminary calculations suggest that peak operating frequencies can exceed 10^6 s^{-1} , although at a substantial (many kT per operation) penalty in energy consumption. Productivity on a per-unit-mass basis will likewise be less than for a mill-style system, but still many orders of magnitude greater than that of conventional macroscopic manufacturing systems.

12.7. Larger-scale assembly

In conventional manufacturing, small components are made separately, then assembled into larger systems. A similar strategy will likely be adopted in molecular manufacturing. The previous chapters and sections have discussed mechanisms that can be combined into systems capable of precise, programmable assembly of structures at the molecular level. In scale, this represents the far extreme from macroscopic manufacturing, yet the similar devices are used: gears, bearings, conveyor belts, motors, digital computers, robotic mechanisms, and so forth. Since these devices can be built on intermediate scales, assembly systems can be built to handle components of intermediate size, permitting the construction of systems of indefinitely large scale from eutaxic components.

12.8. Conclusion

Nanomechanical systems can be used to acquire feedstock molecules from solution, sorting and ordering them to provide a flow of input materials to an essentially deterministic molecule-processing system. Exploiting mechanochemical processes (and mechanical power), these systems can convert feedstock molecules into reactive moieties of the sorts discussed in Chapter 8. These moieties, in turn, can be applied to workpieces in complex patterns to build up complex structures, including nanomechanical systems.

The feasibility of molecular manufacturing systems indicates that, given a suitable system of nanomachines, one can build further nanomachines of various kinds. From this conclusion, it is reasonable to infer that, given the ability to build fairly crude nanomachines, one can eventually build systems capable of building better nanomachines. Chemical synthesis and protein engineering represent steps in this direction; Chapter 13 discusses one of several strategies for making further progress.

Part III

Development strategies

Chapter 13

Positional synthesis exploiting AFM mechanisms*

13.0 Abstract

A class of devices based on the atomic force microscope (Binnig and Quate 1986) is proposed which would enable imaging with tips of atomically-defined structure. These molecular tip array (MTA) systems would enable sequential application of tips with differing structures to a single sample, limited to a small substrate area. MTAs with suitable binding sites can enable nanofabrication *via* positional chemical synthesis exploiting local effective concentration enhancements of $\sim 10^8$. A method for canceling or inverting the net van der Waals attraction between a tip and a substrate in a fluid medium is suggested, and a new analysis of imaging forces for proteins is presented.

13.1. Introduction

The lack of reproducible, well-characterized, atomically-sharp tips in standard AFM systems causes difficulties in image resolution, interpretation, and reproducibility. It has been proposed (Drexler and Foster 1990) that molecular tips could both alleviate these problems and provide an approach to achieving positional control of chemical synthesis. Single molecular tips, however, present problems of fabrication, yield, and damage during use.

Molecular tip arrays as proposed here would permit screening and interchange of tips during operation, reducing sensitivity to yield and damage. They would permit sequential use of differing tips in characterizing a sample and would facilitate positional synthesis.

* Adapted from (Drexler 1991).

A cost of this flexibility is limitation to a small substrate size.

The present work draws on previous experimental results in chemistry and atomic force microscopy. In relatively mature areas of engineering, it is common to present the results of a design analysis in order to suggest goals for development and implementation. This analysis is offered in that spirit.

13.2. Tip-array geometry and forces

The proposed MTA geometry is illustrated in Figs. 1 and 2. On a large scale (Fig. 1), a *bead* can be viewed as an AFM tip, imaging molecules attached to a *flat*. On a small scale (Fig. 2), a single molecular tip on the flat images the bead. The terms “substrate”

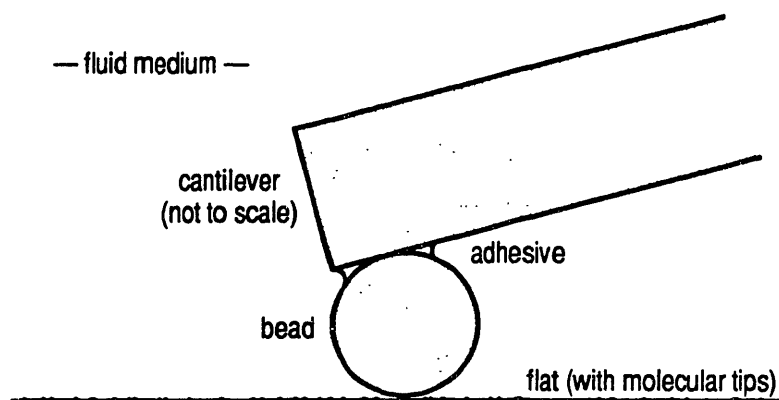


Figure 13.1. Sketch illustrating the relationships among the bead, flat, and cantilever in an AFM using a molecular tip array.

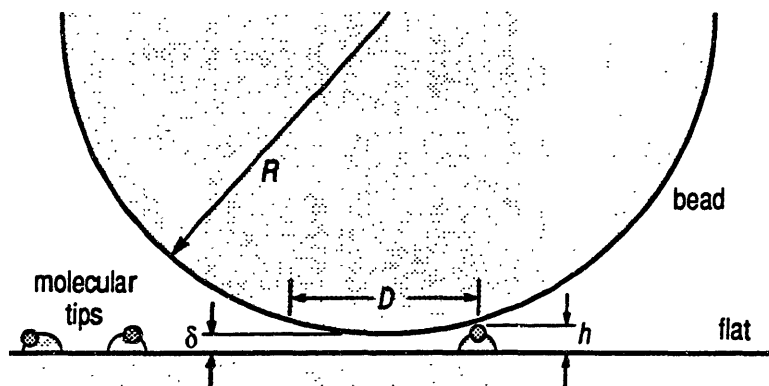


Figure 13.2. Sketch illustrating the relationships among the bead, flat, and molecular tips, including basic geometric parameters.

and “tip” are thus ambiguous, but on the scale of greatest interest, the bead acts as the substrate and a molecule as the tip. Fig. 1 shows the bead on the cantilever side of the AFM mechanism, but this is not essential. Contemplated dimensions (Fig. 2) are $R \approx 100$ nm, $h \approx 4$ nm, and $\delta \approx 3$ nm. Commercial AFMs (Prater, Butt et al. 1990) enable operation in a fluid with atomic resolution and scan ranges of 25 μ ; they seem suitable as basic mechanisms. The following assumes that the fluid medium is an aqueous solution.

AFM imaging using bead-tip forces can proceed if bead-flat forces and stiffnesses are kept small relative to the required imaging forces and stiffnesses. To limit bead-flat interactions, a tip must have a height, h , greater than some minimum imaging separation, δ (discussed below). This permits scanning of a region of diameter D on the bead (Fig. 2). Assuming $R \gg (h - \delta)$,

$$D \approx [8R(h - \delta)]^{1/2}$$

With $R = 100$ nm and $(h - \delta) = 1$ nm, $D \approx 25$ nm. Though small, this area should be adequate for some purposes (e.g., imaging proteins and performing positional synthesis).

Bead-flat interactions result from electrostatic, solvation, and van der Waals forces. Electrostatic interactions between bead and flat can be minimized by ensuring that the surfaces are near electrical neutrality, or by using a solution with an ionic content able to neutralize and screen surface charge. Physiological concentrations of salt yield a Debye length < 0.7 nm, small compared to $\delta \approx 3$ nm. Solvation interactions are strongly distance dependent: in water, the repulsion between hydrophilic surfaces (Israelachvili 1985) falls off exponentially, becoming acceptably small at separations of 1–3 nm. Hydrophobic forces would lead to strong bead-flat adhesion, and are to be avoided. Oscillating solvation forces associated with molecular size effects become small at separations of ~ 2 nm in water (Israelachvili 1985), and will be reduced by surface roughness. Thus, to avoid interference from liquid-structure forces, $\delta \approx 3$ nm will likely prove ample.

Bead-flat van der Waals interactions can be treated using a continuum model based on relationships from (Israelachvili 1985). The force F is related to the radius, R , the Hamaker constant, A , and the bead–flat separation, H

$$F = -AR/6H^2$$

The Hamaker constant for the interaction of insulating materials 1 and 2 across a medium 3 can be calculated from an approximation (Israelachvili 1985) based on the Lifshitz theory:

$$A \approx \frac{3\hbar\omega}{8\sqrt{2}} \frac{(n_1^2 - n_3^2)(n_2^2 - n_3^2)}{(\sqrt{n_1^2 + n_3^2} + \sqrt{n_2^2 + n_3^2})\sqrt{(n_1^2 + n_3^2)(n_2^2 + n_3^2)}} + \frac{3}{4}kT \left(\frac{\epsilon_1 - \epsilon_3}{\epsilon_1 + \epsilon_3} \right) \left(\frac{\epsilon_2 - \epsilon_3}{\epsilon_2 + \epsilon_3} \right)$$

where n is the optical refractive index, ϵ is the zero-frequency dielectric constant, and $\hbar\omega$ is typically $\approx 2 \times 10^{-18}$ J. Given the above device geometry and typical values of the Hamaker constant in liquid media (Israelachvili 1985), bead-flat forces would be attractive and on the order of 0.01 nN at $H = \delta = 3$ nm.

If this proves unacceptable, it can be reduced. The Hamaker constant for interactions between (for example) PTFE and various solids can be adjusted from positive through zero to negative values if the intervening medium is a water solution with a glycerol concentration chosen in the range ~ 15 –65%; see Table I. (These mixtures have a viscosity ~ 1.5 –15 times that of water.) Suitable solids include metals, silicon, mica, optical glasses

Table 13.1. Refractive index and dielectric constant for selected materials, together with the calculated glycerol fraction (in water) required to produce a PTFE/solution/material system with a Hamaker constant, A , equaling zero. No such concentration exists for fused quartz.

Material	n	ϵ	Glycerol fraction
PTFE	1.359	2.1	
Water	1.333	80.0	
Glycerol	1.474	42.5	
PbS	17.4	205.0	0.17
Silicon	3.44	11.7	0.23
BN	2.1	7.1	0.28
High lead glass	1.86	15.0	0.30
Mica	1.60	7.0	0.51
Epoxy	1.58	3.6	0.62
Fused quartz	1.448	3.8	—

of high refractive index, and some polymers, but not quartz. Glycerol has excellent compatibility with biomolecules (living cells tolerate high concentrations); other fluids are also compatible with the construction of systems having a zero Hamaker constant.

PTFE surfaces can be modified to render them hydrophilic or reactive. A low- A material of this sort could be used on either the bead or flat side of the system to minimize interfering van der Waals forces. A net short-range repulsion (whether from van der Waals interactions or steric forces (Israelachvili 1985) from short polymer chains) is desirable to avoid bead-flat adhesion. For $\delta \approx 3$ nm, total bead-flat forces can apparently be limited to < 0.01 nN and stiffnesses to < 0.01 N/m.

The size and shape of the bead are not critical; a sphere is assumed for simplicity. A reasonably smooth surface is desirable, but thermal excitation will roughen amorphous beads formed from a liquid droplet. In a continuum model (Nelson 1989) of thermally-driven interface deformation, the r.m.s. difference in height, σ_Δ , between two points separated by a distance r is

$$\sigma_\Delta \approx [\ln(r/a)kT / \pi\gamma]^{1/2}$$

where γ is the surface energy and a is a characteristic microscopic length. Taking $a = 0.15$ nm (an atomic radius), $r = 15$ nm (see next section), and $\gamma = 18$ mJ/m² (typical of fluorocarbons), $\sigma_\Delta \approx 3.3 \times 10^{-2} T^{1/2}$ nm ≈ 0.75 nm at a formation temperature of 500 K. Substrate corrugations of this magnitude are small compared to many potential specimens (e.g., protein molecules).

13.3. Molecular tips in AFM

Potential tip structures include protein molecules and nanometer-scale particles bearing small adsorbed molecules. (Another approach would use fine particles or surface crystallites alone, to enable imaging with multiple, selected, but non-molecular tips.) A particularly versatile and attractive approach would exploit the broad capabilities of modern organic synthesis and biotechnology by using synthetic ligands as tips and protein molecules as supporting structures. Many natural proteins bind partially-exposed ligands. Ligand analogs could be synthesized with protruding moieties having steric properties suiting them for use as AFM tips. Extensions of monoclonal antibody technology (Huse, Sastry et al. 1989) can rapidly generate proteins able to bind almost any selected small molecule. Single-chain proteins combining antibody V_L and V_H sequences (Bird, Hardman et al. 1988) are compact, ~ 3 nm in height, and lack hinge regions. Use of these

antibody-derived proteins will allow broad freedom in ligand design.

An extensive literature describes the attachment of protein molecules to solid matrices and substrates (Scouten 1987). Potential techniques include direct binding of a protein to a surface (Brash and Horbett 1987) and joining of proteins to surfaces by using crosslinking reagents (Uy and Wold 1977). The use of adapter molecules which bind a surface covalently and a protein noncovalently may prove attractive. Techniques for attaching proteins to surfaces without denaturation but with greater rigidity than that provided by a single covalent tether would be of broad utility in AFM work, but remain to be developed. In the present application, protein engineering techniques can be used to modify exposed protein side chains to facilitate attachment; such modifications are well-tolerated (Bowie, Reidhaar-Olson et al. 1990).

The mechanical stiffness of proteins and ligand complexes is of importance to their use as AFM tips and imaging targets. In protein crystals (which have been taken as a model for bound protein-protein complexes (Finkelstein and Janin 1990)), individual atoms in the protein interior typically experience a 0.03–0.05 nm r.m.s. displacement, σ_{therm} , owing to thermal vibration (Creighton 1984). The atomic-displacement stiffness, k_s , can be derived from the r.m.s. displacement *via* the relationship for a thermally-excited harmonic oscillator

$$\sigma_{\text{therm}} = (kT / k_s)^{1/2}$$

which implies $k_s \approx 1.6\text{--}4.6$ N/m for typical atomic displacements (k_s for surface chains can be far lower). Proteins in crystals are typically anchored to neighbors by a few side chain contacts; comparably stiff attachment of proteins to surfaces seems achievable.

Rigid, polycyclic structures of substantial size can be made by organic synthesis (Webb and Wilcox 1990), and their internal stiffnesses can exceed those of proteins. Such ligands would be anchored with respect to the protein by numerous van der Waals interactions of significant stiffness (Burkert and Allinger 1982), yielding ligand atomic-displacement stiffnesses toward or beyond the upper end of the range characteristic of proteins. The stiffness of the interaction in imaging a protein by a protein/ligand complex will thus be ~ 1 N/m; AFM cantilevers with $k_s < 1$ N/m should display good responsiveness.

Applied forces can destabilize protein folding and ligand binding. The free energy required for unfolding or unbinding (Creighton 1984) is typically $\sim 5\text{--}10 \times 10^{-20}$ J; at the latter energy the unfolding half-life is on the order of 1000 years. From a kinetic perspec-

tive, the destabilizing energy associated with a force is the work it performs as the molecule moves from equilibrium to a transition state for unfolding (where the location of the transition state is affected by the force); estimating this energy requires an estimate of the atomic displacements associated with such a transition state.

In a linear elastic system, a strain energy E_s is associated with a displacement, Δx , related by

$$\Delta x = (2E_s/k_s)^{1/2}$$

Large E_s and low k_s yield a large (adverse) value of Δx . For $E_s = 10 \times 10^{-20}$ J and $k_s = 1.6$ N/m, $\Delta x = 0.35$ nm. Localized displacements of this magnitude (an atomic diameter) might independently be expected to disrupt the tight core packing requisite for stability (Richards 1977). Over this displacement, 0.1 nN performs 3.5×10^{-20} J of work, which is small compared to the energetic differences between more and less stable proteins. Thus, on kinetic stability grounds, this estimate suggests that 0.1 nN forces should be compatible with imaging proteins of moderate stability, and compatible with the use of tips based on proteins incorporating well-bound ligands.

Considering only stiffness and acceptable elastic deformations (taken to be 0.01 nm), and calculating from a continuum model, a maximal tip force of 0.01 nN has been suggested for protein imaging (Persson 1987). Note, however, that with $k_s \approx 1.6$ N/m (a low value from experimental data), a 0.1 nN force would yield $\Delta x \approx 0.06$ nm and $E_s < kT$. These values are compatible with imaging yielding useful structural data.

In AFM systems of standard geometry, repulsive interatomic tip forces (in the presence of net long-range attraction) have been reduced to low values. It was recently suggested that forces as low as 0.01 nN should be within reach of present technology (Weisenhorn, Hansma et al. 1989). This has since been achieved (Hansma 1990). Similarly low forces should be possible in multiple-tip systems, providing a substantial margin of safety when using proteins as tip supports or as imaging targets.

13.4. Imaging with molecular tips

On a large scale, a bead images a flat, showing each molecular tip as a dome. In fine detail, however, each dome is an image of the bead formed by a distinct tip. Maximizing available tips while maintaining nearly-maximal imaging area per tip is achieved with a tip density on the flat on the order of one per disk of diameter D ; for a system with a 25 μ m scan range and $D = 25$ nm, the number of immediately available tips can be $\sim 10^6$.

Tips will almost inevitably have diverse properties. Identical molecules on an isotropic substrate will have diverse azimuthal angles. Substrate irregularities, multiple attachment orientations, and molecular structural differences can all yield diversity in the other two rotational degrees of freedom. A molecular tip array could advantageously include tips with widely differing properties.

Tips can be polar, nonpolar, hydrogen bonding, positively charged, or negatively charged. Information gained by imaging a single molecule with tips of multiple types may help fulfill the goal (Drake, Prater et al. 1989) of determining folded protein structures through AFM and computational modeling. Tips bearing structures with specific biochemical affinities (e.g., enzyme substrate analogs, candidate drug molecule analogs, etc.) would yield data of special biological interest.

Organic synthesis can be used to prepare ligands providing stiff tips of considerable geometric acuity. For example, molecular mechanics parameters (Allinger and Pathiaseril 1987; Burkert and Allinger 1982) indicate that the H atom of $R-C\equiv CH$ will exhibit a transverse bending stiffness of ~ 2 N/m with respect to an sp^3 carbon in the R-group (e.g., in a polycyclic ligand structure); the longitudinal k_s will be ≥ 100 N/m. The effective tip radius for such a probe will be ~ 0.13 nm.

Present tips have variable, poorly characterized structures. Molecular tips will have well-characterized structures but may be of diverse types and orientations. Imaging surfaces with multiple tips should permit discrimination of tip types and orientations. To take a simple example, positive, negative, and neutral tips will have distinct, contrasting responses to a bound charge.

Chemists frequently prepare derivatives of molecules to identify the original structure. Ligands used as molecular tips can include chemically reactive moieties which can be used to map sites of differing reactivity. This mode of operation would typically require prolonged dwell times at a candidate site, as opposed to steady scanning or rapid probing.

13.5. Positional synthesis

Organic synthesis today has a well-developed set of techniques for building molecular structures. Use of maneuverable molecular tips can add a fundamental novelty: flexible, positional control of sequences of synthetic steps (Drexler and Foster 1990). Synthesis would, however, be limited to single-molecule quantities, hence the chief initial product will be information.

MTA systems enable approaches substantially different from those previously proposed (Drexler and Foster 1990). Tip arrays and antibody technologies will enable use of distinct binding sites for each reagent, permitting a series of reactions without cycling the composition of the solution. With rapid, spontaneous dissociation no longer necessary, ligands can be bound tightly; this enables larger effective concentration ratios and hence better site specificity.

Reagent reaction rates are proportional to reagent effective concentrations (by definition). All else being equal, effective concentration is proportional to probability density. Modeling a reagent ligand as a thermally-excited object subject to a force F pressing it against a barrier while constrained by a transverse stiffness k_s (and assuming an additional gaussian jitter in the AFM with a standard deviation $= \sigma_{\text{AFM}}$) yields a probability density having a peak local concentration

$$C_{\text{local}} = \frac{F}{2\pi kT \left(\frac{kT}{k_s} + \sigma_{\text{AFM}}^2 \right)} (1000 N_A)^{-1}$$

(in moles/liter). Neglecting energetic and orientational effects (which can, when favorable, greatly increase the effective concentration), and provided that k_s is substantially less than the transverse stiffness in the reaction transition state (as is the case here), local concentration corresponds to effective concentration. Taking $F = 0.01$ nN, $k_s = 1.6$ N/m, and $\sigma_{\text{AFM}} = 0.01$ nm (displacements of well under 0.01 nm are now routinely measurable in stationary tips (Albrecht 1990), but jitter from ambient vibration is of this order (Drake, Prater et al. 1989)), yields $C_{\text{local}} \approx 240$ M at 300 K. For comparison, mobile surface-residue thiol groups in proteins can exhibit effective concentrations exceeding 100 M; interior residues often exhibit much larger values (Creighton 1984).

Rate constants for ligand binding commonly exceed $10^8 \text{ M}^{-1} \text{ s}^{-1}$. Antibody technologies commonly yield proteins that bind a chosen ligand with equilibrium dissociation constants (Bird, Hardman et al. 1988; Huse, Sastry et al. 1989) on the order of 10^{-9} M. On exposure to a solution with a 10^{-6} M ligand concentration, binding sites with these properties will become occupied in $\sim 10^{-2}$ s, and at equilibrium will remain occupied with probability ~ 0.999 . Under these conditions, the enhancement in effective concentration associated with a reactive moiety on a molecular tip with respect to the background solution concentration is $\sim (100 \text{ M})/(10^{-6} \text{ M}) = 10^8$.

This localized, positionable enhancement in effective concentration can direct site-specific reactions on molecular structures attached to the bead. The transverse probability density distribution is gaussian; with the above parameters, the enhancement falls to background levels at a radius of ~ 0.32 nm, or one atomic diameter. Thus, this mechanism for positional control can produce reaction rate differentials on the order of 10^8 between otherwise chemically equivalent sites separated by two to three bond lengths.

A mature technology of this sort could potentially perform syntheses involving $\sim 10^5$ sequential steps in the presence of ~ 100 chemically equivalent sites with only a modest probability of failure resulting from reactions with uncontrolled reagent molecules in solution. Each reaction step would remove a reagent ligand from a binding site, which would promptly be recharged from ambient ligands in solution. Each reaction step could be tested for success by AFM techniques, to ensure completion. The extension of MTA technology to encompass positional synthesis will require a multidisciplinary collaboration with synthetic organic chemistry playing a prominent role. In the long run, it could provide another approach to the development of general capabilities for molecular manipulation (Drexler 1981).

13.6. Summary

An engineering analysis suggests that molecular tip arrays may have substantial advantages for the intensive study of molecular-scale specimens, including protein molecules. The chief sacrifice is a drastic reduction of the effective substrate area, which is limited to a fraction of the bead diameter. The opportunity to select among $\sim 10^6$ tips, based on imaging results and without interrupting system operation, sidesteps the requirement for high tip yield and high tip stability imposed by single-tip systems. Selection of tips with desirable mechanical and orientational properties from a population comprising well-defined molecular structures will enable more thorough and unambiguous characterization of samples. Use of reactive molecules as tips will enable nanofabrication by positional control of chemical synthesis.

13.7. Acknowledgments

The author thanks T. Albrecht, J. Foster, P. Hansma, T. Kaehler, R. Merkle, and C. Prater for simulating and informative discussions in the course of developing these ideas.

Appendix

Comparison with other work

A.1. Overview

The present volume analyzes the capabilities of systems of molecular machines, including computers and manufacturing systems capable of constructing further molecular machines. There has been an enormous amount of relevant work in chemistry, statistical mechanics, solid-state physics, mechanical engineering, and so forth; that work is cited throughout the previous chapters and makes up the foundation of the present body of design and analysis. Nonetheless, aside from analyses of biological-style mechanisms (i.e., molecules in solution) and distantly-related proposals for molecular electronic systems, there has been a remarkable lack of work that parallels the direction taken here. Accordingly, much of this appendix discusses adjacent fields and why they have taken other directions.

Section A.2 describes how related fields have been divided in their methods and objectives, thereby mapping some of the interdisciplinary crevasses into which promising research topics have fallen. Sections A.3 to A.6 survey the fields of mechanical engineering, microtechnology, chemistry, molecular biology, and protein engineering, describing their relevant work and suggesting reasons why this work has not been extended to embrace theoretical analyses of molecular machinery and manufacturing. Section A.7 discusses recent advances in proximal probe technology that have demonstrated limited forms of molecular manipulation. Finally, Section A.8 quotes and discusses the remarkably foresighted suggestions made by Richard Feynman in 1959.

A.2. How related fields have been divided

Scientific goals vs. technological goals

The construction of systems of molecular machines, like the construction of systems of machines of any sort, is a technological goal. In the U.S. (at least), the study of molecules is chiefly taught as chemistry, which in turn is defined to be a natural science. As integrated circuit design is to solid state physics, and as mechanical engineering is to classical mechanics, so molecular engineering is to chemistry: distinct, but intimately related. It would be remarkable if professional solid state physicists produced computer designs, or if professional physicists produced engine designs: scientific goals differ from technological goals, as do the methods used in their pursuit.

Top-down vs. bottom-up approaches

Engineering systems today range from macroscopic to microscopic, with active research aimed at building electronic and mechanical systems on ever smaller scales. The micron scale, however, is a billion times larger (volumetrically) than the nanometer scale, and existing microtechnologies provide no mechanism whereby one may even attempt to gain precise, molecular control of the surface and interior of a complex, three-dimensional structure. A large gap has separated top-down research in microtechnology (large, complex, and imprecise) from the bottom-up path of chemistry (small, simple, and exact). In one, the challenge is to make imprecise structures smaller; in the other, to make precise structures larger. No engineering discipline has yet formed around the latter goal.

Immediate goals vs. long-term prospects

Science and technology are united by a focus on what can be made and tested in the immediate future. The molecular sciences are centered around the laboratory, and hence around current capabilities. Engineering is centered around the workshop, again binding creative thought to current capabilities. In most sciences, theoretical work receives relatively little effort (the branches of physics are the great exceptions). The study of future prospects in technology, even when they can be adequately modeled using present scientific knowledge, falls outside the usual scope of physics, of laboratory science, and of practical engineering.

A.3. Mechanical engineering and microtechnology

Progress in miniaturization has been viewed as an incremental, top-down process. From this perspective, working at the molecular size scale has been seen as a long-term goal, if it has been considered at all. The implications of precise, molecular control of mechanical structures at the nanometer scale have never been on the research agenda in mechanical engineering or microtechnology. They have accordingly have received no significant attention.

A.4. Chemistry

Chemistry is usually regarded as a laboratory-centered natural science (note, however, that the Department of Synthetic Chemistry at the University of Tokyo is part of the Faculty of Engineering). Lacking any means to synthesize complex molecular machines, and lacking an engineering tradition urging them in that direction, chemists have never had the design and analysis of systems of molecular machinery on their research agenda. Nonetheless, in several areas, chemical research has moved far enough toward the construction of complex molecular systems to merit specific discussion (though the following is far from an exhaustive review).

The early 1980s saw a wave of interest in molecular electronics; funding has since dwindled in the U.S., but remains strong in Europe and Japan. Most research in molecular electronics has focused on the development of molecules that exhibit useful electronic properties when present in microscale aggregates; some proposals, however, have focused on the construction of computational devices in which individual moieties would serve as signal carrying and switching elements (e.g., various papers in (Carter 1982; Carter 1987)). These proposals have suggested various combinations of chemical synthesis, protein engineering, and DNA engineering to make self-assembling systems on a broadly biological model (e.g., (Robinson and Seeman 1987)). This would (if successfully pursued) constitute molecular systems engineering, though not of machines or manufacturing systems.

In his Nobel lecture (Lehn 1988), J.-M. Lehn stated that:

“Components and molecular devices such as molecular wires, channels, resistors, rectifiers, diodes, and photosensitive elements might be assembled into nanocircuits and combined with organized polymolecular assemblies to yield systems capable ultimately

of performing functions of detection, storage, processing, amplification, and transfer of signals and information by means of various mediators (photons, electrons, protons, metal cations, anions, molecules) with coupling and regulation.”

This paragraph closes with a list of references, of which the two earliest are an address by J.-M. Lehn (*Leçon Inaugural*, Collège de France, Paris 1980) and (Drexler 1981).

On a smaller size scale (and with demonstrated experimental successes), chemists have made a variety of molecular devices including sophisticated reagents and catalysts, self-assembling sets of molecules (Cram 1986; Cram 1988; Diederich 1988; Lehn 1988; Rebek 1987), and a molecule that catalyzes the synthesis of copies of itself (Tjivikua, Ballester et al. 1990). Recent years have seen the development of a variety of interesting molecular systems including rod-like structures bearing reactive functional groups at the ends and components that join covalently in steps described as “structure-directed synthesis” (Ashton, Isaacs et al. 1989). Efforts of this sort have on occasion been described as steps toward developing “molecular Tinkertoys,” or “molecular Lego.” These descriptions, however, have not yet been supported by the requisite system-level analysis: with Tinkertoys and Lego blocks, well-known means (e.g., children) are available for joining the components to form complex patterns corresponding to functional designs, but no such means have been described in connection with these molecular structures. Indeed, there appears to have been no systematic exposition in the chemical literature of the requirements for the design and construction of *complex* self-assembling systems, nor of the requirements that must be met if they are to be used in building molecular machines and manufacturing systems.

A.5. Molecular biology

Molecular biologists study and modify systems of molecular machines; genetic engineers reprogram them, sometimes to build novel multi-nanometer-scale molecular objects with complex functions. Molecular biologists, however, work within the traditions of natural science, which are not conducive to the design and construction of engineering systems. Although the Hotani Molecular Dynamic Assembly Project in Japan has studied the bacterial flagellar motor with the objective of building analogous devices through reverse engineering, U.S. researchers have shown little interest in engineering molecular systems in this fashion. Aside from scattered observations at the level of “these molecules might

be useful for something,” there have apparently been no discussions predating (Drexler 1981) suggesting that devices resembling biomolecular motors, actuators, and structural components be combined to build engineering systems. Further, there has been little visible follow-up on these technological suggestions by the scientists in the U.S. biological community. (Proposals to imitate life, coupling catalytic and regulatory molecules by diffusive transport, are in a distinct category of no direct relevance.)

The two chief exceptions to this generalization have been work in engineering protein molecules (typically serving as isolated catalysts in solution or immobilized on surfaces), and work on engineering three-dimensional structures from branched DNA (Robinson and Seeman 1987; Seeman 1982). The use of complementary DNA sequences in the latter work provides one answer to the question of how to form complex patterns from self-assembling molecular objects, and proposals have been advanced for the application of this work to molecular electronics and to mechanical devices.

A.6. Protein engineering

Protein molecules constitute much of the molecular machinery found in living systems, and “protein engineering” has in the last decade become a substantial area of research. It has been argued that protein engineering can provide a path to the development of molecular manufacturing (Drexler 1981), if it were pursued with the objective of constructing self-assembling systems of molecular machines.

The “Instructions to Authors” of the journal *Protein Engineering* give a sense of how the field has developed. Where the MIT School of Engineering (in the MIT Bulletin) defines engineering as “a creative profession concerned with developing and applying scientific knowledge and technology to meet societal needs,” *Protein Engineering* states that “The objectives of those engaged in this area of research are to investigate the principles by which particular structural features in proteins relate to the mechanisms through which biological function is expressed, and to test these principles in an empirical fashion by introduction of specific changes followed by evaluation of any altered structural and/or functional properties.” In short, the stated objective of “protein engineering” is the pursuit of scientific knowledge relevant to biology, rather than the construction of useful new proteins.

This pattern is illustrated by the response to a proposal in (Drexler 1981), which argued that protein engineering could be performed *without* solving the protein-folding problem as understood by biologists, because the engineering objective is to find a

sequence having a given fold, rather than to find the fold of a given sequence. (This proposal has been confirmed by the success reported in (DeGrado, Regan et al. 1987).) After an initial round of citations in *Science* and *Nature* (Pabo 1983; Ulmer 1983), including some discussion of protein engineering for practical ends, subsequent citations of this proposal have described it as defining an “inverse protein folding problem ” (Bowie, Lüthy et al. 1991; Ponder and Richards 1987) and have explored how solutions to this inverse problem can be used to determine how a given sequence will fold (if it in fact folds in a familiar pattern). In short, an insight aimed at engineering objectives has been used as a basis for approaching the original, scientific problem.

Despite this orientation, progress has been made in the engineering of novel objects from proteins. Outstanding examples include what is generally regarded as the first *de novo* structure (DeGrado, Regan et al. 1987), and the development of synthetic, branched protein-like structures that depart substantially from biological models (Mutter, Altmann et al. 1988).

A.7. Proximal probe technologies

The new proximal probe technologies—scanning tunneling microscopes (STMs), atomic force microscopes (AFMs) and their variations—provide a means for positioning and maneuvering tips near surfaces with atomic precision. The possibility of modifying surfaces with these tips is immediately evident. Suggestions for doing so have appeared (Drexler 1986; Farrell and Levinson 1985), and experiments have demonstrated (*inter alia*) atomic-scale surface modifications on germanium (Becker, Golovchenko et al. 1987), pinning of organic molecules to graphite (Foster, Frommer et al. 1988), arrangement of 35 xenon atoms to spell “IBM” on a nickel surface (Eigler and Schweizer 1990) (subsequent work in the same laboratory has demonstrated reliable positioning of carbon monoxide on platinum), and as-yet unreliable removal and deposition of single silicon atoms on a silicon surface (Lyo and Avouris 1991). Much of this work has been conducted at IBM, where the objective of developing atomic-scale mechanisms for data storage and processing has been explicitly articulated (although there appear to be no published designs or analyses of proposed nanometer-scale mechanisms in this connection).

Chapter 13 of this work (Drexler 1991) formed 4 out of the 1007 pages of the Proceedings of the Fifth International Conference on Scanning Tunneling Microscopy/Spectroscopy and First International Conference on Nanometer Scale Science and

Technology, and is the only paper which presents an analysis of a proposed technique for nanofabrication (several papers report on laboratory results). Indeed, most papers reported laboratory results, a minority presented or reviewed theory regarding the physics responsible for those results, and only one other paper made and analyzed a design proposal (for a magnetometer).

A.8. Feynman's 1959 talk

The nearly-perfect absence of design and analysis of nanomachines and molecular manufacturing is all the more remarkable in light of the clear suggestions made by R. Feynman in a talk, "There's Plenty of Room at the Bottom," given at the annual meeting of the American Physical Society in 1959 (Feynman 1961). The body of the talk focuses on miniaturization and microtechnology; this section anticipates capabilities like those that are now basic to the microelectronics industry and proposes an alternative approach to miniaturization (using machines to build smaller machines, which build still smaller machines, and so forth) that has not, in fact, been followed. Near the end of this talk are four paragraphs that form the main prior theoretical work in molecular manufacturing:

"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 a 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 he has got that ring, so he mixes this and that, and 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.”

These remarks clearly point in the direction explored in the present body of work. Why was there little or no follow-up in the literature, even by Feynman? Presumably because these long-term goals appeared to have no near-term consequences: they neither defined a scientific problem nor suggested an engineering project. The work leading to the present volume (which was inspired in 1977 by examples from chemistry and molecular biology) has been motivated chiefly by a concern with understanding future technological possibilities, a discipline which falls between science and engineering and which even now has few serious practitioners.

A.9. Conclusion

Prior work in physics, chemistry, molecular biology, and engineering forms the essential foundation of the design and analysis presented in the previous chapters. Nonetheless, aside from Feynman’s 1959 remarks and preliminary work by the author, the literature appears to contain no substantial examination of the prospects for mechanical engineering and manufacturing in the molecular domain. This gap in the literature can be understood as a result of short-term perspectives in science and technology (reinforced by the strong laboratory orientation of the molecular sciences), and as a reflection of the gap between science and technology. In short, molecular scientists have had the required physical knowledge, and engineers have had the required design methodologies, but these groups have been distinct. There has been no discipline that takes molecular mechanical engineering as its subject matter, and so there has been a dearth of research.

In light of the enormous technological potential of molecular machinery and manufacturing (including nanoelectronic products of molecular manufacturing processes), and in light of the emerging capabilities in chemistry, protein engineering, and proximal probe technologies, the time appears ripe for the vigorous development of systems engineering in the molecular domain. This effort will draw on the knowledge and skills of disciplines as diverse as chemistry, mechanical engineering, and computer science; its

products will contribute to still more diverse fields of science and technology. This volume draws on fundamental principles from several disciplines, attempting to assemble a set of conceptual and mathematical tools adequate to undertake the required design and analysis. If it has helped to show the attractions of this field, while helping those who wish to contribute, then it has achieved its major objective.

References

- Adam, W., P. Hössel, W. Hümmer and H. Platsch. (1987). "UV-Laser Photochemistry: Diffusion-Controlled Trapping of Cyclic 1,3-Diradicals by Molecular Oxygen. Conformation Effects on Triplet Lifetimes." *J. Am. Chem. Soc.* **109**: 7570–7572.
- Albrecht. (1990). Comments on atomic force microscopy, personal communication.
- Allinger, N. L. (1977). "MM2. A Hydrocarbon Force Field Utilizing V1 and V2 Torsional Terms." *J. Am. Chem. Soc.* **99**(25): 8127–8134.
- Allinger, N. L. (1986). *MM2 and MMP2 parameters*.
- Allinger, N. L. and A. Pathiaseril. (1987). "A Force Field for Allenes and for Nonlinear Acetylenes within the MM2 Approximation." *J. Comp. Chem.* **8**(8): 1225–1231.
- Allinger, N. L., Y. H. Yuh and J. Lii. (1989). "Molecular Mechanics. The MM3 Force Field for Hydrocarbons. 1." *J. Am. Chem. Soc.* **111**(23): 8551–8566.
- Alshits, V. I. and L. V. Indenbom. (1986). Mechanisms of Dislocation Drag. *Dislocations in Solids*. Amsterdam, North-Holland.
- Amdur, I. and J. E. Jordan. (1966). Elastic Scattering of High-Energy Beams: Repulsive Forces. *Molecular Beams*. New York, John Wiley and Sons.
- Arnaud, R. and R. Subra. (1982). "Substituent Effects and Radical Reactivity: Addition of *p*-Substituted Phenyl Radicals to Ethylene." *Nou. J. Chim.* **6**: 91–95.

- Ashcroft, N. W. and N. D. Mermin. (1976). *Solid State Physics*. New York, Holt, Rinehart and Winston.
- Ashton, P. R., N. S. Isaacs, F. H. Kohnke, J. P. Mathias and J. F. Stoddart. (1989). "Stereoregular Oligomerization by Repetitive Diels–Alder Reactions." *Angew. Chem. Int. Ed.* **28**: 1258–1261.
- Atkins, P. W. (1970). *Molecular Quantum Mechanics*. Oxford, Clarendon Press.
- Atkins, P. W. (1974). *Quanta: a handbook of concepts*. Oxford, Oxford University Press.
- Badziag, P. and W. S. Verwoerd. (1987). "MNDO Analysis of the Oxidised Diamond (100) Surface." *Surf. Sci.* **183**: 469–483.
- Bates, R. B. and C. A. Ogle. (1983). *Carbanion Chemistry*. Berlin, Springer-Verlag.
- Beauregard, G. and M. Potier. (1985). "Temperature Dependence of the Radiation Inactivation of Proteins." *Anal. Biochem.* **150**: 117–120.
- Becker, R. S., J. A. Golovchenko and B. S. Swartzentruber. (1987). "Atomic-scale surface modifications using a tunnelling microscope." *Nature*. **325**: 419–421.
- Bell, R. P. (1959). "The Tunnel Effect Correction for Parabolic Potential Barriers." *Trans. Farad. Soc.* **55**: 1–4.
- Bennett, C. H. (1982). "The Thermodynamics of Computation—a Review." *Int. J. Theor. Phys.* **21**(12): 905–940.
- Bérces, T. and F. Márta. (1988). Activation Energies for Metathesis Reactions of Radicals. *Chemical Kinetics of Small Organic Radicals*. Boca Raton, CRC Press.
- Bigeleisen, J. (1949). "The Relative Reaction Velocities of Isotopic Molecules." *J. Chem. Phys.* **17**: 675–678.

- Binnig, G. and C. F. Quate. (1986). "Atomic Force Microscope." *Phys. Rev. Letters*. **56** (9): 930–933.
- Bird, R. E., K. D. Hardman, J. W. Jacobson, S. Johnson, B. M. Kaufman, S. Lee, T. Lee, S. H. Pope, G. S. Riordan and M. Whitlow. (1988). "Single-Chain Antigen-Binding Proteins." *Science*. **242**: 423–426.
- Bockris, J. O. and A. K. N. Reddy. (1970). *Modern Electrochemistry*. New York, Plenum.
- Bowie, J. U., J. F. Reidhaar-Olson, W. A. Lim and R. T. Sauer. (1990). "Deciphering the Message in Protein Sequences: Tolerance to Amino Acid Substitutions." *Science*. **247**: 1306–1310.
- Bowie, J. U., R. Lüthy and D. Eisenberg. (1991). "A Method to Identify Protein Sequences That Fold into a Known Three-Dimensional Structure." *Science*. **253**: 164–170.
- Brash, J. L. and T. A. Horbett. (1987). *Proteins at Interfaces*. ACS Symposium Series.
- Brooke, J. H. (1985). *Organic Chemistry. Recent Developments in the History of Chemistry*. London, Royal Society of Chemistry.
- Brookes, C. A., V. R. Howes and A. R. Parry. (1988). "Multiple slip in diamond due to a nominal contact pressure of 10 GPa at 1,000 °C." *Nature*. **332**: 139–141.
- Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus. (1983). "CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations." *J. Comp. Chem.* **4**: 187–217.
- Brown, F. B. and D. G. Truhlar. (1985). "Dissociation Potential for Breaking a C–H Bond in Methane." *Chem. Phys. Lett.* **113**(5): 441–446.

- Brown, R. F. C. (1980). *Pyrolytic Methods in Organic Chemistry*. Organic Chemistry. New York, Academic Press.
- Burkert, U. and N. L. Allinger. (1982). *Molecular Mechanics*. ACS Monograph 177. Washington, DC, American Chemical Society.
- Candlin, J. P., K. A. Taylor and D. T. Thompson. (1968). *Reactions of Transition-Metal Complexes*. Amsterdam, Elsevier.
- Carey, F. A. and R. J. Sundberg. (1983a). *Advanced Organic Chemistry, Part A*. New York, Plenum Press.
- Carey, F. A. and R. J. Sundberg. (1983b). *Advanced Organic Chemistry, Part B*. New York, Plenum Press.
- Carter, F. L. (1982). *Molecular Electronic Devices*. 386.
- Carter, F. L. (1987). *Molecular Electronic Devices II*. 825.
- Caruthers, M. H. (1985). "Gene Synthesis Machines: DNA Chemistry and Its Uses." *Science*. **230**: 281–285.
- Casale, A. and R. S. Porter. (1978). *Polymer Stress Reactions*. New York, Academic Press.
- Ceperley, D. and B. Alder. (1986). "Quantum Monte Carlo." *Science*. **231**: 555–560.
- Clark, T. (1985). *A Handbook of Computational Chemistry*. New York, John Wiley and Sons.
- Corbridge, D. E. C. (1990). *Phosphorous: An Outline of its Chemistry, Biochemistry, and Technology*. Studies in Inorganic Chemistry. Amsterdam, Elsevier.

- Corey, E. J. and X. Cheng. (1989). *The Logic of Chemical Synthesis*. New York, John Wiley & Sons.
- Cowan, D. O. and R. L. Drisko. (1976). *Elements of Organic Photochemistry*. New York, Plenum Press.
- Crabtree, R. H. (1987). *The Organometallic Chemistry of the Transition Metals*. New York, John Wiley & Sons.
- Cram, D. J. (1986). "Preorganization—From Solvents to Spherands." *Angew. Chem. internat. Edit.* **25**(12): 1039–1134.
- Cram, D. J. (1988). "The Design of Molecular Hosts, Guests, and Their Complexes." *Science*. **240**: 760–767.
- Creighton, T. E. (1984). *Proteins*. New York, W. H. Freeman and Company.
- Critchley, J. P., G. J. Knight and W. W. Wright. (1983). *Heat Resistant Polymers*. New York, Plenum Press.
- Davis, J. H., W. A. Goddard III and L. B. Harding. (1977). "Theoretical Studies of the Low-Lying States of Vinylidene." *J. Am. Chem. Soc.* **99**: 2919–2925.
- DeGrado, W. F., L. Regan and S. P. Ho. (1987). The Design of a Four-helix Bundle Protein. *Cold Spring Harbor Symposia on Quantitative Biology, Volume 52*. Cold Spring Harbor Laboratory.
- Diederich, F. (1988). "Complexation of Neutral Molecules by Cyclophane Hosts." *Angew. Chem. internat. Edit.* **27**: 362–386.
- Diederich, F., Y. Rubin, C. B. Knobler, R. L. Whetten, K. E. Schriver, K. N. Houk and Y. Li. (1989). "All-Carbon Molecules: Evidence for the Generation of Cyclo[18]carbon from a Stable Organic Precursor." *Science*. **245**: 1088–1090.

- Dötz, K. H., H. Fischer, P. Hofmann, F. R. Kreissl, U. Schubert and K. Weiss. (1983). *Transition Metal Carbene Complexes*. Weinheim, Verlag Chemie.
- Drake, B., C. B. Prater, A. L. Weisenhorn, S. A. C. Gould, T. R. Albrecht, C. F. Quate, D. S. Cannell, H. G. Hansma and P. K. Hansma. (1989). "Imaging Crystals, Polymers, and Processes in Water with the Atomic Force Microscope." *Science*. **243**: 1586–1589.
- Drexler, K. E. (1981). "Molecular engineering: An approach to the development of general capabilities for molecular manipulation." *Proc. natnl. Acad. Sci. U.S.A.* **78**(9): 5275–5278.
- Drexler, K. E. (1986). *Engines of Creation*. New York, Anchor Press/Doubleday.
- Drexler, K. E. (1987). *Nanomachinery: Atomically precise gears and bearings*. IEEE Micro Robots and Teleoperators Workshop.
- Drexler, K. E. (1988). Rod logic and thermal noise in the mechanical nanocomputer. *Molecular Electronic Devices*. Amsterdam, North-Holland.
- Drexler, K. E. (1991). "Molecular tip arrays for molecular imaging and nanofabrication." *JVST-B*. **9**(2): 1394–1397.
- Drexler, K. E. and J. S. Foster. (1990). "Synthetic tips." *Nature*. **343**: 600.
- Eaton, P. E. and G. Castaldi. (1985). "Systematic Substitution on the Cubane Nucleus. Amide Activation for Metalation of "Saturated" Systems." *J. Am. Chem. Soc.* **107**: 724–726.
- Eaton, P. E. and M. Maggini. (1988). "Cubene (1,2-Dehydrocubane)." *J. Am. Chem. Soc.* **110**: 7230–7332.
- Eggers, D. F., Jr., N. W. Gregory, G. D. Halsey Jr. and B. S. Rabinovitch. (1964). *Physical Chemistry*. New York, John Wiley and Sons.

- Eigler, D. M. and E. K. Schweizer. (1990). "Positioning single atoms with a scanning tunnelling microscope." *Nature*. **344**: 524–526.
- Eisenthal, K. B., R. A. Moss and N. J. Turro. (1984). "Divalent Carbon Intermediates: Laser Photolysis and Spectroscopy." *Science*. **225**: 1439–1445.
- Farrall, G. A. (1980). Electrical Breakdown in Vacuum. *Vacuum Arcs: Theory and Application*. New York, John Wiley & Sons.
- Farrell, H. H. and M. Levinson. (1985). "Scanning tunneling microscope as a structure-modifying tool." *Phys. Rev. B*. **31**: 3593–3598.
- Feynman, R. (1961). There's Plenty of Room at the Bottom. *Miniaturization*. New York, Reinhold.
- Feynman, R. P. (1985). "Quantum Mechanical Computers." *Optics News*. **11**(2): 11–20.
- Field, J. E. (1979). Strength and Fracture Properties of Diamond. *The Properties of Diamond*. London, Academic Press.
- Finkelstein, A. V. and J. Janin. (1990). "The price of lost freedom: entropy of biomolecular complex formation." *Prot. Eng.* **3**: 1–3.
- Fogarasi, G. and J. E. Boggs. (1983). "Theoretical equilibrium geometry, vibrational frequencies and the first electronic transition energy of HCC." *Mol. Phys.* **50**(1): 139–151.
- Foster, J. S., J. E. Frommer and P. C. Arnett. (1988). "Molecular manipulation using a tunnelling microscope." *Nature*. **331**: 324–326.
- Fredkin, E. and T. Toffoli. (1982). "Conservative Logic." *Int. J. Theor. Phys.* **21**(3/4): 219–253.

- Frenklach, M. and K. E. Spear. (1988). "Growth Mechanism of Vapor-Deposited Diamond." *J. Mater. Res.* 3(1): 133–140.
- Gimzewski, J. K. (1988). Scanning tunneling microscopic techniques applied to roughness of silver surfaces. *Surface Measurement and Characterization*. Bellingham, Washington, Society of Photo-Optical Instrumentation Engineers.
- Goodman, F. O. (1980). "Thermal Accomodation Coefficients." *J. Phys. Chem.* 84: 1431–1445.
- Goodman, F. O. and H. Y. Wachman. (1976). *Dynamics of Gas-Surface Scattering*. New York, Academic Press.
- Gray, D. E. (1972). *American Institute of Physics Handbook*.
- Hagler, A. T., J. R. Maple, T. S. Thacher, G. B. Fitzgerald and U. Dinur. (1989). Potential energy functions for organic and biomolecular systems. *Computer Simulations of Biomolecular Systems*. Leiden, Escom Science Publishers.
- Hamza, A. V., G. D. Kubiak and R. H. Stulen. (1988). "The role of hydrogen on the diamond C(111)-(2 × 1) reconstruction." *Surf. Sci.* 206: L833–L844.
- Hansma, P. K. (1990). Comments on atomic force microscopy, personal communication.
- Hehre, W. J., L. Radom, P. v. R. Schleyer and J. A. Pople. (1986). *Ab Initio Molecular Orbital Theory*. New York, John Wiley & Sons.
- Heinicke, G. (1984). *Tribochemistry*. Berlin, Akademie-Verlag.
- Hernley, R. J. and H. K. Mao. (1990). "Critical Behavior in the Hydrogen Insulator-Metal Transition." *Science* . 249: 391–393.
- Hendrickson, J. B. (1990). "Organic Synthesis in the Age of Computers." *Angew. Chem. Int. Ed.* 29: 1286–1295.

- Hiemenz, P. C. (1986). *Principles of Colloid and Surface Chemistry*. New York, Marcel Dekker.
- Huang, D., M. Frenklach and M. Maroncelli. (1988). "Energetics of Acetylene-Addition Mechanism of Diamond Growth." *J. Phys. Chem.* **92**: 6379–6381.
- Hudson, J. A. (1980). *The Excitation and Propagation of Elastic Waves*. Cambridge, Cambridge University Press.
- Huheey, J. E. (1978). *Inorganic Chemistry*. New York, Harper and Row.
- Huisgen, R. (1977). "Tetracyanoethylene and Enol Ethers. A Model for 2 + 2 Cycloadditions via Zwitterionic Intermediates." *Acc. Chem. Res.* **10**: 117–124.
- Huse, W. D., L. Sastry, S. A. Iverson, A. S. Kang, M. Alting-Mees, D. R. Burton, S. J. Benkovic and R. A. Lerner. (1989). "Generation of a Large Combinatorial Library of the Immuglobulin Repertoire in Phage Lambda." *Science*. **246**: 1275–1281.
- Ichikawa, H., Y. Ebisawa and A. Shigihara. (1985). "Potential Energies of Rotation of Double Bond in Ethylene Molecule and Ion." *Bull. Chem. Soc. Jpn.* **58**: 3619–3620.
- Isaacs, N. S. and A. V. George. (1987). "Chemical synthesis at high pressures." *Chem. Br.* **23**: 47–51.
- Israelachvili, J. N. (1985). *Intermolecular and Surface Forces*. New York, Academic Press.
- Israelachvili, J. N., P. M. McGuiggan and A. M. Homola. (1988). "Dynamic Properties of Molecularly Thin Liquid Films." *Science*. **240**: 189–191.
- Jenner, G. (1985). "The Pressure Effect on Strained Transition States." *J. Chem Soc., Faraday Trans. 1*. **81**: 2437–2460.

- Johnson, M. D. (1983). "Bimolecular Homolytic Displacement of Transition-Metal Complexes from Carbon." *Acc. Chem. Res.* **16**: 343–349.
- Karplus, M. and R. N. Porter. (1970). *Atoms and Molecules*. Menlo Park, California, W. A. Benjamin, Inc.
- Kauzman, W. and H. Eyring. (1940). *J. Am. Chem. Soc.* **62**: 3113.
- Kelly, A. (1973). *Strong Solids*. Oxford, Clarendon Press.
- Kent, S. B. H. (1988). "Chemical Synthesis of Peptides and Proteins." *Ann. Rev. Biochem.* **57**: 957–989.
- Kepner, G. R. and R. I. Macey. (1968). "Membrane Enzyme Systems: Molecular Size Determination from Radiation Inactivation." *Biochim. Biophys. Acta.* **163**: 188–203.
- Kerr, J. A. (1973). Rate Processes in the Gas Phase. *Free Radicals*. New York, John Wiley & Sons.
- Kerr, J. A. (1990). Strengths of Chemical Bonds. *CRC Handbook of Chemistry and Physics*. Boca Raton, CRC Press.
- Keyes, R. W. (1985). "What Makes a Good Computer Device?" *Science*. **230**: 138–144.
- Keyes, R. W. (1989). "Physics of digital devices." *Rev. Modern Phys.* **61**(2): 279–287.
- Knight, T. F., Jr. (1991). Personal communication.
- Knox, J. H. (1971). *Molecular Thermodynamics: An Introduction to Statistical Mechanics for Chemists*. New York, Wiley–Interscience.
- Kogan, V. I. and Galitskiy. (1963). *Problems in Quantum Mechanics*. New York, Prentice-Hall.

- Kubiak, G. D. and K. W. Kolasinsky. (1989). "Normally unoccupied states on C(111) (diamond) (2×1): Support for a relaxed π -bonded chain model." *Phys. Rev. B.* **39** (2): 1381–1384.
- Kuntz, P. J., E. M. Nemeth, J. C. Polanyi, S. D. Rosner and C. E. Young. (1966). "Energy Distribution Among Products of Exothermic Reactions. II. Repulsive, Mixed, and Attractive Energy Release." *J. Chem. Phys.* **44**: 1168–1184.
- Kyriacou, D. K. (1981). *Basics of Electroorganic Synthesis*. New York, John Wiley & Sons.
- Lakes, R. (1987). "Foam Structures with Negative Poisson's Ratio." *Science*. **235**: 1038–1040.
- Landauer, R. (1961). "Irreversibility and Heat Generation in the Computing Process." *IBM J. Res. Dev.* **3**: 183–191.
- Landauer, R. (1988). "Dissipation and noise immunity in computation and communication." *Nature*. **335**: 779–784.
- Landman, U., W. D. Luedtke, N. A. Burnham and R. J. Colton. (1990). "Atomistic Mechanisms and Dynamics of Adhesion, Nanoindentation, and Fracture." *Science*. **248**: 454–461.
- Lea, D. (1946). *Actions of Radiation on Living Cells*. London, Cambridge University Press.
- Lee, T. J., A. Bunge and H. F. Schaefer III. (1985). "Toward the Laboratory Identification of Cyclopropenylidene." *J. Am. Chem. Soc.* **107**: 137–142.
- Lehn, J. (1988). "Supramolecular Chemistry—Scope and Perspectives: Molecules, Supermolecules, and Molecular Devices." *Angew. Chem. internat. Edit.* **27**: 89–112.
- Levin, R. L. (1985). *Arynes. Reactive Intermediates*. New York, John Wiley.

- Levine, R. D. and R. D. Bernstein. (1987). *Molecular Reaction Dynamics and Chemical Reactivity*. Oxford, Oxford University Press.
- Lide, D. R. (1990). *CRC Handbook of Chemistry and Physics*.
- Lii, J. and N. L. Allinger. (1989a). "Molecular Mechanics. The MM3 Force Field for Hydrocarbons. 2. Vibration Frequencies and Thermodynamics." *J. Am. Chem. Soc.* **111**(23): 8566–8575.
- Lii, J. and N. L. Allinger. (1989b). "Molecular Mechanics. The MM3 Force Field for Hydrocarbons. 3. The van der Waals' Potentials and Crystal Data for Aliphatic and Aromatic Hydrocarbons." *J. Am. Chem. Soc.* **111**(23): 8576–8582.
- Likharev, K. K. (1982). "Classical and Quantum Limitations on Energy Consumption in Computation." *Int. J. Theor. Phys.* **21**(3/4): 311–325.
- Lothe, J. (1962). "Theory of Dislocation Mobility in Pure Slip." *J. Appl. Phys.* **33**: 2116–2125.
- Lyo, I. and P. Avouris. (1991). "Field-Induced Nanometer- to Atomic-scale Manipulation of Silicon Surfaces with the STM." *Science*. **253**: 173–176.
- Maitland, G. C., M. Rigby, E. B. Smith and W. A. Wakeham. (1981). *Intermolecular forces: their origin and determination*. Oxford, Clarendon Press.
- Mao, H. K., Y. Wu, R. J. Hemley, L. C. Chen, J. F. Shu and L. W. Finger. (1989). "X-ray Diffraction to 302 Gigapascals: High-Pressure Crystal Structure of Cesium Iodide." *Science*. **246**: 649–651.
- Masters, C. (1981). *Homogeneous Transition-metal Catalysis*. London, Chapman and Hall.

- McBride, J. M., B. E. Segmuller, M. D. Hollingsworth, D. E. Mills and B. A. Weber. (1986). "Mechanical Stress and Reactivity in Organic Solids." *Science*. **234**: 830–835.
- McMillen, D. F. and D. M. Golden. (1982). "Hydrocarbon Bond Dissociation Energies." *Ann. Rev. Phys. Chem.* **33**: 493–532.
- Mead, C. and L. Conway. (1980). *Introduction to VLSI Systems*. Reading, Massachusetts, Addison-Wesley.
- Mislow, K. (1989). "Molecular Machinery in Organic Chemistry." *Chemtracts—Org. Chem.* **2**: 151–174.
- Moss, R. A. (1989). "Carbenic Reactivity Revisited." *Acc. Chem. Res.* **22**: 15–21.
- Moss, R. A. and M. J. Chang. (1981). "Intermolecular Chemistry of a Dialkylcarbene: Adamantylidene." *Tetra. Lett.* **22**(38): 3749–3752.
- Mutter, M., Altmann, K.-E., Altmann, R. Hersperger, P. Koziej, K. Nebel, G. Tuchscherer, S. Vuilleumier, H. -. Gremlich and K. Müller. (1988). "The Construction of New Proteins. Artificial Folding Units by Assembly of Amphiphilic Secondary Structures on a Template." *Helv. Chim. Acta.* **71**: 835–847.
- Nabarro, F. R. N. (1987). *Theory of Crystal Dislocations*. New York, Dover.
- Nanis, L. (1984). Field Ion Microscopy. *Experimental Methods in Electrochemistry*.
- Neidlein, R., V. Poignée, W. Kramer and C. Glück. (1986). "Synthesis and Spectroscopic Properties of Triafulvalene Derivatives." *Angew. Chem. internat. Edit.* **25**(8): 731–732.
- Nelson, D. (1989). The Statistical Mechanics of Membranes and Interfaces. *Statistical Mechanics of Membranes and Interfaces*. Singapore, World Scientific.

- Northrup, J. E. and M. L. Cohen. (1982). "Reconstruction Mechanism and Surface-State Dispersion for Si(111)-(2 × 1)." *Phys. Rev. Lett.* **49**: 1349–1352.
- Pabo, C. (1983). "Designing proteins and peptides." *Nature*. **301**: 200.
- Parker, S. P. (1984). *McGraw-Hill Dictionary of Scientific and Technical Terms*.
- Patai, S. (1980). *The Chemistry of Ketenes, Allenes, and Related Compounds, Part 2*. New York, John Wiley & Sons.
- Pate, B. B. (1986). "The Diamond Surface: Atomic and Electronic Structure." *Surf. Sci.* **165**: 83–142.
- Persson, B. N. J. (1987). "The Atomic Force Microscope: Can it be used to study biological molecules?" *Chem. Phys. Lett.* **141**: 366–368.
- Ponder, J. W. and F. M. Richards. (1987). "An Efficient Newton-like Method for Molecular Mechanics Energy Minimization of Large Molecules." *J. Comput. Chem.* **8**(7): 1016–1024.
- Ponder, J. W. and F. M. Richards. (1987). "Tertiary Templates for Proteins." *J. Mol. Biol.* **193**: 775–791.
- Prater, C. B., H. J. Butt and P. K. Hansma. (1990). "Atomic force microscopy." *Nature*. **345**: 839–840.
- Raabe, G. and J. Michl. (1989). Multiple Bonds to Silicon. *The Chemistry of Organic Silicon Compounds, Part 2*. New York, John Wiley and Sons.
- Ranby, B. and J. F. Rabek. (1975). *Photodegradation, Photo-oxidation and Photostabilization of Polymers*. New York, John Wiley & Sons.
- Rebek, J., Jr. (1987). "Model Studies in Molecular Recognition." *Science*. **235**: 1478–1484.

- Reiss, H. and C. S. Fuller. (1959). *Diffusion Processes in Germanium and Silicon. Semiconductors*. New York, Reinhold.
- Richards, F. M. (1977). "Areas, Volumes, Packing, and Protein Structure." *Annu. Rev. Biophys. Bioeng.* **6**: 151–176.
- Rigby, M., B. E. Smith, W. A. Wakeham and G. C. Maitland. (1986). *The Forces Between Molecules*. Oxford, Clarendon Press.
- Robinson, B. H. and N. C. Seeman. (1987). "The design of a biochip: a self-assembling molecular-scale memory device." *Protein Eng.* **1**: 295–300.
- Robinson, J. W. (1974). *Handbook of Spectroscopy*.
- Salem, L. and C. Rowland. (1972). "The Electronic Properties of Diradicals." *Angew. Chem. internat. Edit.* **11**(2): 92–111.
- Sasaki, T., S. Eguchi, M. Tanida, F. Nakata and T. Esaki. (1983). "Convenient Generation of Adamantylidenecarbene from (Bromomethylene)adamantane. An Efficient Method of Adamantylidenecyclopropanation." *J. Org. Chem.* **48**: 1579–1586.
- Sato, S. (1955). "On a new method of drawing the potential energy surface." *J. Chem. Phys.* **23**: 592.
- Saunders, M. and R. M. Jarret. (1986). "A New Method for Molecular Mechanics." *J. Comp. Chem.* **7**: 578–588.
- Saxena, S. C. and R. K. Joshi. (1989). *Thermal Accomodation and Adsorption Coefficients of Gases*. CINDAS Data Series on Material Properties. New York, Hemisphere.
- Schaefer, H. F., III. (1986). "Methylene: A Paradigm for Computational Quantum Chemistry." *Science*. **231**: 1100–1107.

- Schnabel, W. (1981). *Polymer Degradation*. New York, Hanser/MacMillan.
- Schoen, M., C. L. Rhykerd Jr., D. J. Diestler and J. H. Cushman. (1989). "Shear Forces in Molecularly Thin Films." *Science*. **245**: 1223–1225.
- Scouten, W. H. (1987). A Survey of Enzyme Coupling Techniques. *Immobilization Techniques for Enzymes*. New York, Academic Press.
- Seeman, N. C. (1982). "Nucleic Acid Junctions and Lattices." *J. theo. Biol.* **99**: 237–247.
- Shevlin, P. B. and A. P. Wolf. (1970). "The Formation of Carbon Atoms in the Decomposition of a Carbene." *Tet. Lett.* **46**: 3987–3990.
- Singh, U. C. and P. A. Kollman. (1986). "A Combined *Ab Initio* Quantum Mechanical and Molecular Mechanical Method for Carrying out Simulations on Complex Molecular Systems: Applications to the $\text{CH}_3\text{Cl} + \text{Cl}^-$ Exchange Reaction and Gas Phase Protonation of Polyethers." *J. Comp. Chem.* **7**: 718–730.
- Stewart, J. J. P. (1990). "MOPAC: A semiempirical molecular orbital program." *J. Comp.-Aided Mol. Design.* **4**: 1–105.
- Stillinger, F. H. and T. A. Weber. (1984). "Packing Structures and Transitions in Liquids and Solids." *Science*. **225**: 983–989.484
- Swaddle, T. W. (1986). Electron Transfer Reactions. *Inorganic High Pressure Chemistry*. Amsterdam, Elsevier.
- Ternansky, R. J., D. W. Balogh and L. A. Paquette. (1982). "Dodecahedrane." *J. Am. Chem. Soc.* **104**: 4503–4504.
- Thyagarajan, B. S. (1968–1971). *Mechanisms of Molecular Migrations (Vol. 1–4)*.
- Timoshenko, S. P. and J. N. Goodier. (1951). *Theory of Elasticity, Third Edition*. New York, McGraw-Hill.

- Timoshenko, S., D. H. Young and W. Weaver Jr. (1974). *Vibration Problems in Engineering*. New York, John Wiley & Sons.
- Tjivikua, T., P. Ballester and J. Rebek Jr. (1990). "A Self-Replicating System." *J. Am. Chem. Soc.* **112**: 1249–1250.
- Toffoli, T. (1981). "Bicontinuous Extensions of Invertible Combinatorial Functions." *Math. Systems Theory*. **14**: 13–23.
- Trost, B. M. (1985). "Sculpting Horizons in Organic Chemistry." *Science*. **227**: 908–916.
- Truhlar, D. G. and R. Steckler. (1987). "Potential Energy Surfaces for Polyatomic Reaction Dynamics." *Chem. Rev.* **87**: 217–236.
- Tsuda, M., M. Nakajima and S. Oikawa. (1986). "Epitaxial Growth Mechanism of Diamond Crystal in CH₄-H₂ Plasma." *J. Am. Chem. Soc.* **108**: 5780–5783.
- Ulmer, K. M. (1983). "Protein Engineering." *Science*. **219**: 666–671.
- Utsumi, W. and T. Yagi. (1991). "Light-Transparent Phase Formed by Room-Temperature Compression of Graphite." *Science*. **252**: 1542–1544.
- Uy, R. and F. Wold. (1977). Introduction of Artificial Crosslinks into Proteins. *Protein Crosslinking: Biochemical and Molecular Aspects*. New York, Plenum Press.
- Vanderbilt, D. and S. G. Louie. (1985). Energy Minimization Calculations for Diamond (111) Surface Reconstructions. *The Structure of Surfaces*. Berlin, Springer-Verlag.
- Verwoerd, W. S. (1981). "A Study of the Dimer bond on the Reconstructed (100) Surfaces of Diamond and Silicon." *Surf. Sci.* **103**: 404–415.
- Watson, J. D., N. H. Hopkins, J. W. Roberts, J. A. Steitz and A. M. Weiner. (1987). *The Molecular Biology of the Gene*. Menlo Park, Benjamin/Cummings.

- Wayne, R. P. (1988). *Principles and Applications of Photochemistry*. Oxford, Oxford University Press.
- Webb, T. H. and C. S. Wilcox. (1990). "Improved Synthesis of Symmetrical and Unsymmetrical 5,11-Methanodibenzo[b,f][1,5]dioxocines. Readily Available Nanoscale Structural Units." *J. Org. Chem.* **55**: 363–365.
- Weiner, S. J., P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta Jr. and P. Weiner. (1984). "A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins." *J. Am. Chem. Soc.* **106**: 765–784.
- Weiner, S. J., P. A. Kollman, D. T. Nguyen and D. A. Case. (1986). "An All Atom Force Field for Simulations of Proteins and Nucleic Acids." *J. Comp. Chem.* **7**(2): 230–252.
- Weisenhorn, A. L., P. K. Hansma, T. R. Albrecht and C. F. Quate. (1989). "Forces in atomic force microscopy in air and water." *Appl. Phys. Lett.* **54**(26): 2651–2653.
- Weltner, W., Jr. and J. van Zee. (1989). "Carbon Molecules, Ions, and Clusters." *Chem. Rev.* **89**: 1713–1747.
- Wiberg, K. B. and F. H. Walker. (1982). "[1.1.1]Propellane." *J. Am. Chem. Soc.* **104**: 5239–5240.
- Wildman, T. A. (1986). "An *Ab Initio* Quantum Chemical Study of Hydrogen Abstraction from Methane by Methyl." *Chem. Phys. Lett.* **126**: 325–329.
- Williams, F. (1972). Early Processes in Radiation Chemistry and the Reactions of Intermediates. *The Radiation Chemistry of Macromolecules*. New York, Academic Press.

- Xu, J. A., H. K. Mao and P. M. Bell. (1986). "High-Pressure Ruby and Diamond Fluorescence: Observations at 0.21 to 0.55 Terapascal." *Science*. **232**: 1404–1406.
- Zheng, C., C. F. Wong, J. A. McCammon and P. G. Wolynes. (1988). "Quantum simulation of ferrocytochrome c." *Nature*. **334**: 726–728.