NanoScience and Technology

E. Gnecco
E. Meyer (Eds.)

Fundamentals of Friction and Wear on the Nanoscale

Springer

NanoScience and Technology

NanoScience and Technology

Series Editors:
P. Avouris B. Bhushan D. Bimberg K. von Klitzing H. Sakaki R. Wiesendanger

The series NanoScience and Technology is focused on the fascinating nano-world, mesoscopic physics, analysis with atomic resolution, nano and quantum-effect devices, nanomechanics and atomic-scale processes. All the basic aspects and technology-oriented developments in this emerging discipline are covered by comprehensive and timely books. The series constitutes a survey of the relevant special topics, which are presented by leading experts in the field. These books will appeal to researchers, engineers, and advanced students.

Applied Scanning Probe Methods I
Editors: B. Bhushan, H. Fuchs, and S. Hosaka

Nanostructures
Theory and Modeling
By C. Delerue and M. Lannoo

Nanoscale Characterisation of Ferroelectric Materials
Scanning Probe Microscopy Approach
Editors: M. Alexe and A. Gruverman

Magnetic Microscopy of Nanostructures
Editors: H. Hopster and H.P. Oepen

Silicon Quantum Integrated Circuits
Silicon-Germanium Heterostructure Devices: Basics and Realisations
By E. Kasper, D.J. Paul

The Physics of Nanotubes
Fundamentals of Theory, Optics and Transport Devices
Editors: S.V. Rotkin and S. Subramoney

Single Molecule Chemistry and Physics
An Introduction
By C. Wang, C. Bai

Atomic Force Microscopy, Scanning Nearfield Optical Microscopy and Nanoscratching
Application to Rough and Natural Surfaces
By G. Kaupp

Applied Scanning Probe Methods II
Scanning Probe Microscopy Techniques
Editors: B. Bhushan, H. Fuchs

Applied Scanning Probe Methods III
Characterization
Editors: B. Bhushan, H. Fuchs

Applied Scanning Probe Methods IV
Industrial Application
Editors: B. Bhushan, H. Fuchs

Nanocatalysis
Editors: U. Heiz, U. Landman

Roadmap of Scanning Probe Microscopy
Editors: S. Morita

Nanostructures – Fabrication and Analysis
Editor: H. Nejo

Applied Scanning Probe Methods V
Scanning Probe Microscopy Techniques
Editors: B. Bhushan, H. Fuchs, S. Kawata

Applied Scanning Probe Methods VI
Characterization
Editors: B. Bhushan, S. Kawata

Applied Scanning Probe Methods VII
Biomimetics and Industrial Applications
Editors: B. Bhushan, H. Fuchs

Enrico Gnecco
Ernst Meyer

Fundamentals of Friction and Wear

With 300 Figures and 13 Tables

Springer

Editors:

Dr. Enrico Gnecco
Universität Basel
Institut für Physik
Klingelbergstr. 82, 4056 Basel, Switzerland
e-mail: enrico.gnecco@unibas.ch

Professor Dr. Ernst Meyer
Universität Basel
Institut für Physik
Klingelbergstr. 82, 4056 Basel, Switzerland
e-mail: ernst.meyer@unibas.ch

Series Editors:

Professor Dr. Phaedon Avouris
IBM Research Division
Nanometer Scale Science & Technology
Thomas J. Watson Research Center, P.O. Box 218
Yorktown Heights, NY 10598, USA

Professor Bharat Bhushan
Nanotribology Laboratory for Information
Storage and MEMS/NEMS (NLIM)
W 390 Scott Laboratory, 201 W. 19th Avenue
The Ohio State University, Columbus
Ohio 43210-1142, USA

Professor Dr. Dieter Bimberg
TU Berlin, Fakutät Mathematik,
Naturwissenschaften,
Institut für Festkörperphysik
Hardenbergstr. 36, 10623 Berlin, Germany

Professor Dr., Dres. h. c.
Klaus von Klitzing
Max-Planck-Institut für Festkörperforschung
Heisenbergstrasse 1, 70569 Stuttgart, Germany

Professor Hiroyuki Sakaki
University of Tokyo
Institute of Industrial Science,
4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan

Professor Dr. Roland Wiesendanger
Institut für Angewandte Physik
Universität Hamburg
Jungiusstrasse 11, 20355 Hamburg, Germany

ISBN-10 3-540-36806-X Springer Berlin Heidelberg New York
ISBN-13 978-3-540-36806-9 Springer Berlin Heidelberg New York

Library of Congress Control Number: 2006934355

Typesetting: LE-TEX Jelonek, Schmidt & Vöckler GbR, Leipzig
Production: LE-TEX Jelonek, Schmidt & Vöckler GbR, Leipzig
Cover: WMXDesign, Heidelberg

SPIN 11555315 57/3100/YL - 5 4 3 2 1 0 Printed on acid-free paper

Preface

Friction is an old subject of research and is certainly one of the most important ones from a practical point of view. The da Vinci-Amonton laws are common knowledge (1. Friction is independent of apparent contact area, 2. Friction is proportional to the normal load 3. Friction is independent of velocity). Experiments with small contacts have shown that these empirical laws of friction do not always hold. Reasons may be related to the large surface-to-volume ratio and the greater importance of adhesion, surface structure and surface chemistry. Therefore, there is some need to get a better understanding of the phenomenon of friction, to learn how to quantify and eventually control it. In the first half of last century the school of Bowden and Tabor have performed systematic, macroscopic experiments and have related macroscopic friction to small contacting asperities. In the 1990's experiments performed with atomic force microscopy, surface force apparatus and quartz microbalance, revealed interesting new physics on the nanometer scale (atomic-scale stick-slip, confinement of liquid films, determination of electronic and phononic contributions to dissipation). During the same time, theoretical analysis of nanometer-sized contacts has been performed and gave insight into the processes in the buried interface. Strong activities were pursued in the US at Universities and corporate research laboratories. Similar activities were pursued in Japan, where the main focus was on the understanding the tribology of hard disc drives and applications in automobile industries. Europe has a long tradition in mechanical engineering sciences. Activities at the University level were mainly driven by recent developments in nanosciences (scanning probe microscopy, computer modelling). In 2001, the European Science Foundation Programme "Nanotribology" (NATRIBO) was started. The aim of this programme is to bring together experimentalists and theoreticians to improve the understanding of nanometer-sized contacts. The aim of this book is to give an overview of the status of resarch in this field. Members of the NATRIBO-network and a selection of excellent international experts have contributed to this book. They made a strong effort to give a deep insight into the complex phenomena of nanotribology.

The book is divided in seven sections. In the first section the instrumental setups most commonly used in nanotribology are introduced. The first chapter presents the atomic force microscope (AFM), with a special emphasis on the force sensors and the ways to control the contact between tip and

surface. The interrelations between friction, load, material properties, temperature, and the lateral forces detected in dynamic measurements, are also discussed. The second chapter introduces the surface force apparatus (SFA), as an independent tool and in combination with other techniques. A case study of weakly adhesive surface under shear is discussed. The quartz crystal microbalance is treated in Chapt. 3. After the acoustics of the crystal, the driving circuits and the quality of the surface electrodes, the authors present results obtained in ultra-high vacuum (UHV). Chapter 4 describes the effects of normal and shear ultrasonic vibrations in AFM, focusing in particular on friction reduction and adhesion hysteresis. Finally, Chapt. 5 shows how scanning probe microscopes can be combined with transmission electron microscopes to image both tip and sample surface. Contact formation and breaking, adhesion effects, electric conductivity and material transport are consequently discussed.

Section 2 gives a detailed overview on friction phenomena occurring on the atomic scale. Chapter 6 introduces the Tomlinson model and fundamental phenomena observed by AFM (atomic stick-slip, velocity dependence of friction, superlubricity, and nanowear processes). The next chapter shows how the rate theory has been applied to obtain general force-velocity relations. Analytical approximations are compared with precise numerical results. Chapter 8 introduces the important problem of friction control. Mechanical and chemical methods to achieve this goal are discussed from both theoretical and experimental points of view. Superlubricity is the main topic of Chapts. 9–11. Chapter 9 shows how surface incommensurability and thermal effects can lead to a strong reduction of friction, which was recently observed experimentally. Lubrication by graphite, diamond-like carbon, fullerenes and carbon nanotube is discussed within this frame. Chapter 10 presents theoretical studies of superlubricity. Symmetry considerations, role of instabilities, temperature effects, damping in the superlubric regime and long-range elastic deformations are discussed, as well as generic models and applications to layered materials, metal-metal contacts and hydrogen-terminated surfaces. In particular, the presence of hydrogen is proved to be the key factor leading to superlubricity between diamond surfacesas shown in the detailed theoretical study presented in the last chapter of the section.

The third section of the book introduces contact mechanics on the nanoscale. After a brief theoretical introduction, Chapter 12 describes the main experimental methods to investigate elasticity on the nanoscale and recent findings related to inorganic nano-objects and biological samples. Chapter 13 addresses the special case on metallic nanocontacts, whose mechanical properties cannot be separated from electron transport mechanisms. Fabrication, elasticy, fracture, and shape of metal contacts are discussed, as well as chains of gold atoms and metallic adhesion in atomic-sized tunneling junctions. Quasi-crystals are the main subject of Chapt. 14. This leads the authors to describe the surface roughness in relation to friction and adhesion. A par-

ticular emphasis is given on the roughness power spectrum, which is derived from the surface height using optical and scanning probe microscopes. Chapter 15 focuses on the roughness of self-affine fractal surfaces. The contact morphology and the pressure distribution are estimated at different scales, with and without adhesion, using molecular dynamics, and they are compared with analytical contact models based on continuum mechanics. The role of the elastic moduli of the underlying bulk is also treated here. Finally, the last chapter of this section describes how nanoroughness is affected by depth-sensing indentation. A special attention is given to elastomer probes used in AFM investigations.

Section 4 describes dissipative mechanisms at finite separation under different points of view. Chapter 17 deals with the case of amplitude modulation AFM, used to characterize surfaces in air or in liquids. In such case the energy dissipation accompanying the imaging process is given by the phase shift signal acquired while scanning. The next chapter considers dissipation in non-contact AFM. After a review of the experimental data at our disposal, possible mechanisms of atomic-scale damping are discussed, as well as detailed models developed to understand the effect of these mechanisms on the imaging process. The theory of non-contact friction is the subject of Chapt. 19. The fluctuating electromagnetic field which surrounds any solid surface, and is responsible for radiative heat transfer and van der Waals interaction and friction, is examined under semiclassical and quantum theories. At short separations, Van der Waals friction is greatly enhanced. Furthermore, static charges on the surface are responsible of electrostatic friction around a moving body, and possible applications to scanning probe spectroscopy are discussed. This topic is extended in Chapt. 20, where the authors show how the force sensitivity of free cantilevers is limited by thermal fluctuations and material properties and how these problems are reduced by UHV annealing or cooling to cryogenic temperatures.

Wear and fracture are treated in the fifth section of the book. Chapter 21 covers the mechanisms of surface damage down to micro- and nano-scales. Both basic theories and experiments are considered, and a discussion on hardness at different scales is also provided. Chapter 22 examines the relation between stress and chemical reactivity. Examples of single asperity tribochemical wear include dissolution along monolayer steps in calicum carbonates and phosphates, wear of the probing tip on reactive surfaces and tip induced wear of silicate substrates. Chapter 23 gives an overview of stiction, friction and wear phenomena affecting micro- and nano-electromechanical systems. The tribological characterization of these devices is discussed together with various solutions introduced to improve their reliability. The last chapter of the section addresses nanotribological problems in automotive engineering. Wear rates of few nanometers per hours are mandatory in internal combustion engines, which requires exceptional finishing of the sliding surfaces in the engine.

Another growing field of nanotribology is the manipulation of nanoparticles, which is treated in Sect. 6. Chapter 25 shows how the tip of a scanning probe microscope operated in dynamic mode can be alternatively used to image and move particles in a controlled way. With a proper calibration of the excitation amplitude the energy dissipation and the frictional forces involved in the manipulation process can also be estimated. Chapter 26 considers a system of great interest in nanoscience, i. e. carbon nanotubes (CNTs). In such case AFM can be used to test mechanical properties in dynamic and quasi-static ways. Nanotube bundles, catalytically grown CNTs, and diameter dependence of bending moduli are addressed as special cases. The next chapter focuses on the manipulation of fullerene molecules on a silicon surface. After summarizing the experimental results obtained with scanning tunneling microscopes, the authors present a model which successfully interpretes the mechanisms underlying adsorption, diffusion and manipulation of the molecules.

The last section of the book deals with applications of nanotribology to organic materials. Chapter 28 gives a detailed overview of friction on self-assembled monolayers (SAMs). Homogenoues films are first addressed, and the influence of chain length, terminal groups, packing states as well as environmental conditions on friction are discussed. The role of nanoscale heterogeneities on the nanoscale is considered in the second part of the chapter. The next two chapters deal with polymers. In particular, Chapt. 29 considers the influence of hydrophobicity on the frictional forces experienced on two different materials, whereas Chapt. 30 treats the molecular origins of elastomeric friction. Both interfacial adhesion and internal friction are thermally activated processes, and the competition between them gives a correct interpretation of the experimental results. Finally, the last chapter of the book describes the importance of friction and adhesion mechanisms in cell dynamics, with particular emphasis on the adhesive forces experienced on the substrates where the cells can spread and proliferate. This is of great importance in the emerging field of tissue engineering.

In conclusion, we would like to thank all the authors for the time and the energies that they have spent on this project, as well as all the participants to the Nanotribo workshops for the interesting scientific discussions that they have stimulated. A special thanks goes also to Claus Ascheron, Angela Lahee and Steffi Hohensee from Springer-Verlag, who made possible the publication of this book. Financial support from the European Science Foundation, the Pico-Inside project, the Swiss National Center of Competence in Research Nanoscale Science and the Swiss National Science Foundation is gratefully acknowledged.

University of Basel *Enrico Gnecco and Ernst Meyer*

Contents

Wear and Fracture on the Nanoscale

Manipulation of Nanoparticles

Organic Materials

1 Friction Force Microscopy

Roland Bennewitz

Department of Physic, McGill University, Montreal, Quebec, Canada
roland.bennewitz@mcgill.ca

1.1 Introduction

Friction Force Microscopy (FFM) is a sub-field of scanning force microscopy addressing the measurement of lateral forces in small sliding contacts. In line with all scanning probe methods, the basic idea is to exploit the local interactions with a very sharp probe for obtaining microscopic information on surfaces in lateral resolution. In FFM, the apex of a sharp tip is brought into contact with a sample surface, and the lateral forces are recorded while tip and sample slide relative to each other. There are several areas of motivation to study FFM. First, the understanding of friction between sliding surfaces in general is a very complex problem due to multiple points of contact between surfaces and the importance of lubricants and third bodies in the sliding process. By reducing one surface to a single asperity, preparing a well-defined structure of the sample surface, and controlling the normal load on the contact the complexity of friction studies is greatly reduced and basic insights into the relevant processes can be obtained. Furthermore, with the decrease of the size of mechanical devices (MEMS) the friction and adhesion of small contacts becomes a technological issue. Finally, the lateral resolution allows to reveal tribological contrasts caused by material differences on heterogenous surfaces.

The experimental field of FFM has been pioneered by Mate, McClelland, Erlandsson, and Chiang [1]. The group built a scanning force microscope where the lateral deflection of a tungsten wire could be measured through optical interferometry. When the etched tip of the tungsten wire slid over a graphite surface, lateral forces exhibited a modulation with the atomic periodicity of the graphite lattice. Furthermore, a essentially linear load dependence of the lateral force could be established.

In this chapter we will describe aspects of instrumentation and measurement procedures. In the course of this description, a series of critical issues in FFM will be discussed which are summarized in Fig. 1.1.

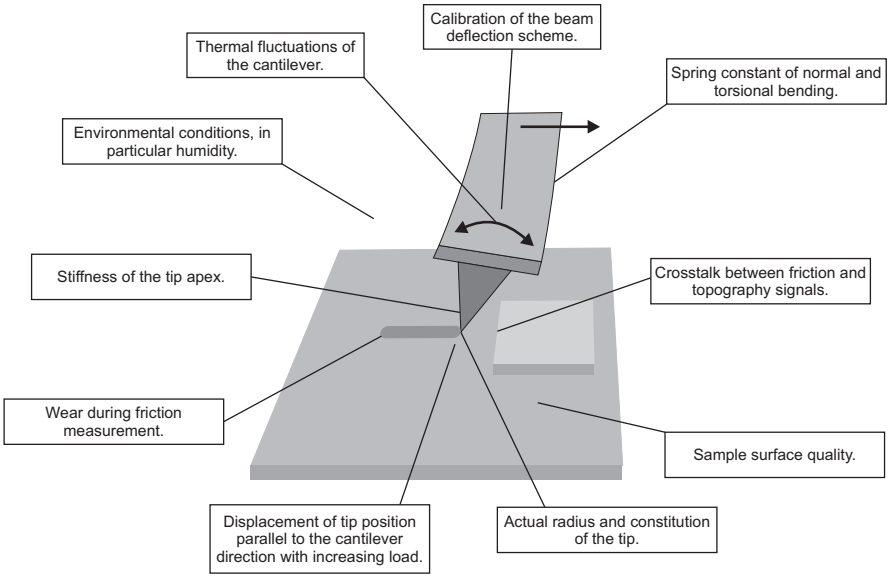


Fig. 1.1. Critical issues in experimental friction force microscopy which are discussed in this chapter

1.2 Instrumentation

1.2.1 Force sensors

The force sensor in the original presentation of FFM by Mate et al. was a tungsten wire [1]. Its deflection was detected by an interferometric scheme where the wire constituted one mirror of the interferometer. A similar concept was later implemented by Hirano et al., who optically detected the deflection of the tungsten wire in a Scanning Tunneling Microscope when scanning the tip in close proximity to the surface [2]. Mate and Hirano report lateral spring constants from 1.5 to 2500 N/m, depending on the wire thickness and length. Etching the wire to form a tip at its end, mounting the wire, aligning of the light beam, and determination of the spring constant comprise some experimental difficulties. These difficulties are greatly reduced by the use of dedicated micro-fabricated force sensors. A very sophisticated instrumental approach to the solution of those problems has been realized by Dienwiebel et al. [3]. The group has attached a stiff tungsten wire to a micro-fabricated force sensor made of silicon. The central part of the sensor is a pyramid holding the tip. The position of the pyramid is detected in all three dimensions by means of four optical interferometers directed towards the faces of the pyramid. It is suspended in four symmetric high-aspect ratio legs which serve as springs with isotropic spring constant in both lateral directions and a higher spring constant in normal direction. The symmetric design of the instrument allows

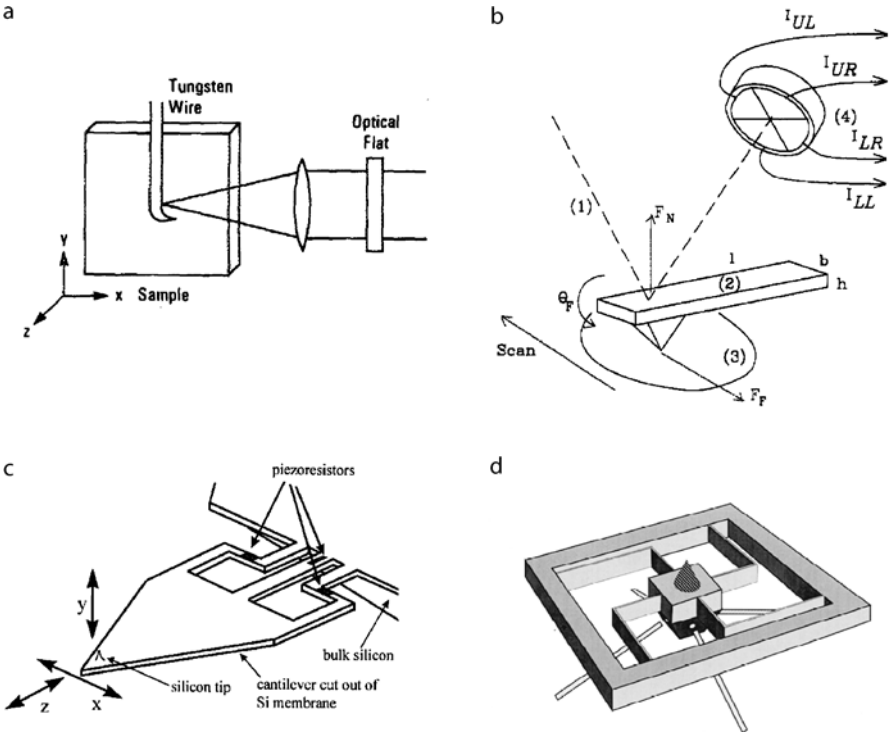


Fig. 1.2. Four design options for Friction Force Microscopy. **a** Concept of the original instrument used by Mate et al. for their pioneering experiments [1]. The deflection of a tungsten wire is detected by optical interferometry. The bent end of the wire is etched into a sharp tip. **b** Beam-deflection scheme as devised by Marti et al. [5]. Normal force F_N and friction force F_F cause bending and twisting of the cantilever. The deflection of a reflected light beam is recorded by comparing currents from four sections of a photodiode. **c** Cantilever device for the measurement of lateral forces with piezoresistive detection [8]. Lateral forces acting on the tip cause a difference in stress across the piezoresistors. **d** Micro-fabricated force detector for isotropic measurements of friction forces. The block in the center holds a tungsten tip, pointing upwards in this figure. The position of the block in all three dimensions is recorded by four interferometric distance sensors which are indicated by the four light beams below the devices [9]

for determination of normal and lateral forces acting on the tip with minimal cross talk. An overview over different experimental realizations of FFM is given in Fig. 1.2.

The most widely used form of micro-fabricated force sensors for FFM is the micro-fabricated cantilever with integrated tip. The cantilever can be either a rectangular beam or a triangular design based on two beams. The lateral force acting on the tip is detected as torsional deflection of the cantilever. This scheme has been implemented in 1990 by Meyer et al. [4] and

Marti et al. [5]. It is interesting to note that the triangular design is more susceptible to deflection by lateral forces than the rectangular beam, contrary to common belief and intuition [6]. However, triangular cantilevers are less prone to the highly unwanted in-plane bending [7].

The deflection of cantilever-type force sensors is usually detected by means of a light beam reflected from the back side of the cantilever at the position of the tip. The reflected light beam is directed towards a position-sensitive photodiode which detects normal and torsional bending of the cantilever as a shift in the position of the light beam in orthogonal directions. Realistically, there is always some cross-talk between the signals for normal and torsional bending. It can be detected by exciting the cantilever to oscillate at the fundamental normal and torsional resonance and measure the oscillation amplitude in the orthogonal channels. The cross-talk can be minimized by rotation of the position-sensitive photodiode or accounted for in the detection electronics or software. Cross-talk can transfer topographic features into the lateral force signal and create topographic artifacts from friction contrast, the latter even amplified by the feedback circuit acting on the sample height.

Calibration of the beam-deflection scheme is not a simple task, however very important in order to compare FFM results from different sources. Many publications in the past have reported on relative changes in frictional properties, without providing any calibration at all. While such relative changes certainly represent important physical findings, it is nevertheless of utmost importance to provide all experimental information available, often allowing for a rough quantitative estimate of the lateral forces. Lateral forces in FFM can easily range from piconewton to micronewton, spanning a range of very different situations in contact mechanics, and knowing at least the order of magnitude of forces helps to sort the results qualitatively into different regimes.

The calibration comprises two steps. First, the spring constant has to be determined for the force sensor. Note that the beam-deflection scheme actually determines the angular deflection of the cantilever. Nevertheless it has become custom to quantify the force constant in N/m, where the length scale refers to the lateral displacement of the tip apex relative to the unbent cantilever. Second, a relation between the deflection of the cantilever and the voltage readout of the instrument has to be established.

For the determination of the spring constant, several methods have been suggested. The easiest is to calculate it from the dimensions of the cantilever. While width and thickness are easily determined by optical or electron microscopy, thickness is better deduced from the cantilevers resonance frequency. Alternatively, the spring constant can be determined from changes in the resonances caused by the addition of masses to the free end of the cantilever. Also, the analysis of a cantilever's resonance structure in air can provide the required quantities. The latter two methods have recently be described and compared by Green et al. [10]. The relation between tip displacement and voltage readout can be established by trapping the tip in

a surface structure and displacing the sample laterally by small distances. For a rough estimate one can also assume that the sensitivity of the position-sensitive photodiode is the same for normal and torsional deflection. Taking into account the geometry of the beam-deflection scheme, the torsional deflection sensitivity can be deduced from the normal deflection sensitivity (See Ref. [11] and page 352 of Ref. [12]).

A method which provides a direct calibration of the lateral force with respect to the readout voltage is the comparison with a calibrated spring standard. Recent implementations of this approach suggest as calibrated standards optical fibers [13] or micro-fabricated spring-suspended stages with spring constants that can be traced to international standards [14]. A particularly elegant method to calibrate FFM experiments is the analysis of friction loops, i. e. lateral force curves from forward and backward scans, recorded across surfaces with well-defined wedges [11, 15].

The torsional deflection of a cantilever can in principle be detected also by optical interferometry, provided that the beam diameter is smaller than the cantilever and the point of reflection is shifted off the torsional axis [16]. However, FFM results including normal and lateral force measurements require the differential reading of multiple interferometers [3, 17].

An alternative to the detection of the cantilever bending via the beam-deflection scheme is the implementation of piezoresistive strain sensors into the cantilever. In order to measure both lateral and normal forces acting on the tip in FFM, two such strain sensors need to be realized on one sensor. Chui et al. have created a piezoresistive sensor which decouples the two degrees of freedom by attaching a normal triangular cantilever to a series of vertical ribs sensing lateral forces [18]. Gotszalk et al. have constructed a U-shaped cantilever with one piezoresistive sensor in each arm, allowing for the the detection of lateral forces at the tip [19]. While the publications presenting these novel instrumental approaches contain experimental proofs of concept, no further use of piezoresistive sensors in FFM experiments has been reported. This is certainly due to a lack of commercial availability. Furthermore, the signal-to-noise ratio in static force measurements using piezoresistive cantilevers seems not to reach that of optical detection schemes.

1.2.2 Control over the contact

The exact knowledge of the atomic configuration in the contact between tip apex and surface is prerequisite for a complete understanding of the results in Friction Force Microscopy. It is the most severe drawback in FFM that this knowledge is not available in most cases. While sample surfaces can often be prepared with atomic precision and cleanliness, the atomic constitution of the tip apex is usually less controlled. Furthermore, in the course of sliding atoms may be transferred from the tip to the surface or vice versa. Such transfer processes occur even for very gentle contact formation, as shown in experiments combining Scanning Probe Microscopy with a mass spectrome-

try analysis of the tip apex [20–22]. The transfer of atoms may quite often not only quantitatively but also qualitatively change the lateral forces encountered. In particular, the occurrence of atomic stick-slip motion can depend on the establishment of a certain degree of structural commensurability between tip and surface in the course of scanning [23, 24]. For atomic stick-slip measurements on graphite surfaces, the role of small graphite flakes attached to the tip has long been discussed and recently confirmed experimentally [1, 25].

The best control over the atomic structure of the tip apex has been achieved for metal tips in vacuum environments. By applying the established procedures of Field Ion Microscopy (FIM), the tip structure can not only be imaged but also conditioned on the atomic scale. Cross et al. have characterized the adhesion between a tungsten tip and a gold surface and proved the conservation of the atomic tip structure by means of FIM [26]. Even with instruments of lower resolution, FIM can at least be used for cleaning procedures and for a determination of the crystalline orientation of the apex cluster [2].

The integrated tips at the end of micro-fabricated silicon cantilevers have a well-defined crystalline orientation, usually pointing with the (100) direction along the tip. However, the tip surface and with it the whole tip apex are at least oxidized and possibly contaminated through packaging, transport, and handling. Furthermore, many tips are sharpened in a oxidation process which introduces large stresses at the apex. While etching in hydrofluoric acid can remove the oxide and for some time passivate silicon surface bonds by hydrogen, a stable formation and reproducible characterization comparable with FIM of metal tips has not yet been reported. Tips integrated into silicon nitride cantilevers are amorphous due to the chemical vapor deposition process and may exhibit an even more complex structure and chemistry at the tip apex.

One way of overcoming the uncertainty of the tip constitution is to use methods of surface chemistry to functionalize the tip [27]. Specific interactions between molecules attached to the tip and molecules on the surface can be sensed by means of FFM [28]. At the same time, very strong adhesion has been reduced by covering the tip with a passivating layer to allow for lateral force imaging for example on silicon [29]. Numerous studies using this method have been published, mainly concentrating on organic monolayers on tip and surface. A recent review of the field has been given by Leggett et al. [30]. Schwarz et al. have prepared well-defined tips for FFM by deposition of carbon from residual gas molecules in a Transmission Electron Microscope, keeping control of the tip radius for a quantitative analysis of a contact mechanics study [31]. Force measurements explicitly aiming at interactions between colloidal particles and a surface have been performed by gluing micrometer-sized spheres of the desired size to the cantilever [32, 33]. As a final note, one should always be aware of the possible occurrence of major tip wear which has been observed to happen in a concerted action of mechanical and chemical polishing [34].

1.3 Measurement procedures

The standard measurement in FFM is the so-called friction loop: The lateral force acting on the tip is recorded for a certain distance of scanning in the direction perpendicular to the long cantilever axis and for the reverse direction. The area in the loop represents the dissipated energy, and the area divided by twice the distance is the mean lateral force. It is always very instructive to record the topography signal of forward and backward scan at the same time, as differences will reveal cross-talk between normal and torsional bending of the cantilever.

Whenever lateral forces are measured as a function of some experimental parameter, the influence of that parameter on adhesion should be studied simultaneously. In order to interpret the experimental results in terms of contact sizes versus dissipation channels the knowledge of adhesion is essential. An excellent example is the jump in lateral forces observed on a C_{60} crystal when cooling to the orientational order-disorder phase transition, which was fully explained by a change in adhesion [35]. For experiments carried out in ambient environment, the dominant contribution to adhesion are usually capillary forces which dependent greatly on the humidity and on the hydrophobicity of the surface [36]. The humidity dependence of FFM results itself can depend again on the temperature [37, 38]. Consequently, an enclosure of FFM experiments for humidity control greatly enhances the reproducibility of results.

1.3.1 Friction as a function of load

One of the central experiments in tribology is the quantification of friction, i. e. the change of lateral force with increasing normal load on the sliding contact. One of the questions to be addressed is whether the relation between lateral and normal force is linear for FFM experiments, i. e. whether Amontons' law extends to the nanometer scale [39]. The number of FFM studies reporting lateral force as a function of load is very large, and the overall physical picture is multifaceted, to express it in a positive way. A collection of results is shown in Fig. 1.3. From a procedural point of view it is extremely important to measure the lateral forces for the full range of small normal forces until the tip jumps out of contact, usually at a negative normal force. In this way the adhesion in the system can be categorized, and possible nonlinear characteristics at minimal loads are not overlooked. A useful way of analyzing load dependence data from FFM experiments is the representation in lateral force histograms, where for example friction on terraces and friction at steps could automatically be distinguished [40].

When the normal load on the tip is varied the position of the contact may be displaced along the long axis of the cantilever. This effect is caused by the tilt of the cantilever with respect to the surface. On heterogeneous surfaces such displacement may distort the friction measurement and, therefore, has

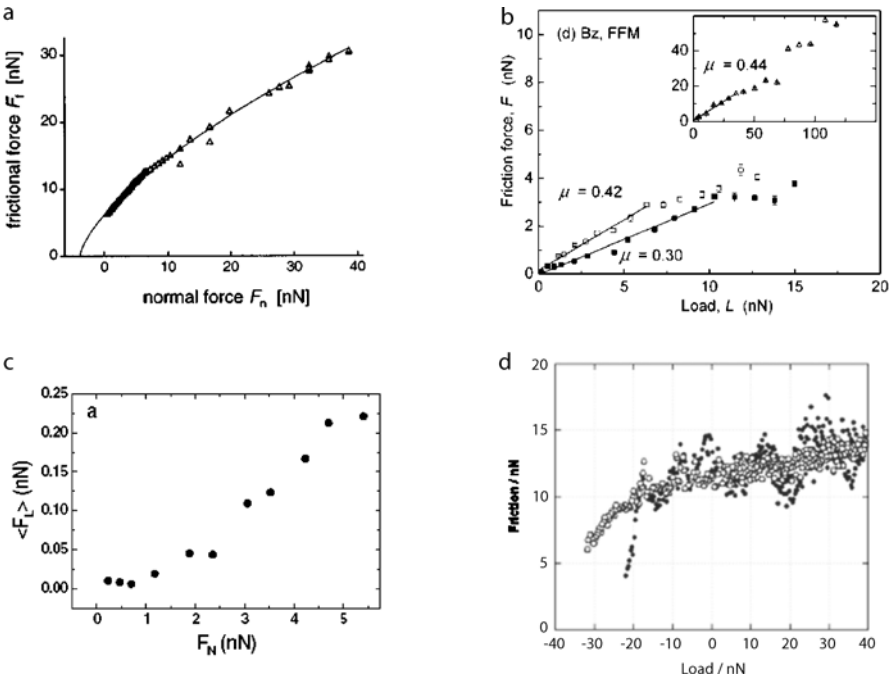


Fig. 1.3. Examples for the diversity of friction vs. load curves measured by FFM. **a** Amorphous carbon measured in an argon atmosphere [31]. The sub-linear characteristic resembles the results of contact mechanics models. **b** Phenyltrichlorosilane monolayer studied in ethanol [41]. A linear dependence is found until the monolayer collapses under the tip pressure. **c** Atomic friction on NaCl(100) recorded in ultra-high vacuum [42]. A regime of vanishing friction is found for low loads. **d** Friction measurement on a hydrogen-terminated diamond surface with nanometer-scale roughness [43]. The closed circles represent the erratic load dependence of FFM results when the lateral displacement of the tip for increasing load is not compensated. The open circles show the expected sub-linear characteristic after activating the compensation

to be compensated [43]. Another effect that can seriously disturb friction experiments is the onset of wear and the concomitant increase of lateral forces. Wear thresholds in FFM can be as low as a few nanonewton normal load, and wear at a constant low load may suddenly start after repeatedly scanning the same area [44].

1.3.2 Friction as a function of material

On inhomogeneous surfaces Friction Force Microscopy can image contrasts between different materials with high lateral resolution. Such contrast has been found to arise from a difference in chemical interactions between different molecular patches at the surface and the tip [45]. As mentioned above,

it is crucial to complement lateral friction contrast with local measurements of adhesion in order to elucidate whether adhesion and contact size or different channels of dissipation are dominating the contrast. Care has to be taken regarding topographical artifacts, as different materials on heterogeneous surfaces are often found at different topographic heights. Interestingly, friction contrast is also found between domains of identical molecular layers with anisotropic lateral orientation [46–48]. Friction anisotropy on a given surface has to be clearly distinguished from friction anisotropy for different azimuthal orientations between the tip and the surface. In order to measure the latter, the sample has to be rotated with respect to the tip [25].

1.3.3 Friction effects in normal force measurements

When the sample is approached towards the tip, the normal force can be determined as a function of distance by measuring the normal bending of the cantilever. In all beam-deflection type FFM the cantilever is tilted with respect to the sample surface to make sure that the tip is the foremost protrusion of the force sensor. Once the tip is in contact, the tilt causes a lateral displacement of the tip position upon further approach. The friction forces arising from this lateral displacement influence the normal force measurement [33]. A detailed analysis of the process proves that one can actually perform a calibrated friction experiment through normal force vs. distance curves, in particular when using extended tips like colloid probes [49]. Even when probing the surface in a dynamic intermittent contact mode these frictional contributions can be detected as a phase shift between excitation and cantilever oscillation [47].

1.3.4 Fluctuations in Friction Force Microscopy

Friction Force Microscopy is naturally subject to thermal fluctuations. Such thermal fluctuations can influence the frictional behavior of sliding contacts, as evident in the logarithmic dependence of friction on velocity at low scanning velocities [50, 51] which has been linked to thermal fluctuations via its temperature dependence [52]. Cantilever-type force sensors have a distinct resonance structure which dominates the thermal noise spectrum. Typically, oscillations at resonances with frequencies of several kHz are averaged out in FFM experiments. However, these resonances influence the experimental result and it is therefore very instructive to study the lateral force signal with high bandwidth [53, 54]. Furthermore, the statistical distribution of lateral forces in FFM experiments can be analyzed to reveal the role of thermal fluctuations [55]. The limited scanning velocity of FFM normally separates the frequency regimes of fast fluctuations and of slower occurrence of topographic or even atomic features. The velocity limitations of FFM have been addressed by new designs combining the force sensor of an FFM with a dedicated sample stage [56, 57].

1.3.5 Friction as a function of temperature

The study of friction as a temperature is an obvious field of great interest. However, the number of groups including a temperature dependence into FFM studies is increasing only recently [35, 37, 52, 58, 59]. Thermal drift is a severe problem in the design of Friction Force Microscopes working at variable temperature, since the optical lever of the beam-deflection scheme needs to have a certain length for sensitivity. Variable-temperature instruments with thermal-expansion compensated design comparable to dedicated Scanning Tunneling Microscopes [60] have not been reported so far. One interesting approach to circumvent drift problems is the local heating of the very tip [61].

1.3.6 Dynamic lateral force measurements

Dynamic friction force microscopy

When the sample is periodically displaced in lateral direction, the lateral force acting on the tip and detected by the cantilever will be modulated with the same periodicity. An early application of such a lateral modulation by Maivald et al. was the enhancement of contrast at step edges [62]. Dynamic Friction Force Microscopy detects the periodic lateral force signal by means of a lock-in amplifier. This idea was implemented by Göddenhenrich et al., who applied the periodic sample displacement along the long axis of the cantilever and detected the lateral force as periodic buckling of the cantilever [63]. Simultaneously, their fiber-interferometric setup could statically measure the deflection of the cantilever caused by normal forces. The same technique was implemented by Colchero et al. for a beam-deflection instrument. The authors provided a detailed analysis for the evaluation of the lateral forces when the sample is displaced in a sinusoidal movement [64]. They also pointed to the fact that using their method of Dynamic Friction Force Microscopy one will obtain quantitative results when taking data, while static experiments need subtraction of forward and backward scan before numbers can be obtained. Carpick et al. have used a similar technique with very small sample displacement amplitudes to avoid any slip of the tip over the surface [65]. In such experiments, the amplitude of the lateral force provides a measure for the contact stiffness. Dynamic friction force microscopy has been combined with sophisticated versions of the pulsed-force mode for a simultaneous measurement of all relevant properties of mechanical contacts [66]. In a recently published study, Haugstad has analyzed the non-linear response of the lateral force to the sinusoidal sample displacement in a Fourier analysis [67]. Using this technique he was able to gain new insights into the transition from static to kinetic sliding on a polymer blend.

Dynamic Friction Force Microscopy can gain sensitivity by tuning the periodic excitation to resonances of the cantilever [68, 69]. However, the coupling between the mechanical properties of the contact and the flexural modes

of the cantilever requires a complex analysis, as provided in a recent review which also references previous work in the field of ultra-sonic force microscopy [70].

Dynamic non-contact lateral force experiments

The success of dynamic non-contact force microscopy in atomic resolution imaging of insulating surfaces and its prospect of measuring dissipation phenomena with the same resolution [71] has initiated projects which aim at a dynamic non-contact microscopy using lateral oscillation of the tip. Jarvis et al. have constructed a novel force sensor which allows to excite and detect oscillations of the tip in normal as well as in lateral direction [72]. The independent oscillations were achieved by suspending the tip holder in hinges at the end of two normally oscillating cantilevers. The group has controlled the tip-sample distance by changes in the normal oscillation frequency, and simultaneously recorded changes in the amplitude of the lateral oscillation pointing to frictional tip-sample interactions.

A standard rectangular cantilever has been employed by Pfeiffer et al. for the dynamic detection of interactions between a laterally oscillating tip and a surface close to but not in contact [73]. In this study, the cantilever was excited to oscillate at its first torsional resonance, making the tip oscillate laterally. The distance between tip and a copper surface was controlled using the tunneling current as feedback quantity. The lateral interaction between tip and monatomic steps or single impurities could be detected as frequency shift in the torsional oscillation. Giessibl et al. attached a tungsten tip to a quartz tuning fork such that it would oscillate laterally over the surface. Again using tunneling as feedback, they were able to study dissipation in the lateral movement with atomic resolution on a Si(111)7×7 surface, thereby tracing friction to a single atom [74]. The damping of the lateral oscillation has been explained in terms of a fast stick-slip process involving one adatom. The same surface has recently been studied in dynamic lateral force microscopy using a standard rectangular cantilever by Kawai et al. [75]. In this study a small frequency shift in the torsional resonance frequency upon approach was used to control the tip-sample distance. The torsional resonance was detected using a heterodyne interferometer scheme, where the focus of the light beam was positioned on one side of the cantilever in order to be sensitive to the torsional bending. This is actually a very informative method to study the resonance structure of cantilevers which can show significant deviations from ideal modeling due to extra masses and asymmetries [16].

The dynamic non-contact experiments introduced in this section are very interesting tools to study conservative and dissipative interactions in lateral motion even before a repulsive contact is established. Their full strength might become evident once they are applied to the manipulation of atoms or molecules on surfaces.

1.4 Outlook

Friction Force Microscopy is now a widely distributed experimental method. The experimental procedures and the calibration have been established to allow for reproducible studies of frictional properties in single-asperity contacts. The biggest drawback within the method is the lack of methods for a reproducible preparation and characterization of tips on atomic scale, as compared to the surface preparation by means of methods of Surface Science. Such control over the atomic constitution of the contact area would greatly advance our understanding of tribological processes on the nanometer scale. Other instrumental challenges in the field include the further improvement of FFM experiments at variable temperatures and in liquid environments.

References

1. C. Mate, G. McClelland, R. Erlandsson, and S. Chiang, Phys. Rev. Lett. **59**, 1942 (1987).
2. M. Hirano, K. Shinjo, R. Kaneko, and Y. Murata, Physical Review Letters **78**, 1448 (1997).
3. M. Dienwiebel, E. de Kuyper, L. Crama, J. Frenken, J. Heimberg, D.-J. Spaanderman, D. van Loon, T. Zijlstra, and E. van der Drift, Rev. Sci. Instr. **76**, 43704 (2005).
4. G. Meyer and N. Amer, Appl. Phys. Lett. **57**, 2089 (1990).
5. O. Marti, J. Colchero, and J. Mlynek, Nanotechnology **1**, 141 (1990).
6. J. Sader and R. Sader, Applied Physics Letters **83**, 3195 (2003).
7. J. Sader and C. Green, Review of Scientific Instruments **75**, 878 (2004).
8. T. Gotszalk, P. Grabiec, and I. Rangelow, Ultramicroscopy **82**, 39 (2000).
9. T. Zijlstra, J. Heimberg, E. van der Drift, D.G. van Loon, M. Dienwiebel, L. de Groot, and J. Frenken, Sensors and Actuators, A: Physical **84**, 18 (2000).
10. C. Green, H. Lioe, J. Cleveland, R. Proksch, P. Mulvaney, and J. Sader, Review of Scientific Instruments **75**, 1988 (2004).
11. D. Ogletree, R. Carpick, and M. Salmeron, Rev. Sci. Instr. **67**, 3298 (1996).
12. E. Meyer, R. Overney, K. Dransfeld, and T. Gyalog, *Nanoscience: Friction and Rheology on the Nanometer Scale* (World Scientific, Singapore, 1998).
13. N. Morel, M. Ramonda, and P. Tordjeman, Applied Physics Letters **86**, 163103 (2005).
14. P. Cumpson, J. Hedley, and C. Clifford, Journal of Vacuum Science & Technology B (Microelectronics and Nanometer Structures) **23**, 1992 (2005).
15. M. Varenberg, I. Etsion, and G. Halperin, Review of Scientific Instruments **74**, 3362 (2003).
16. M. Reinstaedtler, U. Rabe, V. Scherer, J.A. Turner, and W. Arnold, Surface Science **532–535**, 1152 (2003).
17. G. Germann, S. Cohen, G. Neubauer, G. McClelland, and H. Seki, J. Appl. Phys. **73**, 163 (1993).
18. B. Chui, T. Kenny, H. Mamin, B. Terris, and D. Rugar, Appl. Phys. Lett. **72**, 1388 (1998).

19. T. Gotszalk, P. Grabiec, and I. Rangelow, Sensors and Actuators, A: Physical **123–124**, 370 (2005).
20. U. Weierstall and J. Spence, Surface Science **398**, 267 (1998).
21. T. Shimizu, J.-T. Kim, and H. Tokumoto, Appl. Phys. A **66**, S771 (1998).
22. A. Wetzel, A. Socoliuc, E. Meyer, R. Bennewitz, E. Gnecco, and C. Gerber, Review of Scientific Instruments **76**, 103701 (2005).
23. A. Livshits and A. Shluger, Phys. Rev. B **56**, 12482 (1997).
24. R. Bennewitz, M. Bammerlin, M. Guggisberg, C. Loppacher, A. Baratoff, E. Meyer, and H.-J. Güntherodt, Surf. Interface Anal. **27**, 462 (1999).
25. M. Dienwiebel, G. Verhoeven, N. Pradeep, J. Frenken, J. Heimberg, and H. Zandbergen, Phys. Rev. Lett. **92**, 126101 (2004).
26. G. Cross, A. Schirmeisen, A. Stalder, P. Grütter, M. Tschedy, and U. Dürig, Phys. Rev. Lett. **80**, 4685 (1998).
27. T. Nakagawa, K. Ogawa, and T. Kurumizawa, Journal of Vacuum Science & Technology B (Microelectronics and Nanometer Structures) **12**, 2215 (1994).
28. C. Frisbie, L. Rozsnyai, A. Noy, M. Wrighton, and C. Lieber, Science **265**, 2071 (1994).
29. L. Howald, R. Lüthi, E. Meyer, P. Güthner, and H.-J. Güntherodt, Z. Phys. B **93**, 267 (1994).
30. G. Leggett, N. Brewer, and K. Chong, "Phys. Chem. Chem. Phys." **7**, 1107 (2005).
31. U. Schwarz, O. Zwörner, P. Köster, and R. Wiesendanger, Phys. Rev. B **56**, 6987 (1997).
32. W. Ducker, T. Senden, and R. Pashley, Nature **353**, 239 (1991).
33. J. Hoh and A. Engel, Langmuir **9**, 3310 (1993).
34. W. Maw, F. Stevens, S. Langford, and J. Dickinson, Journal of Applied Physics **92**, 5103 (2002).
35. Q. Liang, O. Tsui, Y. Xu, H. Li, and X. Xiao, Physical Review Letters **90**, 146102 (2003).
36. E. Riedo, F. Levy, and H. Brune, Phys. Rev. Lett. **88**, 185505 (2002).
37. F. Tian, X. Xiao, M. Loy, C. Wang, and C. Bai, Langmuir **15**, 244 (1999).
38. R. Szoszkiewicz and E. Riedo, Physical Review Letters **95**, 135502 (23 Sept. 2005).
39. J. Gao, W. Luedtke, D. Gourdon, M. Ruths, J. Israelachvili, and U. Landman, Journal of Physical Chemistry B **108**, 3410 (2004).
40. E. Meyer, R. Lüthi, L. Howald, M. Bammerlin, M. Guggisberg, and H.-J. Güntherodt, J. Vac. Sci. Technol. B **14**, 1285 (1996).
41. M. Ruths, N. Alcantar, and J. Israelachvili, Journal of Physical Chemistry B **107**, 11149 (2003).
42. A. Socoliuc, R. Bennewitz, E. Gnecco, and E. Meyer, Phys. Rev. Lett. **92**, 134301 (2004).
43. R. Cannara, M. Brukman, and R. Carpick, Rev. Sci. Instr. **76**, 53706 (2005).
44. A. Socoliuc, E. Gnecco, R. Bennewitz, and E. Meyer, Physical Review B (Condensed Matter and Materials Physics) **68**, 115416 (2003).
45. R. Overney, E. Meyer, J. Frommer, D. Brodbeck, R. Luethi, L. Howald, H.-J. Guentherodt, M. Fujihira, H. Takano, and Y. Gotoh, Nature **359**, 133 (1992).
46. M. Liley, D. Gourdon, D. Stamou, U. Meseth, T. Fischer, C. Lautz, H. Stahlberg, H. Vogel, N. Burnham, and C. Duschl, Science **280**, 273 (1998).
47. M. Marcus, R. Carpick, D. Sasaki, and M. Eriksson, Physical Review Letters **88**, 226103 (2002).

48. M. Kwak and H. Shindo, Physical Chemistry Chemical Physics **6**, 129 (2004).
49. J. Stiernstedt, M. Rutland, and P. Attard, Review of Scientific Instruments **76**, 83710 (2005).
50. T. Bouhacina, J. Aime, S. Gauthier, and D. Michel, Phys. Rev. B **56**, 7694 (1997).
51. E. Gnecco, R. Bennewitz, T. Gyalog, C. Loppacher, M. Bammerlin, E. Meyer, and H. Güntherodt, Phys. Rev. Lett. **84**, 1172 (2000).
52. S. Sills and R. Overney, Phys. Rev. Lett. **91**, 095501 (2003).
53. T. Kawagishi, A. Kato, Y. Hoshi, and H. Kawakatsu, Ultramicroscopy **91**, 37 (2002).
54. S. Maier, Y. Sang, T. Filleter, M. Grant, R. Bennewitz, E. Gnecco, and E. Meyer, Phys. Rev. B **72**, 245418 (2005).
55. A. Schirmeisen, L. Jansen, and H. Fuchs, Phys. Rev. B **71**, 245403 (2005).
56. N. Tambe and B. Bhushan, Nanotechnology **16**, 2309 (2005).
57. E. Tocha, T. Stefanski, H. Schonherr, and G. Vancso, Review of Scientific Instruments **76**, 83704 (2005).
58. X. Yang and S.S. Perry, Langmuir **19**, 6135 (2003).
59. R.H. Schmidt, G. Haugstad, and W.L. Gladfelter, Langmuir **19**, 10390 (2003).
60. M. Hoogeman, D. van Loon, R. Loos, H. Ficke, E. de Haas, J. van der Linden, H. Zeijlemaker, L. Kuipers, M. Chang, M. Klik, and J. Frenken, Review of Scientific Instruments **69**, 2072 (1998).
61. B. Gotsmann and U. Durig, Langmuir **20**, 1495 (2004).
62. P. Maivald, H. Butt, S. Gould, C. Prater, B. Drake, J. Gurley, and P. Hansma, Nanotecnology **2**, 103 (1991).
63. T. Göddenhenrich, S. Müller, and C. Heiden, Rev. Sci. Instr. **65**, 2870 (1994).
64. J. Colchero, M. Luna, and A. Baro, Appl. Phys. Lett. **68**, 2896 (1996).
65. R. Carpick, D. Ogletree, and M. Salmeron, Appl. Phys. Lett. **70**, 1548 (1997).
66. H.-U. Krotil, T. Stifter, and O. Marti, Applied Physics Letters **77**, 3857 (2000).
67. G. Haugstad, Tribology Letters **19**, 49 (2005).
68. M. Reinstadtler, U. Rabe, V. Scherer, U. Hartmann, A. Goldade, B. Bhushan, and W. Arnold, Applied Physics Letters **82**, 2604 (2003).
69. L. Huang and C. Su, Ultramicroscopy **100**, 277 (2004).
70. M. Reinstadtler, T. Kasai, U. Rabe, B. Bhushan, and W. Arnold, Journal of Physics D: Applied Physics **38**, 269 (2005).
71. S. Morita, R. Wiesendanger, and E. Meyer, *Noncontact Atomic Force Microscopy*, *NanoScience And Technology* (Springer, Berlin, Germany, 2002).
72. S. Jarvis, H. Yamada, K. Kobayashi, A. Toda, and H. Tokumoto, Appl. Surf. Sci. **157**, 314 (2000).
73. O. Pfeiffer, R. Bennewitz, A. Baratoff, E. Meyer, and P. Gruetter, Phys. Rev. B **65**, 161403 (2002).
74. F. Giessibl, M. Herz, and J. Mannhart, Proc. Natl. Acad. Sci. USA **99**, 12006 (2002).
75. S. Kawai, S.-I. Kitamura, D. Kobayashi, and H. Kawakatsu, Applied Physics Letters **87**, 173105 (2005).

2 Surface Forces Apparatus in Nanotribology

Carlos Drummond and Philippe Richetti

Centre de Recherche Paul Pascal. CNRS-Université Bordeaux 1,
Avenue Albert Schweitzer, 33600 Pessac, France

2.1 Introduction

The measurement of the force of interaction between solids dates back to the 1920s, when Tomlinson investigated the interaction between crossed filaments of different metals [1]. Later, research groups in the Netherlands and Russia led by Overbeek and Derjaguin developed different techniques for measuring the force between surfaces of quartz or glass as a function of their separation [2, 3]. The example of these seminal pieces of work was promptly followed by many other groups. Particularly in Cambridge a remarkable body of work was accomplished, leading to the development of the Surface Forces Apparatus, *SFA*, by Tabor, Winterton and Israelachvili [4, 5].

The study of lateral forces between surfaces has a longer history. The problem of friction between surfaces attracted great thinkers as Da Vinci, Coulomb, Euler, Amontons and many others. A fascinating historical account of the history of tribology was compiled by Dowson [6]. The friction phenomena have also been investigated with *SFAs* modified for that purpose; the first friction measurements using this technique date back to the 1970's [7]. A decade later Briscoe and Evans reported extensive results on the study of friction of adsorbed monolayers in air [8]. Nevertheless, it was not until the late 1980's that nanotribology studies with the *SFA* became a very active field of research. Since then, various modifications to the technique have been introduced. In the present chapter we describe the principles of operations and some experimental details of the SFA-nanotribometer.

2.2 Surface Forces Apparatus Technique: Generalities

In a typical *SFA*-nanotribometry experiment molecularly smooth mica surfaces are glued to cylindrically curved silica lenses, and used to confine thin films. The use of mica as a substrate for surface force experiments was originally advanced by Bailey and Courtney-Pratt [9]. The cylindrically shaped silica disks are placed with their axes perpendicular to each other, a configuration that presents several advantages. First, it circumvents the difficult – if not impossible – task of accurately aligning two parallel plates. Unwanted

edge effects are easily avoided by this approach. Second, it allows the investigation of different contact spots on the same pair of surfaces, simply by laterally displacing the crossed cylinders. If wear or contamination of the surfaces appears during the experiment, a fresh contact zone can be readily found. Finally, this geometry is convenient for comparing the results of the measurements with theoretical descriptions, typically sketched for flat surfaces. If the separation between the curved surfaces is much smaller than their radii of curvature, R, the *SFA* cross-cylinder configuration is equivalent to a sphere-on-plate contact. The force between two such surfaces, F, can be related to the energy of interaction between flat surfaces per unit area, E, by using the so-called Derjaguin approximation [10], $E=F/2\pi R$. This provides a normalization method in order to quantitatively compare data from different experiments. The question of the normalization of the measured interaction forces is more involved in friction experiments, as will be discussed below.

One of the major strengths of the *SFA* technique rests on the possibility of imaging the area of contact to determine the distance between the surfaces, the refractive index of the film confined between the surfaces and the geometry of the contact region. The *SFA* is one of few techniques in the field of tribology that allows to image in situ and in real time the geometry of the contact area, and probably the only one with subnanometric resolution. Multiple Beam Interferometry (*MBI*) is used for this purpose [11]. A highly reflective layer is deposited on the back side of the mica surfaces, and white light is passed through this built-in Fabry-Perot interferometer. The intensity of the light transmitted through the stratified media between the two mirrors depends on the optical thickness in a nontrivial way: only wavelengths that interfere constructively after the multiple reflections in the cavity traverse the multilayer system. The emerging beam of light can then be focused on a spectrometer. The resulting constructive interference fringes (Fringes of Equal Chromatic Order, *FECO*) carry with them the information about the thickness and the refractive index of the different layers in the path of the light. Particularly, the thickness and the refractive index of the film confined between the mica surfaces can be determined with an accuracy of 0.1 nm and 0.01 respectively. Israelachvili developed simple explicit expression to calculate these quantities from the wavelength of the *FECO* for a film confined between symmetric mica surfaces [12]. Later, the analysis has been extended to asymmetric, adsorbing, anisotropic or more complicated multilayer systems [13–15]. The potential of extending the analysis to obtain information about the roughness of the surfaces has also been demonstrated [16].

Many different experimental setups for the measurement of the surface forces have been reported. One of the oldest versions, the Mk I, was designed by Israelachvili and Adams for the measurement of forces between liquids and vapours [17]. It was based on the earlier designs of Tabor, Winterton and Israelachvili [4, 5], and was later followed by greatly improved and modified versions, the Mk II and III [18, 19]. Parker and co-workers developed later

a circular steel apparatus (Mk IV) which is simpler to clean and assemble than Mk I or II [20]. The stability and reliability of the apparatus, as well as the simplicity of handling, have been progressively improved on each design. The interested reader is referred to the original publications for the particular details of each apparatus.

For measuring the normal force of interaction in a typical *SFA* experiment, one of the surfaces is displaced using a combination of motors and piezoelectric elements, while the other surface is coupled to a calibrated spring with a fix end. Double cantilever springs are typically used in order to minimize the tilting and/or sliding between the surfaces when the spring is deflected. The interaction force between the surfaces is measured by progressively displacing the fix end of the double cantilever spring and allowing the separation between the surfaces to come to an equilibrium situation, where the surface forces are balanced by the elastic force of the spring. The difference between the displacement imposed to the surface and the actual change in the separation between the surfaces, Δx (measured by *MBI*) will correspond to the deflection of the spring. It will be used to calculate the interaction force simply by means of the Hooke's law, $F = k\Delta x$, where k is the elastic constant of the double cantilever spring. Albeit being conceptually simple, measuring the forces by this procedure is limited by spring instabilities. Quickly decaying forces with a force-distance gradient larger than the spring constant are inaccessible because of mechanical instability of the system under such conditions. Derjaguin and co-workers proposed the use of a force feedback technique to overcome this problem [3]. The idea is to control the force applied to the surfaces independently of the displacement. An external force is applied to the surfaces to maintain the spring undeflected. Effectively, this translates into a continuously changing spring constant, which eliminates the mechanical instability above mentioned. Several implementations of this idea have been reported [21, 22]. Steward and Parker modified a Mark IV by incorporating a magnetic force transducer and a bimorph displacement sensor. Tonck et al. introduced a feedback apparatus with capacitive displacement transducers [23]. An interesting description of the different techniques used for the measurement of the normal force between surfaces was presented by Lodge [24].

2.3 Surface Forces Apparatus Nanotribometer

In a nanotribology experiment with the *SFA*, the mica surfaces are brought to a certain separation, T. By using motors or electomechanical transducers a lateral displacement between the surfaces is imposed, and the force induced by this displacement is measured. Usually a certain normal load is applied, L. If the load is high enough, the glue layer under the surfaces undergoes elastic deformation, and a thin film is confined to a flat circular region of uniform thickness T and area of contact A, as illustrated in the Fig. 2.1. By using

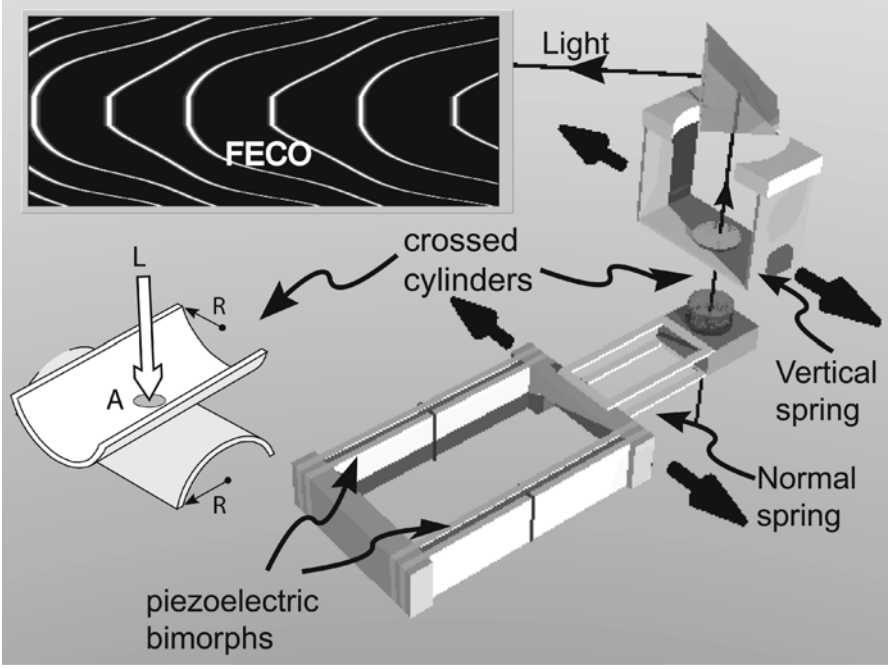


Fig. 2.1. Functional scheme of the *SFA* designed by Israelachvili and coworkers configured for friction experiments. The mica sheets are mounted in crossed-cylinder geometry, and their back surfaces are coated with reflective silver layers to allow for multiple beam interference. The upper and lower surfaces are mounted on cylindrically curved silica discs which are attached to the friction sensing device and the piezoelectric bimorph slider, respectively

MBI an image of the surfaces in contact can be obtained as the surfaces are slid, allowing monitoring of the size and the profile of the contact area and the distance between the surfaces, by observing the flat region on the *FECO*. Shear-induced elastohydrodynamic deformation can also be distinguished. In addition, damage of the surfaces can be easily detected as soon as it occurs, allowing to discriminate between undamaged sliding and friction with wear, and to independently study the two scenarios.

A subject of major importance in the analysis of a *SFA*-nanotribology experiment is to identify the area over which the frictional force takes action. Often the friction force between sliding surfaces will be dominated by the flat contact area. In that case, the sharp edge of the *FECO* allows recognizing the "area of contact", used to normalize the measured force and to calculate the shear stress. This operation is necessary to quantitatively compare the results of different experiments. From this point of view, the customary used friction coefficient is a less fundamental parameter than the shear stress. There is, however, an important caveat to this operation: very often the measured shear

stress depends on the applied pressure. Given that curved surfaces are used in a *SFA* experiment, the normal pressure is not constant over the flatten area. Its value is given by a nonlinear function of the position in the contact area, a problem that has been extensively treated by the contact mechanics community [25]. It is clear then that the shear stress calculated in a *SFA* experiment is an average quantity, to be treated with caution.

An even more complicated scenario is found when there is a significant contribution to the friction force by regions of the surfaces outside the flatten area. This situation can be envisaged, for example, if there is a contribution to the frictional force coming from breaking and reforming bonds of long molecules that are able to bridge the two surfaces together. In that case, there is not an obvious way to identify the effective contact area. One possibility is to adopt a cut-off length, and to assume that the contribution to the frictional force is negligible at larger separations. However, at least two problems persist: the choice of the characteristic length rests somehow arbitrary and the contribution of a given region to the total force will most likely be a function of the local surface separation. It is important to emphasize at this point that the experimental difficulties just outlined are shared by most –if not all– the experimental techniques in nanotribology. Besides, the *SFA*-nanotribometer in its interferometric version is possibly the only technique in nanotribology that allows the observation of the contact geometry while rubbing the surfaces.

2.3.1 Experimental Setup

Several *SFA* experimental setups have been proposed during the last two decades, each with its own capabilities and limitations. In the following we will briefly describe few systems which are broadly used in the field. The reader interested in more complete information is referred to the original papers.

A number of experimental designs have been proposed by the group of Israelachvili [26, 27]; they are all in principle based on the original method proposed by Israelachvili and Tabor [5]. The main features of the most recent version are illustrated in Fig. 2.1. The lower surface is mounted on a bimorph-driven slider [27], which moves laterally in a linear fashion when a constant slope voltage ramp is applied between the two electrodes of sectored piezoelectric bimorphs (electromechanical transducers). Alternatively, a constant frequency sinusoidal input can be imposed to the slider to perform nanorheological experiments. The upper surface is itself attached to a vertical double cantilever spring, whose deflection is monitored using strain gauges connected to form the arms of a Wheatstone bridge. If the displacement of the lower surface induces a viscous or friction force on the upper surface, the vertical spring will deflect. From the deflection of this spring of known spring constant K, the friction force between the surfaces F can be calculated, simply by using Hooke's law of elasticity [26]. The mechanical

properties of the measurement system (e. g., compliance and inertial mass) will influence the results; these factors have to be taken into account in order to obtain meaningful information from the signal measured. This can be done in a straightforward fashion in the *SFA* because of its mechanical simplicity and easy-to-characterize mechanical properties.

The maximal distance that can be slid with this setup depends on the characteristics of the bimorph strips used, being typically of the order of several tens of micrometers. Otherwise, larger displacements can be achieved by mechanically driving the upper surface using a reversible, variable speed motor-driver micrometer shaft that displaces the translation stage holding the vertical double cantilever spring. The detection limit for the friction force of this setup is typically of the order of several μN. By changing the frequency and the amplitude of the input signal to the bimorph slider, the driving speed can be typically varied between several Å/s to 0.1 mm/s. This device has been used to study a large number of systems. Some examples included confined simple liquids [28–30], polymer melt and solutions [26, 27, 31–33], self-assembled surfactant and polymer layers [34–36].

An alternative design conceived for the study of smaller deformations was introduced by Granick and coworkers [37, 38]. The goal of these low amplitude studies is to focus the investigation on the linear response of the confined films. By applying small deformations, the flow of fresh liquid in the contact zone is avoided. This allows the study of long time relaxation process that may be occurring in the contact region. A schematic of this device is illustrated in Fig. 2.2. In this design, the bottom surface remains stationary, while the upper surface is mounted on a holder attached to a double cantilever. Compared with the design of Israelachvili, they replaced the vertical metallic cantilevers by two piezoelectric bimorph strips. One of the bimorphs is used as an actuator and the other as a sensor. In the experiment, a voltage difference is applied to one of the bimorphs to bend it. Typically, a constant frequency sinusoidal signal is used, inducing an oscillating force on the bimorph. Simultaneously, the deformation-induced voltage of the second bimorph is measured. This data is used to determine the actual displacement of the surface. By comparing this response with the one observed when no interaction between the two surfaces is presented, the influence of the confined film on the moving surface can be extracted. The electromechanical characteristics of the system are model as a series of effective masses, springs and dashpots representing the different components of the apparatus. The friction appears as a force acting on the holder of the lower surface, from which an effective viscosity can be extracted [38]. Although mainly conceived for the study of small deformations (of the order of the film thickness) typical displacements range from few nm to few μm. The reported sensitivity on the friction force is around 5 μN. Many different systems have been explored with this device, including simple liquids [37, 39–41] polymer melts [42–44] and solutions [45, 46].

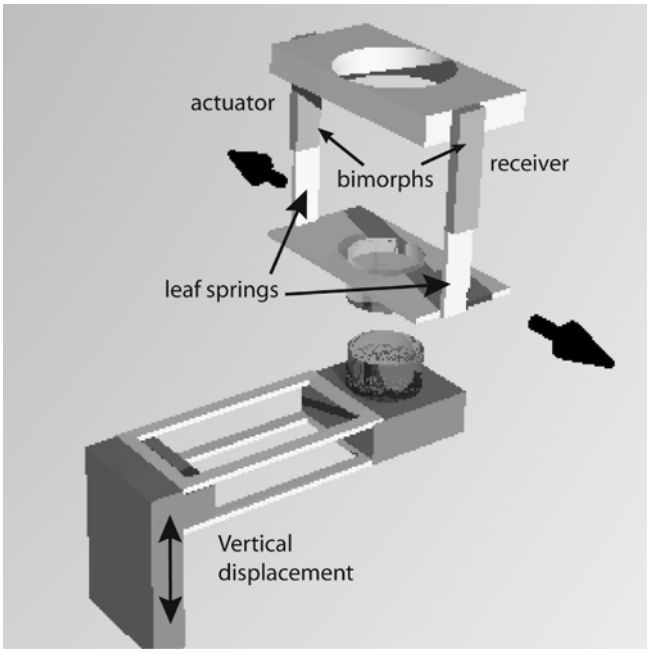


Fig. 2.2. Schematic illustration of the *SFA* designed by Granick and coworkers. The shear force is generated by one bimorph (actuator) and the response of the device induces a voltage across the other bimorph (receiver)

A third experimental setup widely used in the literature has been developed by Klein and co-workers [47,48]. A schematic of this device is presented in Fig. 2.3. In this device the sensitivity to the measured friction forces is greatly improved with respect to the previous designs. Inversely to the configuration adopted by Israelachvili and co-workers, in the most recent version of their design the upper surface is driven and the effect on the lower surface is measured. A sectored piezoelectric tube is use to produce a normal or lateral displacement of the upper surface. An air-gap capacitor is used to measure the lateral displacement of the lower surface, which is coupled to a calibrated double cantilever spring with a fix end. The shear induced frictional force is then directly extracted from this displacement (e. g., the deflection of the spring) by using Hooke's law. The reported sensitivity of the friction force is 50 nN and the maximal displacement of the upper surface is few tens of μm. The improved sensitivity of this device comes from the detection method used. This has proven to be very valuable for the study of polymer melts and solutions [49–52] where small forces are typically observed. Research on water and other simple liquids has also been conducted with this experimental setup [53, 54].

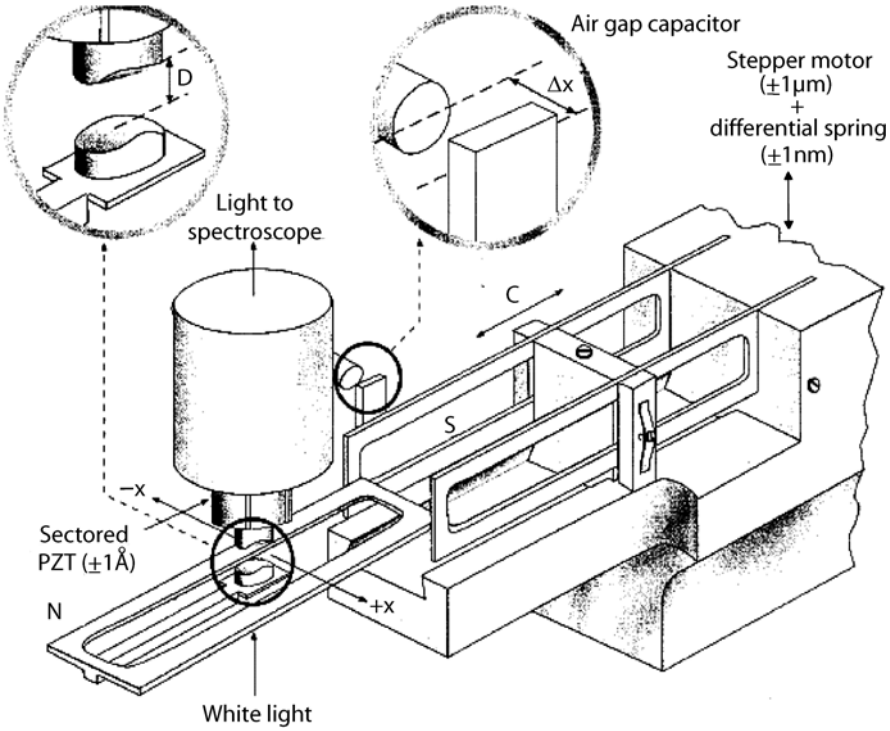


Fig. 2.3. Schematic illustration of the surface force balance (*SFB*) designed by Klein and coworkers. The separation between the surfaces is controlled via a three-stage mechanism with a sectored piezoelectric tube on which the top surface is mounted. The piezoelectric element produces both normal and lateral displacement. The bending of the shear force spring is detected by an air-gap capacitor. Reprinted with permission from The *Journal of Physical Chemistry B* Volume **105(34)**, 8125–8134 (2001). Uri Raviv, Rafael Tadmor and Jacob Klein

Further improvements on different aspects of the experimental technique have been proposed during the last decade. First, substantial efforts have been devoted to automate the procedure of measurement, in order to improve the accuracy and simplicity of the technique. Second, the strategy used to determine the separation between the surfaces has been extended by using non-interferometric techniques. Recently, Qian and co-workers extended the capabilities of the apparatus to include the movement of the surfaces and the measurement of the friction force between them in two orthogonal directions in the plane of contact [55]. This new apparatus should prove to be useful in investigating shear-induced effects (e. g. shear alignment) on the confined thin films.

The automatic detection of the *FECO* have posed some challenges in the past, but increasingly accurate and affordable charge couple devices in the

market are currently used in several laboratories in the world for this purpose. Different strategies for the automation of the measurement process have been extensively described by Quon et al. [56], Grunewald and Helm [57], and more recently by Heuberger and coworkers [58].

As mentioned before, some efforts have also been devoted to determine the surface displacement by noninterferometric techniques. Several groups have proposed to use piezoelectric bimorphs for this purpose [59–61]. This solution is inadequate for long or quasi-static measurements due to the intrinsic drift and leakage of bimorph sensors (electrical drift and decay). A partial solution to these problems was presented by Parker [62], who suggested the use of an ultra-high impedance amplifier to lengthen the decay time of the bimorph sensor. A different method was presented by Frantz and coworkers [63]. They proposed to monitor the capacitance between the silver layers deposited on the back surface of the mica sheets for a fast detection of the surface separation, and described the use of this setup for the study of contact mechanics. Later Stewart proposed to use capacitance dilatometry to measure the separation between the surfaces: one plate of a parallel-plate capacitor was attached to the moving surface, and the other to the chamber of the apparatus [64]. This technique allows a fast and accurate measurement of the displacement of the surface and eliminates the constraint of having to use transparent surfaces. Nevertheless, if opaque surfaces are used it is impossible to obtain an image of the contact region while shearing, which is one of the major strength of the SFA technique.

2.3.2 Local Structural Information: Combination of the SFA with other Techniques

The information gathered in a conventional *SFA* experiment is limited to the average response of the confined film under shear and compression. For achieving a better understanding of the behavior of confined films under shear, it is desirable to obtain structural information at the molecular level. Obtaining this information implies a colossal experimental challenge. On one hand the number of molecules involved in a thin film is relatively small, particularly when a localized area is explored, which inevitably reduces the intensity of any measured signal. On the other hand the investigated thin film is surrounded by layers of different materials that are susceptible to interact with the used probe (e. g. light, x-ray or neutrons) increasing the level of noise of the measured signal. Despite of these difficulties, several experimental groups have reported encouraging results of experiments combining the capabilities of the *SFA* with other techniques in situ. It is reasonable to expect that techniques revealing the local molecular properties of the confined films will improve our understanding about the friction phenomena in the years to come.

The earliest effort in this direction aimed to combine *SFA* with x-ray diffraction (*XSFA*) [65,66]. The second generation of *XSFA* combines the force

measurement capabilities of the *SFA* with in situ small angle x-ray scattering experiments [67]. Films of several liquid crystals have been studied with this technique, and the effect of shear on the molecular alignment in confined films has been directly evidenced. The application of this technique has so far been limited to films thicker than 500 nm, mainly because of the poor signal to noise ratio obtained otherwise. Obviously, the research in nanotribology calls for much thinner films. Moreover, the results reported with this method have been limited to an average investigation of the contact area, because of the size of the x-ray probe used. The possibility of investigating small regions of the contact area by using a micro focused x-ray beam has been suggested, but no results in this direction has been reported so far. Despite of the difficulties encountered, several research groups are actively working in this technique. It has been shown recently that X-ray reflectivity can be used to obtain structural information of ultra confined molecular films only few molecular layers thick [68].

Helm and coworkers [69] showed that *MBI* can be used to obtain structural information of the confined thin films without any modification to the original *SFA* technique. Information about orientation and intermolecular interactions can be extracted from the *FECO* if optically active molecules are investigated. They were able to study ultra thin films, given that the light absorption by the confined molecules is enhanced by the multiple reflections in the optical cavity. Local information in the contact area can also be obtained. Nevertheless, this technique limits the molecules that can be studied to large dye molecules. In addition, for best determination of the adsorption spectra of the confined dyes, relatively thick mica has to be used, reducing the accuracy of the film thickness determination.

In other order of ideas, Salmeron and coworkers suggested to couple second harmonic and sum-frequency generation to the *SFA* to study alignment and relaxation of confined ultra thin films, and showed the potential of the application by investigating self-assembled and langmuir blodget monolayers of several surfactants [70]. However, results with other experimental systems have not been reported so far.

The combination of the *SFA* with other optical techniques has been limited by the reflective silver layer used to determine the surface separation by *MBI*. This layer strongly reduces the intensity of the illumination of the confined films, seriously limiting the in situ performance of other optical methods. Granick and coworkers have overcome this limitation by replacing the reflective silver layer by multilayer dielectric coatings, which are transparent in different regions of the optical spectrum. In that way, they have been able to apply different spectroscopic tools to obtain in situ structural information of ultra thin films under shear [71–73].

By combining the *SFA* nanotribometer with fluorescence correlation spectroscopy they measured the molecular diffusion coefficient in thin films within spots of submicron size, obtaining spatially resolved measurements [71]. This method has the drawback that fluorophore molecules have to be added to the

liquid investigated in order to have a fluorescence signal. However, the authors have shown that the small amount of fluorophores added didn't modify they behavior under shear and compression of the fluids investigated. Thus, diffusion coefficient of rhodamine in 1,2 propane diol was found to decrease by 2 orders of magnitude under confinement. Similar results were observed for the diffusion of cumarin 153 in OMCTS. They also found that the diffusion coefficient decreases from the edges towards the center of the contact region. Their results seem to suggest a heterogeneous dynamic in the confined thin films, where the diffusion appears to involve cooperative rearrangements of many molecules.

They have also reported results on the combination of *SFA* with other techniques, as Confocal Raman Spectroscopy [72]. By using confocal geometry the authors avoided the problem of contribution of the bulk to the scattered signal. By using a multilayer reflective coating transparent to the argon laser and to the scattered Raman signal, they were able to monitor the geometry of the contact area simultaneously with the Raman scattering signal. They reported spatially resolved Raman scattering before and after shear, evidencing the influence of shear on the orientation of the molecules inside the confined film.

A third technique developed in Granick's group is the combination of photoluminescence and absorption dichroism with the *SFA* [73]. The shear-induced alignment of pre-adsorbed polymer molecules on mica was quantified both by photoluminescence and absorption spectra. They found alignment parallel and perpendicular to the shear direction. The alignment direction seems to be extremely sensitive to small changes in the initial conditions that the authors were not able to characterize. Although this technique is limited to the investigation of optical active substances, the information obtained can help to understand the behavior of lubricants with similar molecular structure.

In a different direction, Berg and coworkers recently suggested incorporating a Quartz Crystal Resonator in the *SFA* [74]. Because of the high oscillation frequency of the Quartz Crystal, this configuration allows the study of sliding velocities much higher than typically investigated in a conventional *SFA* nanotribology experiment. Nevertheless, in order to obtain meaningful results extremely thin mica surfaces need to be used, complicating its implementation as a routine technique. In addition, a sphere-on-plate geometry is required, which complicates the procedure of preparation of the mica surfaces.

2.3.3 Beyond Mica: Alternative Substrates

As mentioned previously, mica surfaces are the most popular substrates for *SFA* experiment. It gathers a set of properties seldom observed in other materials. It is transparent and can be prepared in the form of thin sheets of molecularly smooth surfaces over large areas by successive cleaving. The combination of these properties is at the heart of the *SFA* technique: transparent

surfaces are required to determine the geometry of the contact by *MBI*. Besides, mica is a fairly incompressible material, so the forces measured are not flawed by the deformation of the surface. In addition it is inert to chemical reaction, so it is hardly modified during experiments. As a drawback, the process of producing mica surfaces thin enough to fulfill the requirements of the *SFA* technique calls for a skillful experimentalist. This constraint is greatly relaxed by carefully implementation of automatic thickness measurement. On the other hand, thanks to the smoothness of the surfaces the geometry of the contact between the two surfaces can be completely described, simplifying the description and interpretation of the results.

The investigation of substrates other than mica is of interest for obvious reasons. The substrate plays a major role in most of the phenomena investigated by *SFA*, and particularly in tribology. It acts not only as a geometrical barrier, but as a major player: the interaction between the surfaces and with the confined films determines the general frictional behavior. For these reasons a considerable effort has been devoted to investigate the possibility of modification or replacement of the mica surfaces by other substrates, in order to expand the range of applications of the technique. Mica surfaces can be modified by deposition or adsorption of different materials. By properly controlling the modification process, the smoothness of the surfaces can be preserved. In addition, by the deposition of a smooth layer on the mica surfaces it is possible to modify the surface energy of the substrate, preparing surfaces that may be more prone to chemical modifications.

Several groups have investigated the behavior of mica surfaces modified by self-assembly [35,36,75,76] or deposition of Langmuir-blodgett films [35,77,78] of different substances. Mica acquires a negative surface charge when immersed in water, so positively charged species (e. g. cationic surfactants) spontaneously adsorbed on it; the structure of the adsorbed layer and its relationship to the molecular structure of the adsorbed material has been a very active area of research during the last fifteen years [79]. The frictional behavior of the modified surfaces depends strongly on the characteristic of the adsorbed layers: surface properties like the adhesion energy and the morphology of the adsorbed layer will ultimately determine their behavior under shear.

As mentioned before, Mica is in general an inert material. Nevertheless, it can be chemically modified by water vapor plasma treatment, increasing their reactivity to different species, e. g. chlorosilanes, as suggested by Parker and coworkers [80]. In this way, molecularly smooth hydrophobic surfaces can be prepared, given that the chemical structure is modified without increasing the roughness of the substrates. Mica surfaces treated by this procedure have been used in SFA studies [81]. Kessel and Granick modified this procedure to be able to induce the self-assembly of alkoxysilanes on mica, showing that strongly bound monolayers were formed [82].

Several groups have proposed to modify the mica surfaces simply by depositing on them thin films of different materials, including metals and di-

electrics. In order to be able to monitor the geometry of the contact region by *MBI*, it is important for the deposited layers not to be completely opaque. This does not impose a serious limitation for sufficiently thin films. The interpretation of the *FECO* becomes more involved because of the larger number of optical layers in the optical path of the white light, but the information about the thickness and optical properties of the confined film can nonetheless be extracted. Different algorithms which are adequate for the modified experimental conditions have been described in the recent years [83, 84].

Studies of mica modification by deposition of many different materials have been reported in the past. Silver [83, 85], gold [83, 86], platinum [86, 87], silica [87, 88], are only a few of a long list of materials investigated. Horn and coworkers grown single crystals of aluminum oxide [89] by vapor phase condensation. The tribological behavior of these surfaces was later investigated by Berman and coworkers [90]. Vigil and coworkers deposited smooth layers of amorphous silica on mica, and study the behavior under compression and shear of the resulting surfaces [88]. They found that oscillatory structural forces were absent of the interaction between the surfaces. In addition, they observed long time-dependent adhesion and friction of the surfaces in the presence of water. Mc Guiggan and coworkers deposited amorphous carbon by magnetron sputtering on mica, and used these surfaces in the *SFA*-nanotribometer [91]. They found the friction force to be proportional to the area of contact between the surfaces, and the measured shear stress to decrease strongly with increasing relative humidity. Hirz and coworkers sputtered thin films of zirconia and alumina on mica, and investigated the behavior of these surfaces when lubricated with a linear perfluoropolyether [92]. They showed that these metal oxide formed smooth films on mica susceptible of being used as alternative substrates in *SFA* experiments.

Other groups have proposed to simplify the method of substrate preparation by eliminating the use of mica all together. A method of preparing silica surfaces for use in the *SFA* was proposed by Horn and coworkers, although its use has not became widespread [93]. Golan et al. proposed to deposit a thin layer of silicon nitride on rigid silica disks previously coated with a reflective layer to replace the mica substrates [94]. They also reported a succinct tribological study of this generic substrate.

In general, the surface modification processes abovementioned may alter the smoothness of the surfaces at some degree complicating the geometry of the system, changing it from a single-asperity to a multiple-asperity contact. However, they allow the investigation of surfaces of interest in many different fields, extending the range of applications of the *SFA*. In addition, in most of the cases the roughness of the deposited layers can be controlled and/or modified to certain extent, allowing the investigation of the effect of surface roughness on friction, an important field of research on its own. *SFA* studies involving controlled roughness are in progress in several laboratories in the world.

2.4 Case Study: Weakly Adhesive Surfaces under Shear

To illustrate the potential of the *SFA* technique for nanotribology studies, some experimental results obtained with self-assembled surfactant layers are described in this section. Another interesting example is presented later in the book by Mugele.

We investigated the following system: the mica surfaces in the *SFA* are immersed in bulk aqueous surfactant solutions. Cationic surfactants are chosen, so that self assembled layers are formed on the mica surfaces. For surfactant concentrations above the critical micelle concentration (cmc) the adsorbed films show different morphology depending on the surfactant. Some surfactants adsorb as flat bilayers, while others form rather modulated layers, suggesting the adsorption of globular or cylindrical micelles [79]. If two flat bilayers are compressed, eventually the hemifusion of the layers can be induced. In the hemifused region the mica surfaces end up cover by a monolayer of surfactant, and the surfaces are held together by an adhesive interaction, because of the hydrophobic attraction between the hydrophobic chains of the surfactant molecules. The precise measurement of the thickness of the trapped layer allows the clear identification of the hemifusion; an abrupt change of the confined film, corresponding to the expulsion of two monolayers from the contact region, is induced by compression and/or shear [36, 75, 95].

The behavior under shear of these systems is very complex. In general, when two intact bilayers are sheared, we do not detect any frictional resistance at any applied velocity or normal load: the friction force is below the detection limit of our experimental setup, which is similar to the one designed by Israelachvili [26, 27]. On the contrary, after the hemifusion of the layers is induced, a higher friction force can be observed. A typical friction trace measured during the hemifusion process is presented in Fig. 2.4, together with friction traces measured at different driving velocities after the hemifusion has taken place. The general behavior of the measured friction force with the driving velocity after hemifusion is illustrated in Fig. 2.5. At least 5 different regimes can be identified. At low velocities smooth sliding is observed. The force increases first linearly and then logarithmically with the driving velocity, before reaching a plateau. Above a certain critical velocity the movement becomes unstable and stick slip is observed. At even higher velocities the movement becomes again stable and a second smooth sliding regime is observed, when the frictional resistance increases linearly with the driving velocity. As can be observed in the Fig. 2.5, an extensive dynamic regime is necessary to be able to observe the five regimes just described. This exploration of the space of parameters can be readily performed with the *SFA* nanotribometer.

Within the experimental accuracy, the shear stress, defined as $\sigma = F/A$, appears to be independent of the normal load L over the range of load investigated, both along the plateau regime preceding the stick-slip instability, and for the high velocity smooth-sliding regime. This implies that the fric-

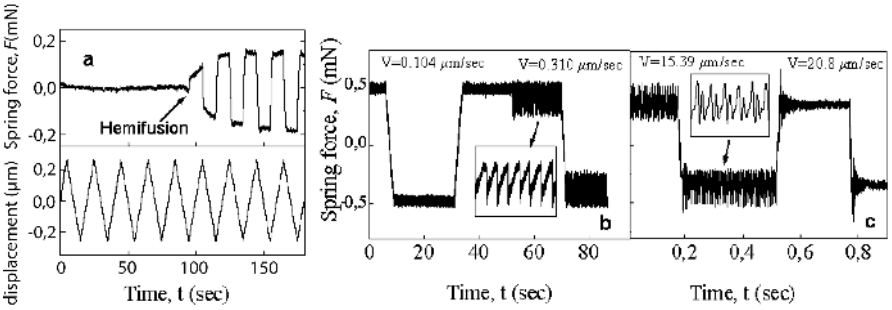


Fig. 2.4. a Friction signal recorded when the lower surface is displaced at constant velocity in a reciprocate mode, at the moment of the shear-induced hemifusion. A dramatic increase in friction force is accompanied by a film thickness reduction from 6.5 nm to 3.5 nm, indicating the hemifusion of the adsorbed bilayers. The normal load remains practically constant **b** A smooth sliding regime is observed at low velocities, $V < V_c$. In the stick-slip regime the friction force oscillates between the kinetic value F_k and a *lower* kinetic value F_{sk}. **c** Increasing the driving velocity the measured spring force changes from an oscillatory state to a smooth steady state. Every time the driving velocity is reversed there is a transient response of few hundredth of a second before the system reaches steady-state sliding

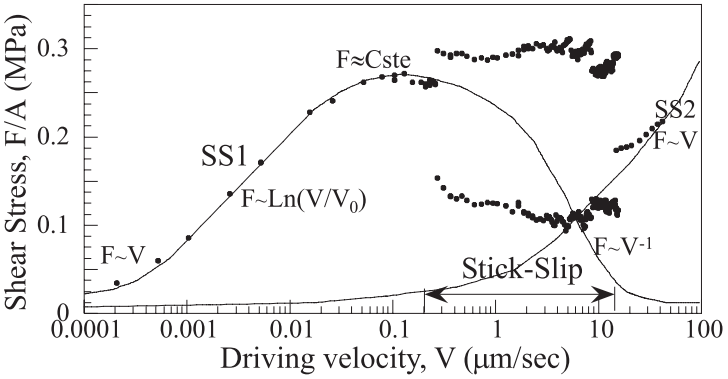


Fig. 2.5. Driving-velocity dependence of the spring force measured while shearing two adsorbed monolayers of the 12-3-12-3-12 surfactant under a load of $L = 4.51$ mN at $T = 20$ C. The smooth sliding to stick-slip transition occurs at $V_c \approx 0.3\,\mu$m/s. Prior to the transition, the kinetic stress σ_k levels off at V_1 after a logarithmic $\sigma - V$ dependence. The quasi-smooth regime persists up to the transition at V_c. At high driving velocities a new transition to a smooth-sliding regime is observed

tion force is proportional to the contact area A, rather than to the contact diameter or the load L. The load independence of the shear stress is no longer verified along the logarithmic regime. It is only due to the possibility of monitoring the real area of contact with the *SFA* (from the flat region on the

FECO) that the shear stress can be univocally calculated at all times during the experiment.

All the trends observed in the sliding curve can be described by a model originally proposed by Schallamach [96] and that we have extensively discussed in the past [36]. The general behavior of the friction force can then be interpreted in terms of a model based on the kinetics of formation and rupture of small adhesive links (bonds) between the two shearing surfaces. Under this scenario, the observed stick-slip regime is just a manifestation of the mechanical instability due to the negative slope of the force vs. velocity curve in a certain range of speeds. This adhesive model is insufficient to account for the steady smooth sliding regime observed at high velocities. A second contribution to the friction force (other than the elastic contribution) must be considered in order to re-stabilize the mechanical system in a kinetic state with finite friction. This extra contribution may be, for instance, the viscous dissipation in the trapped layer. The linear increase of the force at high sliding velocity seems to support this idea [36, 95].

This example illustrates the capabilities of the *SFA* nanotribometer. It shows how the measurement of the parameters pertinent to the friction problem, in particular the capability to monitor the geometry of the rubbing surfaces, greatly improves the understanding of the phenomena involved, allowing a quantitative comparison of the behavior of the system with theoretical models.

References

1. Tomlinson, G.A., *Phil. Mag.* **6**, 695–712 (1928)
2. Overbeek, J.T.G. and Sparnaay, M.J., *Discuss. Faraday Soc.* **18**, 12–24 (1954)
3. Rabinovich, Y.I., Derjaguin, B.V. and Churaev, N.V., *Adv. Colloid Interface Sci.*, **16(1)**, 63–78 (1982)
4. Tabor, D. and Winterton, R.H.S. *Proc. R. Soc. London, Ser. A*, **312**, 435–450 (1969)
5. Israelachvili, J.N. and Tabor, D. *Proc. R. Soc. London, Ser. A* **331**, 19–38 (1972)
6. Dowson, D. History of Tribology; Langman: London and New York, 1979
7. Israelachvili, J.N. and Tabor, D., *Wear* **24**, 386–390 (1973)
8. Briscoe B. J. and Evans, D.C.B., *Proc. R. Soc. London, Ser. A* **380,** 389–407 (1982)
9. Bailey, A.I. and Courtney-Pratt, J.S. *Proc. R. Soc. London, Ser. A* **227,** 500–515 (1955)
10. Derjaguin, B. V., *Kolloid. Z.* **69**, 155–64 (1934)
11. Tolansky, S. *Multiple Beam Interferometry of Surfaces and Films* (University Press, Oxford, 1948)
12. Israelachvili, J.N. *J. Colloid and Interface Sci.* **44(2)**, 259–272 (1973)
13. Horn, R.G. and Smith, D.T., *Applied Optics* **30(1)**, 59–65 (1991)
14. Müller, C., Mächtle, P. and Helm, C.A., *J. Phys. Chem.* **98(43)**, 11119–11125 (1994)

15. Heuberger, M. *Rev. Sci. Instrum.* **72(3)**, 1700–1707 (2001)
16. Heuberger, M.; Luengo, G.; Israelachvili, J.; *Langmuir* **13(14)**, 3839–3848 (1997)
17. Israelachvili, J, N., Adams, G. E. *J. Chem. Soc. Faraday Trans.* **74**:, 975–1001 (1978)
18. Israelachvili, J, N., *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4722–4724 (1987)
19. Israelachvili, J, N., McGuiggan, P.M., *J. Mater. Res.* **5(10)**, 2232–2243 (1990)
20. Parker, J.L., Christenson, H.K. and Ninham, B.W., *Rev. Sci. Instrum.* **60(10)**, 3135–3138 (1989)
21. Parker, J.L. and Stewart, A.M. *Prog. Colloid Polym. Sci.*, **88,** 162–8 (1992)
22. Stewart, A.M. and Parker, J.L., *Rev. Sci. Instrum.* **63(12)** 5626–5633 (1992)
23. Tonck, A., Georges, J.M. and Loubet, J.L., *J. Colloid and Interf. Sci* **126(1)**, 150–163 (1988)
24. Lodge, K.B. *Adv. Colloid Interface Sci.* **19**(1–2), 27–73 (1983)
25. Johnson K.L., *Contact Mechanics* (Cambridge University Press, Cambridge 1985)
26. Homola, A.M., Israelachvili, J.N., Gee, M.L. and McGuiggan, P.M. *J. Tribol.* **111**, 675 (1989)
27. Luengo, G., Schmitt, F.J., Hill, R., Israelachvili, J. *Macromolecules* **30(8)**, 2482–2494 (1997)
28. Gee, M., McGuiggan, P., Israelachvili, J. and Homola, A., *J. Chem. Phys.* **93(3)**, 1895–1906 (1990)
29. Israelachvili, J., McGuiggan, P., Gee, M.L., Homola, A., Robbins, M. And Thompson, P., *J. Phys. Condens. Matter* **2**, SA89-SA98 (1990)
30. Drummond, C. and Israelachvili, J., *Phys. Rev E* **63**, 041506 (2001)
31. Luengo, G., Israelachvili, J., and Granick, S., *Wear* **200(1–2)**, 328–335 (1996)
32. Schorr, P.A., Kwan, T.C.B., Kilbey II, S.M., Shaqfeh, E.S.G. and Tirrel, M., *Macromolecules* **36(2)**, 389–398 (2003)
33. Luengo, G.; Heuberger, M., Israelachvili, J., *J. Phys. Chem. B.* **104(33)**, 7944–7950 (2000)
34. Yoshizawa, H., Chen, Y.-L., Israelachvili, J., *Wear* **168(1–2)**, 161–166 (1993)
35. Yoshizawa, H., Chen, Y. L., Israelachvili, J. *J. Phys. Chem.* **97(16)**, 4128–4140 (1993)
36. Drummond, C., Israelachvili J., Richetti, P. *Phys. Rev. E* **67**, 066110 (2003)
37. Van Alsten, J. and Granick, S., *Phys. Rev. Lett.* **61(21)**, 2570–2573 (1988)
38. Peachey, J., Van Alsten, J. and Granick, S., *Rev. Sci. Instrum.* **62(2)**, 463–473 (1991)
39. Hu, H.- W., Carson, G.A. and Granick, S., *Phys. Rev. Lett.*, **66(21)**, 2758–61 (1991)
40. Zhu, Y. and S. Granick, S., *Phys. Rev. Lett.*, **87(9)**, 096104 (2001)
41. Demirel, A. L., Granick, S., *J. Chem. Phys.* **115(3)**, 1498–1512 (2001)
42. Granick, S.; Hu, H.-W., *Langmuir* **10(10)**, 3857–66 (1994)
43. Granick, S., Hu, H.-W. and Carson, G.A., *Langmuir* **10(10)**, 3867–3873 (1994)
44. Van Alsten, J. and Granick, S., *Macromolecules* **23(22)**, 4856–4862 (1990)
45. Zhu, Y.; Granick, S., *Macromolecules* **36(4)**, 973–976 (2003)
46. Ruths, M, Sukhishvili, S. A, Granick, S., *J. Phys. Chem. B.*, **105(26)**, 6202–6210 (2001)
47. Klein, J., Perahia, D. and Warburg, S., *Nature (London)* **352**, 143–145 (1991)
48. Raviv, U., Tadmor, R. and Klein, J., *J. Phys. Chem. B* **105(34)**, 8125–8134 (2001)

49. Tadmor, R., Janik, J., Klein, J. and Fetters, L.J., *Phys. Rev. Lett.* **91 (11)** 115503 (2003)
50. Raviv, U., Giasson, S., Kampf, N., Gohy, J-F., Jerôme, R. and Klein, J., *Nature* **425**, 163–165 (2003)
51. Eiser, E., Klein, J., Witten, T.A. and Fetters, L.J., *Phys. Rev. Lett.* **82(25)**, 5076–5079 (1999)
52. Klein, J., Kumacheva, E., Perahia, D., Mahalu, D. and Warburg, S., *Faraday Discussions* **98**, 173–188 (1994)
53. Klein, J. and Kumacheva, E, *J. Chem. Phys.* **108(16)**, 6996–7009 (1998)
54. Raviv, U., Laurat, P. and Klein, J., *Nature* **413**, 51–54 (2001)
55. Qian, L., Luengo, G., Douillet, D., Charlot, M., Dollat, X., Perez, E., *Rev. Sci. Instrum.* **72(11)**, 4171–4177 (2001)
56. Quon, R.A., Levins, J.M. and Vanderlick, T.K*., J. Colloid and Interf. Sci* **171(2)**, 474–482 (1995)
57. Grünewald, T. and Helm,C.A., *Langmuir* **12(16)**,3885–3890 (1996)
58. Zäch, M. Vanicek, J. and Heuberger, M., *Rev. Sci. Instrum.* **74(1)**, 260–266 (2003)
59. Van Alsten J and Granick S., *Tribology Transactions* **33**, 436–446 (1990)
60. Israelachvili J.N., Kott, S. J. and Fetters, L.J., *J. Polym. Sci. B* **27(3),** 489–502 (1989)
61. Dhinojwala, A. and Granick, S*., J. Chem. Soc., Faraday Trans.* **92(4)**, 619–623 (1996)
62. Parker, J.L., *Langmuir* **8(2)**,551–556 (1992)
63. Frantz, P., Agrait, N. and Salmeron, M., *Langmuir* **12(13)**, 3289–3294 (1996)
64. Stewart, A.M., *Meas. Sci. Technol.,* **11(3),** 298–304 (2000)
65. Idziak, S.H.J., Koltover, I., Israelachvili, J.N. and Safinya, C.R., *Phys. Rev. Lett.* **76(9)**, 1477–1480 (1996)
66. Idziak, S.H.J., Safinya, C.R., Hill, R.S., Kraiser, K.E., Ruths, M., Warriner, H.E., Steinberg, S., Liang, K.S., Israelachvili, J.N., *Science* **264(5167)**, 1915–1918 (1994)
67. Golan, Y., Martin-Herranz, A. Li, Y. Safinya, C.R. and Israelachvili, J., *Phys. Rev. Lett.* **86(7)** 1263–1266 (2001)
68. Seeck, O. H., Kim, H., Lee, D. R., Shu, D., Kaendler, I. D., Basu, J. K. and Sinha, S. K., *Europhys. Lett.*, **60(3)**, 376–382 (2002)
69. Mächtle, P., Müller, C., Helm, C. A., *J. Phys. II*, **4(3)**, 481–500 (1994)
70. Frantz, P., Wolf, F., Xiao, X.-d., Chen, Y., Bosch, S. and Salmeron, M., *Rev. Sci. Instrum.* **68(6)**, 2499–2504 (1997)
71. Mukhopadhyay, A., Zhao, J., Chul Bae, S. and Granick, S., *Rev. Sci. Instrum.* **74(6)**, 3067–3072 (2003)
72. Chul Bae, S., Lee, H., Lin, Z. and Granick, S., *Langmuir* **21(13)**, 5685–5688 (2005)
73. Chul Bae, S., Lin, Z. and Granick, S., *Macromolecules* **38(22)**, 9275–9279 (2005)
74. Berg, S., Ruths, M., Johannsmann, D., *Phys. Rev. E* **65**, 026119 (2002)
75. Richetti, P., Drummond, C., Israelachvili, J., In, M., Zana, R., *Europhys. Lett.* **55(5)**, 653–659 (2001)
76. Ruths, M.; Alcantar, N. A.; Israelachvili, J. N., *J. Phys. Chem. B* **107(40)**, 11149–57(2003)
77. Helm, C.A., Israelachvili, J.N. and McGuiggan, P.M., *Biochemistry* **31(6)**, 1794–1805 (1992)

78. Yamada, S., Israelachvili, J., *J. Phys. Chem. B* **102(1)**, 234–244 (1988)
79. Warr, G.G., *Curr. Opin. Colloid Interface Sci* **5(1–2)**, 88–94 (2000)
80. Parker, J.L., Cho, D.L. and Claesson, P.M., *J. Phys. Chem.* **93(16)**, 6121–6125 (1989)
81. Parker, J.L., Claesson, P.M., Cho, D.L., Ahlberg, A., Tidblad, J. and Blomberg, E., *J. Colloid and Interf. Sci* **134(2)**, 449–458 (1990)
82. Kessel, C.R. and Granick, S., *Langmuir* **7(3)**, 532–538 (1991)
83. Levins, J.M. and Vanderlick, T.K., *Langmuir* **10(7)**, 2389–94 (1994)
84. Clarkson, M.T., *J. Phys. D: Appl. Phys.* **22(4)**, 475–482 (1989)
85. Parker, J.L., Christenson, H.K., *J. Chem. Phys.* **88(12)**, 8013–8014 (1988)
86. Smith, C.P., Maeda, M., Atanasoska, L., White, H.S. and McClure, D. J., *J. Phys. Chem.* **92(1)**, 199–205 (1988)
87. Alcantar, N.A., Park, C., Pan, J.-M., Israelachvili, J.N., *Acta Materialia* **51(1)**, 31–47 (2003)
88. Vigil, G., Xu, Z., Steinberg, S., Israelachvili, J., *J. Colloid and Interf. Sci* **165(2)**, 367 385(1994)
89. Horn, R.G., Clarke, D.R., Clarkson, M.T., *J. Mater. Res.* **3(3)**, 413–16 (1988)
90. Berman, A., Steinberg, S., Campbell, S., Ulman, A. and Israelachvili, J. *Tribology Letters* **4(1)**, 43–48 (1998)
91. McGuiggan, P.M., Hsu, S.M., Fong, W., Bogy, D. and Bhatia, C.S., *Journal of Tribology* **124(2)**, 239–244 (2002)
92. Hirz, S. J., Homola, A.M., Hadziioannou, G., Frank, C.W., *Langmuir* **8(1)**, 328–333 (1992)
93. Horn, R.G., Smith, D.T., Haller, W., *Chem. Phys. Lett.* **162(4–5)**, 404–408 (1989)
94. Golan, Y., Alcantar, N.A., Kuhl, T.L. and Israelachvili, J., *Langmuir* **16(17)**, 6955–60 (2000)
95. Drummond, C., Elezgaray, J., Richetti, P. *Europhys. Lett.* **58(4)**, 503–509 (2002)
96. Schallamach, A.*Wear* **17(4)**, 301–312 (1971)

3 The quartz crystal microbalance as a nanotribology technique

Lorenzo Bruschi and Giampaolo Mistura

Dipartimento di Fisica G. Galilei, via Marzolo 8, 35131 Padova, Italy

3.1 Introduction

The quartz crystal microbalance technique (QCM) is a powerful probe of interfacial phenomena that has been successfully employed to investigate the sliding friction of objects of nanoscopic size subject to lateral speeds as large as a few m/s [1, 2]. The microbalance is a small quartz disk whose principal faces are optically polished and covered by two metal films, which are used both as electrodes and as adsorption surfaces. By applying an AC voltage across the two electrodes, it is possible to drive the crystal to its own mechanical resonance with the two parallel faces oscillating in a transverse shear motion. The quality factor of these resonances is usually very high ($\gtrsim 10^5$) and this explains why the QCM is quite sensitive to interfacial phenomena. A change in the disk inertia, as caused, for example, by the adsorption of a film on the metal electrodes, is signalled by a shift in the resonant frequency. Similarly, any dissipation taking place in the system determines a decrease in the resonance amplitude.

Thanks to the pioneering work of Krim and coworkers [3], it was found that molecules and atoms weakly bound to the surface of a quartz crystal can slip relative to the oscillating substrate. The slip occurs as a result of the force of inertia F acting on the adsorbates during the vibrational motion of the crystal. Actually, because of its extremely small value, the force F induces a slow, thermally activated motion of the adsorbate along its direction, with a drift velocity proportional to F. Most of the friction experiments carried out so far with the QCM technique have dealt with molecularly thin films of simple gases adsorbed at low temperatures on the metal electrodes, generally gold or silver, evaporated over the faces of a quartz crystal. For such studies, the QCM was implemented in standard cryostats that guarantee very good temperature controls, of the order of a few mK or better [4–7]. In this way, it was possible to achieve very stable quartz resonance curves, an essential requirement to detect the tiny dissipation associated with the friction of very thin films. However, more systematic and quantitative investigations now require the use of very uniform and clean surfaces, well characterized at the microscopic level by techniques like STM or AFM. Furthermore, it is also important to change in situ and in a controlled way the morphology of the

surface. In other words, it is necessary to design a new generation of QCM experiments that combine cryogenics with surface science [8].

In the literature, there are several general reviews on the application of the QCM to nanotribology. However, none of them has covered in detail the experimental aspects of such a technique. Therefore, in this chapter, after a brief description of the acoustics of the QCM, we present the main electronic circuits used to drive the QCM, emphasizing their pro and con. We then discuss the surface quality of the quartz electrodes and conclude with a short summary of the main features of a new apparatus that we have expressly built to fulfil the requirements of low temperatures and an ultra-high-vacuum environment. Finally, we present some preliminary data acquired with set-up that seem to suggest structural depinning of Ne films adsorbed on Pb(111) at a coverage above 0.4–0.5 layers.

3.2 The acoustics of the quartz crystal

The more common quartz crystals used in interfacial physics are of the so called AT-cut. It has been extensively used in the electronics because the temperature dependence of its resonance frequency is very small around room temperature. Other types are the SC-cut (stress-compensated) quartz crystal, which is insensitive to radial stresses, although the minimum of the quartz resonance versus temperature curve occurs close to 200 °C. Furthermore, the SC-cut crystals are much more expensive and require a complicated driving circuit with respect to the more common AT-cut crystals. In practice, QCMs employing SC-cut crystals have been used so far in adsorption studies but not in the field of nanotribology [9].

Let us now consider an AT-cut quartz plate, which is characterized by a shear motion of its two parallel faces. The AT crystal has a natural mechanical resonance when the plate thickness h is half of the transverse mode wavelength λ, or an odd multiple of $\lambda/2$, e. g. $h = n\frac{\lambda}{2}$, where n is called the overtone number ($n = 1$ is the fundamental mode, $n = 3$ is the third overtone...). At room temperature, the resonance frequency of such a plate oscillating in vacuum is related to its thickness h by the simple relation:

$$f_{0,n} = \frac{1.75n}{h} - C \tag{3.1}$$

where $f_{0,n}$ is measured in MHz, h in mm and C is a small correction factor which increases with electrode thickness [10]. (Typical values of $f_{0,1}$ for AT plates employed as QCM sensors lie in the range 1 – 10 MHz).

At a certain frequency f^*, its behavior in vacuum can be described by a complex acoustical impedance:

$$Z_0 \equiv R_0 - jX_0 = R_0 - j\pi n A Z_\mathrm{q} \frac{f^* - f_{0,n}}{f_{0,n}} \tag{3.2}$$

where A is the area of one electrode, $Z_q = 8.862 \cdot 10^5\,\mathrm{g/cm^2 s}$ is the quartz acoustic impedance and the dissipative term R_0, which accounts for all the losses in the plate, is related to the quality factor Q_0 via

$$\frac{1}{Q_0} = \frac{2R_0}{\pi n A Z_q} \tag{3.3}$$

When the quartz plate is immersed in a fluid, its impedance will change because of the adsorption of a film onto the quartz surfaces and of the viscous coupling with the surrounding vapor. The global contribution per unit area can be expressed in terms of a complex impedance $R_{sfv} - jX_{sfv}$, in series with Z_0. If both faces of the quartz plate are exposed to the fluid, the total dissipative and inertial terms become, respectively, $R_0 + 2AR_{sfv}$ and $X_0 + 2AX_{sfv}$. The quality factor will then decrease by an amount $\Delta\frac{1}{Q}$ equal to:

$$\Delta\frac{1}{Q} \equiv \frac{1}{Q} - \frac{1}{Q_0} = \frac{4R_{sfv}}{\pi n Z_q} \tag{3.4}$$

and the resonance frequency f will also be diminished by:

$$\Delta f \equiv f - f_{0,n} = -2X_{sfv}\frac{f_{0,n}}{\pi n Z_q} \tag{3.5}$$

The exact shifts will obviously depend on the explicit forms of R_{sfv} and X_{sfv}. In order to determine them, we have applied the linearized Navier-Stokes equation to the combined system quartz crystal-adsorbed film-bulk vapor [11]. Let d be the thickness of the adsorbed film and ρ_f and η_f its bulk mass density and viscosity, respectively, while η_v and ρ_v represent the viscosity and the density of the bulk vapor. If we assume, as customary, that the transverse velocity field depends only on the vertical distance z from the electrode surface, the general stationary solutions to the Navier-Stokes equations in the vapor and film regions are determined apart from four integration constants.

These can be univocally determined by imposing the following boundary conditions on the velocity fields v_f and v_v: a) $v_v = 0$ very far from the film; b) $v_v = v_f$ at the film-vapor interface (i. e. no slippage at this boundary); c) at this interface, the force exerted by the vapor on the film must be equal to that caused by the film on the vapor, that is $\eta_v \left(\frac{dv_v}{dz}\right)_{z=d} = -\eta_f \left(\frac{dv_f}{dz}\right)_{z=d}$; d) we assume, in general, that there may be slippage at the solid-film interface. Because of this, there will be a frictional force F_{sf} at this boundary. As the last condition, we then impose that the force F_{sf} must be equal and opposite to that due to the film, that is: $F_{sf} = +\eta_f \left(\frac{dv_f}{dz}\right)_{z=0}$. Finally, we make the plausible assumption that F_{sf} depend linearly on the relative velocity between the quartz plate and the film:

$$F_{sf} = -\eta_2[v_0 - v_f(0)] \tag{3.6}$$

where η_2 is called coefficient of sliding friction or interfacial viscosity, v_0 is the velocity of the electrode and $v_\mathrm{f}(0)$ that of the film at the electrode surface. This condition is consistent with recent QCM studies of the velocity dependence of interfacial friction [12]. If there is no slippage at the solid-fluid interface, $\eta_2 = \infty$. The opposite limit, $\eta_2 = 0$, corresponds instead to a superfluid whose motion is totally decoupled from that of the oscillating substrate.

By carrying out the necessary algebra, one finds that the reciprocal of Z_sfv can be easily rewritten as:

$$\frac{1}{Z_\mathrm{sfv}} = \frac{1}{Z_\mathrm{v} + Z_\mathrm{fd}} + \frac{1}{\eta_2} \tag{3.7}$$

which says that the total acoustic impedance Z_sfv of the combined system substrate-film-vapor can be considered as the parallel between the series of the vapor i mpedance, Z_v, and that of the film Z_fd, and the impedance η_2 due to the slippage of the film at the solid boundary.

The formula 3.7 means that it is possible, at least in principle, to measure the friction force of a film adsorbed on a solid surface with a quartz microbalance. In nanotribology one is interested in studying the friction of an adsorbed monolayer. This implies that the acoustic impedance of the film can be simplified as

$$Z_\mathrm{fd} \simeq -j\omega\rho_\mathrm{f} d \tag{3.8}$$

where $\omega = 2\pi f$.

If we solve the Eq. 3.7 in terms of R_sfv and X_sfv we get:

$$\frac{X_\mathrm{sfv}}{R_\mathrm{sfv}^2 + X_\mathrm{sfv}^2} = \frac{\omega\rho_\mathrm{f} d + X_\mathrm{v}}{R_\mathrm{v}^2 + (\omega\rho_\mathrm{f} d + X_\mathrm{v})^2} \tag{3.9}$$

and

$$\frac{R_\mathrm{sfv}}{R_\mathrm{sfv}^2 + X_\mathrm{sfv}^2} = \frac{R_\mathrm{v}}{R_\mathrm{v}^2 + (\omega\rho_\mathrm{f} d + X_\mathrm{v})^2} + \frac{1}{\eta_2} \tag{3.10}$$

The first equation yields the film thickness d as

$$d = \frac{1}{2\omega\rho_\mathrm{f}} \left[R_\mathrm{sfv}^2 + X_\mathrm{sfv}^2 + \sqrt{(R_\mathrm{sfv}^2 + X_\mathrm{sfv}^2)^2 - 4R_\mathrm{v}^2 X_\mathrm{sfv}^2} \right] - \frac{X_\mathrm{v}}{\omega\rho_\mathrm{f}} \tag{3.11}$$

which can be substituted in the second one in order to calculate the interfacial viscosity η_2.

Finally, the slip time τ_s, which represents the time required for the adsorbed film speed to decay to 1/e of its initial value after that the oscillating substrate has been put to rest in the absence of a bulk vapor, can be calculated from the ratio [3]:

$$\tau_\mathrm{s} = \frac{\rho_\mathrm{f} d}{\eta_2} \tag{3.12}$$

In the limit of very low vapor density, this approach yields identical results with the formulas introduced heuristically by other authors [13], according to which the vapor impedance is in series with the parallel of the

film impedance and the interfacial viscosity. As an example, for Kr vapor pressures below 1 Torr, the two approaches yield essentially the same results, while for pressures above 10 Torr this discrepancy can be as high as 15% or more, depending on the amount of sliding observed.

3.3 QCM driving circuits

As we have already mentioned, in order to use the quartz crystal as a nanotribology sensor it is necessary to measure its resonance frequency and its quality factor very accurately. Close to resonance, an AT-plate can in fact be accurately described [10] by the equivalent circuit shown in the enlargement of Fig. 3.1. The capacitor C_0 (of the order of a few pF) represents the static capacitance of the crystal between the two electrodes. The values of L_q and C_q are related to the kinetic and potential energies of the plate. The resistor R_q (of the order of a few tens of Ω) accounts for all the losses of the crystal and thus determines the intrinsic Q_q of the crystal. A quartz plate exhibits a series resonance at $f_s = 1/2\pi\sqrt{L_q C_q}$ and a parallel resonance at f_p where $f_p - f_s \approx f_s C_q/2C_0$. The detailed values of these characteristic parameters are provided by the manufacturer.

Various techniques have been devised to measure f_{res}. The simplest one is that of the oscillator, which uses the quartz crystal (either at its series or parallel resonance) in a positive feedback network in order to obtain an oscillating circuit. Many different configurations of oscillators exist in the literature. For example, Fig. 3.1 shows a very schematic diagram of an oscillator we have built in which the quartz crystal is connected to the inverting

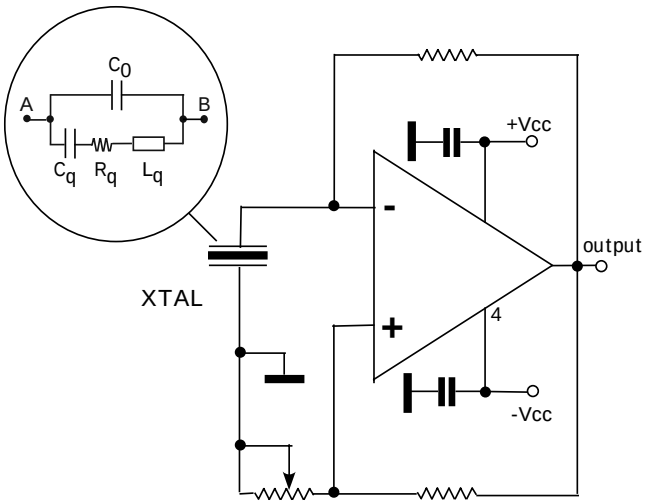


Fig. 3.1. Schematic circuit of a quartz oscillator used in our laboratory

channel of a wide-band operational amplifier. By adjusting the resistence of the trimmer close to R_{q}, the circuit will auto-oscillate to the series resonance frequency of the quartz.

In practice, because of the unavoidable phase shifts introduced by the electronic components employed, $\mathrm{f_{osc}}$ is shifted with respect to f_{res} and this difference depends on the width of the resonance curve, if the overall phase shift does not change. For example, a typical phase error of 10° implies an error in the frequency $\left|\frac{\mathrm{f_{osc}}-\mathrm{f_{res}}}{\mathrm{f_{res}}}\right| \approx 2 \cdot 10^{-6}$ with a quality factor of 40,000. If Q decreases during the measurements by 30%, (a value observed in our studis with multilayers and which may become much bigger if one studies systems like heavy, organic fluids characterized by a large damping), it determines a change in the frequency of the oscillator of about $6 \cdot 10^{-7}$ parts, e. g. an error of about 3 Hz for a crystal of 5 MHz. Another drawback of the oscillator is that it does not allow an easy variation and monitoring of the excitation power of the crystal. Furthermore, the oscillator selects by itself the resonance mode, typically it is the fundamental series, and it is not easy to switch to other modes. In conclusion, such a technique, although it is the simplest one, is not very flexible and might cause serious measurement errors particularly in situations where the Q of the quartz microbalance is small.

Another simple way to measure both the frequency shift and the dissipation taking place in a QCM experiment is the ring-down method shown in Fig. 3.2. The quartz crystal is excited at its resonance frequency either by an high-stability radio frequency generator [14] or by an oscillator [15], which are connected to the electrodes through an electronic switch. This latter one is controlled by a pulse generator that also triggers a digital oscilloscope. In this way, it is possible to turn the excitation on and off at a fixed rate, typically comprised between 10 and 100 Hz, and store the free decaying voltage after the excitation has stopped in the memory of the oscilloscope. The data can then be analyzed and fit according to the function

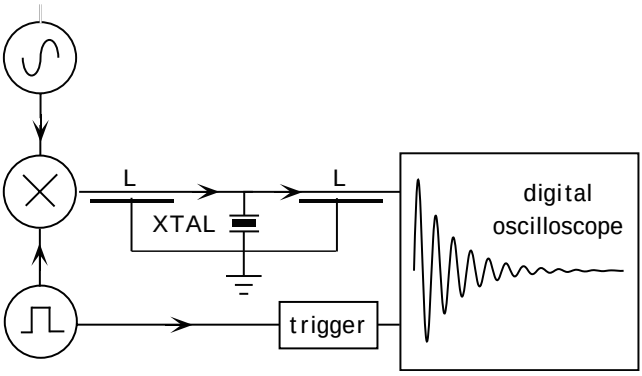


Fig. 3.2. Block diagram of the ring-down technique

$A\exp(-t/\tau)\cos(2\pi ft+\phi)$, where f represents the resonance frequency while the decay time τ is related to the quality factor of the crystal through the simple relation $Q=\pi f\tau$.

In our view, the technique best suited to drive the QCM in nanotribology applications is certainly that of the frequency modulation. In such a technique, $f_{\rm osc}$ corresponds to the maximum (or to the minimum) of the amplitude of the quartz electrical impedance while the quality factor is deduced from the amplitude of the detected signal [16]. Its main advantages can be summarized as: i) it is possible to lock on any resonance mode of the QCM (series or parallel, fundamental or overtone) in a very simple and fast way; ii) the excitation power can be easily varied from a few nW to several μW and, more important for nanotribology studies, its precise value can be accurately determined from the analysis of the crystal circuitry [17]; iii) it is possible to achieve very high sensitivities and time stabilities; iv) most of its main components can be easily found in any laboratory.

Figure 3.3 shows the block diagram of the electronics used in the FM technique. The output of an high stability radio-frequency generator equipped with the external frequency-modulation option (FM) drives the quartz crystal (XTAL) with a frequency $f=f_{\rm car}+\Delta f\sin(2\pi f_{\rm mod}t)$, where $f_{\rm car}$ is the so-called frequency carrier, set by the operator sufficiently close, but not necessarily equal, to the resonance frequency of the quartz mode one wants to lock-on. This $f_{\rm car}$ is modulated at a low-frequency $f_{\rm mod}$ between the extremes $f_{\rm car}-\Delta f$ and $f_{\rm car}+\Delta f$. The frequency of this modulation has to be smaller than the inverse decay time of the crystal $f_{\rm res}/Q_{\rm q}$ [16]. The quartz crystal may either be inserted in a transmission line or have one electrode grounded, depending on the experimental set-up. As a matter of course, the actual configuration affects the choice of the amplifier A. If $\Delta f\ll f_{\rm car}$, the amplified voltage can be well approximated by $A(f)\sim A(f_{\rm car})+\frac{{\rm d}A(f_{\rm car})}{{\rm d}f}\Delta f\sin(2\pi f_{\rm mod}t)$. The high-frequency component of this signal can then be rectified by a diode

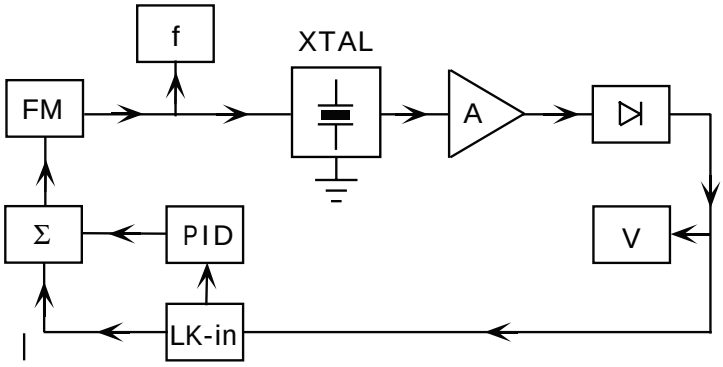


Fig. 3.3. Block diagram of the frequency modulation technique. See text for further details

detector. In our case, to bypass the problems connected with the use of the diode, e. g. periodic calibrations and thermal and time drifts, we have realized a multiplier whose output yields the square of the signal followed by a low-pass filter [17]. The DC amplitude of this signal is read by a high-precision voltmeter. The small component at f_{mod} is instead detected by a lock-in amplifier (Lk-in). The DC output of Lk-in, which changes sign as the frequency passes through quartz resonance, is used to control the value of f_{car} and thus locks the circuit onto the resonance frequency of the quartz. The block Σ adds up the low frequency modulation signal and the output of the Lk-in. The PID sums up the Lk-in output, its integral and its derivative. The integral is needed to perfectly locate the resonance frequency, the derivative guarantees stability to the feedback loop.

Finally, we mention another simple and sensitive circuit to drive a QCM. A generator is locked to the series resonance of the crystal by an high-frequency lock-in amplifier, which detects the reactive component of the transmission signal of the crystal and shifts the synthesizer frequency to null this component, after an offset is applied to cancel the contributions of shunt capacitance [18].

3.4 Quality of the surface electrodes

For a quantitative use of the QCM in nanotribology, it is of paramount importance to have very homogeneous and well characterized surfaces. In fact, surface roughness not only complicates the comparison with theoretical models but can also prevent the sliding of the adsorbed film. Although quartz resonators consist of thin disks of single crystalline silicon dioxide which are optically polished on both sides to roughness of about 1 nm rms, the thermally deposited metal electrodes usually exhibit a roughness of a few nm rms. Unfortunately, this roughness cannot be significantly reduced. To anneal a thermally deposited gold electrode, heating above 600 °C would be required. However, at 573 °C a phase transition occurs from α- to β-quartz. This phase transition is reversible upon cooling, but may introduce regions where the direction of the electrical axis in the quartz is reversed and thus remove or greatly reduce the piezoelectricity.

In a systematic study [19] of the surface morphology of Cu films on quartz in a UHV chamber, it was observed that annealing to 340 °C of the films deposited onto quartz preplated with titanium reduces the rms roughness to about 3 nm over a scan size of 500x500 nm^2, although they maintain a self-affine fractal scaling behavior over the length scale $10 \sim 500$ nm.

More recently, Pb films have been grown by physical deposition using an e-beam heated evaporation source at a rate of 0.5 nm/s [20]. The substrate was a quartz blank polished down to an RMS roughness of about 0.3 nm. Prior to Pb evaporation the quartz substrate was annealed under UHV conditions up to 140 °C in order to remove condensed surface impurities. When

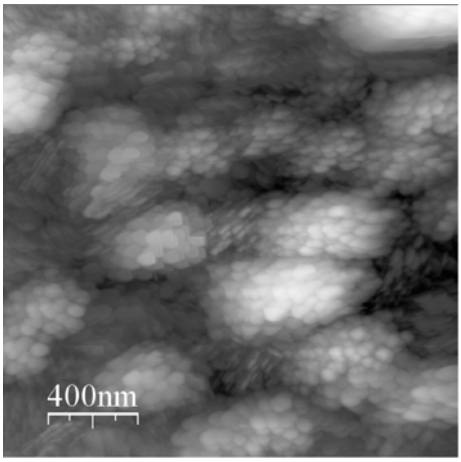


Fig. 3.4. STM topography of a $2000 \times 2000\,\mathrm{nm}^2$ area representative of a 150 nm thick Pb film deposited at 150 K on the surface of an AT-cut quartz. The film is formed by the assembly of 0.5 μm sized domains which are formed by a stacking of platelets. (Photo courtesy of F. Buatier de Mongeot)

Pb deposition is performed at or above room temperature, the thermally activated diffusion of Pb atoms is so high that a non connected percolated network of Pb clusters is formed [20]. Therefore, a substrate temperature of 150 K was chosen to hinder adatom mobility and a connected film is formed. In Fig. 3.4 we can see a large scale STM image ($2 \times 2\,\mu\mathrm{m}$) showing the morphology of a 150 nm thick Pb film deposited at 150 K, followed by annealing at room temperature. The image shows a distribution of domains with lateral dimensions around 0.5 μm . The various domains can be identified by the different orientation of the platelets. The majority of them are stacked parallel to the quartz surface, with an in-plane rotational mismatch, while a minor fraction is stacked with a tilt angle with respect to the substrate.

Another method to obtain very homogenous surfaces consists in glueing very thin layers of mica to the metal electrodes of a QCM [21]. In this way, it is possible to have a quartz sensor with a macroscopic atomically flat area without a significant loss of its sensitivity. These mica covered crystals have been used in studies of contact mechanics [22]. So far, however, no QCM measurements on the sliding friction of monolayers adsorbed on mica have been reported. Following a similar approach, a piece of Grafoil, a type of graphite, was glued uniformly on the Ag electrodes of a quartz crystal in order to study the slippage of ^{4}He films below 3 K [23].

3.5 UHV apparatus

Another problem that affects QCM measurements of the sliding friction of adsorbed monolayers is the contamination of the active surface of the quartz sensor [8]. To overcome these problems, we have recently assembled a new apparatus specifically designed to perform friction experiments on molecularly thin films carried out with the QCM technique in ultra-high-vacuum

and at a temperature as low as 4 K [24]. The main chamber is provided with a stainless steel jacket that allows a quick change of the various temperature inserts (cryocooler head, liquid nitrogen insert, Peltier-cell stage...) that span the working range 4 – 400 K, without ever breaking the vacuum. The crystal mounting is compatible with UHV conditions and with the Omicron standard and guarantees a good thermal and electrical contact over the entire temperature range. The quartz is inserted in a specially designed copper sample holder, housing the wiring for the temperature control and signal detection systems in a volume separated by that of the chamber. In this way, we can control the QCM temperature within 5 mK or less at any temperature below 10 K. Furthermore, the quartz frequency and amplitude stabilities are found to be as good as the best recorded values achieved with QCMs mounted in standard high-vacuum, liquid bath cryostats. The system is provided with a sputtering ion gun mounted in the fast-entry load section to clean the QCM electrodes from surface contaminants. The movements of the quartz sample inside the vacuum volume of the system are performed through the combination of a magnetic translator and a wobble stick.

Facing the circular hole in the crystal mounting, there is a thin-wall stainless steel tube whose end is attached to a sapphire variable leak valve. The high pressure side of the valve is connected to a high-purity gas cylinder and another port allows to purge the system effectively. A film is condensed onto the QCM, kept at low temperature, by slowly leaking gas through this nozzle. Depending onto the vapor pressure, the film may cover either only one or both electrodes. In the former case, which typically occurs at very low temperatures, the deposition of a Ne film can be controlled within 0.1 layers working at a fundamental frequency of 5 MHz. Obviously, this figure improves significantly using an heavier adsorbate and/or an higher overtone. For instance, the dosing of a Kr film onto a QCM running at 15 MHz (and third overtone) is done with a resolution better than 0.05 layers. If necessary, the adsobed film can be annealed to higher temperatures by simply turning off the cryocooler.

In the following, we present some preliminary data acquired with this setup at temperatures below 10 K. They refer to Ne deposited on a Pb(111) electrode grown and characterized in the group of Prof. Valbusa following the procedure described in the previous section. The data have been acquired at the third overtone of a 5 MHz quartz plate characterized by a quality factor of 380,000 at low temperatures. The two top graphs in Fig. 3.5 show the variations in the measured QCM resonance parameters during Ne evaporation. By acting on the leak valve, the film was slowly grown in steps of about 0.2 layers and then we waited for the system to equilibrate. At low coverages, there is no change in the quartz amplitude and accordingly the slip time is zero. Close to 0.4–0.5 layers, dissipation starts to appear and the slip time reaches values close to 0.3 ns, which are typical for rare gases sliding on metal surfaces. The slip times described in the bottom of Fig. 3.5 have been normalized with coverage according to the formulas reported in [8].

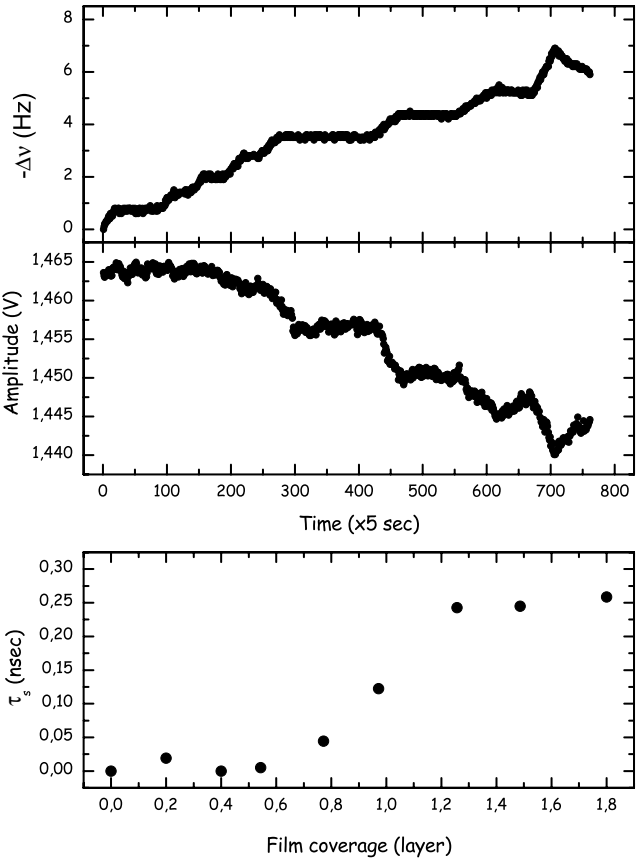


Fig. 3.5. Raw data of the resonance frequency shift (*top*) and amplitude (*middle*) during an adsorption isotherm of Ne on lead at 6.5 K. *Bottom*: calculated slip time as a function of Ne film coverage

Although there is no data available in the literature on the 2D phase diagram of Ne adsorbed on Pb(111), it is tempting to interpret our data in terms of a structural depinning of the film. At low coverages, the Ne film is in a fluid phase that at such low temperatures is locked to the substrate. Close to about 0.4 layers, the film enters an incommensurate solid phase which is weakly bound to the substrate and can easily slides. This interpretation is consistent with the structural phase diagrams of heavy rare gases adsorbed on Ag(111) [25], systems that are very similar to Ne/Pb(111). Our measurements are also in very good qualitative agreement with the results of extensive computer simulations of a model system carried out by Persson [26] in the case of a low-corrugated susbtrate .

A systematic study of the dependence of the slip time on film coverage, driving amplitude and temperature is currently under way in order to estab-

lish the phase diagram of this intrinsic depinning [27]. In a previous QCM study of the sliding friction of Kr monolayers adsorbed on Au(111) at 85 K we have also observed a nonlinear behavior, but in that case the depinning was induced by increasing the driving amplitude above a certain threshold [7]. More recently, highly sensitive AFM experiments have found changes in the friction behavior that are based on similar structural effects. Dienwiebel et al. observed that friction between graphite surfaces is significantly reduced when the surfaces are rotated out of the commensurate locking angle [28]. Socoliuc et al., by varying the normal load on the contact between tip and substrate, have also observed a new regime of very low friction in which negative and positive lateral forces sum up to a vanishing average force in the time average instead of the spatial average [29].

It is a pleasure to acknowledge several clarifying explanations of the theoretical aspects of nanofriction we have had over the past few years with Francesco Ancilotto, Bo Persson and Erio Tosatti. We also wish to thank our experimental partners Francesco Buatier de Mongeot, Renato Buzio, Bruno Torre, Corrado Boragno and Ugo Valbusa for many interesting discussions and suggestions. We have greatly benefitted from daily interactions with our students who have been involved with these studies: Alessandro Carlin, Moira Ferrari, Luca Stringher, Francesco dalla Longa, Giovanni Fois and Alberto Pontarollo, and we must also thank Giorgio Delfitto for his technical mastery. Finally, funding from INFM, PRA *Nanorub*, and MIUR, FIRB *Carbon based micro and nanostructures* and PRIN *Nanotribologia,* is kindly acknowledged.

References

1. See e. g., B.N.J. Persson, *Sliding Friction* (Springer-Verlag, Berlin, 1998), chapter 1.
2. J. Krim, Sci. Am. 275, **74** (1996).
3. J. Krim and A. Widom, Phys. Rev. B **38**, 12184 (1988).
4. J. Krim, D.H. Solina, and R. Chiarello, Phys. Rev. Lett. **66**, 181 (1991).
5. C. Daly and J. Krim, Phys. Rev. Lett. **76**, 803 (1996).
6. R.L. Renner, P. Taborek, and J.E. Rutledge, Phys. Rev. B **63**, 233405 (2001).
7. L. Bruschi, A. Carlin, and G. Mistura, Phys. Rev. Lett. **88**, 046105 (2002).
8. A. Carlin, L. Bruschi and G. Mistura, Phys. Rev. B **68**, 045420 (2003).
9. G. Hayderer, M. Schmid, P. Varga, H.P. Winter, and F. Aumayr, Rev. Sci. Instrum. **70**, 3696 (1999).
10. J.C. Brice, Rev. Mod. Phys. **57**, 105 (1985).
11. L. Bruschi and G. Mistura, Phys. Rev. B **63**, 235411 (2001).
12. C. Mak and J. Krim, Phys. Rev. B **58**, 5157 (1998).
13. E.T. Watts, J. Krim and A. Widom, Phys. Rev. B **41**, 3466 (1990).
14. S. Berg, T. Prellberg and D. Johannsmann, Rev. Sci. Instrum. **74**, 118 (2003).
15. M. Rodahl, F. Höök, A. Krozer, P. Brzezinski and B. Kasemo, Rev. Sci. Instrum. **66**, 3924 (1995).
16. M.J. Lea, P. Fozooni and P.W. Retz, J. Low Temp. Phys. **54**, 303 (1984).

17. L. Bruschi, G. Delfitto and G. Mistura, Rev. Sci. Instrum. **70**, 153 (1999).
18. G.B. Hess, M.J. Sabatini, and M.H.W. Chan, Phys. Rev. Lett. **78**, 1739 (1997).
19. S.M. Lee and J. Krim, Thin Solid Films, **489**, 325 (2005).
20. B. Torre, F. Buatier de Mongeot, F. Krok, R. Alessio, C. Boragno, U. Valbusa, to be published.
21. See e. g., S. Berg, M. Ruths and D. Johannsmann, Rev. Sci. Instrum. **74**, 3845 (2003).
22. S. Berg and D. Johannsmann, Phys. Rev. Lett. **91**, 145505 (2003).
23. N. Hosomi, A. Tanabe, M. Hieda, and M. Suzuki, J. Low Temp. Phys. **138**, 361 (2005).
24. L. Bruschi, A. Carlin, F. Buatier de Mongeot, F. dalla Longa, L. Stringher and G. Mistura, Rev. Sci. Instrum. **76**, 023904 (2005).
25. J. Unguris, L.W. Bruch, M.B. Webb and J.M. Phillips, Surf. Sci. **114**, 219 (1982).
26. B. Persson, Phys. Rev. B **48**, 18140 (1993).
27. L. Bruschi, G. Fois, A. Pontarollo, G. Mistura, B. Torre, F. Buatier de Mongeot, C. Boragno, R. Buzio, and U. Valbus, Phys. Rev. Lett. **96**, 216101 (2006).
28. M. Dienwiebel, G. Verhoeven, N. Pradeep, J. Frenken, J. Heimberg, and H. Zandbergen, Phys. Rev. Lett. **92**, 126101 (2004).
29. A. Socoliuc, R. Bennewitz, E. Gnecco and E. Meyer, Phys. Rev. Lett. **92**, 134301 (2004).

4 Nanoscale Friction and Ultrasonics

M. Teresa Cuberes

Dpto. Mecánica Aplicada e Ingeníera de Proyectos, Universidad de Castilla-La Mancha, Plaza Manuel de Meca 1, 13400 Almadén, Spain

4.1 Introduction

Ultrasonic technology finds many applications in our society. It is used in chemistry, biology and medicine, i. e. for preparation of colloids or emulsions, the pregermination of seeds, for imaging of biological tissues, etc. Also, it is used in nondestructive testing (NDT), for measurement of materials properties, in metrology, etc. Ultrasonic vibrations are commonly employed in mechanical machining of materials [1]. Procedures such as ultrasonic cutting of metals, ultrasonically assisted wire-drawing, ultrasonically assisted drilling, etc., take advantage of a modification of friction by ultrasonic vibration. Macroscopically, it is well known that friction and acoustics are very much related [2]. The development of *nanoscale ultrasonics* can be of interest in nanotechnology. Nevertheless, studies related to the emission of ultrasound from nanoscale contacts or to the influence of ultrasonic vibrations on nanofriction are still scarce [3].

The investigation of friction at the nanometer scale can be realized with an atomic force microscope (AFM). A specific AFM-mode, friction force microscopy (FFM), has been developed for this purpose [4]. FFM monitors the torsion of a microcantilever as a sample is laterally displaced by means of piezoelectric actuators, being the cantilever tip in contact with the sample surface. Typically, the deformation of the cantilever is sensed by optical beam deflection, and both bending in normal direction and torsion are simultaneously recorded with a four-quadrant photodiode detector [5]. The measurement of the lateral forces that act upon the tip–sample contact during forward and backward scans allows us to distinguish frictional forces, which reverse when reversing the scanning direction, from the lateral forces that stem from topographical features. The lateral resolution in FFM depends on the tip–sample contact area, which is typically $10-100\,\mathrm{nm}$ in diameter, in ambient conditions.

Ultrasound refers to mechanical vibrations of frequencies ranging from $20\,\mathrm{kHz}$ up to GHz. Typical ultrasound propagation velocities in solid materials are of the order of $10^3\,\mathrm{m\,s^{-1}}$. Hence, ultrasonic wavelengths in solid materials are of the order of mm, much larger than the diameter of the mean tip–sample contact area. The actuation of ultrasonic vibration at a nanocontact is always accomplished in the near-field regime. The understanding of

whether it is possible to detect ultrasonic vibration at the contact of an AFM cantilever tip and a sample surface is not trivial at first sight. A cantilever tip in contact with a surface will certainly be subjected to forces when the surface atoms displace due to ultrasound excitation, but if the ultrasonic frequency is sufficiently high, considering the cantilever tip as a point mass, it is clear that it will not be able to follow the surface motion due to its inertia.

Starting from 1992, different procedures to monitor ultrasonic vibrations at a sample surface using an AFM cantilever tip have been explored, which will be described in this chapter [6–23]. A first motivation for most of those studies was to implement a near-field approach that provided the kind of information that is obtained with the acoustic microscope, i. e. information about the elasticity and viscoelasticity of materials, but with a lateral resolution on the nanometer scale. To this aim, different AFM-based techniques such as ultrasonic force microscopy (UFM) [7, 9], atomic force acoustic microscopy (AFAM) [10], and heterodyne force microscopy (HFM) [21] have been quite successfully implemented. The different methods and their main opportunities for the characterization of nanoscale materials properties will be briefly outlined in Sect. 4.2.

Shear ultrasonic vibration excited at a sample surface can also be detected with the tip of an AFM cantilever [24–36]. Experiments that monitor the cantilever response to shear ultrasonic vibration excited at the tip–sample interface, with the tip in contact with the sample surface, provide novel methods to study nanoscale friction. Some interesting results concerning the response of nanocontacts to shear ultrasonic vibration will be introduced in Sect. 4.3.

In Sect. 4.4, experimental evidence of the reduction and/or elimination of friction at nanometer-sized contacts by means of ultrasonic vibration will be considered. The opportunity to control friction at the nanometer scale is of tremendous significance in nanotechnology. By now, it has been unambiguously demonstrated that ultrasound of sufficiently high amplitude can act as a lubricant in nanoscale contacts [38, 43–45]. Nevertheless, only a few experiments that address this topic have been performed to date, and hence the opportunities of ultrasonic vibration to modify the mechanisms of friction at a nanometer scale are still an open question.

In Sect. 4.5, some attempts to obtain information about adhesion and/or the adhesion hysteresis using ultrasonic AFM techniques will be summarized [21, 51–57]. Procedures for the measurement of adhesion hysteresis from UFM have been investigated, and a relationship between adhesion hysteresis and friction has been formally established [54]. Phase-HFM provides information about dynamic relaxation processes related to adhesion hysteresis nanoscale contacts with an extremely high time sensitivity, superior to any other ultrasonic-AFM procedure [21]. In view of a comparison of phase-HFM and friction data, the opportunities to take advantage of the time resolution of HFM for the study of nanoscale friction processes will be discussed.

4.2 Normal Ultrasonic Vibration at Nanocontacts

In the following, we will consider the nanocontact formed by the tip of an AFM cantilever in contact with a sample surface. Normal ultrasonic vibrations at the tip–sample interface can be excited using, for instance, an appropriate piezoelectric element attached to the back of the sample; longitudinal acoustic waves originated by mechanical vibrations of the piezo will propagate through the sample, and reach the surface–tip contact area.

As indicated in the introduction, in the limit of high ultrasonic frequencies (hundreds of MHz for instance), it is not expected that the cantilever tip in contact with the sample surface can move fast enough to keep up with surface atomic vibrations at ultrasonic frequencies, due to its inertia. Nevertheless, the displacement of the surface atoms will lead to modification of the tip–sample interaction forces. In the absence of ultrasound, with the tip in contact with the sample surface, in the repulsive interaction force regime, the cantilever is bent to compensate for the sample surface repulsive interactions, so that the net force at the tip–sample interface is zero, and the tip is indented into the sample to a certain extent, which depends on both the cantilever and the tip–sample contact stiffness. In the presence of normal ultrasonic vibration the tip–sample distance is varied at ultrasonic frequencies between minimum and maximum values, which depend upon the amplitude of ultrasound excitation and the initial set-point force (see Fig. 4.1a). If the amplitude of the ultrasound is small, the tip–sample distance sweeps a linear part of the tip–sample interaction force curve. The net average force that acts upon the cantilever during an ultrasonic time period will be in this case the initial set-point force. However, if the amplitude of ultrasound is increased, and the tip–sample distance is swept over the nonlinear part of the force curve, the average force will then include an additional force. If the ultrasonic amplitude is sufficiently high, the cantilever experiences an additional displacement due to this force, which can be easily detected with the optical lever technique [7]. This additional force constitutes the so-called *ultrasonic force* and it is the physical parameter evaluated in *ultrasonic force microscopy (UFM)* [7, 9]. The ultrasonic force induces a static cantilever displacement (UFM signal) as long as vertical ultrasonic vibration of sufficiently high amplitude is present at the tip–sample contact. In this sense, the cantilever behaves as a mechanical diode, and UFM has also received the name of mechanical-diode ultrasonic mode.

The ultrasonic force is hence understood as the averaged force experienced by the tip during each ultrasonic period. Its magnitude depends upon the part of the tip–sample force regime over which the tip–sample distance varies while being modulated at ultrasonic frequencies, i. e. on the initial tip–sample distance (the initial indentation or set-point force) and on the ultrasonic amplitude. The ultrasonic response will be dependent on the details of the tip–sample interaction force, and hence on sample materials properties such as local elasticity and adhesion. Figure 4.1a and 4.1b illustrates the

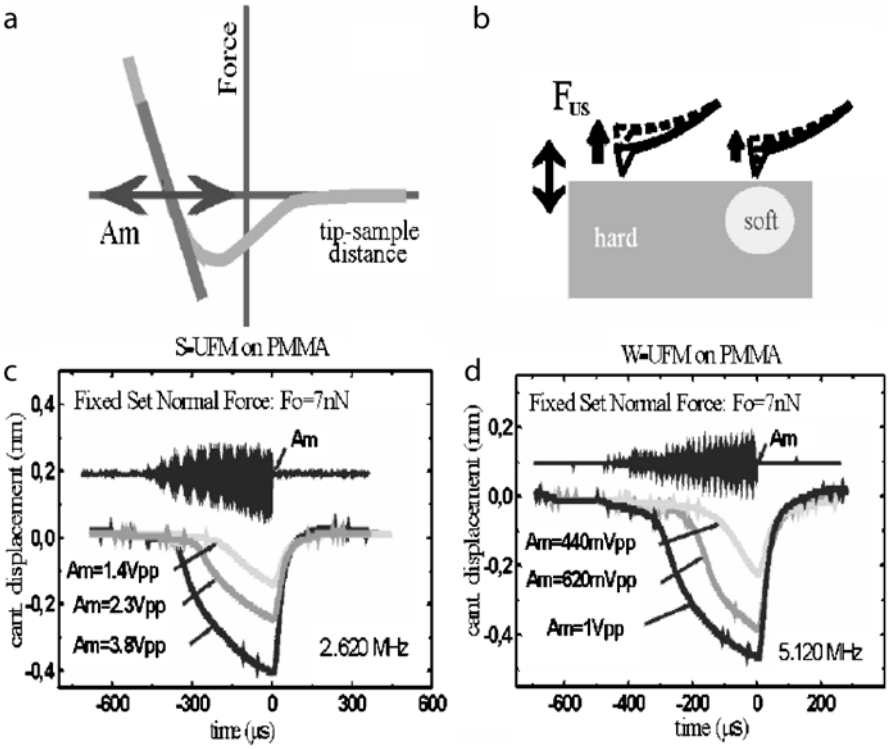


Fig. 4.1. a,**b** The physical principle of UFM measurements (see text). The ultrasonic excitation may be introduced through the sample (S-UFM) (**c**) or through the tip using the cantilever as a waveguide (W-UFM) (**d**). The piezo excitation is given a triangular modulation, with maximum amplitude A_{m}. The effect of varying the static force F_{o} (set-point force) is similar for S-UFM and W-UFM (from [22])

physical principle of the UFM measurements. Softer surface or near-surface regions of nanoscale dimensions at the sample under consideration will be easily distinguished from harder regions because of a smaller UFM signal at the former (Fig. 4.1b). Fig. 4.1c and d displays UFM responses of a sample of poly(methylmethacrylate) about 3 mm thick (see [22] for more details about these measurements). As shown in the figure, the piezo excitation is given a triangular modulation, with maximum amplitude A_{m}. In Fig. 4.1c, the piezo is located at the back of the sample, and works at a frequency of 2.620 MHz (the way ultrasound is excited at the tip–sample contact in Fig. 4.1d will be discussed below). The set-point force is kept constant at 7 nN. UFM responses for different maximum ultrasonic amplitudes are shown. As it is noticeable from the figure, the UFM response is zero until the amplitude of ultrasound excitation reaches a threshold value, and it then increases as the ultrasonic amplitude is increased. If the ultrasonic excitation amplitude is periodically varied at some low kHz frequencies, the UFM response will change accord-

ingly, and by monitoring its magnitude at every surface point by means of a lock-in amplifier, UFM images can be measured. To date, it has already been demonstrated that UFM is a useful technique to map the nanoscale elasticity and adhesive properties of surface and subsurface regions in a variety of both stiff and compliant samples [9, 19].

When working in the UFM mode, the high-frequency cantilever vibration is not directly monitored. If the cantilever is regarded as a simple point mass, the amplitude of vibration at the driving frequency should vanish in the limit of very high frequencies [7]. Nevertheless, the cantilever is not a point mass, but a tiny elastic beam that can support high-frequency resonant modes. *Atomic acoustic force microscopy (AFAM)* [10, 13] monitors the resonance frequencies of the high-order bending modes of the cantilever, being the tip of the AFM cantilever in contact with the sample surface, in the presence of normal ultrasonic vibration at the tip–surface interface. According to the wave theory of elastic beams, the flexural resonance frequencies of a rectangular cantilever are the solutions of a fourth-order differential equation, which can be analytically solved for a clamped-free cantilever, and for a clamped spring-coupled cantilever with the tip in contact with a sample surface [13]. In the latter case, the resonances are shifted in frequency and the vibration amplitudes along the cantilever changes. Using a linear approximation for the tip–sample interaction forces, the frequency shift can be calculated. Figure 4.2 shows the resonance frequencies of the clamped spring-coupled

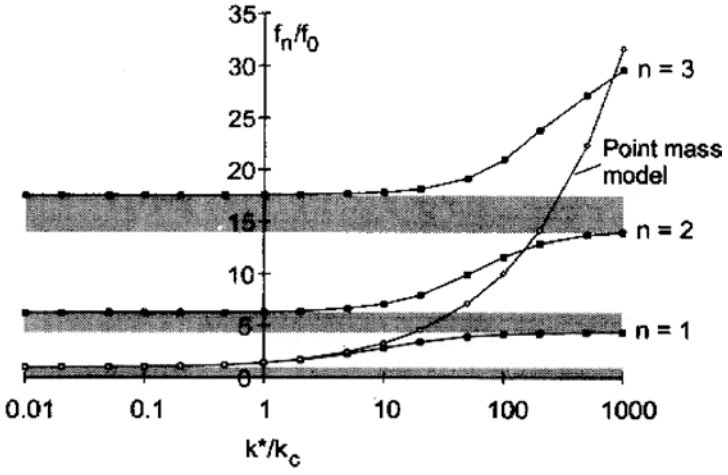


Fig. 4.2. Resonance frequencies f_n of the clamped spring-coupled cantilever with the tip in contact with a sample surface (black squares) normalized to the first resonance frequency of the clamped-free cantilever f_o. K^* and K_c are the tip–sample contact stiffness and the cantilever stiffness, respectively. A comparison with the point-mass model for the cantilever (open circles) shows that this model predicts too large frequency shifts for $K^*/K_c > 1$ (from [13])

cantilever as a function of the stiffness of the tip–sample contact normalized to the cantilever stiffness for the first three modes. The experimental determination of the shift of the resonance frequencies of the high-order flexural cantilever modes provides a measurement of the tip–sample contact stiffness, with lateral resolution in the nanometer scale. From the contact stiffness, the sample indentation modulus can be derived using, for instance, Hertz contact theory [13].

In UFM, it is assumed that the cantilever is *dynamically frozen*, and does not vibrate at ultrasonic frequencies [7]. Even though resonant modes can certainly be excited at a microcantilever, the point-mass picture for the AFM cantilever tip allows us to understand certain peculiarities of its high-frequency dynamic behavior. Thus, the inertia of the cantilever "explains" that in ultrasonic-AFM techniques soft cantilevers can indent hard samples, and yield information about surface and subsurface elastic inhomogeneities. In the limit of high ultrasonic frequencies, the amplitude of vibration at the crests of the resonant modes of a clamped spring-coupled cantilever is expected to be very small, and extremely difficult, if possible, to detect. Hence, UFM appears as the most appropriate technique for measurements at higher ultrasonic frequencies. Typically, in AFAM, the tip–sample distance is kept sufficiently small that the tip–sample interactions remain in the linear regime. In contrast, UFM relies on the nonlinearity of the tip–sample interaction force; if the tip–sample interactions are in the linear regime, no ultrasonic force is expected to set in at the tip–sample contact.

The detection of surface ultrasonic vibration with the tip of an AFM cantilever was first demonstrated in [6] by exciting surface acoustic waves (SAWs) at slightly different frequencies, and using a cantilever tip in contact with the sample surface to detect the surface vibration at the difference frequency. SAWs are acoustic modes that are confined within a wavelength to the surface of a solid, and propagate along specific crystalline directions. They can be excited using interdigital transducers (IDTs) on appropriate substrates. *Scanning acoustic force microscopy (SAFM)* was particularly implemented for the characterization of SAW field amplitudes [11] and phase velocities [18]. The procedure in SAFM is actually equivalent to this in UFM: the superposition of two SAWs of slightly different frequencies leads to surface high-frequency vibration that is modulated in amplitude at the (lower) difference frequency. When the surface vibration amplitude is sufficiently high, a cantilever tip detects the signal via the mechanical diode effect, due to the nonlinearity of the tip–sample force curve.

In *scanning local acceleration microscopy (SLAM)* [14], the cantilever tip is considered a point mass. Three different working modes are distinguished: the contact-mode, the mechanical-diode mode and the subharmonic mode. In contact-mode SLAM, the sample is vibrated at high frequency, with the tip in contact with the sample surface, and the tip displacement, which yields the contact stiffness, is monitored at the excitation frequency; the high-frequency

surface vibration amplitude is kept sufficiently low that the tip–sample interaction remains in the linear regime. The mechanical-diode SLAM mode is equivalent to UFM. In subharmonic SLAM, the sample surface is excited at very high ultrasonic vibration amplitudes. According to interesting reported data [12], the analysis of the generation of subharmonics and chaos may provide information about the local coefficient of restitution of a tip bouncing on a sample surface.

Scanning microdeformation microscopy (SMM) [8] uses a piezoelectric element to both excite ultrasonic vibration at a sample, and detect the acoustic wave generated by the microdeformations caused by a tip in contact with a sample surface. The technique can operate in transmission mode, with the piezo located at the back of the sample. In this way, contrast of local elastic constants, inhomogeneities and/or subsurface features is obtained with a lateral resolution essentially related to the tip diameter.

It is worth remarking at this stage that most of the different ultrasonic-AFM approaches discussed so far have capabilities of *subsurface imaging* [8, 9,14]. Nevertheless, so far the resolved buried feature sizes are typically much smaller than the used acoustic waves, the sensitivity to subsurface features does not appear straightforwardly related to acoustic wave propagation, but rather to a near-field effect.

The development of AFAM has proved that in the presence of ultrasound, with the tip is in contact with a sample surface, flexural resonant modes are excited at typical AFM cantilevers at frequencies of some MHz. Nevertheless, UFM usually also works quite well in the frequency range of some MHz. In principle, the ultrasonic frequency selected for UFM measurements should not be coincident with the cantilever contact resonances in order that the high-frequency displacements of the tip are as small as possible. However, it has additionally been demonstrated that ultrasound can be excited at a sample surface from a piezoelement located at the cantilever base. In this case, the cantilever acts as an acoustic waveguide that propagates the ultrasonic signal to the sample. As in AFAM, the measurement of the amplitude and resonant frequency of the high-order resonances of a cantilever in contact with the sample surface when ultrasound is excited from the cantilever base provides information of the sample elasticity with nanoscale resolution [15, 16]. SMM has also been implemented in the so-called "reflexion mode", with a piezoelement located at the cantilever base that is used for both the excitation and the detection of ultrasound [17]. And even though the propagation of ultrasound from the cantilever base to the sample surface necessarily requires that the cantilever tip vibrates at the excitation frequency, it has been experimentally demonstrated that UFM works in this configuration, renamed as waveguide-UFM (W-UFM) for distinction. As in the case when ultrasound is excited at the tip–sample contact from the back of the sample (sample-UFM, S-UFM) [22, 23], in W-UFM the ultrasonic excitation is input at the tip–sample contact via tip displacements. W-UFM and S-UFM signals recorded

on PMMA can be compared in Fig. 4.1c and d. In Fig. 4.1d, a piezo located at the cantilever base is excited at 5.120 MHz. As it is apparent from the figure, both procedures lead to remarkably similar qualitative responses. In principle, excitation of ultrasound from the cantilever base in ultrasonic-AFM techniques is potentially advantageous as there are by far fewer restrictions on the sample shape or its internal structure (e. g. porous or hollow samples can be studied). In addition, the use of same piezo–cantilever–tip assembly for different samples simplifies a quantitative comparison of nanoscale mechanical data.

In *heterodyne force microscopy (HFM)* [21], ultrasound is excited both at the tip (from a transducer at the cantilever base) and at the sample surface (from a transducer at the back of the sample) at adjacent frequencies, and mixed at the tip–sample gap (see Fig. 4.3). The physical principle of HFM is described in Fig. 4.3. As the sample vibrates at a frequency ω_1 and the tip at a frequency ω_2, the maximum tip–sample distance, is modulated at $\omega_1 - \omega_2$ (beat frequency). Provided that the total amplitude is large enough to cover the nonlinear range of the tip–sample interaction force, an ultrasonic force (stronger for larger amplitudes) will act upon the cantilever and displace it

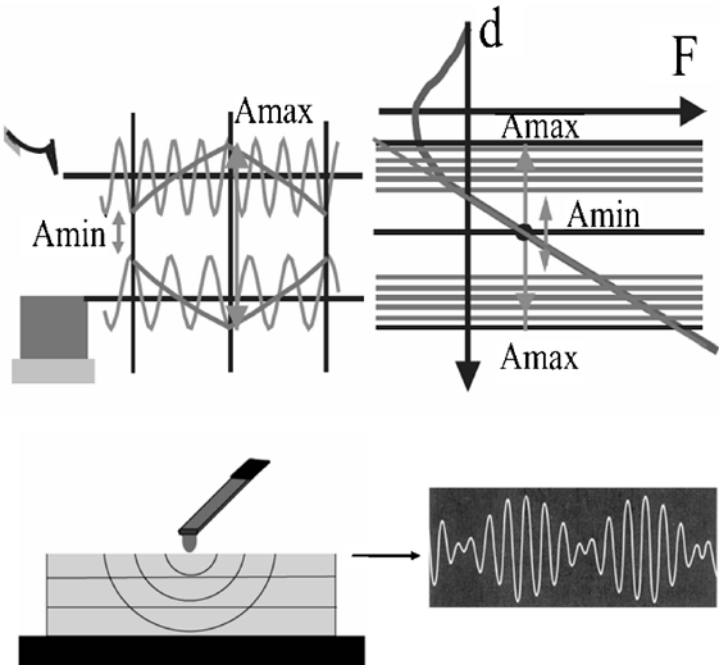


Fig. 4.3. A schematic diagram illustrating HFM. Small phase delays between tip and sample vibration (at ω_1 and ω_2, respectively) will cause a phase variation of the cantilever vibration at the difference frequency $\omega_1 - \omega_2$. This is detected as the HFM response

from its initial position. Owing to the varying ultrasonic force, the cantilever vibrates at the difference mixed frequency. In HFM, this vibration is monitored in amplitude and phase with a lock-in amplifier, using the (externally) electronically mixed signal as a reference. The information provided by the amplitude-HFM (A-HFM) response is very similar to that obtained by UFM. Nanoscale lateral variations in sample elasticity and/or adhesive properties will give rise to A-HFM contrast. A unique feature of HFM is its ability to monitor phase shifts between tip and sample ultrasonic vibrations with an extremely high temporal sensitivity, i. e. fractions of an ultrasonic time period. Small differences in the sample dynamic viscoelastic and/or adhesive response to the tip interaction result in a shift in phase of the beat signal that is easily monitored in phase-HFM (ph-HFM). In this way, HFM makes it possible to study dynamic relaxation processes in nanometer volumes with a time-sensitivity of nanoseconds.

Recently, *scanning near-field ultrasound holography (SNFUH)* [23] has been proposed as a nondestructive imaging method. The technique is implemented in a similar way to HFM, save that here the difference frequency is chosen in the range of hundreds of kHz whereas in [21] difference frequencies of some kHz are used. The experimental data obtained by SNFUH demonstrate its capability to provide elastic information of buried features with great sensitivity. Interestingly, in phase-HFM most of the contrast apparently stems from surface effects, as will be discussed in Sect. 4.5 of this chapter.

4.3 Shear Ultrasonic Vibration at Nanocontacts

If we consider the nanocontact formed by the tip of an AFM cantilever in contact with a sample surface, shear ultrasonic vibrations at the tip–sample interface can be excited using, for instance, a shear piezoelectric element attached to the back of the sample; shear acoustic waves originated by mechanical vibrations of the piezo will propagate through the sample, and reach the surface–tip contact area.

With a shear-wave transducer oriented in such a way that the surface in-plane vibrations are polarized perpendicular to the long axis of the cantilever, torsional resonant modes of a cantilever with the tip in contact with the sample surface are excited. *Lateral-acoustic friction force microscopy (L-AFAM) (or resonant friction force microscopy (R-FFM))* [24–27] monitors the vibration amplitudes of the cantilever torsional resonant modes at different surface points. In this technique, the sample is typically laterally vibrated at MHz frequencies, and the torsional vibration amplitudes provide information about the lateral forces between tip and sample. Apparently, L-AFAM images are independent of the scanning direction, i. e. not influenced by topography-induced lateral forces [25]. When scanning in the presence of shear ultrasonic vibration at the tip–sample contact, the relative tip–sample velocities are of the order of $1\,\mathrm{mm\,s^{-1}}$, much larger than those in conventional FFM (about

$100-250\,\mu ms^{-1}$), and nearer to the sliding operating velocities in MEMs and NEMs (in the range of tens of $mm\,s^{-1}$ to a few ms^{-1}) [37].

The analysis of the torsional contact resonances of AFM cantilevers in contact with a sample surface provides a novel means to study friction and stick–slip phenomena at the nanometer scale [26,27]. At low shear-excitation voltages, the resonance curve torsional cantilever vibration amplitude versus excitation frequency is a Lorentzian with a well-defined maximum; the cantilever with the AFM tip stuck to the sample surface following the surface motion, behaves like a linear oscillator with viscous damping. Above a critical shear excitation amplitude, which depends on the static cantilever load, and is of the order of 0.2 nm for bare and lubricated silicon samples [26], the shape of the resonance curve exhibits a characteristic flattening, attributable to the onset of sliding friction at the tip–sample contact. Experimental evidence of energy dissipation before sliding friction sets in has been related to microslip, i. e. slipping of an annulus at the tip–sample contact before the whole contact starts to slide (see Ref. [26] for further details).

The local vibration amplitudes and phases of the torsional resonances of clamped-free AFM cantilevers have been studied using optical interferometry [28]. The finite size of the cantilever beam and asymmetries in its shape leads to coupling between flexural and torsional vibrations. Lateral resonant modes of AFM cantilevers, which consist in flexural vibration modes in the cantilever width direction parallel to the sample surface, have also been experimentally observed [29]; asymmetries in the cantilever thickness lead to a z component of the displacement that can be monitored by optical beam deflection with an AFM.

The torsional resonant modes of a cantilever tip in contact with a sample surface have also been excited using a shear piezo located at the cantilever base [30,31]. In the *torsional resonance dynamic-AFM mode (TR mode)* [32] torsional vibrations of the cantilever are excited via two piezoelectric elements mounted beneath the holder of the chip, which vibrate out-of-phase, in such a way that they generate a rotation at the length axis of the cantilever. Using this procedure, the torsional resonances of the cantilever can be monitored in both near-contact and contact modes. In ultrahigh vacuum (UHV), torsional cantilever resonances can be excited via vertical vibrations, due to their high quality factors. Lateral forces between a cantilever tip and objects on surfaces have been measured in UHV by monitoring the induced change of the frequency of the fundamental cantilever torsional resonant mode [33]. In the *torsional overtone microscopy* [34], torsional cantilever resonances excited by thermal noise are used to obtain information about the shear stiffness of the tip–sample contact.

In the limit of high ultrasonic frequencies, it is questionable if high-order torsional resonances will be excited at the cantilever. Nevertheless, in *lateral scanning acoustic force microscopy (LFM-SAFM)* [35,36] SAWs with inplane oscillations components such as Love waves have been detected by modulating the rf signal's amplitude at some kHz. When the tip is in contact with

the sample surface, in the presence of shear ultrasonic vibration at the tip–sample contact, the cantilever experiences an additional amplitude-dependent torsion or lateral mechanical-diode effect. From the ultrasound-induced additional torsion, information about the amplitude and phase velocity of in-plane polarized SAWs can be obtained.

In *lateral ultrasonic force microscopy (L-UFM)* [9] lateral vibrations of the sample surface at a relatively low frequency of some kHz, polarized perpendicular to the length axis of the cantilever, are superimposed on a continuous vertical ultrasonic surface vibration. The measurement of the amplitude of torsion of the cantilever at the lateral low-frequency surface vibration provides information about the sample shear elastic properties with subsurface sensitivity.

4.4 Reduction of Friction by Ultrasonic Vibration

The reduction of friction by ultrasound is a well-known macroscopic effect [1, 2]. Its occurrence at the nanometer scale is only recently being investigated.

Dinelli et al. [38] studied the influence of out-of-plane ultrasonic vibration on the frictional response of a Si sample in ambient conditions, using FFM and UFM. Their results clearly demonstrated that dynamic friction vanishes in the presence of ultrasound when the tip–surface contact breaks for part of the out-of-plane vibration cycle (see Fig. 4.4). Figure 4.4 shows the friction force and the cantilever deflection measured at different surface ultrasonic vibration amplitudes. The friction force in Fig. 4.4 was independently deter-

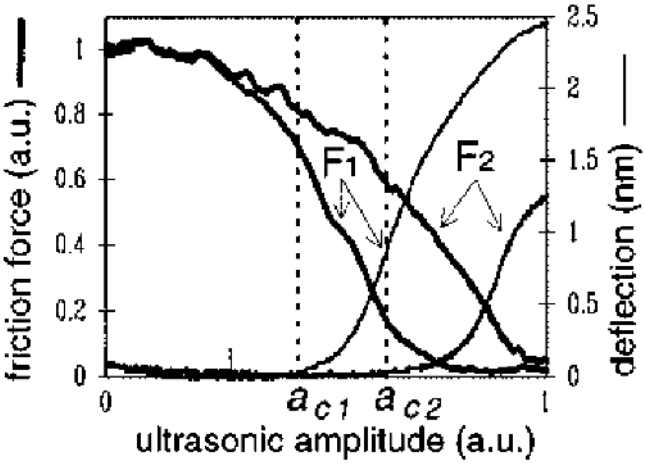


Fig. 4.4. Experimental measurements of dynamic friction (thick line) and cantilever deflection (thin line) dependencies on the ultrasonic amplitude, for two different applied loads $F_1 = 0\,\mathrm{N}$ and $F_2 = 2\,\mathrm{nN}$ on a Si sample (from [38])

mined for each of the different amplitudes of surface ultrasonic vibrations by laterally scanning the sample back and forth in the direction perpendicular to the cantilever axis, using a lock-in amplifier (see Ref. [38] for further details). The cantilever deflection signal in Fig. 4.4 corresponds to the cantilever response to the ultrasonic force, i. e. the UFM signal, which depends on the ultrasonic amplitude (see Fig. 4.1). The onset of an UFM response for a given set-point force roughly indicates the ultrasonic amplitude needed for the tip to detach from the sample surface at part of the surface ultrasonic vibration cycle.

The breaking of the tip–sample contact at each ultrasonic cycle explains the reduction or elimination of friction because of a reduction of slippage during sliding. Interestingly, it is apparent from Fig. 4.4 that, for a given applied load, the friction force considerably decreases well before the onset of the UFM response, i. e. while the tip remains in "linear contact" with the sample surface during the ultrasonic vibration cycle. For the case of F_2 in Fig. 4.4, the reduction of friction already amounts to about 60% when the UFM cantilever response sets off.

The influence of normal ultrasonic vibration on the static friction force was studied by keeping the amplitude of the lateral displacement small enough that the tip remained stick to a surface point without sliding, see Ref. [38] for details. It was demonstrated that the static friction force begins to decrease at very low ultrasonic amplitudes, and that the onset of friction reduction does not depend on the applied shear force. Evidence on this latter point ruled out the possibility that the reduction of friction is due to slippage during the part of the period that the tip–sample forces are the lowest.

In order to explain a reduction of friction at low ultrasonic amplitudes, the presence of a surface layer at the tip–sample gap, i. e. a liquid layer formed by water and possibly organic contaminants, has been considered [38]. In the absence of ultrasonic vibration, such a layer might organize in a solid-like configuration between the tip and the sample and partially sustain the load. As the tip–sample distance is varied at ultrasonic frequencies, the viscosity of the layer would hinder its rearrangement, thereby reducing the probability of tip stick–slip processes, and hence friction.

Using molecular dynamics (MD) simulations, Gao *et al.* [39] demonstrated that small amplitude (of the order of 0.1 nm) oscillatory motion of two confining interfaces in the normal direction to the shear plane can lead to transitions of a lubricant from a high-friction stick-slip shear dynamics to an ultralow kinetic friction state (superkinetic friction regime), provided that the characteristic relaxation time for molecular flow and ordering processes in the confined region is larger than the time constant of the out-of-plane mechanical oscillations.

Heuberger et al. [40] observed load- and frequency-dependent transitions between a number of dynamic friction states of a lubricant using a surface forces apparatus, modified for measuring friction forces, while simultaneously

inducing normal vibrations between two boundary-lubricated sliding surfaces. In particular, they found regimes of vanishingly small friction at interfacial oscillation amplitudes below 0.1 nm, and demonstrated that they originate due to the dynamics of the relaxation processes of the lubricant at the molecular level.

Recently, Socoliuc et al. [41] have demonstrated that mechanical vibrations normal to the plane of sliding at cantilever resonance frequencies in the range of hundreds of kHz in ultrahigh-vacuum (UHV) conditions lead to an ultralow friction regime in atomic scale friction even when the amplitude is not sufficiently high that the tip detaches from the sample during the vibration cycle. Previously [42], the authors had reported on the observation of an ultralow dissipation state in atomic friction related to the absence of mechanical instabilities, attained by varying the normal force. Such a state may exist because a modification of the tip–sample normal load leads to changes in the lateral surface corrugation felt by the tip without significantly altering the stiffness of the tip–sample contact. In the case that the tip–sample force is periodically varied at high frequencies, it is feasible that the tip slides through ultralow dissipation atomic friction states when being laterally displaced.

The effect of in-plane ultrasonic vibration in nanoscale friction has also been considered. Scherer et al. [25] observed that when lateral ultrasonic vibrations are excited at a sample surface at ambient conditions using a shear piezo bonded to the back of the sample, friction nearly vanishes at certain frequency bands, whereas it remains as high as on a nonvibrating surface at other frequencies. However, they verified that the near-zero friction bands coincided with frequencies at which a lift-off (vertical displacement) of the AFM cantilever occurred. As discussed by the authors [25] such "lift-off" might be attributed to the set in of a vertical ultrasonic force due to parasitic out-of-plane motions of the sample surface or to mode coupling in the cantilever. Nevertheless, the buildup of an elastohydrodynamic lubrication film whose viscosity and hence thickness is dependent on the lateral tip–sample relative velocity was proposed as a reasonable hypothesis that could account for a vertical cantilever displacement in the absence or in the case of low-amplitude out-of-plane surface vibrations.

Behme et al. [43–45] studied the influence of surface acoustic waves (SAWs) on nanoscale friction. SAWs constitute a precise source of acoustic vibration, with well-defined surface oscillations in a perfectly determined polarization, whereas when working with bulk shear-wave transducers parasitic surface displacements due to the existence of boundaries, etc. can hardly be avoided. LFM and multimode SAFM were used to measure and distinguish the influence of inplane and vertical surface oscillations components on the cantilever torsion and bending. To this aim, the authors [43–45] excited a standing Rayleigh-wave field, and considered the dependence of friction on the acoustic excitation amplitude. In Rayleigh waves, the atoms oscillate on elliptical trajectories with a large vertical and a smaller lateral os-

cillation component. The experiments showed that by increasing the rf amplitude, friction is locally reduced an eventually suppressed. In addition, it was clearly demonstrated that at the point at which friction disappears, the lateral-SAFM signal breaks down. Hence, it was concluded that the effect of friction reduction is essentially due to the vertical mechanical-diode effect that leads to an effective shift of the cantilever, whereas inplane oscillations do not play a significant role. This hypothesis was further reinforced by the fact that apparently in-plane polarized Love-type SAWs did not significantly alter the frictional behavior. When using the in-plane polarized Love-type SAWs, no cantilever lift-off induced by a lateral oscillation of the sample was observed [25]. At very high Rayleigh-wave amplitudes a lateral force rectification of the longitudinal component of the standing-wave field was apparent, which resulted in a scan-direction-independent appearance of the LFM traces.

Ultrasonic vibration covers a broad range of frequencies, and the processes involved in a reduction of friction by ultrasound can vary at different relative tip–sample velocities. De Hosson and Kessermakers [46] studied the influence on nanoscale friction of lateral high-frequency vibration of the cantilever, up to frequencies of 1 MHz, on a NbS_2 sample at ambient conditions, and observed gaps of lowered or eliminated friction at specific frequencies, presumed to be around torsional and/or lateral cantilever resonances. In these experiments a Au-coated cantilever was used, and the oscillating lateral cantilever vibration was applied by means of an electrostatic field. At a particular friction-gap frequency, a slow increase in driving field amplitude caused a gradual increase in friction, and above a certain threshold level of driving amplitude, a partial stick-slip behaviour with the tip periodically alternating between a zero friction an a nonzero-friction state was apparent.

Riedo et al. [47] also reported about a reduction of friction when lateral oscillations around a frequency of 19.5 kHz were applied to an AFM cantilever sliding on mica. In the range of scanning velocities they used, the thermally activated hopping of contact atoms over the effective lateral interatomic potential led to increased energy dissipation when increasing the sliding velocity. By superimposing a lateral oscillation on the cantilever and sweeping its frequency between about 20 to 300 kHz, and a clear peak of friction-reduction was observed around 19.5 kHz, independently of the applied load. This friction reduction peak was attributed to the excitation of a cantilever torsional contact resonance, which increased the attempt frequency for thermally activated jumps during sliding. The effect did not occur above a certain critical value of the sliding velocity.

In recent experiments performed by Socoliuc et al. [41] on KBr samples in UHV no reduction-of-friction effect was apparent upon the excitation of torsional cantilever contact resonances in the frequency range from 40 kHz up to 200 kHz, even though friction was strongly reduced when the excitation frequency matched one of the normal resonance frequencies of the pinned lever or half its value.

Other studies that have considered the possibility to control nanoscale friction by mechanical action at high frequencies on the system motion are described in [48, 49] and Ref. therein.

4.5 Adhesion Hysteresis at Ultrasonic Frequencies

On the nanoscale, adhesion phenomena become decisive to the performance of nanodevices, and surface properties acquire a particular relevance. Usually, the work of adhesion is defined as the energy needed to separate two surfaces, assuming that this is reversible [50]. The adhesion hysteresis is defined as the difference between the work needed to separate two surfaces and that gained when bringing them together. The fact that those two works are different in magnitude, i. e. the adhesion hysteresis is different from zero, can be attributed to elastic, viscoelastic and plastic deformations in the contact zone, reconfiguration of surface molecules during contact, chemical reactions, etc.

Recently, novel methods to obtain information about the work of adhesion and the adhesion hysteresis at the tip–sample contact using UFM have been proposed [51–55]. Essentially, they take advantage of the fact that the ultrasonic amplitude at which an UFM response sets off when increasing the excitation is different from this at which it falls down when decreasing the excitation. This is illustrated in Fig. 4.5 [51], in which both experimental and simulated UFM signal versus ultrasonic excitation amplitude curves have been drawn. In UFM, with the tip in contact with the sample, when

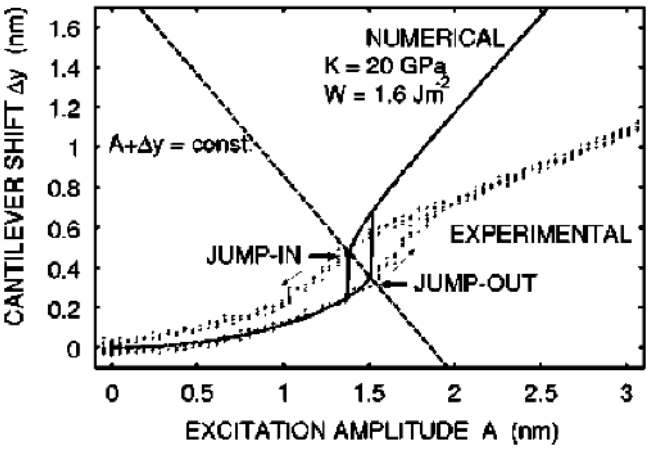


Fig. 4.5. UFM signals recorded when increasing and decreasing the ultrasonic excitation amplitude (see arrows to distinguish each case) on an aluminum thin film. The continuous lines correspond to a numerical evaluation of the UFM responses according to the model detailed in [51] (from [51])

increasing the normal ultrasonic amplitude at the tip–sample contact, at certain amplitude the tip detaches from the surface at part of the ultrasonic period, and the ultrasonic force (see Sect. 4.2 of this chapter) experiences a sudden increase that gives rise to a "jump-out" of the cantilever (see Fig. 4.5). When decreasing the ultrasonic amplitude, at certain amplitude the tip can no longer separate from the surface, and the ultrasonic force experiences a sudden decrease that gives rise to a "jump-in" of the cantilever (see Fig. 4.5). For the evaluation of the ultrasonic force, it is considered that mechanical hystereses, i. e. snap-in and -out of the cantilever when approaching or separating from the sample surface do not occur. In the absence of ultrasound, compliant cantilevers are subjected to large mechanical hysteresis when approaching or separating from a sample surface due to the force gradient being larger than the cantilever spring constant. However, at ultrasonic frequencies, the inertia of the cantilever leads to an effectively much larger cantilever stiffness, and the cantilever can probe the hysteretic cycle of tip–sample in-and-out interactions without a decrease of its sensitivity for force-field detection.

In [51] a method for quantitative analysis of the UFM signal is proposed in order to determine both the sample elastic modulus and the work of adhesion by monitoring the cantilever jumps such as those in Fig. 4.5. In UFM, both elasticity and adhesion contribute to the ultrasonic force. Dinelli et al. [56] evaluated the contact stiffness by comparing the jump-in positions in ultrasonic amplitude for different applied loads. Using the Johnson–Kendall–Roberts–Sperling (JKRS) model to account for both elastic and adhesive forces between tip and sample, the authors in [51] evaluated both the stiffness and the work of adhesion as defined in JKRS by calculating the jump-in and jump-out cantilever shifts. According to their modeling, the normalized cantilever jump-in shift turns out to be constant and effectively independent of the set-point force, the stiffness and the work of adhesion. Hence, they derived a universal relation between the work of adhesion, the stiffness and the cantilever shift at jump-in, the latter being easily measured from the experimental data (see Ref. [51] for further details).

In [52] the area between experimental curves such as those in Fig. 4.5 is measured and defined as the UFM hysteresis area (UH), and it is assumed that UH scales with the local adhesion hysteresis. A detail procedure to obtain quantitative information about the adhesion hysteresis from UFM signal versus ultrasonic excitation amplitude curves is discussed in [55]. The correlations between adhesion hysteresis and local friction have been theoretically and experimentally investigated [54]. According to a model based on the classical theory of adhesional friction and contact mechanics, which includes the effects of capillary hysteresis and nanoscale roughness and assumes an adhesive, elastic and wearless tip–sample contact, a relationship between adhesion hysteresis and friction has been derived, which depends on the varying ratio of the tip–sample work of adhesion over the reduced

Young's modulus (see Ref. [54] for further details). In the model, the adhesion hysteresis is estimated as the pull-off force times the critical separation at which the tip–sample contact is about to be broken. Measurements on a wide range of engineering samples with varying adhesive and elastic properties have confirmed the model [52, 54]. The aforementioned ratio does not vary much between typical metallic samples, and for a limited number of specimen's adhesion hysteresis and friction the experimental relationship may appear linear. In addition, it is found that capillary hysteresis offsets the measured adhesion hysteresis from the friction force, and that roughness reduces both friction and adhesion hysteresis: friction decreases because of a smaller area of a real contact, and adhesion hysteresis drops due to a smaller pull-off force at rough surfaces. Recently, it has been demonstrated that the study of the dependence of local adhesion hysteresis on relative humidity using UFM may provide information about protein–water binding capacity with molecular-scale resolution [53].

Procedures to obtain information about the work of adhesion using AFAM are also being considered [54]. In AFAM, the tip–sample contact stiffness can be determined by monitoring the resonance frequency of an AFM cantilever tip in contact with the sample surface (see Sect. 4.2 of this chapter). Strictly, the contact stiffness is influenced by both the tip–sample elastic properties and the work of adhesion. Typically, the tip–sample distance in AFAM is kept sufficiently small that the tip–sample interactions remain in the linear regime. Recently, a method has been proposed to evaluate both these properties quantitatively from the analysis of the nonlinear AFAM cantilever response excited when the tip–sample distance sweeps the nonlinear part of the tip–sample interaction in such a way that the tip always remains in contact with the sample surface, considering the case of a perfect contact. To this aim, the dependence of the resonance frequency on the vibration amplitude is studied; the elastic properties and the work of adhesion are separately determined by finding the optimal set of values that minimizes the difference between the theoretical and empirical relationship of cantilever resonance frequency versus ultrasonic excitation amplitude (see Ref. [56] for further details).

In HFM, the phase signal provides information of the adhesion hysteresis related to the formation and breaking of the tip–surface contact [21]. Contrast in phase-HFM mostly stems from dissipative processes. An exceptional feature of this technique is its ability to probe a local response in extremely short time. HFM may test effects that take place at nanoseconds in nanometer-scale volumes. Hence, phase-HFM can reveal dissipation due to extremely quick transitions that otherwise remains unresolved from other dissipative effects occurring at larger time scales. For instance, using phase-HFM, it has been possible to distinguish differences in contrast at identical thin polymer layers with different boundary constraints on the nanometer scale. These layers, however, exhibited the same FFM contrast, which confirms the ability of phase-HFM to resolve dynamic dissipative processes in a much shorter time

scale than conventional FFM. In the following, the results presented in [21] relative to those experiments will be summarized here, with a main focus in understanding the opportunities of phase-HFM to provide information about adhesion hysteresis with extremely high time sensitivity.

In metals, anelastic or viscoelastic contributions are expected to be small. In contrast, in polymeric materials, intra- or intermolecular perturbations induced by tip actuation, and/or dissipative effects of the molecules due to adhesion to the tip or to other neighboring molecules will play a significant role in the phase-HFM contrast. Phase-HFM has been applied to PMMA/rubber nanocomposites that consist in an acrylic matrix, a copolymer based upon PMMA and toughening particles composed of a core of acrylic enclosed with rubber with a bonded acrylic outer shell to ensure good bonding to the matrix (see Fig. 4.6).

Figure 4.6a–c shows contact-mode AFM (a), phase-AFM (b) and LFM images recorded over the same surface region of a PMMA/rubber sample. The topographic protrusions in Fig. 4.6a indicate the presence of core-shell PMMA particles in the surface and/or near surface region. Two different kinds of topographic protrusions may be distinguished from those and other images recorded on the PMMA/rubber sample surface: (i) some that give rise to a lower Ph-HFM contrast than the PMMA matrix, and (ii) others that show a Ph-HFM contrast similar to that of the PMMA matrix. Such

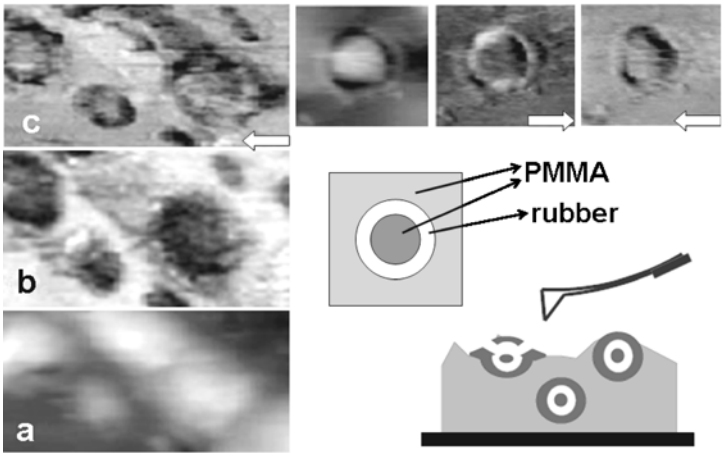


Fig. 4.6. a–c AFM contact-mode topography (**a**), Phase-AFM (**b**) and LFM (**c**) images recorded over a same surface region of a PMMA/rubber sample. The images at the top right-hand side correspond to AFM contact-mode topography, and LFM images recorded scanning from left to right, and vice versa respectively (see arrows), over a same surface region of the sample, different from that in (**a–c**). Below, schematic drawings illustrate the apparent structure at the PMMA/rubber sample surface

different protrusions are apparent from the comparison of Fig. 4.6a and b. The drawings in Fig. 4.6 illustrate a model for the two different protrusions: at some particles, the PMMA particle shell is well-bonded and indistinguishable from the PMMA matrix, whereas in others the rubber particle is still capped with the PMMA layer, but this is detached from the matrix material. Such a picture is corroborated when considering FFM images (see Fig. 4.6c) as well as UFM and A-HFM images recorded in the same surface region (not shown here, see Ref. [21]). Both UFM and A-HFM reveal the presence of the toughening particles by a darker contrast, indicative of the presence of a softer material in the surface or near-surface region. The aforementioned different particles cannot be distinguished from the UFM and A-HFM measurements [21]. However, they are clearly differentiated in Ph-HFM, and discernible by the presence or absence of a kind of halo contrast in FFM.

At the top right-hand side of Fig. 4.6, contact-mode AFM and FFM images recorded over a particular PMMA/rubber particle scanning from left to right (forward scan), and vice versa (backward scan, see arrows in the figure) are shown. This particle is representative of those that typically give rise to Ph-HFM contrast, and the image quality is a little better than this in Fig. 4.6c. From those images it is apparent that the particle is characterized by a halo-shaped frictional contrast, in both forward (bright halo) and reversed (dark halo) FFM scans, which can be attributed to the presence of rubber directly exposed at the sample surface. Notice that *the PMMA layer on top of the rubber exhibits the same frictional contrast as the PMMA matrix, being indistinguishable from that in both forward and backward FFM scans*. In contrast, Ph-HFM resolves small differences in viscoelastic and/or adhesion hysteresis response time of the PMMA on top of the rubber that is not linked to the PMMA rubber matrix. Relaxation processes of polymeric materials are strongly dependent on the constraints for molecular movement. A different molecular density, entanglement density and/or molecular weight in the PMMA layer on top of the rubber that is detached from the PMMA matrix may lead to differences in the PMMA viscoelastic and/or adhesion hysteresis response. In addition differences in interfacial bonding between the rubber and the PMMA on top depending on whether the PMMA is well adhered to the PMMA matrix or not, may also modify the PMMA dynamic behavior. According to the obtained experimental results, the contrast provided by Ph-HFM allows us to distinguish differences in the locally probed dynamical response of PMMA on top of rubber depending on whether the PMMA is well adhered to the matrix or not, in spite of the fact that no difference can be resolved in conventional FFM. Hence, Ph-HFM allows us to study quick dissipative transitions not resolved by FFM that, however, may play an important role in MEM/NEMs devices working at much higher sliding velocities than those typically used in AFM/FFM measurements.

It is also worth noting that, when probed with extreme sensitivity, a locally measured response might be strongly affected by small dissipative ef-

fects induced by long-range interactions (via molecular entanglements) at molecules outside the immediate contact region. The possibility that those kinds of interactions might be detected in an extremely short time scale can be of interest in the implementation of dynamic mechanical procedures for communications in nanodevices.

Acknowledgement. E. Gnecco is gratefully acknowledged for scientific discussions, and careful reading of the manuscript. Financial support from the Spanish Ministerio de Educación y Ciencia (MEC) under project MAT2002-00076), the Junta de Comunidades de Castilla-La Mancha (JCCM) under projects PBI-02-003 and PBI-05-018, and the European Science Foundation (ESRF), under the ESRF Scientific Programme NATRIBO, is also gratefully acknowledged.

References

1. See, e. g. *Review on ultrasonic machining*. T.B. Thoe, D.K. Aspinwall, and M.L.H. Wise, Int. J. Mach. Tools Manufact. 38 (1998) 239 and Ref. therein.
2. See, e. g. *Acoustics of friction* A. Akay, J. Acoust. Soc. Am. 111 (2002) 1525 and Ref. therein.
3. See, e. g. K. Dransfeld, *Generation of ultrasonic waves in sliding friction*, Chap. 7 in *Nanoscience: Friction and Rheology on the Nanometer Scale*, ed. E. Meyer, R.M. Overney, K. Dransfeld, and T. Gyalong, World Scientific (1998) and refs. therein; this chapter and refs. therein.
4. *Atomic scale friction of a tungsten tip on a graphite surface*, C.M. Mate, G.M. McClelland, R. Erlandsson, and S. Chiang, Phys. Rev. Lett. 59 (1987) 942.
5. *Simultaneous measurement of lateral and normal forces with an optical-beam-deflection atomic force microscope*, G. Meyer and N. Amer, Appl. Phys. Lett. 57 (1990) 2089.
6. *Detection of Surface Acoustic Waves by Scanning Force Microscopy*, W. Rohrbeck and E. Chilla, Phys. Stat. Sol. (a) 131 (1992) 69.
7. *Nonlinear detection of ultrasonic vibrations in an Atomic Force Microscope*, O. Kolosov and K. Yamanaka Jpn. J. Appl. Phys. 32 (1993) L1095.
8. *Scanning microdeformation microscopy*, B. Cretin and F. Sthal, Appl. Phys. Lett. 62 (1993) 829.
9. *Ultrasonic Force Microscopy for nanometer resolution subsurface imaging*, K. Yamanaka, H. Ogiso, and O. Kolosov, Appl. Phys. Lett. 64 (1994) 178.
10. *Acoustic Microscopy by Atomic Force Microscopy*, U. Rabe and W. Arnold, Appl. Phys. Lett. 64 (1994) 1493.
11. *Scanning acoustic force microscopy measurements in grating-like electrodes*, T. Hesjedal, E. Chilla and H.-J. Froehlich, Appl. Phys. A 61 (1995) 237.
12. *Nanosubharmonics: the dynamics of small nonlinear contacts*, N.A. Burnham, A.J. Kulik, G. Gremaud, and G.A.D. Briggs, Phys. Rev. Lett. 74 (1995) 5092.
13. *Vibrations of free and surface-coupled atomic force microscope cantilevers: theory and experiments*, U. Rabe, K. Janser, and W. Arnold, Rev. Sci. Instr. 67 (1996) 3281.

14. *Scanning local-acceleration microscopy*, N.A. Burnham, A.J. Kulik, G. Gremaud, P.J. Gallo, and F. Oulevey, J. Vac. Sci. Technol. B 14 (1996) 794.
15. *Ultrasonic Atomic Force Microscopy with overtone excitation of the cantilever*, K. Yamanaka and S. Nakano, Jpn. J. Appl. Phys. 35 (1996) 3787.
16. *Contact imaging in the AFM using a high order flexural mode combined with a new sensor*, S.C. Minne, S.R. Manalis, A. Atalar and C.F. Quate, Appl. Phys. Lett. 68 (1996) 1427.
17. *Scanning microdeformation microscopy in reflexion mode*, P. Variac and B. Cretin, Appl. Phys. Lett. 68 (1996) 461.
18. *Nanoscale determination of phase velocity by scanning acoustic force microscopy*, E. Chilla, T. Hesjedal, and H.-J. Fröhlich, Phys. Rev. B 55 (1997) 15852.
19. *Mapping surface elastic properties of stiff and compliant materials on the nanoscale using ultrasonic force microscopy*, F. Dinelli, M.R. Castell, D.A. Ritchie, N.J. Mason, G.A.D. Briggs, and O.V. Kolosov, Philos. Mag. A 80 (2000) 2299.
20. *Waveguide ultrasonic force microscopy at 60 MHz*, K. Inagaki, O. Kolosov, A. Briggs, and O. Wright, Appl. Phys. Lett. 76 (2000) 1836.
21. *Heterodyne force microscopy of PMMA/rubber nanocomposites: nanomapping of viscoelastic response at ultrasonic frequencies*, M.T. Cuberes, H.E. Assender, G.A.D. Briggs, and O.V. Kolosov, J. Phys. D.: Appl. Phys. 33 (2000) 2347.
22. *Nonlinear detection of ultrasonic vibration of AFM cantilevers in and out of contact with the sample*, M.T. Cuberes, G.A.D. Briggs, and O. Kolosov, Nanotechnology 12 (2001) 53.
23. *Nanoscale imaging of buried structures via Scanning Near-Field Ultrasound Holography*, G.S. Shekhawat and V.P. Dravid, Science 310 (2005) 89.
24. *Local Elasticity and Lubrication Measurements Using Atomic Force and Friction Force Microscopy at Ultrasonic Frequencies*, V. Scherer, B. Bhushan, U. Rabe, and W. Arnold, IEEE Trans. Magn. 33 (1997) 4077.
25. *Lateral Force Microscopy Using Acoustic Force Microscopy*, V. Scherer, W. Arnold, and B. Bhushan, Surf. Interface Anal. 27 (1999) 578.
26. *On the nanoscale measurement of friction using atomic-force microscopy cantilever torsional resonances*, M. Reinstädtler, U. Rabe, V. Scherer, U. Hartnann, A. Goldade, B. Bhushan and W. Arnold, Appl. Phys. Lett. 82 (2003) 2604.
27. *Investigating ultra-thin lubricant layers using resonant friction force microscopy,* M. Reinstädtler, U. Rabe, A. Goldade, B. Bhushan and W. Arnold, Tribol. Int. 38 (2005) 533.
28. *Imaging of flexural and torsional resonance modes of atomic force microscopy cantilevers using optical interferometry,* M. Reinstädtler, U. Rabe, V. Scherer, J.A. Turner, and W. Arnold, Surf. Sci. 532 (2003) 1152.
29. *Imaging using lateral bending modes of atomic force microscopy cantilevers*, A. Caron, U. Rabe, M. Reinstätler, J.A. Turner, and W. Arnold, Appl. Phys. Lett. 85.
30. *Quantitative elasticity evaluation by contact resonance in an atomic force microscope*, K. Yamanaka and S. Nakano, Appl. Phys. A 66 (1998) S313.
31. *Mapping of lateral vibration of the tip in atomic force microscopy at the torsional resonance of the cantilever*, T. Kawagishi, A. Kato, U. Hoshi, H. Kawakatsu, Ultramicroscopy 91 (2002) 37.

32. *Imaging and measurement of elasticity and friction using the TRmode*, M. Reinstädtler, T. Kasai, U. Rabe, B. Bhushan, and W. Arnold, J. Phys. D: Appl. Phys. 38 (2005) R269.
33. *Lateral-force measurements in dynamic force microscopy*, O. Pfeiffer, R. Bennewitz, A. Baratoff, E. Meyer and P. Grütter, Phys. Rev. B 65 (2002) 161403.
34. *Determination of shear stiffness based on thermal noise analysis in atomic force microscopy: passive overtone microscopy*, T. Drobek, R.W. Stark, and W.M. Heck, Phys. Rev B 64 (2001) 0454001.
35. *Transverse surface acoustic wave detection by scanning acoustic force microscopy*, G. Bheme, T. Hesjedal, E. Chilla, and H.-J. Fröhlich, Appl. Phys. Lett. 73 (1998) 882.
36. *Simultaneous bimodal surface acoustic-wave velocity measurements by scanning acoustic force microscopy*, G. Behme and T. Hesjedal, Appl. Phys. Lett. 77 (2000) 759.
37. Recently, a novel AFM-based technique for studying nanoscale friction at velocities near to 10 nm s^{-1} has been implemented; see *A new atomic force microscopy based technique for studying nanoscale friction at high sliding velocities*, N.S. Tambe and B. Bhushan, J. Phys. D: Appl. Phys. 38 (2005) 764.
38. *Ultrasound induced lubricity in microscopic contact*, F. Dinelli, S.K. Biswas, G.A.D. Briggs, and O.V. Kolosov, Appl. Phys. Lett. 71 (1997) 1177.
39. *Friction control in thin-film lubrication*, J. Gao, W.D. Luedtke, and U. Landman, J. Phys. Chem. B 102 (1998) 5033.
40. *Coupling of normal and transverse motions during frictional sliding*, M. Heuberger, C. Drummond, and J. Israelachvili, J. Phys. Chem. B 102 (1998) 5038.
41. A. Socoliuc, E. Gnecco et al., submitted.
42. *Transition from stick-slip to continuous sliding in atomic friction: entering a new regime of ultralow friction*, A. Socoliuc, R. Bennewitz, E. Gnecco, and E. Meyer, Phys. Rev. Lett. 92.134301 (2004).
43. *Influence of ultrasonic surface acoustic waves on local friction studied by lateral force microscopy*, G. Behme and T. Hesjedal, Appl. Phys. A 70 (2000) 361.
44. *Influence of surface acoustic waves on lateral forces in scanning force microscopies*, G. Behme and T. Hesjedal, J. Appl. Phys. 89 (2001) 4850.
45. *The origin of ultrasound-induced friction reduction in microscopic mechanical contacts*, T. Hesjedal and G. Behme, IEEE Trans. Ultrason. Ferroelec. Freq. Cont. 49 (2002) 356.
46. *Probing the interface potential in stick/slip friction by a lateral force modulation technique*, J. Kerssemakers and J.T.M. De Hosson, Surf. Sci. 417 (1998) 281.
47. *Interaction potential and hopping dynamics governing sliding friction*, E. Riedo, E. Gnecco, R. Bennewitz, E. Meyer, and H. Brune, Phys. Rev. Lett. 91 (2003) 084502-1.
48. *The nonlinear nature of friction*, M. Urbakh, J. Klafter, D. Gourdon, and J. Israelachvili, Nature 430 (2004) 523 and Ref. therein.
49. *Tuning diffusion and friction in microscopic contacts by mechanical excitations*, Z. Tshiprut, A.E. Filippov, and M. Urbakh, Phys. Rev. Lett. 95 (2005) 0166101.
50. Chapter 15 in *Intermolecular and surface forces*, J. Israelachvili, Academic Press, Elsevier Ltd., 2nd edn (1992).
51. *Hysteresis of the cantilever shift in ultrasonic force microscopy*, K. Inagaki, O. Matsuda, and O.B. Wright, Appl. Phys. Lett. 80 (2002) 2386.

52. *Tribology and ultrasonic hysteresis at local scales*, R. Szoszkiewicz, B.D. Huey, O.V. Kolosov, G.A.D. Briggs, G. Gremaud, A.J. Kulik, Appl. Surf. Sci. 219 (2003) 54.
53. *Probing local water contents of in vitro protein films by ultrasonic force microscopy*, R. Szoszkiewicz, A.J. Kulik, G. Gremaud and M. Lekka, Appl. Phys. Lett. 86 (2005) 123901.
54. *Correlations between adhesion hysteresis and friction at molecular scales*, R. Szoszkiewicz, B. Bhushan, B.D. Huey, A.J. Kulik, and G. Gremaud, J. Chem. Phys. 122 (2005) 144708.
55. *Quantitative measure of nanoscale adhesion hysteresis by ultrasonic force microscopy*, R. Szoszkiewicz, A.J. Kulik, and G. Gremaud, J. Chem. Phys. 122 (2005) 134706.
56. *Measurements of stiff-material compliance on the nanoscale using ultrasonic force microscopy*, F. Dinelli, S.K. Biswas, G.A.D. Briggs, and O.V. Kolosov, Phys. Rev. B 61 (2000) 13995.
57. *A method of evaluating local elasticity and adhesion energy from the nonlinear response of AFM cantilever vibrations*, M. Muraoka and W. Arnold, JSME Int. J. Series A 44 (2001) 396.

5 Probing of Nanocontacts Inside a Transmission Electron Microscope

Donats Erts[1], Ants Lõhmus[2], Justin D. Holmes[3], and Håkan Olin[4]

[1] Institute of Chemical Physics, University of Latvia, Riga, Latvia
[2] Institute of Physics, University of Tartu, Tartu, Estonia
[3] Department of Chemistry, National University of Ireland, Cork, Ireland
[4] Engeneering Physics, Mid Sweden University, Sundsvall, Sweden

5.1 Introduction

This chapter reviews a recent new method, which is a combination of the scanning probe microscope (SPM) and the transmission electron microscope (TEM), with important applications in nanotribological investigations of contact properties. In these TEM-SPM instruments, the electron microscope is used for imaging and analysis of the sample as well as the SPM tip, while the SPM is used for probing of the electrical and mechanical properties, measurements of force interaction or for manipulation of the sample at the nanometer scale. One advantage with the TEM-SPM instrument, compared with standard SPM, is the direct imaging of both tip and sample giving important information such as tip-sample distance, tip and sample radius and shape, which is lacking or only available indirectly using standard SPM.

The first TEM-SPM instrument, introduced by Spence [1], was a combination of the scanning tunnelling microscope (STM) and a TEM. While this early instrument was used for STM imaging, other groups later used TEM-STM instruments for probing and manipulations. Examples of their use can be illustrated by the creation of atomically thin gold nanowires and simultaneous conductance measurements by the group of Takayanagi [2], studies of the mechanical properties of carbon nanotubes by Poncharal et al. [3], the low friction linear bearing realized by carbon nanotubes by Cumings and Zettl [4], the investigations of conductance of gold point contacts by Erts et al. [5], and studies of electromigration of metal inside and outside carbon nanotubes [6,7]. The family of in situ probe instruments is growing and work is progressing on the atomic force microscopes (AFM) [8–11] and nanoindenters [12–14]. The body of applications of this young technique is expected to grow with the appearance of commercially available instruments [15].

5.2 TEM-SPM Instruments

One challenge in the design of a TEM in situ probe instrument is the short distance between the pole pieces (2 – 10 mm) of the objective lens where the sample is placed. The standard TEM side entry sample holder, roughly a 20 cm

long cylinder with a diameter of less than 1 cm with the sample placed at one end, is inserted into the TEM column.

The simplest, single directional TEM-SPM prototype device can be designed on the base of commercially available TEM specimen goniometre type holder. It is improved by a specimen support disc with slit perpendicular to main axes [16]. One part of the disc is fixed to the base cartridge and other to the movable part. The sample, laser-ablated nanoparticle aggregates, were deposited on the specimen disc. Due to a big plastic deformation of the disc material the slit can be enlargened without breaking the disc when tension is applied. The origin of the elastic behaviour of the nanoparticle chain aggregates were studied using this device.

In the TEM-SPM designs the scanning probe part is placed at the sample position. The most common sample is a tip-formed sample in order not to shadow the electron beam. The two main parts in a scanning probe microscope are the piezo- scanner with a range of a few micrometers and a rough positioning system, which is necessary to reduce the probe-sample distance within the reach of the piezo-scanner. While the standard piezo scanner is a tube with a diameter of a few mm, the rough positioning system is usually bulkier.

5.2.1 TEM-STM

TEM-SPM, based on micromechanical techniques [17], has significant advantages in achieving ultra-low noise measurements in comparison to standard SPM. Due to very small dimensions, only 2.5 by 2.5 by 0.5 mm, the resonance frequency is high. Such dimensions allow the use of very high resolution pole pairs in TEM. However due to the difficulties of making such an instrument in common laboratories and complications arising from manipulation of the sample under study this device has not found wide use.

One common way for TEM-STM design [1,8,18] is based on an electrical motor with a long shaft to displace a piezo element into operating range. Such a design [8] with a piezo tube scanner and a geared stepper motor is shown in Fig. 5.1a. This design is simple and robust; however, one disadvantage is the lack of rough motion in lateral directions. The lack of lateral positioning capability results in frequent misalignments between the tip-shaped sample and the probing tip during the coarse approach, leading to necessary manual readjustment outside the TEM.

An elegant design, using a single piezo tube both for fine motion and for 3-dimensional rough adjustment, is shown in Fig. 5.1b [8,19]. The rough motion is based on an inertial slider mechanism. A saw-tooth voltage applied simultaneously on piezotube electrodes enables a shift along a straight line. For lateral motions only a pair of electrodes will be used resulting in a tilt of the tip. The approach of the tips can be carried out step by step: by shifting the movable element- tip holder and by correcting the direction under TEM observation.

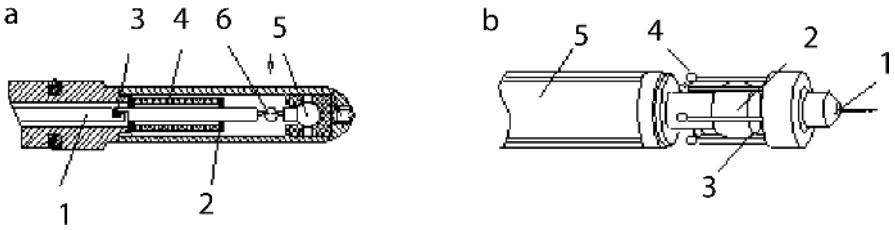


Fig. 5.1. TEM-SPM designs **a** Stepper motor based TEM-holder (not to scale): 1 – shifting rod by stepper motor, 2 – graphite rings, 3 – clutch, 4 – piezo tube, 5 – preadjustment ball, 6 – electron beam. **b** Ball-type of the inertial slider: 1 – tip holder, 2 – sapphire ball, 3 – sliding rods, 4 – counter weight, 5 – piezo tube

5.2.2 TEM-AFM

In the simplest version of TEM-AFM one tip, opposite to the piezo one is replaced with the AFM cantilever with a tip on it. If the standard AFM technique involves a cantilever and optical system for detecting the cantilever deflection then in TEM instead of the optical system the cantilever image is used for detecting the displacement of the tip [8, 9]. Kizuka [10, 11] used also optical detector for cantilever position control.

We have developed a TEM-AFM for shear force measurements based on a quartz resonator tuning fork inside the TEM. Friction increases rapidly as the distance between objects is reduced below 10 nm [20]. The instrument allows simultaneous turning and measurement of several parameters (frequency, amplitude, phase, sample potential, distance between nanoobjects, shear force damping, contact current, etc) while directly observing the system behaviour in TEM. An electronic module has been developed and integrated in the SPM controller. It consists of a frequency synthesizer, with frequencies tunable in 0.02 Hz steps, which can operate up to 10 MHz, and a lock-in amplifier for measurements of dissipated energy in the resonator.

5.2.3 In Situ Nanoindenter

The nanoindentation technique utilizes an actuator to press a sharp diamond tip a few nanometers into the sample while measuring the applied force, typically giving information about the hardness or elastic modulus of the material in the sub-micron regime. The data obtained by the nanoindentation method has been limited to load - displacement data and, by the lack of direct observation of the induced plastic deformation, to ex situ studies of the analysis of the indentation mark. Recently, however, an extension of the nanoindentation method has been demonstrated, using a TEM for in situ imaging of the entire indentation process [12–14]. To take full advantage of the TEM-nanoindentation method a proper force sensor is needed. Such a force sensor, fabricated by micromachining methods, is described in [13].

5.3 Force Interactions

TEM-AFM has a great potential for in situ observations of all force interaction. Nevertheless only few investigations of force interactions have been measured between Si, Cu and Au surfaces and between semiconductor nanowires and gold [9–11,21]. Figure 5.2 shows a typical sequence of TEM images where the Au sample with a protrusion of radius of 7.6 nm is moved towards and then away from the Au coated AFM tip of radius of 30 nm [9]. The corresponding force curve is found in Fig. 5.3.

5.3.1 Van der Waals Forces

Van der Waals forces can be determined by visualisation of the AFM tip positions before and after the jump to contact (Figs. 5.2 and 5.4) and calculating the force using the cantilever spring constant (Figs. 5.3 and 5.5).

A simple expression for the jump to contact, using a model with van der Waals forces between two spheres is [22]:

$$k \leq \frac{dF}{ds} = \frac{AR}{3s^3}\,, \tag{5.1}$$

where A is the Hamaker constant, s is the distance between the spheres, and R is the reduced radius of the spheres R_1 and R_2 : $R = ((R_1 + R_2)/R_1R_2)^{-1}$.

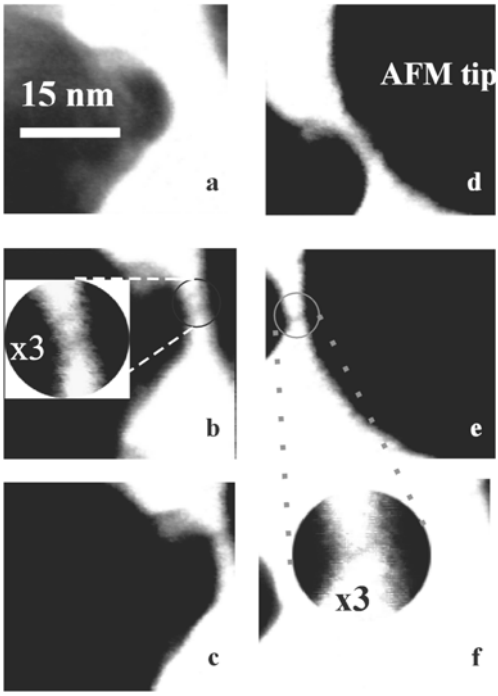


Fig. 5.2. A set of TEM images of a gold tip and gold coated AFM tip: **a** No contact. **b** Jump-to-contact. (Inset: neck formed in the gap in connection during the jump-in-contact event) **c** Moving further in. **d** Withdrawal of the sample. **e** By lateral motion of the sample a small nanowire is formed between the sample and tip. Just before breaking, the diameter of the wire is about 1 nm, which corresponds to about 10 atoms. (*Inset*: nanowire area magnified 3 times). **f** After jump-off-contact. [9]

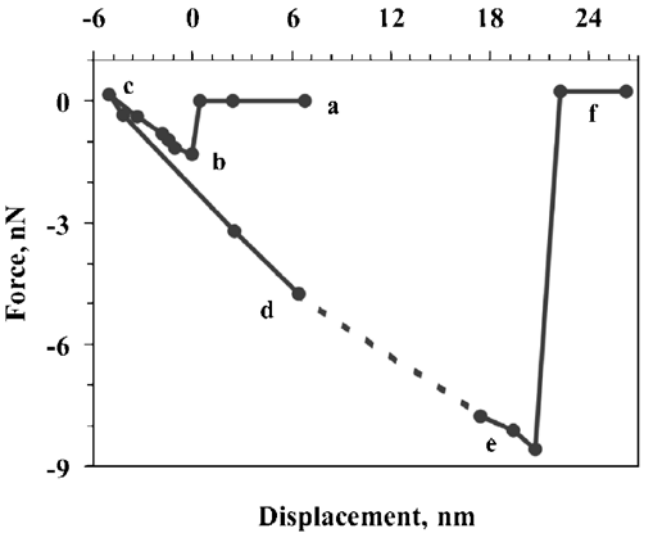


Fig. 5.3. Experimental force-distance curve where the labels a–f corresponds to the TEM images in Fig. 5.4a–f. [9]

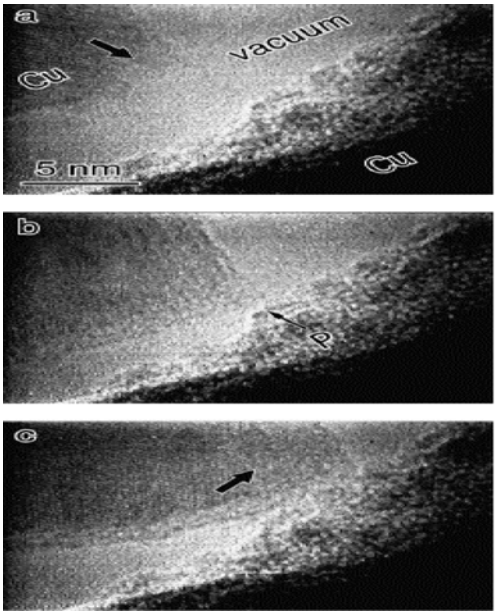


Fig. 5.4. TEM image snapshots of tip and surface, **a** before the tip contacts the surface; **b** just as the tip contacts the surface; **c** after the tip contacts, showing lateral displacement on the surface [11]

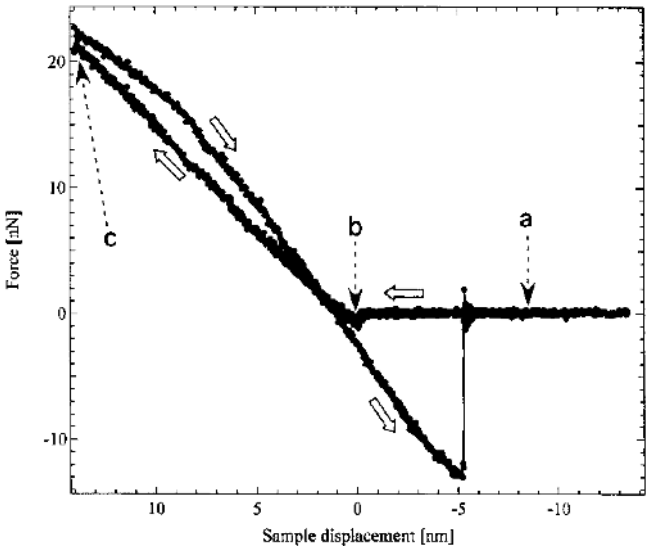


Fig. 5.5. Force-displacement curves for Cu tip and Cu surface [11]

Jump-to-contact distances shown in Fig. 5.2 are more than 20 times lower in comparison to those calculated using the cantilever spring constant. Such difference may be explained by the high mobility of the gold atoms in the gap between the two tips at room temperature when the static view might not be relevant.

One of conclusions from measurements of nanowires and nanotubes by TEM-STM is that both of them can be used as sensitive force sensors. The force constant of free standing Si and Ge nanowires is compatible to the AFM cantilever force constants. Measurements of the jump to contact for Si and Ge nanowires [21] have been used to determine force constant of nanowires.

5.3.2 Pull off Forces

With great success TEM-AFM can be used for the characterization of adhesion forces by visualization of contacts during the retraction process [9–11]. Figures 5.2c–f and 5.3 shows an example of the force measurement and TEM image of a nanowire with a diameter of less than 1 nm and length of 2 nm (Fig. 5.2e and inset) [9]. The nanowire was created from the larger diameter contact by shear force by a lateral motion of the sample (Fig. 5.2d,e). The nanowire broke at a retraction of 22 nm (Fig. 5.2e,f) which corresponds to an attractive force of about 9 nN (Fig. 5.3). The calculated cohesion force was approximately 1 nN per atom which is in agreement with gold nanowire force experiments (1.6 nN for the last atom) [23] as well as theoretical calculations (from 1 to 2.2 nN) [24].

Kizuka [10] measured the tensile strength, strain-stress force and conductivity at the same time for a gold nanowire breaking. The strain-stress curve exhibits a swath edge curve, indicating that the deformation proceeds by the repeated process of elastic elongation and slip. Oscillations were observed in the strain-stress curve before fracture. The yield stress is determined and is ≈ 8 GPa.

One more application is measurement of the strength of a junction between a carbon nanotube and gold created by Jeoule heating [25].The strength of contact is estimated from the critical bending of the cantilever at the fracture and is approximately 0.6 MPa. This strength is similar to the exfoliation strength of the graphite layers 0.4 MPa.

5.3.3 Shear Forces, Friction

Kizuka [10, 11] observed a frictional movement of the AFM tip during approach and the first contact to the surface between Au, Cu and oxide coated Si surfaces (Fig. 5.6). A hysteresis loop in the force-distance curves caused by frictional movement of the tip at the contact region is observed [11] (Fig. 5.5). Lateral frictional displacement of the tip on the sample surface along the cantilever in the contact state (Fig. 5.6) is observed by the TEM (Fig. 5.4). Frictional displacements on the surface just after the contact may achieve 5 nm and more and no contacts without sliding were observed. The tip slides in along the opposite direction due to the retraction and the tip position returns to its initial contact position [10]. It is shown that the lateral displacement is not constant but changes with normal load, which is very important for interpretation of friction data, especially on an atomic scale [11]. For oxide coated Si surfaces the contact boundary of silicon oxide was hardly deformed

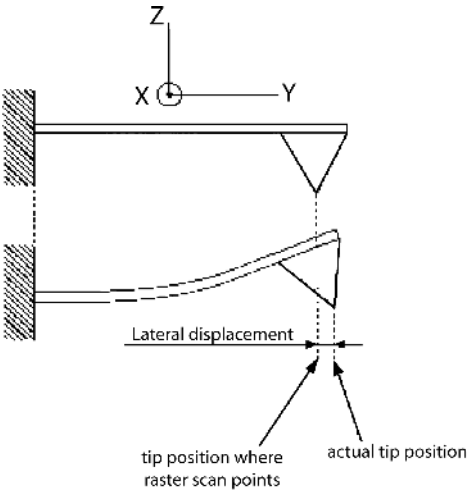


Fig. 5.6. Lateral displacement along Y direction, generated by the Z direction force [11]

when the load force is less than 10 nN [10]. At a greater load force viscous–flow-like deformation was observed [26]. Tip sliding causes torque motion and this motion contributes to the deflection and torsion of the cantilever in addition to displacement of the cantilever along the y direction. The ratio of the deflection component can be estimated from in situ measurements [10].

Shear behaviour in contacts was investigated in [27]. Figure 5.7 shows a time-sequence series of a shear deformation process in a gold contact. By shear displacement twinning occurs in the upper part (Fig. 5.7b) with subsequent twinning in the middle (Fig. 5.7) and lower parts (Fig. 5.7) when the tip is moved left. The twinning gradually recovers by displacement of the tip to the right and finally tips become a different single crystalline structure in comparison to Fig. 5.7. This shows that slip is also attributed to the deformation. A slip process with atomic resolution is visualized in Fig. 5.8

Stick slip motion was also observed when the gold tip is scanned over a gold surface with zero distance [27] (Fig. 5.9). This means that the displacement is disturbed by a kinetic friction. Several layers at the two surfaces and the contact boundary are responsible for the contact-type surface scanning process. The strength of the boundary is attributed to a static friction when two tips bond and fix and is responsible for a kinetic friction during the displacement.

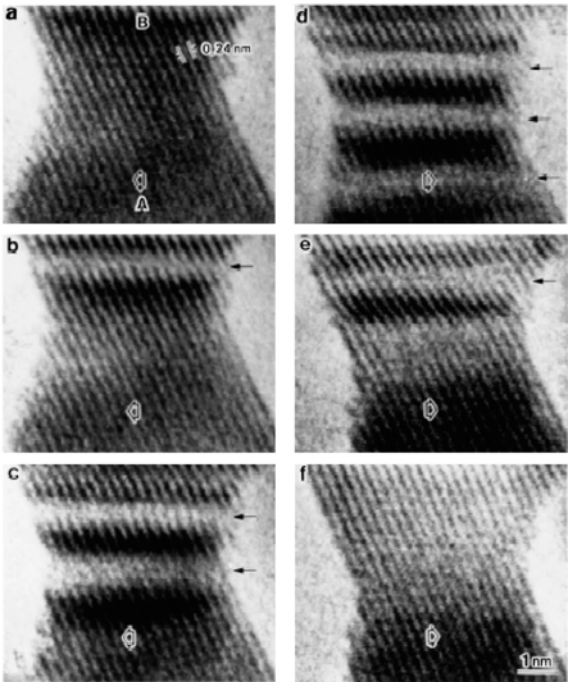


Fig. 5.7. Images of the process of shear tests in nanometer-sized gold. Bold arrays show the direction of displacement of the mobile side. Arrows show twinning [27]

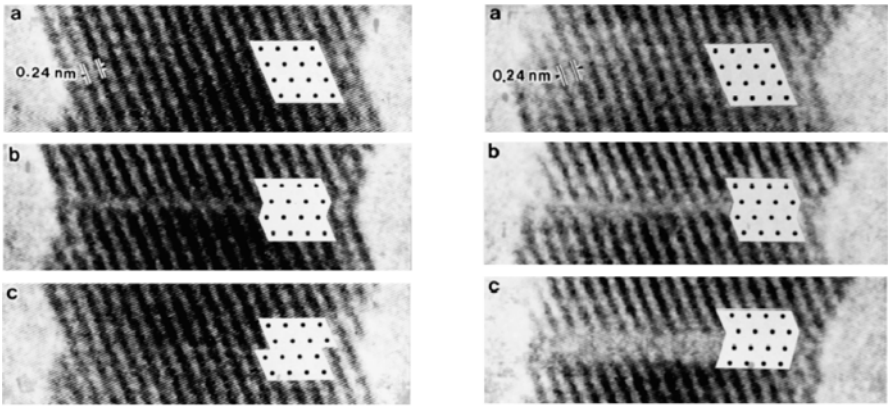


Fig. 5.8. Images of the process of slip (*left side column*) and twinning (*right side column*) during the shear deformation. An atomic arrangement projected along the [110] axis is inserted [27]

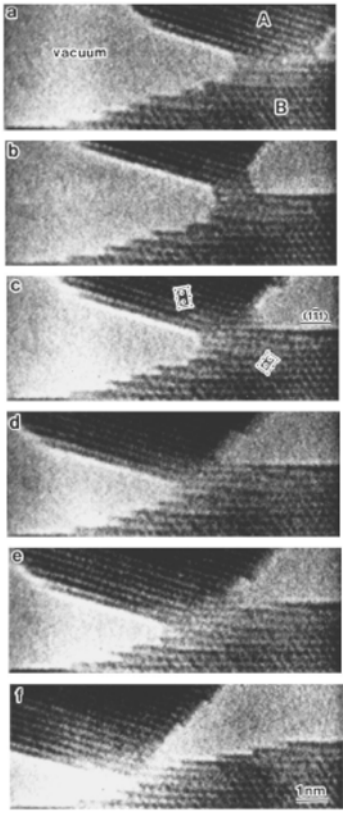


Fig. 5.9. Images of the scanning of a gold tip when the distance between the tip and a gold tip in a fixed side is 0 nm. The two tips bond by a boundary of a few atomic columns width. Frames show the unit cells of gold with face-centred cubic structure projected along the [110] direction [18]

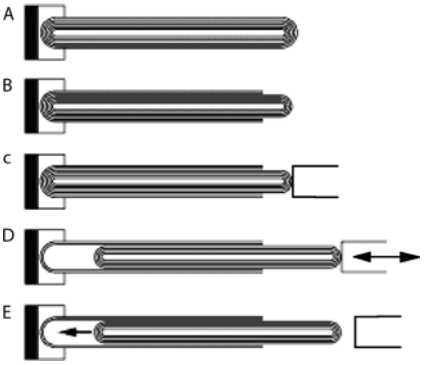


Fig. 5.10. Schematic representation of the experiments performed inside TEM. **A** to **C** The process of opening the end of MWNT (**A**), exposing the core tubes (**B**), and attaching the nanomanipulator to the core tubes (**C**). **D** and **E** Two different classes of subsequent experiments performed. In **D**, the nanotube is repeatedly telescoped while observations for wear are performed. In **E**, the core is relased and pulled into the outer shell housing by attractive van der Waals force [4]

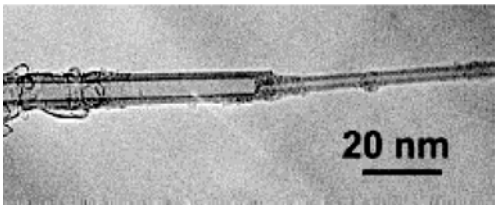


Fig. 5.11. A TEM image of a telescoped nanotube. This particular nanotube originally had nine shells, but upon telescoping a four shell core has been nearly completly extracted [4]

In [4] low friction nanoscale linear bearing from multiwall carbon nanotubes was realized as shown in schematics in Figs. 5.10 and 5.11. The manipulator is contacting to the inner shells of nanotube and controlled and reversible telescoping of inner shells according to outer shells is realized. After 20 telescopic processes with different nanotubes no evidence for sliding induced wear on active surfaces was found. Telescopic nanotube is expected to act as constant force spring and the telescoped part is moving back by vander Waals driven forces. Static and dynamic friction forces are calculated.

5.3.4 Electrostatic forces

Force curves can be changed by applying an electrical field between the surface and the tip. Electrostatic forces have been investigated between Si and Ge nanowires and a gold electrode in [21] and carbon nanotube bundles [28]. In [29, 30] electrostatic forces between the electrode and a carbon nanotube were used for exiting carbon nanotubes to resonance frequencies and determination of the Young modulus.

Forces acting on a Ge nanowire – gold tip (Fig. 5.12a) are schematically shown in Fig. 5.12b. Figure 5.12c shows the comparison between the calculated and measured forces for the interactions between a Si nanowire – gold electrode at a bias of 1 V. There is good agreement with the experimental results of the jump-to-contact and jump-off-contact distance at low voltages. Higher voltages tend to result in longer jump-to-contact and shorten jump-off-contact distances than calculated. This suggests that the electrostatic at-

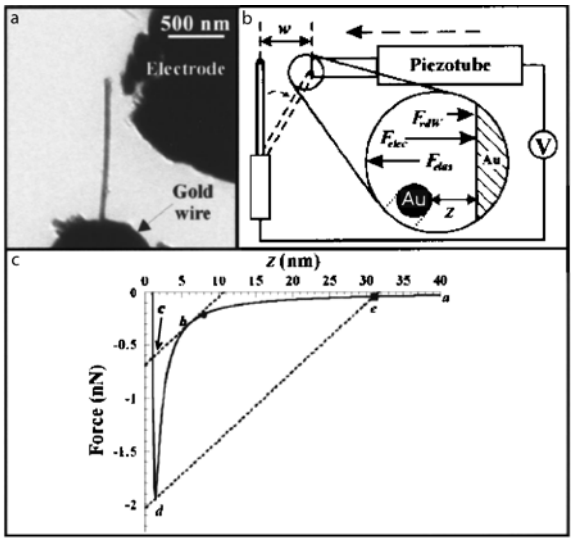


Fig. 5.12. a TEM image of a Ge nanowire utilized for TEM-STM measurements. **b** Schematic representation of TEM-STM studies. The electrode is positioned by movement of the piezotube. The zoom-in picture demonstrates the force interactions between the nanowire tip and the electrode where z is the distance of separation between the nanowire tip and the electrode with w being the initial separation distance. The attractive vdW (F_{vdW}) and electrostatic (F_{elec}) forces are countered by the elastic force exerted by the nanowire (F_{elas}). With applied electrostatic voltages, the total force acting on the nanowire tip is $F_{\mathrm{T}} = F_{\mathrm{vdW}} + F_{\mathrm{elec}}$. **c** Force-distance plot calculated for the interactions of a Si nanowire ($d = 90\,\mathrm{nm}$) with an applied voltage of 1 V. The dotted lines represent the spring constant of the nanowire. Measured jump-to-contact (*circle*) and jump-off-contact (*square*) distances are plotted for comparison [21]

tractive interaction forces are stronger than the sphere-plane interactions calculated at high potentials which could be due to the breakdown of the electrostatic potential equation at high voltages or large distances. Movement of the nanowire during withdrawal results in shearing forces at the contact point and shorter jump-off-contact distances are observed.

Electrostatic force driven jump-to-contact from a distance of 30 nm was observed between carbon nanotube bundles [28]. Forces acting on the nanowire were calculated by a capacitator model but were not directly measured.

Electrostatic forces are applied for the realization of fast nanoelectromechanical devices. The switching behaviour of a bistable nanowire-based nanoelectromechanical erasable programmable read-only memory (NEMPROM) device is shown in Fig. 5.13. It can be seen by the calculated energy diagrams in Fig. 5.13a, there are two local minima at low voltages and the circuit is OFF due to an energy minimum at the device separation distance (w) where the elastic energy of the nanowire is zero. The other minimum is due to van

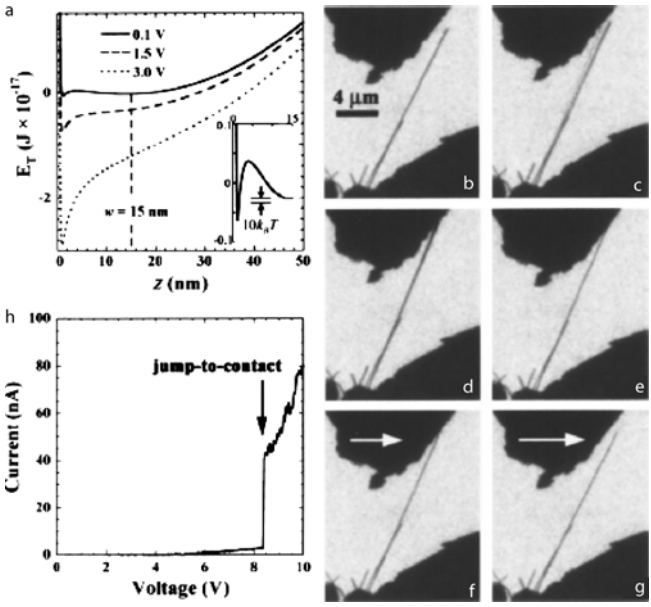


Fig. 5.13. a NEMPROM device calculations at different electrostatic potentials for Ge nanowire ($d = 50\,\text{nm}$; $l = 1.5\,\mu\text{m}$). Inset shows the energy barrier between two stable (ON/OFF) minima in relation to $10\,k_BT$. **b–d** TEM sequence showing the jump-to-contact of a Ge nanowire as the voltage is increased. **e** TEM image demonstrating the stability of device after removal of the electrostatic potential. **f,g** The resetting behaviour of the device. Note that the device is indefinitely stable but reset with the slight amount of shearing motion. **h** I(V) of NEMPROM device showing no conductivity until after contact is made at a potential of 8.4 V [21]

der Waals interactions when the wire and electrode are in contact. To switch between these two minima, an electrostatic field of 3 V is applied which alters the interaction energy resulting in a new energy minimum at shorter separation distances and deflection of the nanowire into contact with the gold electrode resulting in an ON state. Removal of the electrostatic potential does not allow the nanowire to switch to the OFF position due to the energy barrier and is stable when the barrier $\gg 10k_BT$. An NEWPROM device made from a Ge nanowire can be seen in the TEM sequence shown in Fig. 5.13b–f. By applying a voltage, the resulting electrostatic field deflects the nanowire into contact with the gold electrode. The nanowire does not jump-to-contact until the attractive electrostatic potential is greater than the elastic potential energy of the nanowire. Figure 5.13h verifies that there is no conductivity until the jump-to-contact is made at 8.4 V. The nanowire remains in contact with the electrode even when the electrostatic field is removed due to the minimum in the potential energy curve. Although these devices are stable, these NEWPROM devices can be easily switched OFF by mechanical motion

or by heating the device above the stability limit ($\gg 10k_{\mathrm{B}}T$). Figure 5.13e through 5.13g demonstrates that very little shearing motion is required to overcome the van der Waals attractive forces.

5.4 Nanocontacts

5.4.1 Contact Formation

Using TEM-SPM as manipulator, contact formation between two tips can be observed. For example contact formation between gold surfaces [18,31] and Si surfaces coated with an oxide layer [32] and without an oxide layer [33] have been investigated with atomic resolution. For an Au tip approach to distance of 0.3 nm, a few atomic columns emerge (Fig. 5.14) and two tips are contacted with a boundary of a few atomic columns width [18, 31]. Atom diffusion at nanometer sized contacts occurs due to the decrease in barrier height in addition to atomic force, and/or atomic emission in an electric field [31]. Material jump in contact between gold nanoasperities was also observed from larger distances of 0.9 – 1 nm (Fig. 5.15) [34]. This distance is two times larger than the value obtained by molecular dynamics simulations by Landman et al. [35].

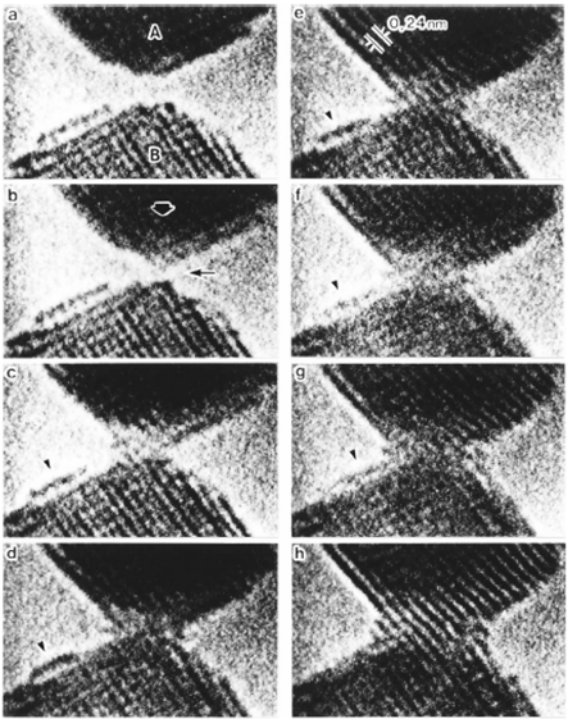


Fig. 5.14. Images of the formation of contact boundary and neck growth in the point contact of gold [31]

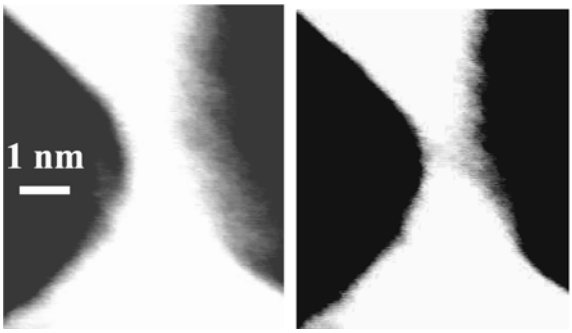


Fig. 5.15. Gold bride organization between two nanoasperities at distance around 1 nm [34]

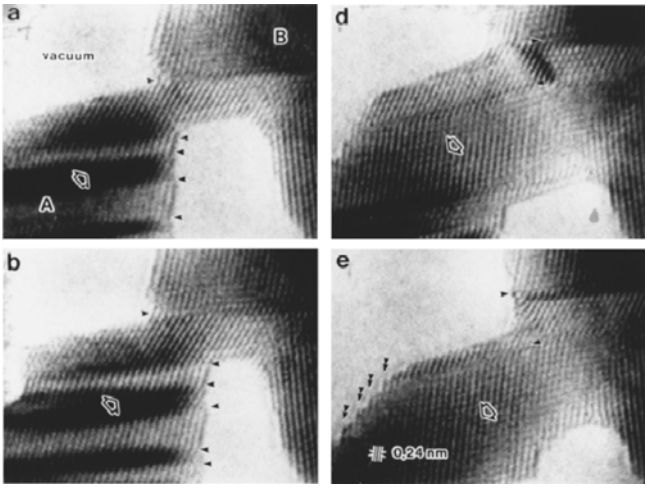


Fig. 5.16. Images of the process of compressing and tensile deformation in nanometer-sized gold. Bold arrays show the direction of displacement of the mobile side. Triangles show $\{111\{/\{111\}\ \Sigma = 3$ twin boundaries. Double triangles show atomic-scale slip steps [27]

By tip movement further growth of the diameter of the neck occurs due to the compressive deformation and contribution of gold surface diffusion [31]. By contact compression twin boundaries (the bright bands between the dark bands are twins) are generated, migrated, and annihilated [27] (Fig. 5.16a–c).

Figure 5.17 shows contact formation by pressing together two Si tips coated with amorphous oxide coated silicon with a thickness of 2 nm [32]. The contact boundary is located in the centre of amorphous interphase layer. For contact between clean Si surfaces elastic deformation occurs near the contact boundary [33]. Contacts created are shown to be crystalline (Fig. 5.18). Depending on tip orientations and the mismatch angle, the tip rotation and defect formation is observed at the contact boundary (Fig. 5.18). The defect formation shows that atomic diffusion to annihilate defects does not occur at

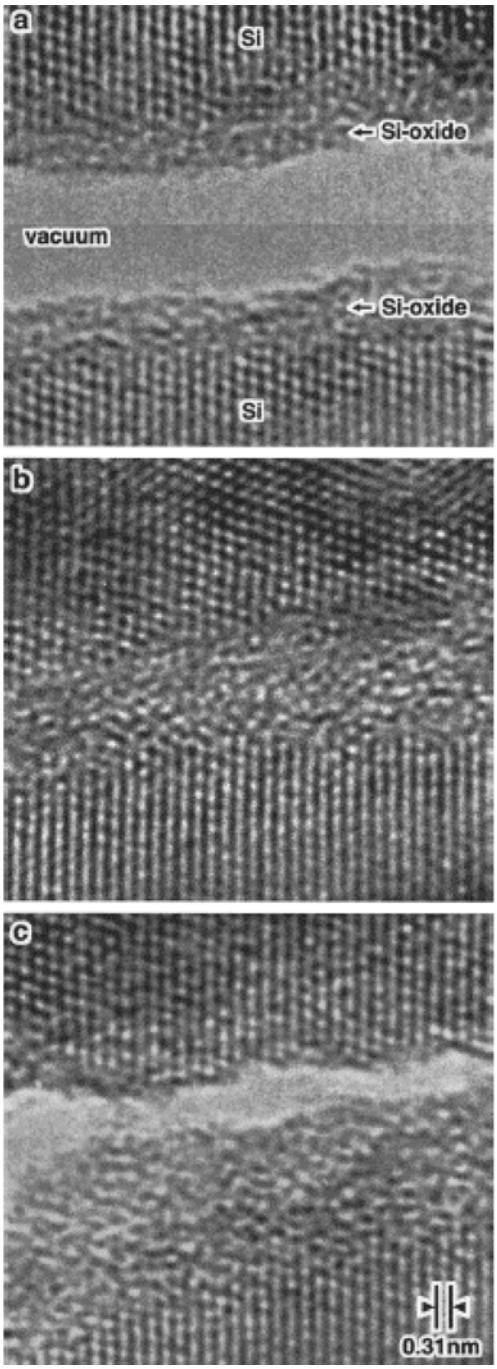


Fig. 5.17. Images of contact and the subsequent retraction process of two Si tips coated with amorphys Si oxide of about 1 nm thickness [32]

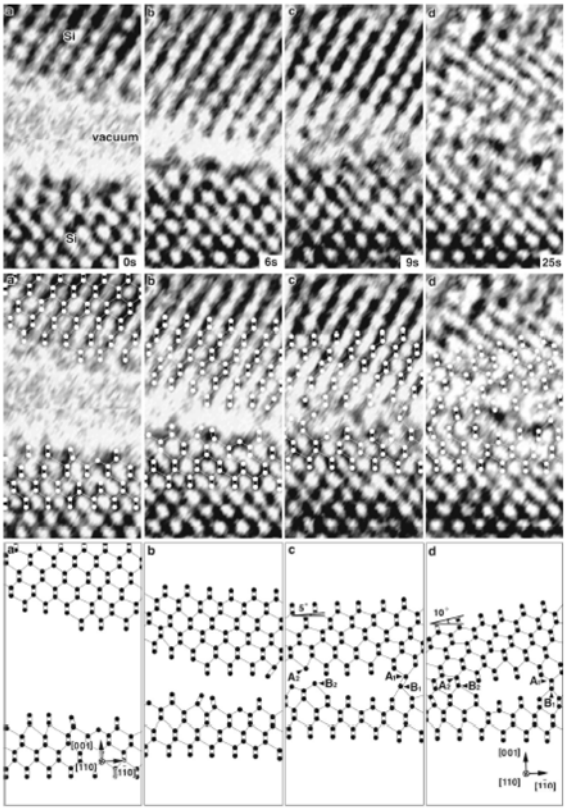


Fig. 5.18. TEM image of Si surfaces before in situ contact and after. White filled circles and lines show the positions of the atomic columns along [110] and Si–Si atomic bonding, respectively [33]

room temperature. In the gold contacts, contact boundaries are relaxed due to atomic diffusion at room temperature; localized defect structures including dislocations are not stable in the gold contact boundaries [31].

5.4.2 Contact breaking

TEM-SPM has been used to investigate contact behaviour under tensile stress for gold, silicon, and carbon nanotubes [2,4,27,32]. When large gold contacts are deformed by tensile force (Fig. 5.16) twin boundaries are generated, migrated, and annihilated [27] (Fig. 5.16c–f) similar to that seen during contact compression (Fig. 5.16a–c). Slip steps on an atomic scale form edges (shown by triangles in Figs. 5.16a–c) even twin boundaries are not observed. This shows that slip is also attributed to the deformation. Stick slip motion and stacking fault formation were investigated together with measurements of tensile stress and strain-stress curves in [10]. It was shown that between slips the point contact is deformed elastically. Structural relaxation due to atomic flow is observed after the slip particularly when the width of the point contact is less than 1 nm [31].

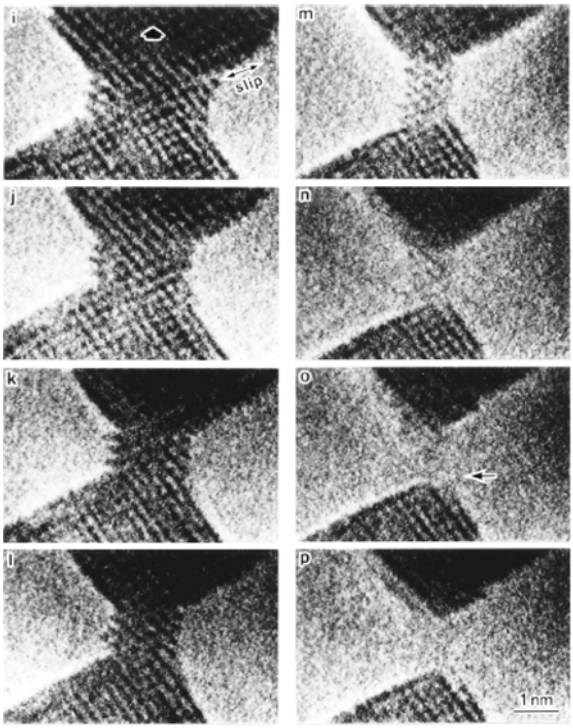


Fig. 5.19. Images of the formation of an atomistic pillarlike neck of gold during retraction. The width of the pillarlike neck is shown by the number of the (002) atomic layers in upper tip (A); the numbers is 14 (**i**), 11 (**j**), 9 (**k**), 8 (**l**), and 5 (**m**) [31]

The formation of atomistic pillar like necks by the slip deformation and structural relaxation during retraction was observed by [31] (Fig. 5.19). During the fracture and disappearance of such necks the introduction of a dislocation or dislocation-like localized strain is could not be confirmed. Finally pillar like neck breaks and disappears. The shape of tips becomes sharp as compared with that before contact. The top of the tips elongates by a few atomic layers after contact breaking.

Figure 5.20 demonstrates the breaking gold contact diameter from only a 6 atom line to 1 atom line [2]. It can be seen that the gold chains break one by one. The distance between individual atoms in the last gold atom chain is larger in comparison to gold atomic distances (Fig. 5.21).

The contact between oxidized Si surfaces and the amorphous interphase layers is viscously elongated during retracting and the crystalline Si regions are also deformed (Fig. 5.17) [32]. No slip of the lattices or dislocation motion was observed during the deformation in the crystalline regions and they deformed elastically not plastically. The bonding boundary is fractured at the boundary between the amorphous Si oxide and the crystalline Si (Fig. 5.17c). The thickness of the amorphous layer along the retraction direction increases by about 1 nm due to a viscous flow-like deformation. Such deformations were observed for surfaces with other orientations.

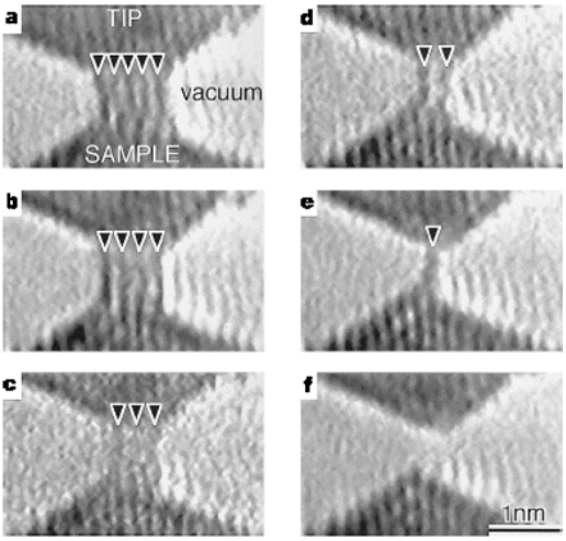


Fig. 5.20. Images of a contact while withdrawing the tip. A gold bridge thinned from **a** to **e** and ruptured at **f**. Dark lines indicated by arrowheads are rows of gold atoms. The conductance of the contact is 0 at **f** and in units of quantum conductance $\approx (13\,\mathrm{k\Omega})^{-1}$ at $V = -10\,\mathrm{mV}$, $R = 10\,\mathrm{k\Omega}$ [2]

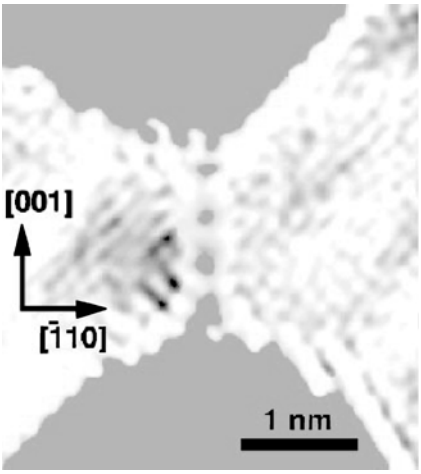


Fig. 5.21. Image of a linear strand of gold atoms (four colored dots) forming a bridge between two gold films (colored areas). The spacings of the four gold atoms are 0.35 – 0.40 nm. The strand is oriented along the direction of the gold (110) film. This image was processed to highlight the linear strand, where the lattice fringes of the gold film in the original electron microscope image were filtered out by Fourier transform [2]

Stretching and breaking weaker contacts formed by nanoparticle chain aggregates composed of carbon, titanium, alumina, and iron oxide have been performed by [16,36] using a breaking device inside a TEM. Figure 5.22 shows one example of stretching, breaking and recoiling of such chains observed in situ.

5.4.3 Adhesion

In [9] TEM-AFM was used to compare adhesion theories. The contact area between two gold electrodes was measured by a zero applied load when the contact area is determined by adhesion forces only. At zero loads it is not

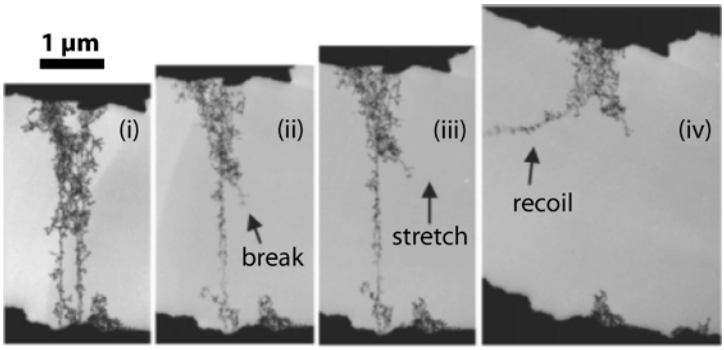


Fig. 5.22. TEM photo showing the sequential stretching and breaking of single nanoparticle chain aggregate (NCA) chains: **i** NCA deposited between two separating surfaces; **ii** stretching and breaking of one of the chains; **iii** stretching and **iv** breaking of second chain. The lower broken part disappeared, probably because it recoiled to the specimen support. In both **ii** and **iv**, the chain broke somewhere along its length and did not detach at the support surfaces [36]

necessary to know the cantilever spring constant and uncertainties in the cantilever force constant are not important.

The limiting cases in contact adhesion are based on Johnson–Kendall–Roberts (JKR) [37], Derjaguin–Müller–Toporov (DMT) [38] theories, and the transition between them, can be described by a dimensionless transition parameters called the Tabor parameter μ [39] and the Maugis parameter λ [40]. Both parameters are related and for contact between identical materials can be expressed as:

$$\lambda = 1.157\mu = \left(\frac{64R\gamma^2}{9K^2 z_0^3}\right)^{\frac{1}{3}} , \tag{5.2}$$

where z_0 is a typical atomic dimension, γ is surface energy and the reduced Young modulus K is given by $K = 4/3((1-\nu_1^2)/E_1 + (1-\nu_2^2)/E_2)^{-1}$, where ν_1 and ν_2 are Poissons ratio and E_1 and E_2 are Young modulus for two contacting spheres.

The DMT theory is valid when $\mu < 0.1$ and the JKR theory is valid when $\mu > 5$. Maugis [40] provides a more general theory suitable for the full range, where λ is a transition parameter. A dimensionless contact radius at zero applied loads, $\breve{\alpha}_0$ is described by:

$$\breve{\alpha}_0(\lambda) = \alpha_0 \left(\frac{K}{2\pi\gamma R^2}\right)^{\frac{1}{3}} , \tag{5.3}$$

where α_0 is the real contact radius. Maugis solution is analytical, but here we use a simplified fitting version given by Carpick et al. [41]:

$$\breve{\alpha}_0(\lambda) = 1.54 + 0.279 \left(\frac{2.28\lambda^{1.3} - 1}{2.28\lambda^{1.3} + 1}\right) . \tag{5.4}$$

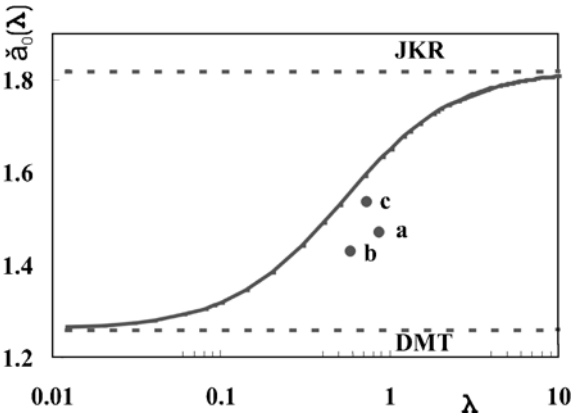


Fig. 5.23. Dimensionless contact radius at zero applied load for three contacts: **a** corresponds to the contact in Fig. 5.2, **b** and **c** are not shown). The solid line is the Maugis theory (Eq. 5.4) and the JKR and DMT limits are shown with dotted lines. [9]

To calculate λ (and μ), the following values for gold were used: $\gamma = 1.37\,\mathrm{J/m^2}$ [42], $E = 117\,\mathrm{GPa}$, $\nu = 0.42$ [43], $z_0 = 0.28\,\mathrm{nm}$. These values are dependent on the lattice orientation and the reported values in the literature have a spread of up to 50%, which could change the picture quite a bit. The theoretical and experimental values of the contact radius at zero applied loads are shown in Fig. 5.23. Our experimental results were in the transition region between the DMT and JKR models.

This TEM-AFM method, if extended to include the critical load, could mean that this question could be addressed experimentally in an adequate way. In this way one could measure γ independently, and arrive with a safer λ.

5.5 Conductivity of Nanocontacts

Analysis of conductivity dependence on contact size by TEM-STM was the main task in the beginning leading to the invention of the TEM-SPM. [2, 5, 10]. Previously conductivity quantization was observed by different contact breaking techniques, for example, STM and break junctions where the main problem was the real observation of contact size. The TEM-SPM technique gives clear verification that conductivity quantization occurs for contacts with atomic dimensions. Conductivity for last atom chain as shown in Fig. 5.20 is equal to $2 \times 13\,\mathrm{k\Omega}$ where $13\,\mathrm{k\Omega}$ corresponds to the quantum resistance $R_e = h/2e^2$ [44], where h is Plank constant and e is electron charge. Figure 5.24 shows the same figure in conductivity units (Fig. 5.24a) and schematics of the atom arrangements just before breaking (Fig. 5.24d)

In [5] conductivity was also measured for larger sized gold nanocontacts and results were compared with the Sharvin [45] and Wexler [46] theories. It was shown that ballistic electron transport is observed for gold nanowires with diameter around 1 nm (straight line in Fig. 5.25). At larger diameters experimental points were fitted with the Wexler formulae for the mixed

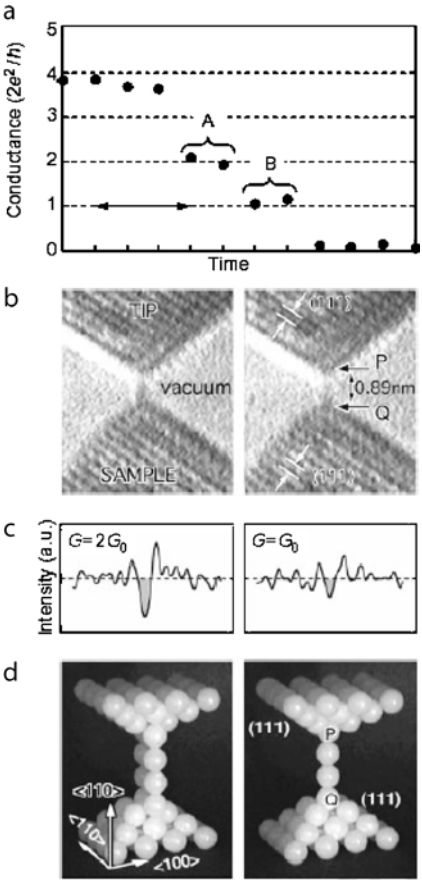


Fig. 5.24. Quantized conductance of a single and double strand of gold atoms. **a** Conductance change of a contact while wuthdrawing the tip. Conductance is shown in units of quantum conductance $G_0 = 2e^2/h \approx (13\,\mathrm{k\Omega})^{-1}$. $V = 13\,\mathrm{mV}$, $R = 100\,\mathrm{k\Omega}$. **b** images of gold bridges obtained simultaneously with the conductance measurements in (**a**). Left bridge at *step A*; right bridge at *step B*. **c** Intensity profiles of the left and right bridges shown in (**b**). The shaded area is the intensity from the bridge after subtraction of the background noise. **d** Models of the left and right bridges. The bridge at *step A* has two rows of atoms; the bridge at *step B* has only one row of atoms. The distance from P to Q (see **b**) is about 0.89 nm, wide enough to have two gold atoms in a bridge if the gold atoms have the nearest-neighbour spacing of the bulk crystal (0.288 nm) [2]

Sharvin-Maxwell regime. From this data, the electron mean free path was calculated and it was found to be only 4 nm which is 10 times lower in comparison to bulk gold and can be explained by the presence of defects in the gold nanowires [5].

Recently TEM-SPM was also used to determine conductivity through individual nanowires and carbon nanotubes. The $I(V)$ characteristics of Si and Ge nanowires indicated that ohmic contacts could be made with silicon nanowires whereas germanium nanowires displayed $I(V)$ that were dependent on the point of contact (Fig. 5.26) [21]. The observed nonconductive gaps in $I(V)$ characteristics for Ge nanowires were explained by the presence of different thickness oxide layers on the nanowires.

Although the contact resistances cannot be adequately determined, it was found that the resistivities of the Si nanowires were approximated in the order of $10^{-2}\,\Omega\,\mathrm{m}$ which are indicative of a highly doped nanowires with an impurity (most probably gold).

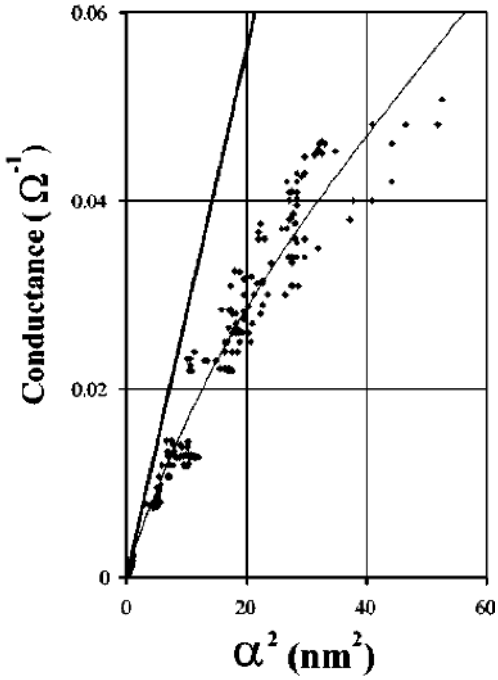


Fig. 5.25. Measured point contact conductance (Ω^{-1}) vs. radius squared (α^2), at bias 10 mV. The Wexler interpolation formula is plotted using a mean free path value of 3.8 nm and $\Gamma = 0.7$. Sharvin conductance (*straight thick line*) is added for comparison. Reproduced with permission from *Phys. Rev. B* **2000**, 61, 12725, Copyright 2000 American Physical Society

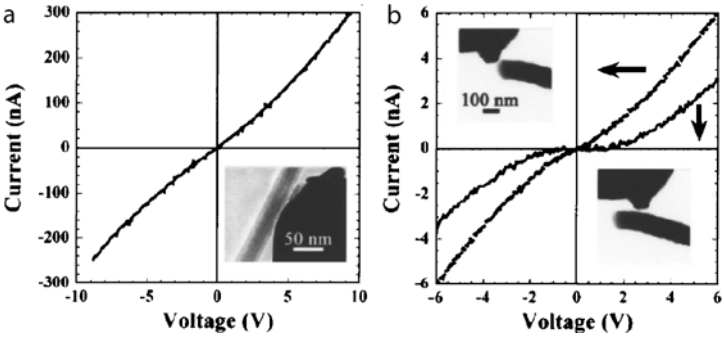


Fig. 5.26. I(V) behavior for an individual **a** Si and **b** Ge nanowire. I(V) for Ge nanowires are contact dependent [21]

For precise measurements of conductivity in carbon nanotubes in [47] contacts with liquid metal surface are created and contact resistances in the range 0.1 to 1 kΩ µm were obtained. Ballistic electron transport in carbon nanotubes was measured and an electron mean free path longer than 65 µm was found in the nanotubes.

Resistivity between the ends of multiwall carbon nanotubes during telescopic extension of the nanotube was investigated in [48]. Nonlinear resistance between the ends of the multiwall nanotubes during telescopic extension of

nanotube was observed and a one-dimensional localized system with characteristic localization length 1000 – 1500 nm predicted.

Besides conductivity TEM-STM can be applied for the characterization of tunnelling and field emission. Field emission properties of boron nitride and carbon nanotubes were measured in situ by Zettl [49] and our groups. Current-voltage characteristics were measured and work-function determined. In [50] electron holography of carbon nanotubes has been performed in situ. This method gave information about inner electric fields of materials. Measurements of the phase shift and phase gradient maps (Fig. 5.27) show that the electric field is concentrated precisely at the end of nanotubes and not at other nanotube defects such as sidewall imperfections.

One specific topic in conductivity measurements is contact behaviour under high current densities which can be achieved by relatively low bias voltages. By applying voltage between the graphite coated tip and graphite surface occasionally nanotube growth was observed (Fig. 5.28) [51]. It is shown that the bias voltage plays a key role in triggering the formation of

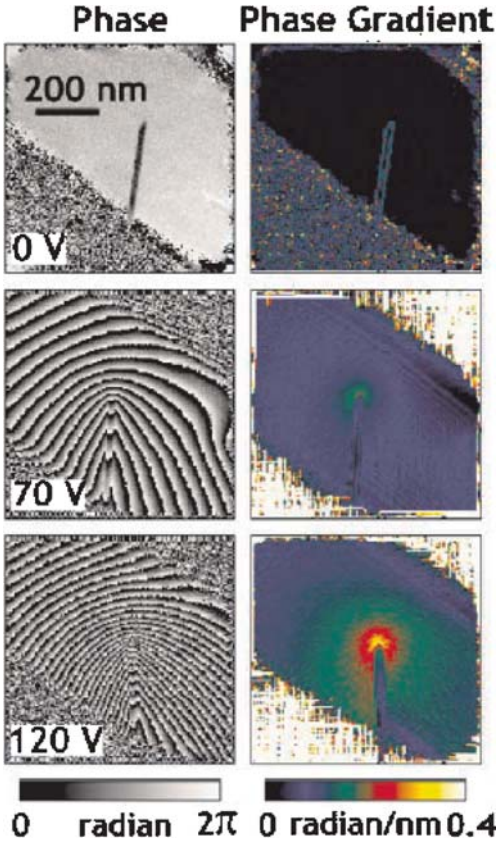


Fig. 5.27. Phase shift and phase gradient maps extracted from holograms of the same nanotube at bias voltages 0, 70, and 120 V. The phase gradient indicates where the electric field is the strongest; note the concentration of the electric field at the nanotube tip for 70 and 120 V [50]

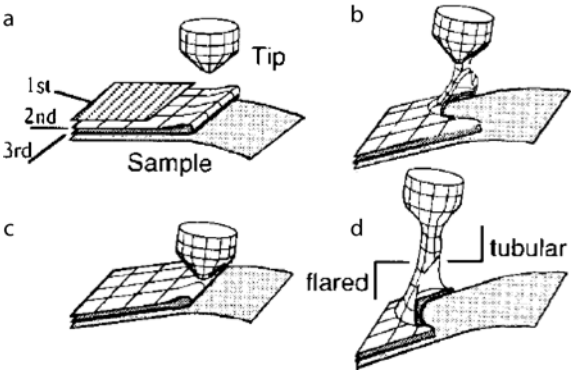


Fig. 5.28. Model of the growth of a nanobridge. **a** Before contact. **b** The tip touches edge. **c** The tip is retracted. **d** Nanotube growth by further retraction [51]

a nanobridge at the initial stage. The voltage can drive the formation through Joule heating which may achieve at least 3700 K at which rearrangements of graphite layer into hexagons, heptagons, and pentagons can be achieved. The tubular parts of the carbon bridges grow in length during the retraction of the tip which may be driven by Joule heating.

Welding of nanotubes to Au and Si surfaces was realized [25]. Thermal heating at a bias voltage in the range of 2 V is sufficient for contact formation with Au and 5 V with Si. Diffusion of Au into the nanotube interlayer is observed after bonding. Si and Au surface melting was observed in contact areas, which means that the temperature reaches at least 1687 K. In [28] welding of two carbon nanotube bundles was observed using bias voltages 2 – 3 V.

High current densities can be used for controlled nanotube cutting as was realized by contact with amorphous carbon. In each contact the cut part of the nanotube achieve 5 – 10 nm at a bias of 5 V [25].

5.6 Electromigration

Material transport in nanotubes [6] and on nanotube surfaces [7,52] has been observed. Electromigration forces, created at high electron current densities, have been shown to enable the transport iron inside carbon nanotubes [6] (Fig. 5.29).

In [7] carbon nanotubes were used for controllable, reversible atomic scale mass transport of indium metal along the nanotubes. Surface driven nanoelectromechanical relaxation oscillator has been proposed (Fig. 5.30) [52].

Recently nanocrystal powered nanomotors have been realized by using a nomanipulator inside a TEM (Fig. 5.31) [53]. Movement is achieved by applying an electric field.

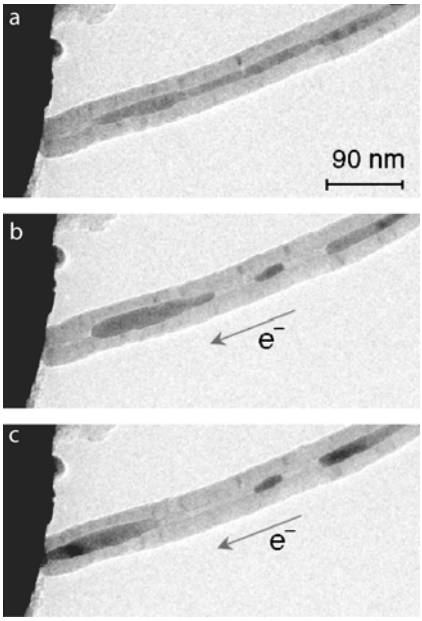


Fig. 5.29. Sequential TEM images showing the induced movement of iron inside carbon nanotubes at time **a** – 0, **b** – 2, **c** – 3 min. Iron migrates in the same direction as the electron flow [6]

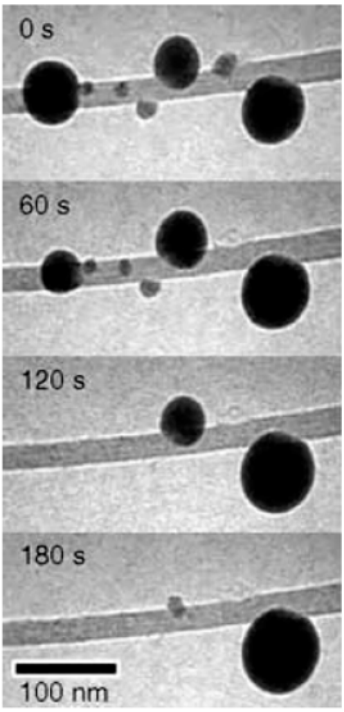


Fig. 5.30. Four TEM images, spaced by one-minute increments, left to right indium transport on a single MWNT [7]

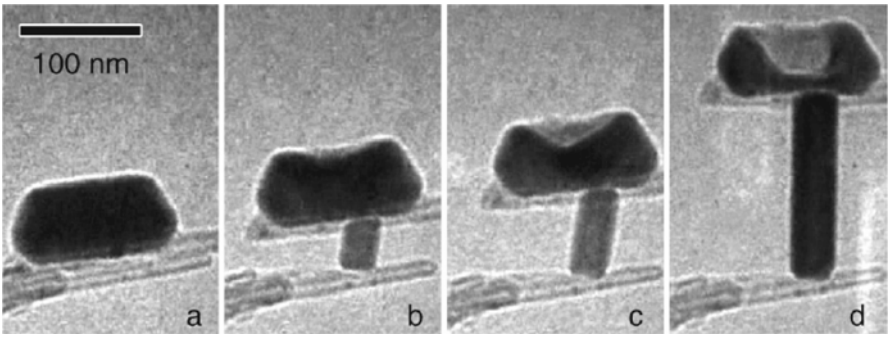


Fig. 5.31. Images showing nanocrystal ram extension. **a** Two multiwall nanotubes (MWNT) lie in contact with one another. A reservoir of indium atoms rests on the top nanotube. **b** Driving 2.1 μA through the circuit creates the nanocrystal ram, which begins to push the MWNT apart. **c** The nanocrystal ram has growth to 75 nm long. **d** At full extension the nanocrystal ram is more than 150 nm long [53]

5.7 Conclusions

The new methods of probing inside transmission electron microscopes have proved to be of use tool for investigation of nanocontacts. However, many of the examples discussed above only demonstrate the potential of in situ probing to address problems in nanotribology. There is one single parameter in friction that is of importance in almost all studies: the contact area, which is directly visible using the TEM-AFM. If future nanotribology work, using TEM-AFM, is contact area alone makes it an important new tool. Almost all AFM tribology studies done, except for the ones in liquid environments, can be repeated using in situ methods and will provide new information, and will more than ones be with unexpected results. For example be interesting to make more experiments along the original slip-stick experiment [18, 54], the force and conductance AFM measurements on atomic-scale metallic contacts [10, 23], or studies of lubricated samples, for example the squeezing out of lubrication fluids between two surfaces [55]. The power of imaging the contacts in nanotribology experiments will, if pursued, give new insight to this interesting field.

References

1. J.C.H. Spence: Ultramicroscopy 25, 165 (1988); J.C.H. Spence, W. Lo, M. Kuwabara: Ultramicroscopy 33, 69 (1990); W. Lo and J.C.H. Spence: Ultramicroscopy 48, 433 (1993)
2. H. Ohnishi, Y. Kondo, K. Takayanagi: Nature 395, 780 (1998)
3. P. Poncharal, Z.L. Wang, D. Ugarte, W.A. de Heer:Science 283, 1513 (1999)
4. J. Cumings and A. Zettl: Science 289, 602 (2000)

5. D. Erts, H. Olin, L. Ryen, E. Olsson, A. Thölén: Phys. Rev. B 61,12725 (2000)
6. K. Svensson, H. Olin, E. Olsson: Phys. Rev. Lett 93, 14590 (2004)
7. B.C. Regan, S. Aloni, R.O. Ritchie, U. Dahmen, A. Zettl. Nature, 428, 924 (2004)
8. D. Erts, A. Lõhmus, R. Lõhmus, H. Olin: Appl. Phys. A 72, 71 (2001)
9. D. Erts, A. Lõhmus, R. Lõhmus, H. Olin, A.V. Pokropivny, L. Ryen, K. Svensson: Appl. Surf. Sci. 188, 460 (2002)
10. T. Kizuka, H. Ohmi, T. Sumi, K. Kumazawa, S. Deguchi, M. Naruse, S. Fujisawa, S. Sasaki, A. Yabe, Y. Enemoto: Jpn. J. Appl. Phys. 40 L170 (2001)
11. S. Fujisawa and T. Kizuka: Tribol. Lett. 15, 163 (2003)
12. A.M. Minor, E.A. Stach, J.W. Morris: Appl. Phys. Lett. 79, 1625 (2001)
13. A. Nafari, P. Enoksson, H. Olin, A. Danilov, and H. Rödjegrd: Sens. Actuators A 123-124, 44 (2005)
14. M.S. Bobji, J.B. Pethica, B.J. Inkson: J. Mat. Res. 20, 2726 (2005)
15. Nanofactory Instruments, www.nanofactory.com
16. Y J. Suh, S.K. Friedlander: J. Appl. Phys. 93, 3515 (2003)
17. M.I. Lutwyche, Y. Wada: Sens. Actuators, A 48, 127 (1995)
18. T. Kizuka, K. Yamada, S. Degushi, M. Naruse, N. Tanaka: Phys. Rev. B 55, R7398 (1997)
19. K. Svensson, Y. Jompol, H. Olin, E. Olsson: Rev. Sci. Instr. 74, 4475 (2003)
20. A. Debarre,A. Richard, Techenio: Rew. Sci. Instrum, 68, 4120 (1997).
21. K.J. Ziegler, D.M. Lyons, J.D. Holmes, D. Erts, B. Polyakov, H. Olin, K. Svensson, E. Olsson: Appl. Phys. Lett., 84, 4074 (2004)
22. J.N. Israelachvili: *Intermolecular and surface forces* (2nd ed. Academic Press, Inc.: San Diego, CA 1992)
23. G. Rubio, N. Agrait, S. Vieira: Phys. Rev. Lett. 76, 2302 (1996)
24. S. Blom, H. Olin, J.L. Costa-Krämer, N. Garcia, M. Jonson, P.A. Serena, R. Shekhter: Phys. Rev. B 57, 8830 (1998); C.A. Stafford, D. Baeriswyl, J. Bürki: Phys. Rev. Lett. 79, 2863 (1997); D. Sánchez-Portal, E. Artacho, J. Junquera, P. Ordejón, A. Garcia, J.M. Soler: Phys. Rev. Lett. 83 3884(1999)
25. T. Kuzumaki, H. Sawada, H. Ichinose, Y. Horiike, T. Kizuka: Apl. Phys. Lett. 9, 4580 (2001)
26. T. Kizuka, K. Hosoki, S. Deguchi, M. Naruse: Mat. Sci. Forum 551, 304-306 (1999)
27. T. Kizuka: Phys. Rev. B 57, 11158 (1998)
28. H. Hirayama, Y. Kawaamoto, Y. Ohshima, K. Takayanagi: Appl. Phys. Lett. 79, 1169 (2001)
29. P. Poncharal, Z.L. Wang, D. Ugarte, W.A. de Heer: Science 283, 1513 (1999)
30. Z.L. Wang, R.P. Gao, P. Poncharal, W.A. de Heer, Z.R. Dai, Z.W. Pan: Mater. Sci. Eng. C 16, 3 (2001)
31. T. Kizuka: Phys. Rev. Lett. 81, 4448 (1998)
32. T. Kizuka, K. Hosoki: Appl. Phys. Lett. 75, 2743 (1999)
33. T. Kizuka: Phys. Rev. B 63, 033309 (2001)
34. R. Lõhmus, D. Erts, A. Lõhmus, A, Svensson K, Jompol, H. Olin: Phys. Low-Dimens. Str. 3-4, 81 (2001)
35. U. Landman, W.D. Luedtke, N.A. Burnham, R.J. Colton: Science 248, 454 (1990)
36. W. Rong, A.E. Pelling, A. Ryan, J.K. Gimzevski, S.K. Friedlander: Nano Lett. 4, 2287 (2004).

37. K.L. Johnson, K. Kendall, and A.D. Roberts: Proc. Roy. Soc. London A 324, 301 (1971)
38. B.V. Derjaguin, V.M. Muller, Yu.P. Toporov: J. Colloid Interface Sci. 53, 314 (1975)
39. D. Tabor, R.H.S. Winterton: Proc. R. Soc. London Ser. A 312, 435 (1975)
40. D. Maugis. J. Colloid Interface Sci. 150, 243 (1992)
41. R.W. Carpick, D.F. Ogletree, M. Salmeron: J. Coll. Interf. Sci. 211, 395 (1999)
42. D. Maugis: *Contact, Adhesion and Rupture of Elastic Solids* (Springer: Berlin 2000)
43. N. Agrait, G. Rubio, S. Vieira: Langmuir 12, 4505 (1996)
44. R. Landauer: IBM J. Rev. Dev. 1, 223 (1957)
45. Y.V. Sharvin: Zh. Exp.Teor. Fiz 48, 984, [Sov. Phys. JETP 21, 655 (1965)
46. G. Wexler. Proc. Phys. Soc. London, 89, 927 (1966)
47. P. Poncharal, C. Berger, Y. Yi, Z.L. Wang, W.A. de Heer: J. Phys.Chem. B 106, 12104 (2002)
48. J. Cumings, A. Zettl: Phys. Rev. Lett. 93, 086801-1 (2004)
49. J. Cummings, A. Zettl: Sol. St. Communications 129, 661-664 (2004)
50. J. Cumings, A. Zettl: Phys. Rev. Lett. 88, 056804-1 (2002)
51. J. Yamashita, H. Hirayama, Y. Ohshima, K. Takayanagi: Appl. Phys. Lett. 74, 2450 (1999)
52. B.C. Regan, S. Aloni, K. Jensen, A. Zettl: Appl. Phys. Lett. 86, 123119 (2005)
53. B.C. Regan, S. Aloni, K. Jensen, R.O. Ritchie, A. Zettl: Nano Lett. 5, 1730 (2005)
54. C.M. Mate, G.M. McClelland, R. Erlandsson, S. Chiang: Phys. Rev. Lett. 59, 1942 (1987)
55. B.N.J. Persson: Surf. Sci.Rep. 33, 85 (1999)

6 Stick-Slip Motion on the Atomic Scale

Tibor Gyalog, Enrico Gnecco, Ernst Meyer

Institute of Physics, University of Basel, Klingelbergstrasse 82, CH-4056 Basel, Switzerland

6.1 Introduction

In the mid-eighties the newly developed friction force microscopy (FFM) opened the feasibility to investigate friction processes in a single asperity contact. FFM delivered interesting results, which could not be explained by simply calculating the energy needed to deform the surface asperities. There are good reasons to assume that during a very slow friction experiment in a very well described setup, where every atom after scanning remains the same place where it was in the beginning, no energy would dissipate. Since all interatomic potentials we know are conservative, and due to the very small velocities no kinetic energy appears anywhere, the change in the position of all the atoms defines the change of energy of the whole system.

This chapter is organized as follows. In Sect. 6.2 we introduce the Prandtl–Tomlinson model, which explains the main features observed in atomic friction measurements. In Sect. 6.3 we describe some significant experiments in details, and discuss the effects of unusually small loads, finite temperature, and atomic-scale abrasion on friction.

6.2 The Tomlinson Model

6.2.1 A Qualitative Description of Tomlinson's Mechanism

Interatomic forces are conservative and do not give rise to the typical dissipative character of friction. A wearless scan with infinitely small velocity therefore should result in non-dissipative reversible process. In 1929 G.A. Tomlinson computed the amount of plastic deformation in a locomotive and concluded that every locomotive has to be completely damaged after a few kilometers of travel if plastic deformation is responsible for the loss of energy [1]. He therefore proposed the existence of irreversible stages in a friction process:

> *"To explain friction it is necessary to suppose the existence of some irreversible stage in the passage of one atom past another, in which heat energy is developped at the expense of external work"*

He presented a basic mechanism to explain irreversible jumps observed during a friction process, nowadays referred to as the *Tomlinson mechanism.*

Instabilities and irreversible jumps

When dragging a particle very slowly by an elastic coupling over a surface, the particle does not always follow the support continuously. For a soft coupling it might become impossible to place the particle on top of certain "hills". Furthermore, it is possible that the particle's velocity becomes high also when the support is dragged with an infinitely small velocity. Adapted to a typical FFM situation, Tomlinson's mechanism reads as follows: Dragging the tip of the FFM through the elastic cantilever at an infinitely slow velocity may result in sudden irreversible jumps of the tip, giving rise to hysteresis and friction.

The particle's jumps are not controllable and are, according to G.A. Tomlinson, the reason for energy dissipation in a wearless friction process. Since the resulting motion of a particle would look quite similar to the motion of a piece of rubber dragged over a table, the atomic process is often referred to as *atomic stick-slip* or just *stick-slip*, although the origin of the two phenomenona is not quite the same.

Relation to phenomenological friction laws

Classically, friction is described through a material-dependent friction coefficient μ, which mainly depends on the material of the two bodies in motion, but also on the environmental conditions. This picture falls down when dealing with single asperity contact. In such a case the dependence of friction on the applied load is usually not linear, and introducing a friction coefficient has not a well-defined meaning.

Nevertheless, in Tomlinson's model a dimensionless parameter η can be introduced, which describes the ratio between elastic and chemical forces acting on the contact asperity. Depending on the values of η, two different friction regimes can be observed, i. e. *stick-slip* or *superlubricated motion.*

6.2.2 Quantitative description

The following analysis describes in detail the frictional force experienced by a single particle sliding on a surface in the framework of the one-dimensional Tomlinson model. This approach has been used by Mc Clelland et al. [2], Zhong and Tomanek et al. [3, 4], Colchero and Marti *et al.* [5] and others to model the FFM.

We consider the tip of a FFM in a one-dimensional periodic potential, which represents the interaction with an atomically flat surface. We will concentrate on the quasistatic limit, of vanishing relative velocity of the two bodies. In the Tomlinson model, a tip with coordinate x is coupled through a spring of stiffness k to the support with coordinate X. Neglecting inertia, the total energy of the system, consisting of the potential ($V(x)$) at the

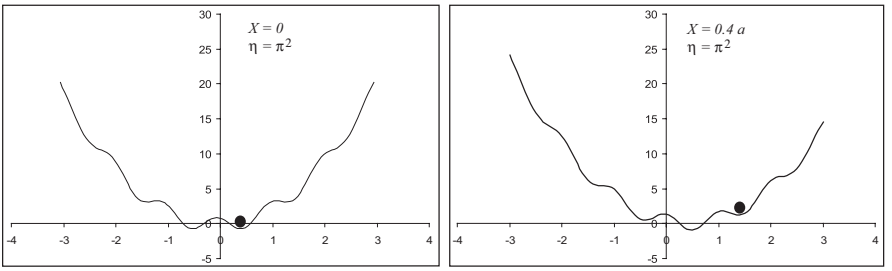


Fig. 6.1. Total potential experienced by the FFM tip for given support positions: **a** Support on top of the potential hill, **b** Support at an arbitrary position

position of the tip and the energy which is stored in the spring, is given by

$$E = V(x) + \frac{1}{2}k(x - X)^2 \,. \tag{6.1}$$

In what follows, we will consider a periodic potential of the form ($V(x) = V_0 \cos(2\pi x/a)$, and introduce a dimensionless parameter

$$\eta = \frac{4\pi^2 V_0}{ka^2} \tag{6.2}$$

The total energy (6.1) is shown in Fig. 6.1 for two different support positions (X) and a parameter value $\eta = \pi^2$.

In a quasistatic motion, the tip position remains in a local energy minimum, since there is always enough time for the system to relax. In order to understand the time evolution of the system, it is therefore sufficient to follow the evolution of the local minima. The latter are given by the solutions of the equation ($\partial E/\partial x = 0$),

$$V'(x) + k(x - X) = 0 \,, \tag{6.3}$$

which also satisfy the stability condition $\partial' 2E/\partial x^2 = V''(x) + k > 0$.

It is obvious that Eq. (6.3) might have more than one solutions, especially for small values of (k). In Fig. 6.2 these solutions are indicated graphically for the potential $V(x) = V_0 \cos(2\pi x/a)$ and for $\eta = \pi^2$.

A change of the support position (X) may lead to a change of sign of the second derivative $\partial^2 E/\partial x^2$, which occurs for our sample potential when

$$\eta \cos\left(\frac{2\pi x}{a}\right) = 1 \tag{6.4}$$

At this point, the lateral force is

$$F^* = \frac{2\pi V_0}{a}\sqrt{\eta^2 - 1} \tag{6.5}$$

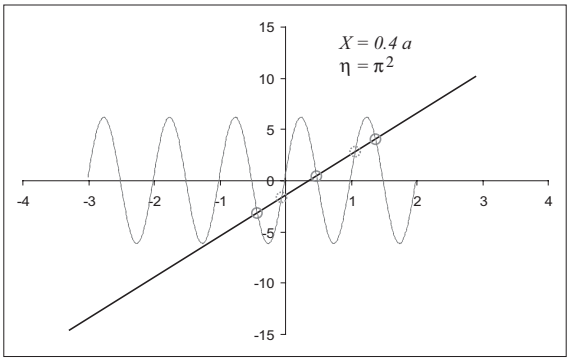


Fig. 6.2. Graphical solution of the equilibrium equation (6.3) for a potential ($V(x) = V_0 \cos(2\pi x/a)$.)

It can be easily seen that Eq. (6.4) has no solution in x for $\eta < 1$. In such case the tip may be placed at any position on the surface. If $\eta > 1$, Eq. (6.4) has two solutions x^* in each lattice cell, describing the borders of the unstable positions, later referred to as the *critical curve*. For $\eta = 1$ the critical positions $x^*_{1,2} = 0$, whereas they diverge towards $x^*_{1,2} = \pm a/4$ when η increases towards infinity. For any value of $\eta > 1$ there exist certain areas where the tip position is unstable and where it is impossible to keep the tip in rest whatever support position is chosen. By making use of the equilibrium condition (6.3) the corresponding support positions X, where irreversible jumps occur, can be computed from the solutions of (6.3).

When sliding starts, the lateral force F_x increases with the support and tip positions, X and x, according to the law

$$F_\mathrm{x} = k_\mathrm{exp}(X - x)\,, \quad k_\mathrm{exp} = \frac{\eta}{\eta + 1} k \tag{6.6}$$

The relation between the "slope" k_exp and k is derived in [6]. Note that, when $\eta \gg 1$, the quantity k_exp is close to the effective spring constant k. The maximum value of the lateral force, $F_\mathrm{x}^\mathrm{max}$ is obtained when $x = a/4$, and it is related to the energy amplitude by the relation

$$F_\mathrm{x}^\mathrm{max} = \frac{2\pi V_0}{a} \tag{6.7}$$

At this point, the lateral force F_x starts to decrease, as the tip apex moves faster than the support. Finally, when x reaches one of the critical values $x^*_{1,2}$ (depending on the moving direction of the support), the tip jumps.

Experimentally, the frictional parameter η can be estimated from the relation

$$\eta = \frac{2\pi F_\mathrm{x}^\mathrm{max}}{k_\mathrm{exp} a} - 1 \tag{6.8}$$

which is easily obtained from Eqs. (6.6) and (6.7).

Other chapters of this book deal with a more detailed description of the FFM-tip taking into account the fact that several atoms are usually in contact

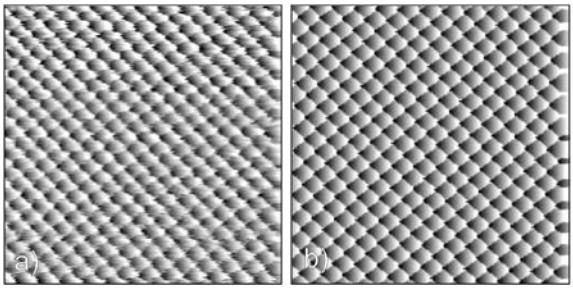


Fig. 6.3. **a** A typical Friction Force Microscope Image on KBr and **b** a computed image within Tomlinson's model. From [10] and [11]

with the sample. Theoretical investigations on the Frenkel–Kontorova and on the Frenkel–Kontorova–Tomlinson model are given in Refs. [7–9].

6.2.3 Two dimensional effects

The Tomlinson model can be extended to the two-dimensional case, where it reproduces the basic features of the experimental friction maps acquired by FFM. Furthermore, higher dimensionality considerations may explain additional interesting phenomena such as the "spike like" character of the atoms that are imaged through an FFM experiment, as shown in Fig. 6.3.

Critical Curve

The set of support positions where irreversible jumps occur are referred to as the *critical curve*. In the one-dimensional case they were computed through the mapping from the solutions x of equation $(\partial^2 E/\partial x^2 = 0)$ into the corresponding support position by making use of the equilibrium condition. In two dimensions, due to the complex character of the mapping from support

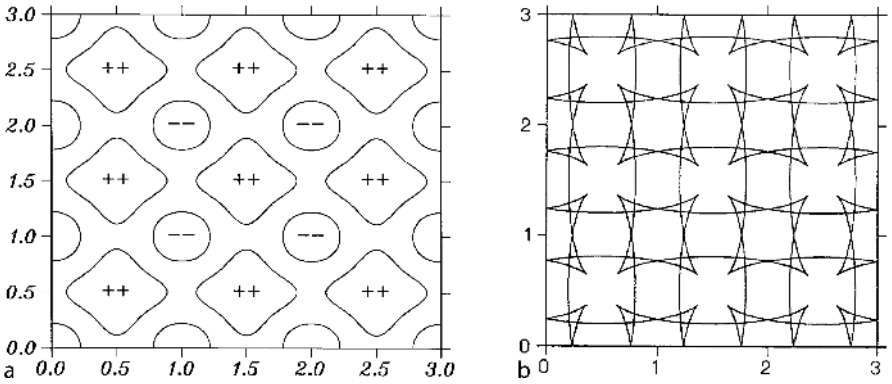


Fig. 6.4. Critical curves for a non-separable two-dimensional potential in **a** the plane of tip positions and in **b** the plane of the corresponding support positions

positions to tip positions, the critical curves look quite complicated. Some examples are shown in Fig. 6.4.

6.3 Friction experiments on the atomic scale

The FFM technique is described in details in Chapt. 1. As the zero value of the lateral force F_x is always affected by a certain offset, the so-called *friction loops* are usually acquired. Figure 6.5 shows the frictional force detected when a FFM tip slides forwards and backwards across an alkali halide surface as a function of the support position X. The average friction is given by half the difference of the two curves. The total energy dissipated in the sliding process corresponds to the area delimited by the friction loop. Dissipation does not occur continuously while scanning, but only when the tip jumps from one equilibrium position to the next one, releasing phonons into the underlying sample.

The first FFM measurements were performed in 1987 by Mate et al. with a tungsten tip sliding on a graphite surface (Fig. 6.6). The average frictional

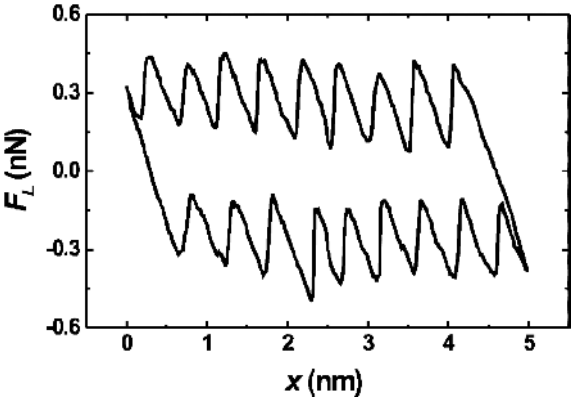


Fig. 6.5. "Friction loop" detected on a NaCl surface in ultra-high vacuum. From [12]

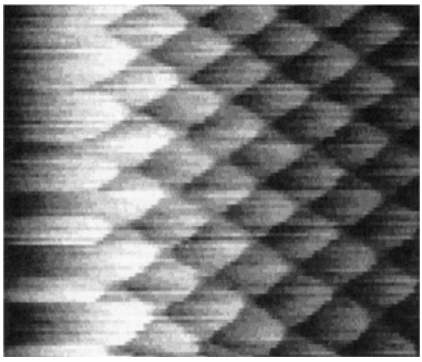


Fig. 6.6. First frictional map on the atomic scale obtained by a tungsten tip sliding on graphite. Frame size: 2 nm. From [13]

force was roughly proportional to the applied load, with a low friction coefficient $\mu = 0.012$. The atomic structure of the surface lattice was observed in the frictional maps with normal forces up to $56\,\mu$N, and a simple interpretation of these results, based on the Tomlinson model, was also proposed. Friction on graphite was later observed by other groups, each of them reporting different features [14–19]. Fujisawa et al., for instance, investigated the load dependence of friction by means of a two-dimensional FFM. Tip jumps were observed in both x and y directions on the surface lattice, due to a zig-zag motion predicted by the Tomlinson model in 2D [15]. Miura et al. compared friction maps recorded with a sharp tip and a graphite flake [16]. In the second case the stacking of graphite layers was maintained while scanning, and anisotropy effects were observed. Friction reached a minimum value when the flake was moved parallel to well-defined pulling directions. If the direction of motion was not parallel to the pulling direction, the flake could not move below a certain thresold. A rotation of the flakes around a pivot point, due to the zig-zag motion, was also recognized. On other layered materials, like mica and MoS_2, atomic stick-slip was also observed [17, 20].

The first FFM measurements in ultra-high vacuum (UHV) were reported by McClelland and coworkers, who investigated friction of a diamond tip, prepared by chemical vapor deposition, against a diamond surface [21]. A few years later, Flipse and coworkers repeated the experiment with a standard silicon tip [22]. Atomic stick-slip could be observed only in presence of hydrogen.

Several studies of atomic friction on ionic crystals were also performed in UHV. KBr and NaCl have been investigated by Meyer and coworkers [11, 12, 23–25], whereas Carpick et al. studied the KF surface [26]. In most cases, the periodicity of the frictional maps on these surfaces corresponds to the distance between equally charged ions. NaF represents an exception, as both positive and negative species could be distinguished [27]. The load dependence of friction is easily evaluated with a 2D-histogram technique, in which the load is increased or decreased stepwise along each scan line when acquiring the frictional force [23]. On the KBr surface UHV friction depends linearly on the applied load F_N, when F_N is below a few nanonewtons. In such a case, a low friction coefficient $\mu < 0.04$ was found [11]. With higher loads the friction coefficient increased, and the corresponding topography images showed the occurrence of wear on the surface.

Recently, Maier et al. investigated the statistical distribution of the slip durations on KBr [25]. A wide variation of values up to several milliseconds was reported, by far longer than expected for a relaxation process on atomic scale. The experimental results were compared with computer simulations, based on a two-spring model of the sliding system, and a certain correlation between the duration of the atomic slip events and the atomic structure of the contact was found.

Atomic stick-slip is not peculiar of insulating surfaces. Howald et al. observed stick-slip on the reconstructed Si(111)7×7 surface using a tip coated

by polytetrafluoroethylene [28]. Due to its lubricant properties, this coating did not react with the dangling bonds of the surface, which made possible the imaging process. Bennewitz et al. observed reproducible stick-slip also on the Cu(111) surface, whereas sliding on Cu(100) resulted in irregular patterns with some atomic features [29, 30]. These results are in a certain agreement with molecular dynamics simulations by Sørensen et al., who predicted that wear occurs easier on Cu(100) than on the closed-packed Cu(111) surface [31]. By using passivated tips, atomic stick-slip was also observed on Pt(111) [32].

6.3.1 Contact stiffness and contact area

The slope of the sticking part of the F_x vs. X curve is related to the effective lateral stiffness of the contact k, according to Eq. (6.6). In FFM experiments, k is given by a combination of three springs in series, each of them corresponding respectively to the cantilever torsion, k_{tors}, the lateral bending of the probing tip, k_{tip}, and the lateral deformation of the contact region k_{con} [33–35]:

$$\frac{1}{k} = \frac{1}{k_{\mathrm{tors}}} + \frac{1}{k_{\mathrm{tip}}} + \frac{1}{k_{\mathrm{con}}}$$

The values of k_{tors} and k_{tip} are usually in the order of 100 N/m, whereas the lateral stiffness of the contact, k_{con}, is two order of magnitude less [24]. Thus, when interpreting FFM experiments, we can reasonably assume that $k \simeq k_{\mathrm{con}}$.

If continuum mechanics would be applicable down to the nanometer scale, the radius of the contact area, a_{con} could be estimated from the contact stiffness k_{con} using the relation [36]

$$a_{\mathrm{con}} = \frac{k_{\mathrm{con}}}{8G^*} \tag{6.9}$$

In Eq. (6.9) the effective shear modulus G^* is related to the shear moduli G_1 and G_2 and the Poisson numbers ν_1 and ν_2 of sample and tip by

$$\frac{1}{G^*} = \frac{2-\nu_1^2}{G_1} + \frac{2-\nu_1^2}{G_1}$$

However, the contact radius estimated from the experimental values of k_{con} is usually well below the lattice constant a). The breaking of continuum models on the nanometer scale has been recently discussed by Luan and Robbins by means of molecular dynamics simulations [37].

6.3.2 Friction at finite temperature

The finite temperature of the sliding systems introduces interesting statistical effects in the stick-slip process. A detailed theoretical analysis of these

effects is given in Chap. 7. Here we summarize the basic ideas and discuss the experiments which proved the occurrence of thermal effects on friction.

The jump of the tip apex from one equilibrium position to the next one on the surface lattice is prevented by a certain energy barrier ΔE (Fig. 6.1b). At zero temperature the tip does not jump until $\Delta E = 0$, i. e. when the condition (6.4) is satisfied. At a finite temperature T, the tip can jump even if $\Delta E \neq 0$. The reason for that is that the tip apex oscillates in the potential well where it is confined with a characteristic frequency f_0. The probability p that the tip does *not* jump changes with time according to the master equation [12]

$$\frac{\mathrm{d}p}{\mathrm{d}t} = -f_0 \exp\left(-\frac{\Delta E}{k_\mathrm{B}T}\right) p(t)$$

Here, the energy barrier ΔE is a function of time t or, equivalently, of the lateral force $F_\mathrm{x}(t)$. Assuming that the energy barrier ΔE decreases linearly with the lateral force F_x (*linear creep approximation*), and noting that $\mathrm{d}F_\mathrm{x}/\mathrm{d}t \simeq kv$, where k is the effective stiffness of the system and v is the sliding velocity, a logarithmic dependence of friction on velocity is obtained:

$$F_\mathrm{x}(v) = \text{const.} - \frac{k_\mathrm{B}T}{\lambda} \log \frac{v}{v_0} \tag{6.10}$$

(λ is in the order of the lattice constant a, and the velocity v_0 is arbitrarily chosen in the range of applicability of the linear creep approximation). Thus, temperature effects can be experimentally found measuring the velocity dependence of friction. Experimental data in agreement with Eq. (6.10) were reported by Gnecco et al. on a NaCl crystal in UHV in a velocity range between 5 nm/s and 3 µm/s. However, Eq. (6.10) cannot be applied at high velocities. The reason for that is that the linear creep approximation is not valid when the jumps occur close to the critical position x^*, which is the case if the sliding speed is high enough. From a formal analogy with magnetic-flux fluctuations in superconducting quantum interference devices, Sang et al. suggested that the *ramped creep approximation* $\Delta E \sim (\text{const.} - F_\mathrm{L})^{3/2}$ has to be used in such case [41]. The master equation leads then to an implicit relation between the lateral force F_x and the sliding velocity v, as discussed in [42]. Two different regimes are observed depending on the *critical velocity*

$$v_\mathrm{c} = \frac{\pi\sqrt{2}}{2} \frac{f_0 k_\mathrm{B} T}{ka}$$

If $v \ll v_\mathrm{c}$ the logarithmic relation

$$F_\mathrm{x}(v) = \text{const.} - \left(\frac{k_\mathrm{B}T}{\mu}\right)^{2/3} \left(\ln \frac{v_\mathrm{c}}{v}\right)^{2/3} \tag{6.11}$$

holds (which, experimentally, cannot be easily distinguished from (6.10). The constant in Eq. (6.11) corresponds to the critical value of the lateral force F^*.

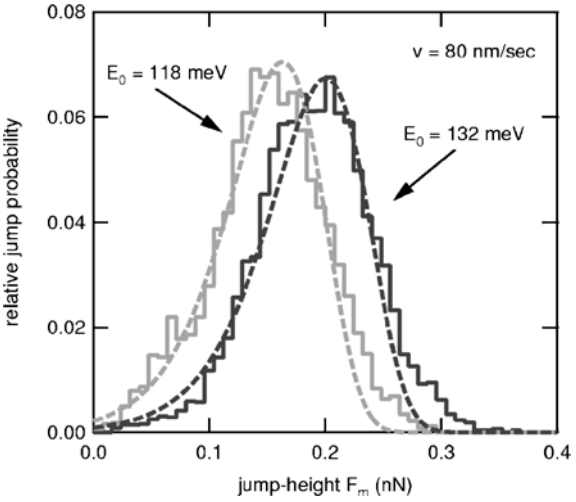


Fig. 6.7. Experimental and theoretical jump-curve distribution functions of atomic stick-slip for a scan speed $v = 80\,\mathrm{nm/s}$ (*solid and dashed lines* respectively). From [43]

In the opposite case $v \gg v_\mathrm{c}$, the lateral force tends to F^* according to the law

$$F_\mathrm{x}(v) = F^* \left(1 - \frac{v_\mathrm{c}}{v}\right)^2$$

Using these relations, Riedo et al. estimated characteristic frequencies $f_0 \sim 50\,\mathrm{kHZ}$ and tip-surface interaction energies of a few eV for a Si tip sliding on mica (in air) [42].

Sang et al. found also that the statistical distribution of the jump heights, due to the finite temperature, has the following shape [41]:

$$P(f^*) = \frac{3}{2}\frac{\sqrt{f^*}}{v^*} \exp\left(-f^{*3/2} - \frac{\mathrm{e}^{-f^{3/2}}}{v^*}\right) \tag{6.12}$$

In Eq. (6.12) the dimensionless variables v^* and f^* are directly related to the sliding velocity v and to the critical force F^*. The statistical distribution (6.12) resembles a Gaussian distribution, slightly distorted towards high values of the lateral force. The expression (6.12) is used in Fig. 6.7 to fit two experimental distributions acquired by Schirmeisen et al. on highly-oriented graphite in UHV [43].

6.3.3 Superlubricity

If the frictional parameter η is less than 1, friction tends to vanish. This transition was observed by Socoliuc et al. on a NaCl surface in UHV. In Fig. 6.8 frictional loops detected on this surface are compared with theoretical curves obtained with the Tomlinson model. The area enclosed by the loops is reduced when the normal load F_N, or, equivalently, the parameter η decreases, until the backward and forward scan lines overlap at the critical threshold

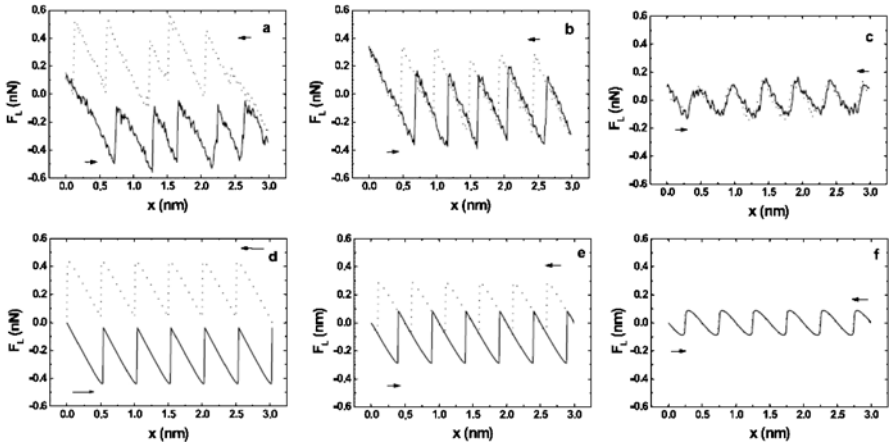


Fig. 6.8. a–c Experimental and **e–f** theoretical friction loops observed when the frictional parameter η is reduced down to $\eta = 1$. From [24]

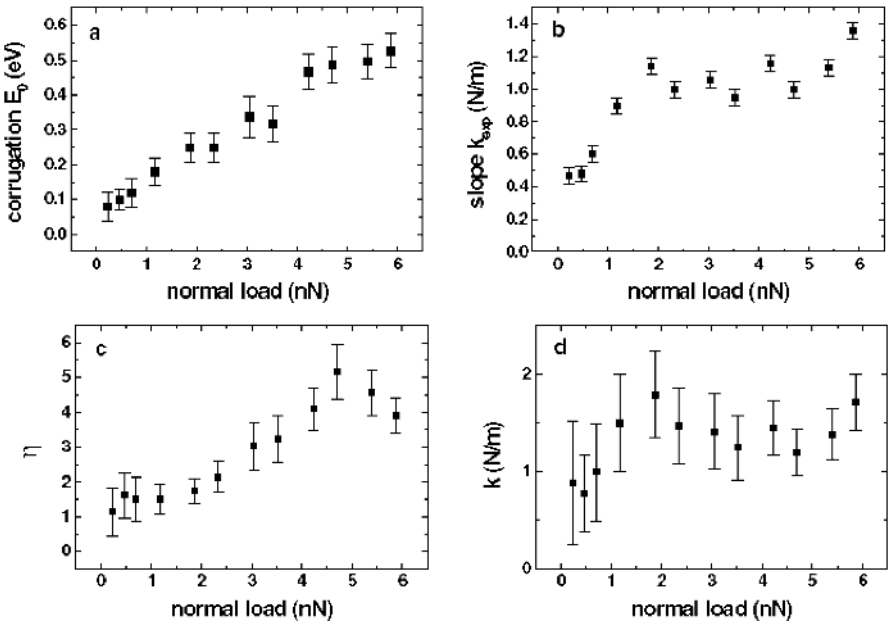


Fig. 6.9. Corrugation energy $E_0 = 2V_0$, experimental stiffness $k_{\rm exp}$, frictional parameter η and effective stiffness k as a function on the normal force $F_{\rm N}$. From [24]

$\eta = 1$. Figure 6.9 shows how the corrugation energy $E_0 = 2V_0$, the slope $k_{\rm exp}$, the frictional parameter η, and the effective spring constant k changes with the normal force $F_{\rm N}$. The parameters E_0, η and k were derived from the experimental data using the relations (6.6, 6.7, 6.8). The corrugation

energy E_0 between silicon and NaCl was found to increase linearly with F_N. A similar conclusion was also reported by Riedo et al. in their experiments on mica. A reasonable explanation is the following [38]. The quantity E_0 is the difference between the maximum and minimum values assumed by the tip-surface interaction energy on a unit cell of a surface. These values, apart from the sign, are given by the integral of the normal force *vs.* distance curves taken along the normal direction z. In the elastic regime, these curves are straight lines with constant slope k_z, which leads to the linear dependence experimentally observed.

The concept of vanishing friction, or *superlubricity*, goes well beyond the experiment discussed here. Dienwiebel et al. observed vanishing friction while dragging a graphite flake out of registry over a graphite surface [39]. This is related to the lateral stiffness and to the incommensurability of the two surfaces. Dry friction decreases also when the sliding speed is reduced down to a few nm/s or less. This effect is due to thermally activated jumps occurring in the contact area, and it is called *thermolubricity* [40]. A detailed discussion of superlubricity and related issues is given in Chapters 8.4–10.4.

6.3.4 Wear on the atomic scale

If the normal force applied on the FFM tip exceeds a critical value, dependent on the elastic properties of the materials in contact, wear occurs. E. Gnecco et al. investigated the initial stage of damage on alkali halide surfaces [44]. Figure 6.10 shows a rupture event in the stick phase, in which the FFM tip picked up some couples of K^+ and Br^- ions. Starting from this moment, a continuous exchange of ions between tip and surface is established. The stick-slip mechanism, even if complicated by the exchange of "debris", was observed during the whole wear experiment. On long time scales, the mean friction force extracted from the friction loops increases asymptotically with the number of scans towards an equilibrium value, where the applied load is balanced by the normal reaction of the sample without further damage.

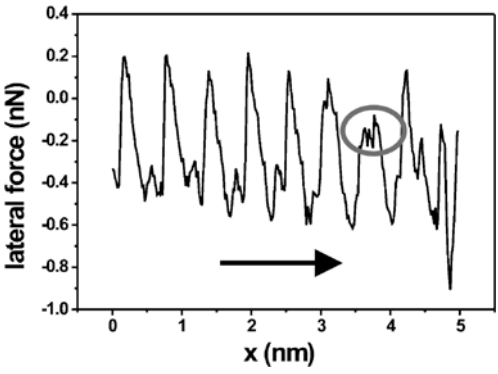


Fig. 6.10. Wear onset on the atomic scale while scanning a KBr surface

A comparison between friction loops and topography images acquired before and after wearing off the surface gives important information, as the energy dissipated in the wear process. On KBr only a minor part (30%) of the total energy dissipation went into wear.

When micrometer size areas were scanned, the formation of quasiperiodic patterns of mounds and pits was observed on alkali halide surfaces, as well as on metals [45]. These structures result from a delicate interplay of friction-induced strain and erosion, material transport operated by the tip, and diffusion, which presents some similarities with features observed in beam cutting [46], ion-beam sputtering [47] or even wind-blown sand rearrangement [48].

Wear on layered material was studied by the group of Salmeron. Assuming that wear is initiated by atomically defects, and it occurs only when these accumulate beyond a critical concentration, the following relationship was derived [49]:

$$F_{\mathrm{L,wear}} = AF_{\mathrm{N}}^{2/3} \exp(BF_{\mathrm{N}}^{2/3})$$

where A and B are constant. Other issues related to tribochemical wear are discussed in Chapt. 21.6.

6.4 Conclusions

In conclusion, the mechanism of atomic stick-slip revealed in FFM experiments in well described by the Tomlinson-Prandtl model. In the quasi-static limit of scan velocities $v < 10\,\mu\mathrm{m/s}$ the nanocontact formed by the microscope tip and a crystal surface is elastically deformed, suddenly broken and completely reestablished on a different site. The frictional force can be analytically related to the interaction between the two surfaces and to the effective stiffness of the contact. The ratio of these two quantities, η, is an important parameter, which allows to distinguish between a stick-slip dissipative motion and a superlubricated regime, where dissipation falls down to negligible values. When the elasticity limit of the surface is overcome, wear occurs. The initial stages of wear can also be investigated by FFM down to the atomic scale. Finally, we have also discussed the important role of temperature in atomic friction. Tip jumps are favored by thermal activation, and a logarithmic dependence of friction on velocity is consequently built up.

References

1. G.A. Tomlinson, Philosoph. Mag. Ser. 7 (1929) 905
2. G.M. McClelland and J.N. Glosli, "Friction at the atomic scale", in *Fundamentals of Friction: Macroscopic and Microscopic Processes* edited by I.L. Singer and H.M. Pollock, p. 405-425, NATO ASI Series E: Applied Sciences, Vol. 220, Kluwer Academic Publishers (1992).

3. W. Zhong and D. Tomanek, Phys. Rev. Lett. 64 (1990) 3054
4. D. Tomanek, W. Zhong and H. Thomas, Europhys. Lett. 15 (1991) 887 D. Tomanek, p. 269 in *Scanning Tunneling Microscopy III*, Eds. R. Wiesendanger and H.-J. Güntherodt, Springer Berlin (1993)
5. J. Colchero, O. Marti and J. Mlynek, p. 345 in *Forces in Scanning Probe Methods*, Eds. H.-J. Güntherodt, D. Anselmetti and E. Meyer, NATO ASI Series E, Vol. 286, Kluwer Academic Publishers (1995)
6. E. Gnecco, R. Bennewitz, T. Gyalog, and E. Meyer, J. Phys.: Condens. Matt. 13 (2001) R619
7. T. Strunz, Sliding dynamics of the Frenkel–Kontorova model, *Z. Naturforsch.* 50a (1995) 1108 and T. Strunz and F.J. Elmer, On the sliding dynamics of the Frenkel–Kontorova model, in *Physics of Sliding Friction*, Eds. B.N.J. Persson and E. Tosatti, Kluwer Academic Publishers (1996)
8. T. Gyalog and H. Thomas, Europhys. Lett.37 (1997) 195
9. M. Weiss and F.J. Elmer, Phys. Rev. B 53 (1996) 7539
10. T. Gyalog, M. Bammerlin, R. Lüthi, E. Meyer and H. Thomas, Europhys. Lett.31 (1995) 5
11. R. Lüthi, E. Meyer, M. Bammerlin, L. Howald, H. Haefke, T, Lehmann, C. Loppacher, H.J. Güntherodt, T. Gyalog, and H. Thomas, J. Vac. Sci. Technol. B 14 (1996) 1280
12. E. Gnecco, R. Bennewitz, T. Gyalog, Ch. Loppacher, M. Bammerlin, E. Meyer, and H.J. Güntherodt, Phys. Rev. Lett. 84 (2000) 1172
13. C.M. Mate, G.M. McClelland, R. Erlandsson, and S. Chiang, Phys. Rev. Lett. 59 (1987) 1942
14. J. Ruan and B. Bhushan, J. Appl. Phys. 76 (1994) 5022
15. S. Fujisawa, K. Yokoyama, Y. Sugawara, and S. Morita, Phys. Rev. B 58 (1998) 4909
16. K. Miura, N. Sasaki, and S. Kamiya, Phys. Rev. Lett. 69 (2004) 075420
17. S. Morita, S. Fujisawa, and Y. Sugawara, Surf. Sci. Rep. 23 (1996) 1
18. N. Sasaki, K. Kobayashi, and M. Tsukada, Phys. Rev. B 54 (1996) 2138
19. H. Hölscher, U.D. Schwarz, O. Zworner, and R. Wiesendanger, Phys. Rev. B 57 (1998) 2477
20. S. Fujisawa, Y. Sugawara, S. Ito, S. Mishima, T. Okada, and S. Morita, Nanotechnology 4 (1993) 138
21. G.J. Germann, S.R. Cohen, G. Neubauer, G.M. McClelland, and H. Seki, J. Appl. Phys. 73 (1993) 163
22. R.J.A. van der Oetelaar and C.F.J. Flipse, Surf. Sci. 384 (1997) L828
23. R. Lüthi, E. Meyer, H. Haefke, L. Howald, W. Gutmannsbauer, M. Guggisberg, M. Bammerlin, and H.J. Güntherodt, Surf. Sci. 338 (1995) 247
24. A. Socoliuc, R. Bennewitz, E. Gnecco, and E. Meyer, Phys. Rev. Lett. 92 (2004) 134301
25. S. Maier, Yi Sang, T. Filleter, M. Grant, R. Bennewitz, E. Gnecco, and E. Meyer, Phys. Rev. B 72 (2005) 245418
26. R. Carpick, Q. Dai, D. Ogletree, and M. Salmeron, Trib. Lett. 5 (1998) 91
27. M. Ishikawa, S. Okita, N. Minami, and K. Miura, Surf. Sci. 445 (2000) 488
28. L. Howald, R. Lüthi, E. Meyer, and H.J. Güntherodt, Phys. Rev. B 51 (1995) 5484
29. R. Bennewitz, T. Gyalog, M. Guggisberg, M. Bammerlin, E. Meyer, and H.J. Güntherodt, Phys. Rev. B 60 (1999) R11301

30. R. Bennewitz, E. Gnecco, T. Gyalog, and E. Meyer, Tribol. Lett. 10 (2001) 51
31. M.R. Sørensen, K.W. Jacobsen, and P. Stoltze, Phys. Rev. B 53 (1996) 2101
32. M. Enachescu, R. Carpick, D.F. Ogletree, and M. Salmeron, J. Appl. Phys. 95 (2004) 7694
33. M.A. Lantz, S.J. O'Shea, A.C.F. Hoole, and M.E. Welland, Appl. Phys. Lett. 70 (1997) 970
34. M.A. Lantz, S.J. O'Shea, M.E. Welland, and K.L. Johnson, Phys. Rev. B 55 (1997) 10776
35. R.W. Carpick, D.F. Ogletree, and M. Salmeron, Appl. Phys. Lett. 70 (1997) 1548
36. K.L. Johnson, *Contact Mechanics*, Cambridge University Press, Cambridge, U.K. (1985)
37. B. Luan and M.O. Robbins, Nature 435 (2005) 929
38. A. Baratoff, private communication (2005)
39. M. Dienwiebel, G.S. Verhoeven, N. Pradeep, J.W.M. Frenken, J.A. Heimberg, H.W. Zandbergen, Phys. Rev. Lett. 92 (2004) 126101
40. S. Yu Krylov, K.B. Jinesh, H. Valk, M. Dienwiebel, and J.W.M. Frenken, Phys. Rev. E 71 (2005) R65101
41. Yi Sang, M. Dube, and M. Grant, Phys. Rev. Lett. 87 (2001) 174301
42. E. Riedo, E. Gnecco, R. Bennewitz, E. Meyer, and H. Brune, Phys. Rev. Lett. 91 (2003) 085402
43. A. Schirmeisen, L. Jansen, and H. Fuchs, Phys. Rev. B 71 (2005) 245403
44. E. Gnecco, R. Bennewitz, and E. Meyer, Phys. Rev. Lett. 88 (2002) 215501
45. A. Socoliuc, E. Gnecco, R. Bennewitz, and E. Meyer, Phys. Rev. B 68 (2003) 115416
46. R. Friedrich, G. Radons, T. Ditzinger, and A. Henning, Phys. Rev. Lett. 85 (2000) 4884
47. U. Valbusa, C. Boragno, and F.B. de Mongeot, J. Phys.: Condens. Matter 14 (2002) 8153
48. H. Nishimori and N. Ouchi, Phys. Rev. Lett. 71 (1993) 197
49. S. Kopta and M. Salmeron, J. Chem. Phys. 113 (2000) 8249

7 Velocity dependence of atomic friction: Rate theory and beyond

Mykhaylo Evstigneev and Peter Reimann

Universität Bielefeld, Fakultät für Physik, 33615 Bielefeld, Germany
mykhaylo@physik.uni-bielefeld.de

7.1 Introduction

Macroscopic friction between solids is well known to be both of paramount practical importance and of notorious difficulty regarding its theoretical understanding [1, 2]. Here, we restrict ourselves to the simpler case of a microscopic contact in the form of a single asperity. Such studies of frictional forces between nanoscale objects are vital both for engineering of micromechanical devices and advancement of our understanding of the laws of nature acting in the nanoworld. While macroscopic friction involves interactions between numerous asperities of the two contacting surfaces, employing an atomic force microscope (AFM) offers a unique opportunity to probe the frictional forces between a single asperity – the tip of an AFM cantilever – and an atomically flat surface. Therefore the research direction of friction force microscopy (FFM) [3] had been initiated only a year after the invention of the AFM in 1986 [4] and became a subject of intensive studies since then (see the reviews [5–7] and references therein).

The laws of nanofriction differ drastically from those of macroscopic friction. In particular, it has been known from the time of Coulomb that the force of friction between two macroscopic bodies in contact is independent of their relative velocity. In contrast, friction force on the nanoscale exhibits a non-trivial velocity dependence, which will be the subject of the present contribution.

Though simpler than macroscopic friction, the adequate interpretation and modeling of microscopic friction experiments still represents a formidable challenge. In particular, direct molecular dynamics simulations are still very far from reaching experimentally realistic conditions [8, 9]. The reason is the enormous time scale separation between molecular vibrations and the very rare slip-events of the AFM tip which still cannot be bridged by today's computer facilities. Hence, non-trivial theoretical modeling steps are indispensable, in particular the concepts of non-linear stochastic processes [10–14]; the above-mentioned time-scale separation will greatly facilitate the calculations within such a model.

The behavior of an atomic force microscope tip in contact with a uniformly moving atomically clean surface is modeled as one-dimensional Brownian mo-

tion in a potential of the tip-surface interaction and of the elastic forces resulting from the deformation of the cantilever, the tip, and the surface in the contact region. A theoretical description of friction force microscopy experiments within such a model is derived on the basis of microscopic considerations. At the focus of this review is the relation between the pulling velocity and the time-averaged lateral force developed in the cantilever, which equals in magnitude the force of friction.

An exact analytical force-velocity relation can be found for asymptotically small cantilever stiffness and high damping. For an arbitrary stiffness, one needs to resort to an approximate treatment. A particularly successful approximation is possible in the stick-slip regime of the tip motion, when the elastic force exhibits a random sawtooth-like time-dependence resulting from the thermally activated transitions of the tip from one surface site to the next. This regime can be treated within the framework of Kramers' rate theory of thermally activated transitions. The range of validity of such a rate approach in the context of friction force microscopy is discussed. An approximate analytic formula relating the pulling velocity and the average elastic force is derived and its high accuracy is demonstrated numerically. Within the stick-slip regime, the average lateral force increases approximately logarithmically with velocity. While the rate description is applicable when the pulling velocity is not too high, going beyond the stick-slip regime results in a maximum of the average force as a function of velocity, followed by a subsequent decrease. This theoretically predicted non-monotonic force-velocity relation should be observable under realistic experimental conditions.

7.2 Experimental Set-Up

In a typical FFM experiment [3], the tip of an AFM is brought in contact with an atomically clean surface moving at a constant velocity v by means of a normal load F_{N} (see Fig. 7.1a). The interaction between the tip and the surface leads to a torsional deformation of the cantilever. One can determine the magnitude of this deformation by optical means and thus deduce the resulting elastic force $f(t)$, which, by Newton's third law, equals the instantaneous force of friction. As a rule, the temporal evolution of the friction force proceeds in a sawtooth-like pattern (see Fig. 7.1b showing the results of our numerical simulations; the experimentally observed force evolution is similar, see e. g. [15]). This type of motion of the AFM cantilever is called *stick-slip motion*. The central quantity of interest is the behavior of the time-averaged friction force

$$\bar{f} := \lim_{t\to\infty} \frac{1}{t} \int_0^t \mathrm{d}t'\, f(t') \tag{7.1}$$

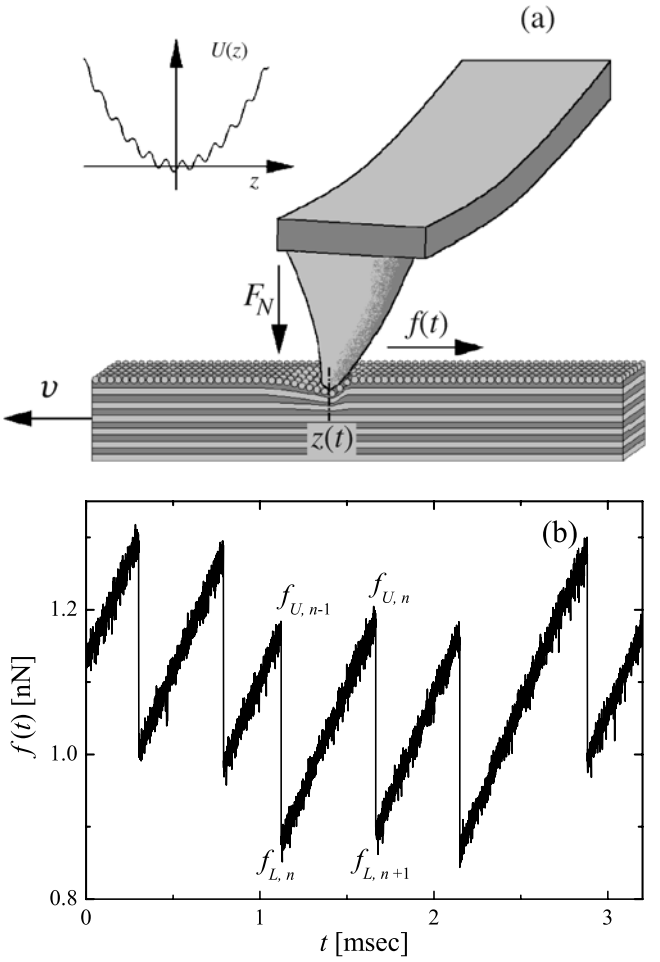


Fig. 7.1. a Schematic illustration of an FFM experiment. The inset depicts the instantaneous potential (7.2), in which the AFM tip finds itself at a given instant of time t. **b** A typical example of the temporal evolution of the elastic force (7.6) in the stick-slip regime obtained from simulations of the Langevin equation (7.11) in the overdamped ($m \to 0$) limit with pulling velocity $v = 1\,\mathrm{nm/msec}$, viscosity $\eta = 1\,\mathrm{pN\,msec/nm}$ [26, 28], and $\eta_\mathrm{s} = 0$. The thermal energy kT is $4.04\,\mathrm{pN\,nm}$ (room temperature), the spring constant κ (cf. (7.2), (7.5)) is $0.5\,\mathrm{N/m}$ [42], and the lattice potential is given by Eq. (7.58) with parameters $\Delta U_0 = 250\,\mathrm{pN\,nm}$, and $a = 0.52\,\mathrm{nm}$

as a function of the pulling velocity, v. From this dependence of $\bar{f}$ on v one tries to gain insight into the specific molecular properties of the probed surface and into the general microscopic principles of surface friction within an exceptionally simple "minimal" model. The interpretation of the measurements along these lines is the subject of our present contribution.

7.3 Modeling

The system from Fig. 7.1 gives rise to a paradigmatic applications of the general principles of stochastic modeling. Starting from microscopic considerations, basically one relevant (slow) state variable (collective degree of freedom) can be identified, while the effect of all the remaining (fast) degrees of freedom essentially boils down to friction, thermal noise, and a renormalization of the relevant potential and inertia within an effective model dynamics for the slow variable alone. While the general mathematical framework for eliminating the fast "thermal bath" variables along these lines is described in detail e. g. in [10–14], here we mainly restrict ourselves to the basic physical picture behind those calculations. We remark that we are using here the notion "thermal bath" instead of e. g. "electron-phonon subsystem" [16] since in general the entire microscopic dynamics is not exhausted by electrons and phonon modes.

7.3.1 Langevin Equation

The system studied – though small – still involves a huge number of microscopic degrees of freedom, which we denote collectively as $\phi(t)$. The experimentally observable lateral force $f(t)$ can be deduced from the torsional deformation of the cantilever and is directly related to the displacement $z(t)$ of the AFM tip (more precisely: the center of mass of the tip apex) from its equilibrium position at a moment of time t, cf. Eq. (7.6) below. To obtain the evolution equation for this relevant collective degree of freedom $z(t)$, one may proceed as follows. First, one writes down the equation of motion for all coordinates $\phi(t)$ of the system, and then projects the system's microscopic state onto the subspace characterized by a given value of z by averaging out the microscopic degrees of freedom [10]. As a result of this procedure, an equation of motion for $z(t)$ is obtained, in which the effect of the molecular degrees of freedom is accounted for by introduction of the following objects: (i) a free-energy type potential $U(z,t)$ of mean force, (ii) memory-dependent friction, and (iii) a random force (noise) of finite correlation time. In view of the fact that the characteristic frequency associated with the collective variable $z(t)$ is of the order of $10^5\,\mathrm{s}^{-1}$, i. e. many orders of magnitude lower than the Debye frequencies describing the time-scale of the molecular degrees of freedom, both the memory effects in friction and the finite noise correlation time can be neglected. In other words, the relevant coordinate $z(t)$ can be regarded as slow compared to all the other molecular degrees of freedom of the system. The possibility of the presence of other slow collective coordinates (describing, e. g., the elastic deformation of the substrate) is discussed in [17].

The potential of mean force $U(z,t)$ consists of three contributions. The first one accounts for the elastic deformations of AFM and substrate, the second for the tip-substrate interaction, and the third for the entropy of the

microscopic degrees of freedom. Since the elastic deformations are typically small [18], we may neglect anharmonic effects in the elastic energy. Furthermore, we can assume that interaction and entropy only depend on the relative tip-substrate position $z + vt$. We thus arrive at the approximation

$$U(z,t) = \frac{\kappa z^2}{2} + U_0(z + vt) \,. \tag{7.2}$$

The argument in the second term indicates that the surface moves at a constant velocity v to the left with sign convention $v > 0$ (cf. Fig. 7.1). Furthermore, focusing on an ideally flat atomic surface with lattice constant a in z-direction, we conclude that $U_0(z)$ is invariant under a displacement of the substrate by one period:

$$U_0(z + a) = U_0(z) \,. \tag{7.3}$$

Without loss of generality, we can assume that the tip-surface interaction potential $U_0(z)$ has minima located at integer multiples of the lattice constant, na. The position of the nth minimum of the moving surface potential $U_0(z + vt)$ is, then,

$$z_n^{(0)}(t) = na - vt \,. \tag{7.4}$$

The spring constant κ describes the combined effect of the elastic deformation of the cantilever, the tip and the elastically deformed surface in the contact region [18–23]:

$$\frac{1}{\kappa} = \frac{1}{\kappa_{\text{cantilever}}} + \frac{1}{\kappa_{\text{tip}}} + \frac{1}{\kappa_{\text{surface}}} \,. \tag{7.5}$$

The experimentally observable lateral force $f(t)$ can be identified, according to Newton's third law, with the negative of the force caused by the elastic deformations, i. e.

$$f(t) = -\kappa\, z(t) \,. \tag{7.6}$$

With this in mind, we consider the elastic deformations of the cantilever, and, in particular, those of the tip apex (see Fig. 7.1). If these deformations, or equivalently, the state variable z, are changing adiabatically slowly, then the system is at every instance of time in a thermal equilibrium state, i. e., we are dealing with a reversible process. Next we consider the case when these changes are taking place at a finite speed, but still slowly enough that the thermal bath due to the fast molecular degrees of freedom of cantilever and tip always remains close to the instantaneous accompanying thermal equilibrium. The remaining "small amount of disequilibrium" renders the process under consideration "slightly irreversible" and hence gives rise to a linear-response type dissipative force

$$F_{\text{c}}(t) = -\eta_{\text{c}} \dot{z}(t) \tag{7.7}$$

with an effective coupling strength $\eta_{\text{c}} > 0$ between the collective coordinate z and the close to equilibrium "cantilever and tip bath" (subscript "c"). In fact,

this is nothing else than Onsager's theory for close to equilibrium processes, associating currents, $\dot{z}(t)$, with corresponding dissipative forces, $F_{\rm c}(t)$, via linear response/Onsager coefficients, $\eta_{\rm c}$. In particular, since the cantilever and tip deformations are typically small [18], the implicitly assumed independence of the Onsager coefficient $\eta_{\rm c}$ on the state z of the system is well justified.

In a similar manner, the influence of the microscopic degrees of freedom of the substrate will result in a dissipative force $F_{\rm s}(t)$, which is proportional to the *relative* velocity of the tip with respect to the substrate with the proportionality coefficient $\eta_{\rm s}$:

$$F_{\rm s}(t) = -\eta_{\rm s}(\dot{z}(t) - v)\,. \tag{7.8}$$

Finally, we come to the randomly fluctuating forces acting on the slow state variable z. They have the same origin as the dissipative forces, namely, the large number of fast degrees of freedom of the cantilever, tip and substrate baths. Due to this common origin and the fact that the baths always remain close to thermal equilibrium, one can show that those randomly fluctuating forces are completely fixed (in the statistical sense) by the functional form of the dissipative forces via the fluctuation-dissipation theorem [24,25]. Namely, the thermal "cantilever-and-tip-noise" acts on $z(t)$ in the usual form [10–14] of a fluctuating force $\sqrt{2\eta_{\rm c}kT}\xi_{\rm c}(t)$ with temperature T, Boltzmann constant k, and unbiased δ-correlated Gaussian noise $\xi_{\rm c}(t)$. Similarly, the substrate gives rise to thermal fluctuations of the form $\sqrt{2\eta_{\rm s}kT}\xi_{\rm s}(t)$ with an unbiased δ-correlated Gaussian noise $\xi_{\rm s}(t)$ independent of $\xi_{\rm c}(t)$. Essentially, the uniqueness of these thermal noises follows from the fact that any deviation from the above specified statistical properties could be exploited to construct a perpetuum mobile of the second kind [25]. Their independence is an approximation which is well justified by the fact that the contact between the two baths consists of comparatively few atoms.

Collecting all acting forces, we arrive at the following equation of motion [17,26]:

$$\begin{aligned} m\ddot{z}(t) = &- U_0'(z(t) + vt) - \kappa z(t) - \eta_{\rm c}\dot{z}(t) \\ &+ \sqrt{2\eta_{\rm c}kT}\xi_{\rm c}(t) - \eta_{\rm s}(\dot{z}(t) - v) + \sqrt{2\eta_{\rm s}kT}\,\xi_{\rm s}(t)\,, \end{aligned} \tag{7.9}$$

where m is the relevant effective mass associated with inertia effects of cantilever, tip, and substrate. To better understand its meaning, we consider an arbitrary point within any of these three objects. Then a small displacement Δz of the tip apex yields a displacement of our reference point proportional to Δz with some proportionality factor q. Hence, inertia forces against an acceleration $\ddot{z}(t)$ acquire a local weighting factor q, and m follows by integrating over the local (mass-) density times q.

We further simplify the Langevin equation by introducing the total friction coefficient

$$\eta := \eta_{\rm s} + \eta_{\rm c}\,, \tag{7.10}$$

allowing us to rewrite it as [17, 26]

$$m\,\ddot{z}(t) + \eta\dot{z}(t) = -U_0'(z(t) + vt) - \kappa z(t) - \eta_s v + \sqrt{2\eta kT}\xi(t)\,. \qquad (7.11)$$

Such an equation (or its noise-free version) has been considered in a number of publications [27–32], the important difference being that it has been assumed there that $\eta = \eta_s$, and thus $\eta_c = 0$, while, in general, there is no reason to expect that any of the coefficients of viscous drag, η_s or η_c, is zero.

7.3.2 Noise-Free Prandtl–Tomlinson Model

To understand the effect of noise on the dynamics of the tip apex $z(t)$, it is instructive to consider the equation of motion (7.11) without the last term, i. e. at $T = 0$. Therefore, for now, we will focus only on the first two effects mentioned in the beginning of the previous section and neglect the thermal fluctuations. The zero-noise limiting case is essentially equivalent to the early model of friction due to Prandtl [33] and Tomlinson [34]. The qualitative picture of the stick-slip process within this model is as follows. In the stick phase, the cantilever is trapped within one of the minima, $z_n(t)$, of the total potential (7.2), and therefore moves together with the surface. This results in the growth of the elastic force due to the deformation of the cantilever spring, the tip, and the surface in the contact region. At some moment, this elastic force becomes sufficient to render the cantilever jump from the current lattice site into the next lattice site. During this slip phase, the redundant elastic energy is quickly dissipated into the thermal baths (electron-phonon subsystem [16]). After the cantilever settles into the new minimum, the process repeats itself.

To simulate this process numerically, a Newtonian equation of motion of the cantilever tip has been investigated numerically and analytically in the works [30–32], which had the form of Eq. (7.11), but without the noise term and with $\eta_s = \eta$, $\eta_c = 0$. The two-dimensional generalization of this model allowing for the lateral motion of the cantilever in the direction perpendicular to the pulling velocity is presented in [32, 35–37]. The resulting force-velocity relation was shown by Fusco and Fasolino [32] to obey the power law

$$\bar{f} \sim f_0 + cv^{2/3} \quad \text{for} \quad v \to 0\,, \qquad (7.12)$$

regardless of the dimensionality of the model and the value of damping.

7.3.3 Effect of Noise

Experimental results reveal, however, that the effects of thermal noise play an important role in the stick-slip motion. The specific indications of the importance of thermal effects are the randomness of slip events, see Fig. 7.1 and Ref. [38], the temperature dependence of atomic friction [27], and the approximately *logarithmic* – as opposed to the power-law (7.12) – dependence

of the friction force for asymptotically small pulling velocity, which has been interpreted using a model based on the assumption that the tip transitions (slips) from one lattice site to the next are due to thermal activation [7, 27, 39–42].

To generalize the Prandtl–Tomlinson model by incorporating the effects of thermal noise, Sang et al. [28] and independently Dudko et al. [29] have proposed to include the random force due to the thermal vibrations of the surface into the equation of motion of the tip, and thus arrived at Eq. (7.11), but, again, with $\eta_s = \eta$ and $\eta_c = 0$. This equation was simulated numerically [28, 29] using a random number generator. Recently, the Langevin analysis has been extended to the two-dimensional case in [32]. The ensuing relation between the friction force (7.1) and pulling velocity was found in all cases to be logarithmic, namely,

$$\bar{f} \propto |\ln v|^{\alpha} \tag{7.13}$$

with the exponent α close to one [27].

7.4 Rate Theory

7.4.1 Preliminary Remarks

One of the consequences of Eq. (7.11) is that at high velocities the third term in the right-hand side, which describes the viscous drag, exceeds all the other acting forces, and the force of friction behaves as (cf. 7.12)

$$\bar{f} \to \eta_s v \quad \text{for } v \to \infty\,, \tag{7.14}$$

allowing us, at least in principle, to experimentally determine the coefficient η_s associated with the substrate from the slope of the force-velocity plot at high v. On the other hand, the term $\eta_s v$ can be eliminated from the equation of motion by a change of variables

$$\tilde{z} = z - \eta_s v/\kappa\,, \quad \tilde{t} = t + \eta_s/\kappa\,, \tag{7.15}$$

which allows us to rewrite (7.11) (omitting the tildes) as

$$m\,\ddot{z}(t) + \eta\dot{z}(t) = -U_0'(z(t) + vt) - \kappa z(t) + \sqrt{2\eta kT}\xi(t)\,. \tag{7.16}$$

The experimentally observed friction force can thus be decomposed into two contributions:

$$\bar{f}_{\rm exp} = -\kappa\bar{z} + \eta_s v\,, \tag{7.17}$$

where $\bar{z}$ is the time-averaged value of coordinate z from Eq. (7.16). The first term in Eq. (7.17) describes the effect of interaction with the lattice, and the second one the effect of the viscous drag due to molecular degrees of freedom. Since we are interested in the former contribution to the friction force, we will calculate the friction force $\bar{f}$ from Eqs. (7.1), (7.6) and (7.16). To analyze

the experimental data, a term linear in the velocity can simply be added to the final result, yielding Eq. (7.17).

7.4.2 Conditions of Validity of the Rate Approximation

The qualitative difference between the Prandtl–Tomlinson model at a finite temperature and at zero temperature, which leads to the logarithmic (7.13) force-velocity relation, as opposed to the power-law behavior (7.12), is, evidently, that the transitions of the tip from one well of the total potential (7.2) to the other are random thermally activated events at $T > 0$. In contrast to the noise-free model, they occur *before* the potential barrier separating one well from the other disappears completely due to the action of the elastic force.

To develop an understanding of the logarithmic relation (7.13), a rate approximation can be used [13]. This approximation is valid if the typical heights of the barriers separating the adjacent minima is much greater than the thermal energy. Then, the system possesses two very different time scales: the one describing the relaxation of the tip within a given potential well and the much longer one describing the thermally activated interwell transitions of the tip.

The condition that the relevant barrier heights be much greater than the thermal energy kT [13] imposes certain restrictions on the system studied. In particular, it implies that the combined stiffness κ of the cantilever, the tip, and the surface in the contact region must not exceed some value, at which the potential (7.2) becomes monostable. This upper value can be estimated [15] as a second derivative of the surface potential at one of its minima, resulting in the requirement

$$\kappa \ll U_0''(na)\,. \tag{7.18}$$

For realistic parameter values (see caption to Fig. 7.1b), this implies that the rate description is valid when $\kappa \ll 20\,\mathrm{N/m}$. Even though the typical stiffness of the cantilever $\kappa_{\mathrm{cantilever}}$ is of the order of 70 N/m, the combined spring constant (7.5) is usually about 1 N/m, because the tip and the substrate in the contact region are usually quite soft. Furthermore, pulling must proceed sufficiently slowly to allow the transitions to occur before the respective time scales become comparable. Both conditions are well satisfied for most of the so far reported experimental studies [7, 23, 27, 41, 42], in which the stick-slip motion is observed, whereas violation of these conditions leads to the onset of the opposite regime of steady sliding [15, 26, 31] characterized by low friction forces.

7.4.3 Effective Spring Constant

The time-scale separation condition allows one to deal not with the coordinate of the AFM tip, $z(t)$, but rather with the probability to find the AFM tip in

a given (say, nth) potential well, and the rates $\omega(f_n(t))$ of transitions of the tip from one lattice site to the next. The argument of the rate is the elastic force

$$f_n(t) = -\kappa z_n(t) \tag{7.19}$$

corresponding to the given nth minimum of the total potential $U(z,t)$ from Eq. (7.2). Approximating the position of the minimum $z_n(t)$ with that of the moving interaction potential, Eq. (7.4), we find that the force corresponding to the nth minimum increases roughly as $\kappa v t$, allowing one to determine the spring constant from the rate of force increase during the stick phase [7, 21, 23, 41].

While experimentally this seems to be the most reasonable procedure to measure the spring constant, it yields only an approximation to the "true" spring constant, κ. We will denote such an approximation as $\tilde{\kappa}$. To estimate the difference between the "bare" or "true" spring constant, κ, from Eq. (7.5) and the experimentally determined, "effective" or "apparent" one, $\tilde{\kappa}$, we need to find the position of the nth minimum of the combined potential (7.2) more precisely. This minimum is found by differentiating the potential $U(z,t)$ and setting the derivative to zero, leading to

$$U_0'(z_n + vt) = -\kappa z_n \,. \tag{7.20}$$

As substantiated by both the experimental findings and numerical results, the elastic force increases almost linearly during the stick phase. With this in mind, we expand the function in the left-hand side of Eq. (7.20) around some point, b_n, yielding to the first order

$$U_0'(b_n) + U_0''(b_n)(z_n + vt - b_n) + \ldots = -\kappa z_n \,, \tag{7.21}$$

from which an approximation for z_n follows:

$$z_n = \frac{U_0''(b_n)(b_n - vt) - U_0'(b_n)}{\kappa + U_0''(b_n)} \,. \tag{7.22}$$

Then, by combining Eqs. (7.19) and (7.22), we find the elastic force corresponding to the nth minimum

$$f_n(t) \simeq -\tilde{\kappa}(na - vt) + \tilde{\kappa}/U_0''(b_n) \,, \tag{7.23}$$

where the effective spring constant is [7, 15, 43]

$$\tilde{\kappa} = \frac{\kappa U_0''(b_n)}{\kappa + U_0''(b_n)} \,. \tag{7.24}$$

What is the best choice for the expansion point b_n in this expression? The average value of the right-hand side of Eq. (7.20) during the stick-slip motion

is $-\kappa\bar{z} = \bar{f}$, and thus the optimal choice of the expansion point b_n is given by the equation

$$U_0'(b_n) = \bar{f}\,. \tag{7.25}$$

In view of the periodicity (7.3) of the potential $U_0(z)$, the solution of this equation is defined up to an integer multiple of the lattice constant. This uncertainty, however, does not affect the numerical value of the effective spring constant (7.24). Denoting the smallest root of Eq. (7.25) as b_0, we adopt the convention that

$$b_n = b_0 + na\,. \tag{7.26}$$

The important conclusion of this discussion is that the effective spring constant (7.24), which is experimentally deduced from the slope of the stick sections on the f–t curves [7,21,23,41], actually includes not only the effects of the elastic deformations of the cantilever, the AFM tip, and the surface in the contact region, but also the effect of the curvature of the tip-substrate interaction potential. Moreover, since the average force $\bar{f}$ depends on the pulling velocity, we conclude that so do the expansion point b_n and thus the effective spring constant $\tilde{\kappa}$ given by Eq. (7.24). The often used approximation $\kappa \simeq \tilde{\kappa}$, however, seems to be well justified in view of the condition (7.18).

7.4.4 Calculation of the Transition Rate

According to the Kramers' theory of thermally activated escape [13], the force-dependent transition rate behaves as

$$\omega(f) = \omega_0(f)\,\mathrm{e}^{-\Delta U(f)/kT}\,, \tag{7.27}$$

where $\Delta U(f)$ is the force-dependent height of the energy barrier separating the current minimum from the next one, and the preexponential factor $\omega_0(f)$ depends weakly on the force. When the inertia effects are negligibly small, this prefactor is given by [13]

$$\omega_0(f) = \frac{\sqrt{U''(z_{\min}(f))\,U''(z_{\max}(f))}}{2\pi\eta}\,, \tag{7.28}$$

where $z_{\min}$ and $z_{\max}$ denote the positions of the potential minimum and of the barrier corresponding to the force value f, see Fig. 7.2. In particular, according to Eq. (7.19), the position of the minimum is given by

$$z_{\min} = -f/\kappa\,. \tag{7.29}$$

The condition that one of the minima of the potential (7.2) coincides with $-f/\kappa$ is met not at arbitrary, but at specific moments of time, t_f, which can

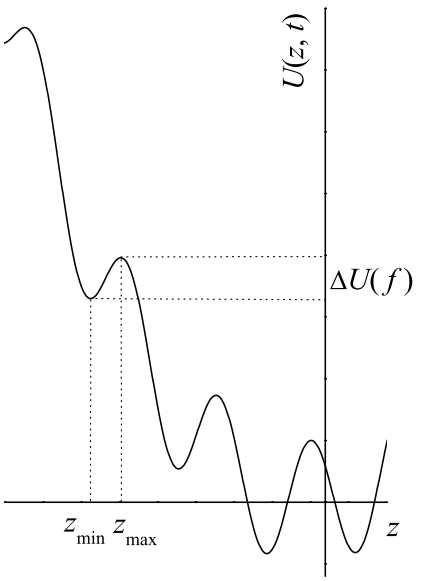


Fig. 7.2. To the formulae (7.27, 7.28)

be found from the requirement that the derivative of the total potential (7.2) at $z = -f/\kappa$ and $t = t_f$ be zero, i. e. from the equation

$$U_0'\big(vt_f - (f/\kappa)\big) = f\,, \quad U_0''\big(vt_f - (f/\kappa)\big) > 0\,. \tag{7.30}$$

In view of the periodicity of the potential $U_0(z)$, Eq. (7.3), this equation defines t_f only up to an integer multiple of a/v. This ambiguity, however, does not affect any of the parameters entering the rate formula (7.27, 7.28).

Next, we determine the position of the maximum by setting the derivative of the potential (7.2) to zero,

$$U'(z_{\max}, t_f) = 0\,, \tag{7.31}$$

and choosing that solution of this equation, which is the closest to $z_{\min}$. Once we have solved Eqs. (7.30, 7.31) analytically or numerically, we find the barrier height

$$\Delta U(f) = U(z_{\max}(f), t_f) - U(z_{\min}(f), t_f)\,, \tag{7.32}$$

thus completing the definition of all parameters entering the rate formula (7.28).

7.4.5 Asymptotic Cases of Low and High Spring Constants

To find the force-velocity relation in the stick-slip mode, two viewpoints have been adopted in the literature.

(i). The average number of interstitial transitions per unit time is approximately $\omega(\bar{f})$. As a result of each such transition, the tip gets displaced by the distance a, and hence the force-velocity relation is [39, 40, 44, 45]

$$v = a\omega(\bar{f}) \,. \tag{7.33}$$

(ii). Within an alternative approach [28, 29, 41, 42], one deals with the probability $\mathcal{P}(t|t_{\rm L})$ of staying within the same lattice site up to the moment of time t, provided that the tip entered this site at the initial time $t_{\rm L}$, i. e. $\mathcal{P}(t_{\rm L}|t_{\rm L}) = 1$. The time evolution of $\mathcal{P}(t|t_{\rm L})$ for $t \geq t_{\rm L}$ is governed by the rate equation

$$\frac{\partial \mathcal{P}(t|t_{\rm L})}{\partial t} = -\omega(f(t))\, \mathcal{P}(t|t_{\rm L}) \,. \tag{7.34}$$

With the help of the transformation of variables (7.23), we find from the rate equation the probability that the transition into the next site occurs at a force value between f and $f + {\rm d}f$, provided that the initial lower force value for a given stick phase was $f_{\rm L}$,

$$-\frac{\partial \mathcal{P}(f|f_{\rm L})}{\partial f} = \frac{1}{\tilde{\kappa} v} \omega(f)\, \mathcal{P}(f|f_{\rm L}) \,. \tag{7.35}$$

The most probable force f_*, at which the transition into the next site occurs, is evaluated by setting the derivative of this function to zero. This results in the relation between the pulling velocity and the most probable force at the moment of slip:

$$v = \frac{\omega^2(f_*)}{\tilde{\kappa}\omega'(f_*)} \,. \tag{7.36}$$

This equation in various forms has been presented in Refs. [28, 29, 41, 42], the difference between these works stemming from different assumptions regarding the functional dependence $\omega(f)$. A further difference of Eq. (7.36) from the corresponding equations of [28, 29, 41, 42] is that the "bare" stiffness κ was used in these works, not its effective value.

At low forces, we may assume that

$$\Delta U(f) \simeq \Delta U(0) - af/2 \,, \tag{7.37}$$

as the distance between the two adjacent extrema of the potential (7.2) is about $a/2$. At high forces close to the critical tilt, the potential can be shown to behave as [28, 29]

$$\Delta U(f) \propto -(f - f_0)^{3/2} \,, \tag{7.38}$$

where f_0 is the critical force, at which the potential barrier disappears. Then it is easily shown that both approximations (7.33) and (7.36) yield the logarithmic force-velocity relation (7.13) with the exponent $\alpha = 1$ at low forces and $\alpha = 2/3$ at high forces (i. e. f slightly below f_0).

Nevertheless, the two force-velocity relations, (7.33) and (7.36), are in an apparent contradiction with each other. Indeed, the latter predicts that at

vanishingly small $\tilde{\kappa}$ the velocity v must diverge, while the former is independent on $\tilde{\kappa}$. This contradiction is due to the fact that the two expressions (7.33) and (7.36) are valid in the opposite ranges of $\tilde{\kappa}$, respectively.

Indeed, as a result of each slip event, the force drops approximately by the amount $\tilde{\kappa}a$ (see Eq. (7.23)) which, therefore, can be taken as an estimate of the magnitude of force fluctuations during the stick-slip motion. Only when it is much smaller than the average force itself – i. e. $\tilde{\kappa} \ll \bar{f}/a$ – is the statement that the transitions occur at about the same frequency $\omega(\bar{f})$ justified, leading to Eq. (7.33).

On the other hand, the jump probability distribution (7.35) depends on the value of the initial lower force f_{L}, at which a given stick phase begins, while the equation (7.36) for f_* is independent on f_{L}. Therefore, it may turn out that the most probable transition force f_* defined by Eq. (7.36) is smaller than the initial force f_{L}; in this case, the observed jump probability distribution (7.35) is a monotonically decreasing function peaked at the initial force f_{L}, and the actual most probable force of transition is just f_{L}, not f_*. In order for this not to be the case – i. e. in order for the relation (7.36) to hold – the separation between the typical forces at which the stick phase begins and ends must be sufficiently high. Since this separation is given by $\tilde{\kappa}a$, we conclude that the force-velocity relation (7.36) is valid at a relatively high combined stiffness $\tilde{\kappa}$. A more precise condition of validity of the relation (7.36) is obtained in the next section, where we derive the general force-velocity relation valid both at high and low values of $\tilde{\kappa}$ which unites the relations (7.33) and (7.36).

7.4.6 A Unified Force-Velocity Relation from the Rate Theory

Solving Eq. (7.35), we find the no-jump conditional probability for a given initial force f_{L}

$$\mathcal{P}(f|f_{\mathrm{L}}) = \exp\Big(-\frac{1}{\tilde{\kappa}v} \int_{f_{\mathrm{L}}}^{f} \mathrm{d}f'\, \omega(f') \Big) . \tag{7.39}$$

For further convenience, we introduce the long-time probability distribution $W(f_{\mathrm{L}})$ that a given stick phase of motion begins with the force value in a small interval around f_{L}. It obeys the following integral equation:

$$W(f_{\mathrm{L}}) = \int_{-\infty}^{f_{\mathrm{L}}+\tilde{\kappa}a} \mathrm{d}f'_{\mathrm{L}} \left(-\frac{\partial \mathcal{P}(f_{\mathrm{L}}+\tilde{\kappa}a|f'_{\mathrm{L}})}{\partial f_{\mathrm{L}}} \right) W(f'_{\mathrm{L}}) . \tag{7.40}$$

The physical meaning of this equation is as follows: the probability $W(f_{\mathrm{L}})$ that a given stick phase starts around the force value f_{L} is nothing else but the probability $-\partial\mathcal{P}(f_{\mathrm{L}}+\tilde{\kappa}a|f'_{\mathrm{L}})/\partial f_{\mathrm{L}}$ that the previous phase ends at a value $f_{\mathrm{L}} + \tilde{\kappa}a$ averaged over all initial forces $f'_{\mathrm{L}} < f_{\mathrm{L}} + \tilde{\kappa}a$ of the previous phase with the weighting function $W(f'_{\mathrm{L}})$.

If we find the distribution of lower forces by solving Eq. (7.40), the time-averaged force follows immediately [46, 47]:

$$\bar{f} = \bar{f}_{\mathrm{L}} + \frac{\tilde{\kappa}a}{2} \equiv \int_{-\infty}^{\infty} \mathrm{d}f_{\mathrm{L}}\, f_{\mathrm{L}}\, W(f_{\mathrm{L}}) + \frac{\tilde{\kappa}a}{2}\ . \tag{7.41}$$

While the relation (7.41) between the time-averaged force and the first moment of the distribution $W(f_{\mathrm{L}})$ is simple, the iterative determination of the latter distribution according to Eq. (7.40) is only possible by means of a time-consuming numerical procedure. Therefore, our next goal is to obtain an approximate analytic relation between the average force $\bar{f}$ and velocity v without the knowledge of the distribution $W(f_{\mathrm{L}})$.

Multiplication of both sides of Eq. (7.40) by f_{L} and integration yield

$$\begin{aligned} \tilde{\kappa}a &= \int_{-\infty}^{\infty} \mathrm{d}f_{\mathrm{L}} \int_{-\infty}^{f_{\mathrm{L}}} \mathrm{d}f'_{\mathrm{L}}\, \mathcal{P}(f_{\mathrm{L}}|f'_{\mathrm{L}})\, W(f'_{\mathrm{L}}) \\ &= \int_{-\infty}^{\infty} \mathrm{d}f_{\mathrm{L}}\, W(f_{\mathrm{L}}) \int_{f_{\mathrm{L}}}^{\infty} \mathrm{d}f'_{\mathrm{L}}\, \mathcal{P}(f'_{\mathrm{L}}|f_{\mathrm{L}})\ , \end{aligned} \tag{7.42}$$

where we used integration by parts with subsequent change of the variables of integration from f_{L} to $f_{\mathrm{L}} + \tilde{\kappa}a$ to obtain the first equality, and replaced the order of integration with subsequent interchange of the variables $f'_{\mathrm{L}} \leftrightarrow f_{\mathrm{L}}$ to obtain the second one.

It can be shown [46] that the inner integral in the rightmost expression represents the average force increment $\Delta F(f_{\mathrm{L}})$ during a given stick phase, provided that the initial force value was f_{L}. The double integral, therefore, is the force increment during the stick phase averaged over all initial forces; this quantity equals the force drop $\tilde{\kappa}a$ during the slip to the next potential well.

The behavior of the function $\Delta F(f_{\mathrm{L}}) := \int_{f_{\mathrm{L}}}^{\infty} \mathrm{d}f'_{\mathrm{L}}\, \mathcal{P}(f'_{\mathrm{L}}|f_{\mathrm{L}})$ does not deviate strongly from linearity in that force interval around $\bar{f}_{\mathrm{L}}$, where the distribution $W(f_{\mathrm{L}})$ is significantly different from zero. To see that this is indeed so, let us consider two cases of high and low $\tilde{\kappa}$.

(i) At high combined stiffness, the probability of staying within the same well (7.39) is close to one in a rather extended force interval above f_{L} (because of a large factor $\tilde{\kappa}$ in the denominator of the expression in the exponent in Eq. (7.39)). This means that the statistics of jump events, and hence the average force at the moment of transition, is practically independent of the initial force f_{L}. Correspondingly, the average force increment for a fixed initial force f_{L} indeed behaves linearly with f_{L}, i. e., as $\Delta F(f_{\mathrm{L}}) = \bar{f}_U - f_{\mathrm{L}}$ at high $\tilde{\kappa}$, where the average upper force $\bar{f}_U$ is practically independent of f_{L}.

(ii) At low values of $\tilde{\kappa}$ the magnitude of force fluctuations, $\tilde{\kappa}a$, and hence the width of the distribution $W(f_{\mathrm{L}})$ is also small. Then, the deviations of the function $\Delta F(f_{\mathrm{L}})$ from linearity can be neglected within the relevant force interval.

Since the function $\Delta F(f_{\mathrm{L}})$ is approximately linear in the physically important range of f_{L} both at low and high $\tilde{\kappa}$, it can reasonably be expected that it is also almost linear in this range for intermediate combined stiffness. This allows us to replace the average value of this function in the inner integral of the rightmost expression (7.42) with the function evaluated at the average value of its argument, $\Delta F(\bar{f}_{\mathrm{L}})$, resulting in the following implicit force-velocity relation [43, 46, 47]:

$$\int_{\bar{f}-\tilde{\kappa}a/2}^{\infty} \mathrm{d}f \exp\left(-\frac{1}{\tilde{\kappa}v} \int_{\bar{f}-\tilde{\kappa}a/2}^{f} \mathrm{d}f' \omega(f') \right) = \tilde{\kappa}a \,. \tag{7.43}$$

Here, we wrote an explicit expression (7.39) for $\mathcal{P}(f'_{\mathrm{L}}|f_{\mathrm{L}})$ and used the relation (7.41) between $\bar{f}_{\mathrm{L}}$ and $\bar{f}$.

For an analytic solution of this equation, we need to further approximate the transition rate so as to be able to evaluate the integral in the left-hand side. According to Eq. (7.27), the rate depends exponentially strongly on force, $\omega(f) \propto \mathrm{e}^{-\Delta U(f)/kT}$. On the other hand, the energy barrier $\Delta U(f)$ is a much weaker function, see Eqs. (7.37) and (7.38). This observation suggests to expand the *logarithm* of the transition rate about some force value f_0, i. e. to take, to the first order

$$\omega(f) \simeq \omega(f_0)\, \mathrm{e}^{\alpha(f_0)(f-f_0)} \,, \quad \alpha(f) := \omega'(f)/\omega(f) \,. \tag{7.44}$$

The next question is how to choose the force f_0, about which the expansion is performed. To answer this question, let us examine Eq. (7.43) more closely. Depending on the value of $\tilde{\kappa}$, the integrand, $\exp(\ldots)$, may exhibit two kinds of behavior:

(i) At high $\tilde{\kappa}$, there is a rather wide region of forces between the average lower force, $\bar{f}_{\mathrm{L}}$, and the average upper force, $\bar{f}_U \equiv \bar{f}_{\mathrm{L}} + \tilde{\kappa}a$, where the integrand has the value 1, followed by an abrupt drop to zero in the immediate vicinity of $\bar{f}_U$. The nature of the approximation (7.44) is such that if we choose f_0 to lie in the region of the steepest descent of the integrand, we will correctly reproduce its behavior not only in this region, but also outside of it, where the integrand is very close to 0 (at higher forces) or 1 (at lower forces).

(ii) At low $\tilde{\kappa}$, the integrand drops to zero in a rather narrow interval above $\bar{f} - \tilde{\kappa}a/2$. Therefore, we expect that the integral will not be very sensitive to the choice of f_0, provided that f_0 belongs to that narrow region where the integrand is notably different from zero. This region extends from $\bar{f}_{\mathrm{L}}$ to a value slightly higher than the average force at the moment of transition, $\bar{f}_U = \bar{f} + \tilde{\kappa}a/2$.

Thus, the choice of the expansion point in Eq. (7.44), which applies to both cases equally well, is simply

$$f_0 = \bar{f}_U \equiv \bar{f} + \tilde{\kappa}a/2 \,. \tag{7.45}$$

Making a change of variables according to $x = \mathrm{e}^{\alpha(\bar{f}_U)(f-\bar{f}_L)}$, we have from Eqs. (7.43, 7.44, 7.45):

$$\alpha(\bar{f}_U)\tilde{\kappa}a = \mathrm{e}^{g(\bar{f}_U)}\, E_1\big(g(\bar{f}_U)\big)\,, \quad g(f) := \frac{\omega(f)\,\mathrm{e}^{-\alpha(f)\tilde{\kappa}a}}{\tilde{\kappa}v\alpha(f)}\,. \tag{7.46}$$

Here $E_1(g) = \int_1^\infty \mathrm{d}x\,\mathrm{e}^{-gx}/x$ is the exponential integral, which can be evaluated numerically using a standard algorithm [48].

It follows from Eq. (7.46) that the sought relation between force and velocity has the form

$$v(\bar{f}) = a\omega(\bar{f} + \tilde{\kappa}a/2)\, Q\left(\frac{\omega'(\bar{f} + \tilde{\kappa}a/2)}{\omega(\bar{f} + \tilde{\kappa}a/2)}\tilde{\kappa}a\right), \tag{7.47}$$

where the function $Q(x)$ is defined implicitly by the relation

$$E_1\left((x\,\mathrm{e}^x Q(x))^{-1}\right)\mathrm{e}^{(x\,\mathrm{e}^x Q(x))^{-1}} = x\,. \tag{7.48}$$

From the asymptotic properties of the exponential integral [49] it can be inferred that $Q(x)$ is a monotonically decreasing function with $Q(0) = 1$ and $Q(x) \sim \mathrm{e}^\gamma/x$ at $x \to \infty$, where $\gamma = 0.5772156649\ldots$ is Euler's constant. We further approximate this function by

$$Q(x) \simeq 1/\sqrt{1 + (\mathrm{e}^{-\gamma}x)^2}\,. \tag{7.49}$$

The high accuracy of this approximation is demonstrated by Fig. 7.3.

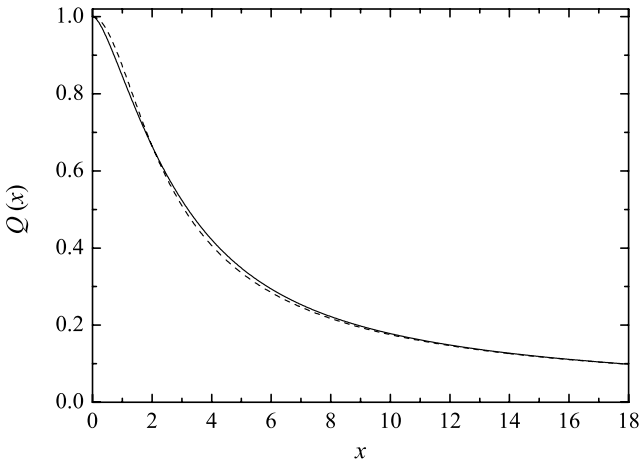


Fig. 7.3. The function $Q(x)$ appearing in the force-velocity relation (7.47), as calculated numerically from Eq. (7.48) (*solid line*) and the approximation (7.49) (*dashed line*)

In contrast to the experimental procedure, in which one imposes the pulling velocity v and measures the friction force $\bar{f}$, the practical implementation of Eq. (7.47) involves (i) setting the value of $\bar{f}$, (ii) determining the effective stiffness according to Eqs. (7.24, 7.25), and (iii) calculating the pulling velocity v, at which the given value of $\bar{f}$ is realized.

In the limit $\tilde{\kappa} \to 0$, our force-velocity relation (7.47) simplifies to Eq. (7.33). In the opposite limit of large $\tilde{\kappa}$ (but not large enough to render the combined potential (7.2) lose its multistable character, so that the condition (7.18) still holds) we have from (7.47, 7.49) the relation

$$v(\bar{f}) = \frac{\omega^2(\bar{f} + \tilde{\kappa}a/2)}{\tilde{\kappa}\omega'(\bar{f} + \tilde{\kappa}a/2)}\,\mathrm{e}^{\gamma}\,, \quad \text{large } \tilde{\kappa}\,, \tag{7.50}$$

which is equivalent to Eq. (7.36), apart from the factor $\mathrm{e}^{\gamma} \simeq 1.78$. By analyzing the problem of escape from a metastable potential well under the action of a steadily increasing force, it can be shown [50, 51] that this factor stems from the fact that the argument in Eq. (7.50) is the average force at transition, $\bar{f}_U \equiv \bar{f} + \tilde{\kappa}a$, while in Eq. (7.36) it is the most probable force f_* of transition.

It is possible within the present approach to find the parameter range in which the high-$\tilde{\kappa}$ approximation (7.50) is applicable. It follows from Eqs. (7.47, 7.49) that for this to be the case, the argument of the function $Q(x)$ must be much greater than one, that is

$$\frac{\omega'(\bar{f} + \tilde{\kappa}a/2)}{\omega(\bar{f} + \tilde{\kappa}a/2)}\tilde{\kappa}a \gg 1\,. \tag{7.51}$$

In the lowest non-vanishing order, the Kramers' rate (7.27) can be approximated as $\omega(f) \simeq \omega(0)\,\mathrm{e}^{af/2kT}$, as the distance between the minimum and maximum is about $a/2$. The condition of validity of the high-$\tilde{\kappa}$ limit (7.50) is thus seen to be

$$\tilde{\kappa}a^2/2 \gg kT\,. \tag{7.52}$$

Keeping in mind that at room temperature $kT \simeq 4\,\mathrm{pN\,nm}$ and $a \simeq 0.5\,\mathrm{nm}$, we conclude that Eq. (7.50) is a good approximation for $\tilde{\kappa} \gg 0.03\,\mathrm{N/m}$.

7.5 Zero-Stiffness Asymptotic Limit

An exact asymptotic force-velocity relation can be obtained in the limit of vanishingly small stiffness κ. If the magnitude of force fluctuations, which is of the order of κa, is much smaller than the average force itself, $\bar{f} = -\kappa\bar{z}$, we can replace the elastic force $-\kappa z$ in Eq. (7.16) with its average value, leading to

$$m\ddot{z}(t) + \eta\dot{z}(t) = -U_0'(z(t) + vt) + \bar{f} + \sqrt{2\eta kT}\xi(t)\,,\ \kappa \to 0\,. \tag{7.53}$$

By switching to the moving frame of reference of the substrate,

$$x = z + vt \tag{7.54}$$

we rewrite this equation as

$$m\ddot{x}(t) + \eta\dot{x}(t) = -U_0'(x) + \bar{f} + \eta v + \sqrt{2\,\eta kT}\xi(t)\,. \tag{7.55}$$

This equation describes the diffusion of a Brownian particle in a tilted periodic potential $U_{tilted}(x) = U_0(x) - (\bar{f} + \eta v)\,x$. In the general case of arbitrary damping, this system has been extensively investigated by Risken, see [12] and references therein.

The problem of finding the average velocity of such a particle, $\langle\dot{x}\rangle \equiv v$, in the overdamped limit $m \to 0$ has been solved *exactly* by Stratonovich [52], who derived the analytic formula

$$v = \frac{akT(1 - \mathrm{e}^{-a(\bar{f}+\eta v)/kT})}{\eta \int_0^a \mathrm{d}x_1 \int_{x_1}^{x_1+a} \mathrm{d}x_2\, \mathrm{e}^{[U(x_1)-U(x_2)-(x_1-x_2)(\bar{f}+\eta v)]/kT}}\,. \tag{7.56}$$

The argument of the function in the right-hand side is not the average force $\bar{f}$, but rather the combination $\bar{f} + \eta v$. In order to plot the $\bar{f}$-v relation, one can, first, for each given value of the combined force $\bar{f} + \eta v$ calculate the corresponding velocity v using the Stratonovich formula (7.56), and then deduce the average friction force $\bar{f}$ corresponding to this velocity value by subtracting the value of ηv from the combined force.

7.6 Numerical Results

In order to demonstrate the high accuracy of the analytic force-velocity relation (7.47), we will focus on the overdamped ($m \to 0$) version of Eq. (7.16). The overdamped limit is justified by the fact the main contribution to the effective mass m comes from only a small number of atoms in the tip-substrate contact region, which take part in the interaction [17, 26]. The experimental justification for the overdamped limit [15] is that if the inertia effects were to play an important role in the stick-slip motion, one would be able to observe multiple jumps of the AFM tip over several lattice sites. However, such multiple jumps are not seen in actual experiments.

The Fokker-Planck equation corresponding to the overdamped version of the Langevin equation (7.16) reads [12]:

$$\frac{\partial W(z,t)}{\partial t} = \frac{1}{\eta}\frac{\partial}{\partial z}\left(kT\frac{\partial W(z,t)}{\partial z} + \Big(U_0'(z,t) + \kappa z\Big)W(z,t)\right)\,, \tag{7.57}$$

where $W(z,t)$ is the probability to find the AFM tip at the position z at the moment of time t. For concreteness, we assume the substrate potential to be

a trigonometric function

$$U_0(z) = -\frac{\Delta U_0}{2} \cos \frac{2\pi z}{a} \tag{7.58}$$

with parameters $\Delta U_0 = 250\,\mathrm{pN\,nm}$, $a = 0.52\,\mathrm{nm}$. Furthermore, the thermal energy at room temperature is $kT = 4.04\,\mathrm{pN\,nm}$, and the coefficient of viscous friction is assumed to take the representative value $\eta = 1\,\mathrm{pN\,msec/nm}$. We have propagated the solution of this equation in time numerically until it became time-periodic, i. e.

$$W(z, t + a/v) = W(z, t)\,. \tag{7.59}$$

Then, we calculated the mean time-dependent force

$$\langle f(t) \rangle = -\kappa \int_{-\infty}^{\infty} \mathrm{d}z\, z W(z, t)\,, \tag{7.60}$$

and its time-average value

$$\langle \bar{f} \rangle = \frac{v}{a} \int_0^{a/v} \mathrm{d}t \langle f(t) \rangle\,, \tag{7.61}$$

which coincides with the expression (7.1) for ergodicity reasons.

For implementation of the relation (7.47), we first need to determine the parameters entering the expressions for the rate (7.27), (7.28), (7.32). Solving Eq. (7.30) with the potential (7.58), we find the moments of time, at which the elastic force f is realized:

$$t_f = \frac{1}{v} \left(\frac{f}{\varkappa} + \frac{a}{2\pi} \sin^{-1} \left(\frac{af}{\pi \Delta U_0} \right) + na \right). \tag{7.62}$$

The position of the maximum, $z_{\max}$, nearest to $z_{\min} = -f/\kappa$ at the moment of time t_f was found from Eq. (7.31) numerically, although a rather accurate analytical approximation is also possible [47]. Finally, to calculate the effective spring constant $\tilde{\kappa}$ from Eq. (7.24), we need to determine the expansion point b_n from Eq. (7.25). For the potential (7.58), the solution of this equation reads

$$b_n = \frac{a}{2\pi} \sin^{-1} \frac{a\bar{f}}{\pi U_0} + na\,. \tag{7.63}$$

Presented in Fig. 7.4 is the force-velocity relation obtained from the numerical solution of the Fokker-Planck equation (7.57) and from the analytic formula (7.47). Within the range of validity of the rate theory (low velocities), the average friction force indeed increases in proportion to $\ln v$. We note that the approximate relation (7.47) is in a very good agreement with the numerically exact results in the increasing part of the graph. On the other hand, at higher velocities, the average force starts to decrease to zero [17,26].

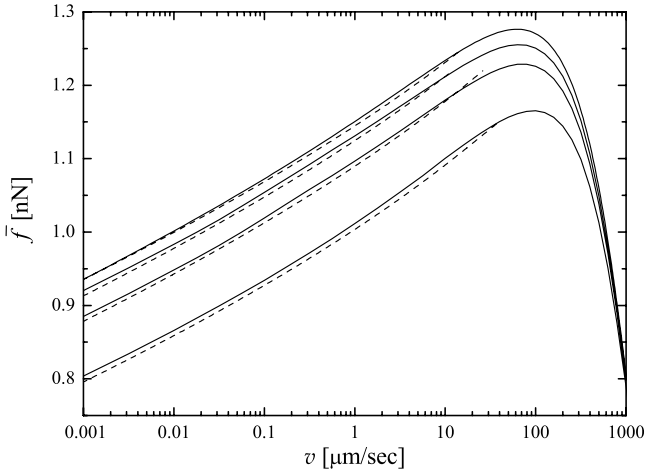


Fig. 7.4. The force-velocity relation for the FFM model (7.16, 7.58) with the parameters specified in caption to Fig. 7.1, but with different spring constants κ from Eq. (7.5), namely $\kappa \to 0$, 0.3, 0.6, and 1.2 N/m (*from top to bottom*). The results are obtained from the numerical solution of the Fokker-Planck equation (7.57) corresponding to the Langevin equation (7.16) (*solid lines*), and the rate approximation (7.47) (*dashed lines*). The zero-κ limit is given by Stratonovich's formula (7.56)

This occurs because at high v, the AFM tip cannot follow the fast temporal variations of the underlying surface potential in view of the tip's finite intrinsic relaxation time. Since the rate theory is valid under the condition that this relaxation time be extremely short, the effect of force non-monotonicity cannot be captured by Eq. (7.47).

It should be borne in mind that the calculations presented in Fig. 7.4 omit the effect of the substrate viscous drag, while the true force-velocity relation is given by Eq. (7.17). This means that at still higher velocities, the experimentally observed friction force should pass through a minimum and then start increasing again in proportion to v.

A force-velocity relation qualitatively similar to that depicted in Fig. 7.4 has been observed experimentally by Riedo et al. [42], who report a logarithmic increase of the lateral force, followed by its leveling-off as a function of velocity; apparently, this plateau corresponds to the maxima on the curves of Fig. 7.3. As far as we know, the regime of higher velocities, where the force decreases, has yet to be explored experimentally.

As mentioned earlier, the effective stiffness $\tilde{\kappa}$ is expected to depend on velocity, but how strongly? Presented in Fig. 7.5 is the ratio of the effective to the "bare" stiffness, $\tilde{\kappa}/\kappa$, vs. v, for different values of κ. In agreement with the discussion above, this ratio tends to one for small values of κ, and decreases monotonically as κ increases. We note that even for moderately high κ, the experimentally determined spring constant may be smaller than

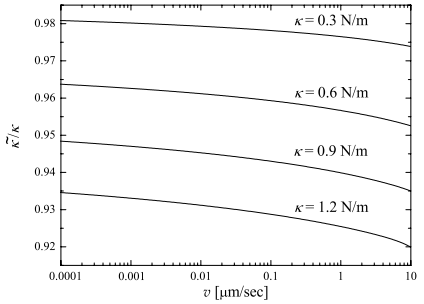


Fig. 7.5. The dependence of the effective spring constant $\tilde{\kappa}$ on the pulling velocity for the model parameters specified in the caption to Fig. 7.1, but for different "bare" spring constants κ from (7.5), as indicated on each curve

the "bare" value by as much as 10%. However, the overall velocity dependence of $\tilde{\kappa}$ is rather weak.

7.7 Conclusions and Outlook

A first result of our present review is the approximative analytical relation (7.47) between the pulling velocity v and the resulting average friction force $\bar{f}$, supplemented by the approximation (7.49) for $Q(x)$, the connection (7.24), (7.25) between bare and effective spring constants κ and $\tilde{\kappa}$, respectively, and the transition rates $\omega(f)$ according to (7.27)–(7.32). This relation (7.47) covers a rather extended range of pulling velocities v and spring constants κ, namely all cases for which the stick-slip motion of the cantilever tip can be approximately captured by means of a rate theoretical approach. Previously known, apparently contradicting approximations (7.33) and (7.36) are recovered as special cases of the general solution (7.47) in the asymptotic regimes of small and large spring constants κ, respectively, supplemented by a previously missed constant factor in (7.50) and a new validity condition (7.52).

Our second main point is the exact analytical force-velocity relation for asymptotically small cantilever stiffness and high damping in (7.56), but in turn without any of the further restrictions as required e. g. in the rate approach discussed above, especially with respect to the pulling velocity. In the regime of high pulling velocities, where the rate description is no longer applicable and the stick-slip motion is significantly smeared out, the analytic result (7.56) predicts a maximum of the average force as a function of velocity, followed by a subsequent decrease. This theoretically predicted non-monotonic force-velocity relation should be observable under realistic experimental conditions, especially pulling velocities only slightly larger than those achieved so far.

By comparison with numerically exact solutions we found that the analytical approximation (7.47) is very accurate within the parameter range typically explored by experiments and hence appears to be an adequate tool

for interpretation of the experimental data. However, further improvements and refinements are possible in several directions.

The analytical treatment of the fast pulling regime, i. e. beyond the validity of a rate description, in combination with an arbitrary cantilever stiffness remains a challenging open problem. Also unsolved, but probably of less practical relevance, is the analytical treatment of finite inertia effects beyond the validity of a rate approach.

One of the most drastic simplifications made in this work is that the motion of the AFM tip is essentially one-dimensional, whereas in reality the tip also moves in the direction perpendicular to that of pulling. An extension of the Tomlinson model to the more general two-dimensional case is treated numerically in [32,35,36]. The qualitative difference from the one-dimensional case studied here is that the AFM tip has several choices of the next minima to jump into from the current minimum of the potential. It follows from Langer's generalization of Kramers' theory to many dimensions [13] that the tip will follow the path of the "least resistance" with the highest probability, i. e. it will jump to that minimum which is separated by the smallest barrier. Thus, the motion of the tip will no longer proceed along a straight line, but rather follow a zig-zag path. Clearly, the one-dimensional model involving a periodic substrate potential as treated in the present work is no longer applicable for arbitrary pulling directions relative to the substrate surface, but only in the special cases when pulling proceeds along such a crystallographic direction that the inter-node transition rates remain the same along the path; e. g., for a square underlying lattice, these are the $\langle 01 \rangle$ and $\langle 11 \rangle$ directions. Although the tip then still moves along a one-dimensional manifold, the shape of the zig-zag path and hence the rate of transitions from one node to the next will vary upon changing the lateral position of AFM relative to the substrate surface, i. e. perpendicularly to the pulling direction. While such variations may in principle still be captured by an effective one-dimensional model, the same is in general no longer true regarding the variations of the most probable escape path from one node to the next upon variation of the the longitudinal position of AFM relative to the substrate surface, i. e. parallel to the pulling direction. While the one-dimensional model treated in the present work thus strictly applies only in the special case that the zig-zag path actually degenerates to a straight line, the approximation in the more general case is expected to be still very good. For all these reasons, the one-dimensional model studied in this work is a special case of the more general two-dimensional model, whose development is an interesting subject for future research.

For technological applications, one naturally looks for conditions characterized by as small friction as possible. From this point of view, the following improvements of the theory may be of importance.

In the derivation of the relation (7.47), we have taken into account only the transitions of the AFM tip into the next minimum of the potential (7.2), see Eq. (7.34), and neglected the exponentially disadvantaged back-transitions.

Such an approximation is justified when the pulling is sufficiently fast, so that the corresponding potential barrier becomes very high before any such back-transition takes place. Therefore, the relation (7.47) does not apply to the case of extremely slow pulling, when the back-transitions do play a role [53]. As is evident from Fig. 7.4, this regime is characterized by low friction forces.

The low-friction regime of high velocities (after the force maximum in Fig. 7.3) can so far be covered analytically only in the limit of zero stiffness κ, see Eq. (7.56). To describe this regime for arbitrary κ, one needs to go beyond the rate description in an attempt to generalize the Stratonovich's formula (7.56).

A further promising research direction is related to the control and reduction of friction by means of lateral or normal vibrations [42]. The possibility of such a friction reduction is discussed theoretically in the work [54] and references therein. We expect that the rate description can still be applicable in this case, but with a suitable modification to account for the temporal oscillations of the potential [55].

Acknowledgement. The authors are grateful to Alexander von Humboldt-Stiftung and the Deutsche Forschungsgemeinschaft (RE 1344/3-1 and RE 1344/4-1) for financial support of this work.

References

1. M. Urbakh, J. Klafter, D. Gourdon, and J. Israelachvili, Nature **430**, 525 (2004)
2. N.J. Mosey, M.H. Müser, and T.K. Woo, Science **307**, 1612 (2005)
3. C.M. Mate, G.M. McClelland, R. Erlandsson, and S. Chiang, Phys. Rev. Lett. **59**, 1942 (1987)
4. G. Binning, C.F. Quate, and Ch. Gerber, Phys. Rev. Lett. **56**, 930 (1986)
5. G.V. Dedkov, Physics Uspekhi **43**, 541 (2000)
6. B.N.J. Persson, *Sliding Friction*, Springer, Berlin, 2000
7. E. Gnecco, R. Bennewitz, T. Gyalog, and E. Meyer, J. Phys.: Condens. Matter **13**, R619 (2001)
8. M.R. Sorensen, K.W. Jacobsen, and P. Stoltze, Phys. Rev. B **53**, 2101 (1996)
9. A.I. Livshits and A.L. Shluger, Phys. Rev. B **56**, 12482 (1997)
10. H. Grabert, *Projection operator techniques in nonequilibrium statistical mechanics*, Springer, Berlin 1982
11. P. Hänggi and H. Thomas, Phys. Rep. **88**, 207 (1982)
12. H. Risken, *The Fokker-Planck equation*, Springer, Berlin 1984
13. P. Hänggi, P. Talkner, and M. Borkovec, Rev. Mod. Phys. **62**, 251 (1990)
14. P. Reimann, Phys. Rep. **361**, 57 (2002)
15. A. Socoliuc, R. Bennewitz, E. Gnecco, and E. Meyer, Phys. Rev. Lett. **92**, 134301 (2004)
16. D. Tománek, W. Zhong, and H. Thomas, Europhys. Lett. **15**, 887 (1991)
17. P. Reimann and M. Evstigneev, New Journ. Phys. **7**, 25 (2005)

18. M.A. Lantz, S. J. O'Shea, M.E. Welland, and K.L. Johnson, Phys. Rev. B **55**, 10776 (1997)
19. M.A. Lantz, S. J. O'Shea, A.C.F. Hoole, and M.E. Welland, Appl. Phys. Lett. **70**, 970 (1997)
20. J. Colchero, A. M. Baró, and O. Marti, Tribology Lett. **2**, 327 (1996)
21. R.W. Carpick, D.F. Ogletree, and M. Salmeron, Appl. Phys. Lett. **70**, 1548 (1997)
22. K.L. Johnson and J. Woodhouse, Tribology Lett. **5**, 155 (1998)
23. R. Bennewitz, T. Gyalog, M. Guggisberg, M. Bammerlin, E. Meyer, and H.-J. Güntherodt, Phys. Rev. B **60**, R11301 (1999)
24. R. Kubo, Rep. Progr. Phys. **29**, 255 (1966)
25. P. Reimann, Chem. Phys. **268**, 337 (2001)
26. P. Reimann and M. Evstigneev, Phys. Rev. Lett. **93**, 230802 (2004)
27. S. Sills and R. Overney, Phys. Rev. Lett. **91**, 095501 (2003)
28. Y. Sang, M. Dubé, and M. Grant, Phys. Rev. Lett. **87**, 174301 (2001)
29. O.K. Dudko, A.E. Filippov, J. Klafter, and M. Urbakh, Chem. Phys. Lett. **352**, 499 (2002)
30. O. Zwörner, H. Hölscher, U.D. Schwarz, and R. Wiesendanger, Appl. Phys. A **66**, S263 (1998)
31. T. Baumberger and C. Caroli, Eur. Phys. Journ. B **4**, 13 (1998)
32. C. Fusco and A. Fasolino, Phys. Rev. B **71**, 045413 (2005)
33. L. Prandtl, Zeitschr. f. Angew. Math. u. Mech. **8**, 85 (1928)
34. G.A. Tomlinson, Philos. Mag. **7**, 905 (1929)
35. T. Gyalog, M. Bammerlin, R. Lüthi, E. Meyer, and H. Thomas, Europhys. Lett. **31**, 269 (1995)
36. H. Hölscher, U.D. Schwarz, and R. Wiesendanger, Europhys. Lett. **36**, 19 (1996)
37. R. Prioli, A.M.F. Rivas, R.L. Freire Jr., and A.O. Caride, Appl. Phys. A **76**, 565 2003
38. A. Schirmeisen, L. Jansen, and H. Fuchs, Phys. Rev. B **71**, 245403 (2005)
39. F. Heslot, T. Baumberger, B. Perrin, B. Caroli, and C. Caroli, Phys. Rev. E **49**, 4973 (1994)
40. T. Bouhacina, J.P. Aimé, S. Gauthier, D. Michel, and V. Heroguez, Phys. Rev. B **56**, 7694 (1997)
41. E. Gnecco, R. Bennewitz, T. Gyalog, Ch. Loppacher, M. Bammerlin, E. Meyer, and H.-J. Güntherodt, Phys. Rev. Lett. **84**, 1172 (2000)
42. E. Riedo, E. Gnecco, R. Bennewitz, E. Meyer, and H. Brune, Phys. Rev. Lett. **91**, 084502 (2003)
43. M. Evstigneev and P. Reimann, Phys. Rev. B **73**, 113401 (2006)
44. B.J. Briscoe and D.C.B. Evans, Proc. R. Soc. London A **380**, 389 (1982)
45. J.N. Glosli and G.M. McClelland, Phys. Rev. Lett. **70**, 1960 (1993)
46. M. Evstigneev and P. Reimann, Europhys. Lett. **67**, 907 (2004)
47. M. Evstigneev and P. Reimann, Phys. Rev. E **71**, 056119 (2005)
48. W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical recipes in C* (Cambridge, 1999)
49. M. Abramowitz and I. Stegun, ed., *Handbook of mathematical functions* (Dover, 1965)
50. P.W. Williams, Analytica Chimica Acta **479**, 107 (2003)
51. A. Garg, Phys. Rev. B **51**, 15592 (1995)

52. R.L. Stratonovich, Radiotekhn. Elektron. (Moscow) **3**, 497 (1958). English translation in: P.I. Kuznetsov, R.L. Stratonovich, and V.I. Tikhonov, ed., *Nonlinear Transformations of Stochastic Processes*, Pergamon, Oxford, 1965
53. S.Yu. Krylov, K.B. Jinesh, H. Valk, M. Dienwiebel, and J.W.M. Frenken, Phys. Rev. E **71**, 065101 (2005)
54. Z. Tshiprut, A.E. Filippov, and M. Urbakh, Phys. Rev. Lett. **95**, 016101 (2005)
55. J. Lehmann, P. Reimann, and P. Hänggi, Phys. Rev. E **62**, 6282 (2000)

8 The Basic of Nanoscale Friction and Ways to Control it

Joseph Klafter and Michael Urbakh

School of Chemistry, Raymond and Beverley Sackler Faculty of Exact Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel

8.1 Introduction

Friction plays a central role in a large number of systems and phenomena which are seemingly unrelated, but which on closer scrutiny are found to display common features shared by tribological processes whether in technological, biological or geological areas. These systems encompass a vast spectrum of length and time scales, ranging from Ångstroms to kilometres (i. e. from atomic scale friction to motion of geological plates) and from picoseconds to centuries.

Due to its practical importance and the relevance to basic scientific questions there has been major increase in activity during the last decade in the study of interfacial friction on the microscopic level [1–9]. The development of durable and/or low friction surfaces and thin lubricating films has become an important factor in the miniaturization of moving components in many technological devices. These include micro-electro-mechanical systems (MEMS), computer recording systems, miniature motors, and more [10, 11]. In particular, there has been a steadily growing interest in the control of friction in such miniature systems with emphasis on methods other than the conventional chemical control. When dealing with micro-mechanical components and small loads, the simple old empirical laws of friction do not always hold. This breakdown stems from the high surface-to-volume ratio and from neglecting the important role of surface chemistry, adhesion and surface structure or roughness. The conventional tribological and lubrication techniques used for large objects can be ineffective in the nano-world, where different criteria and novel methods for control are needed. The difficulties in realizing an efficient control of friction are related to the lack of fundamental understanding of many of the underlying physical processes which take place within the microscopic separation between two materials, and to the presence of many degrees of freedom in the problem under a strict size confinement.

In order to understand the behavior of two real surfaces in relative motion while still in contact, one needs to look first into what goes on at the 'single asperity' level. With the advent of the atomic force microscope (AFM) [12] and the surface forces apparatus (SFA) [13] it became possible to study individual sliding junctions at the molecular level. AFM and SFA are nowadays powerfull tools in nano-, micro- and macroscopic tribological experiments when

measuring the normal and lateral forces, and wear between a nanometre-radius tip or micrometre-sized colloidal particle against a substrate surface (AFM), and between two macroscopic molecularly smooth or rough surfaces of measurable molecular contact area that confine a lubricant film of measurable thickness (SFA). These experiments revealed some new structural and dynamical phenomena which occur in nanoscale liquids confined between two atomically smooth solid surfaces, which include: (a) structural transitions in thin liquid films induced by confining surfaces (both layering and orientational ordering) [14, 15], (b) periodic and chaotic stick-slip motion whose frequency and amplitude change with the driving velocity [16–19], (c) transitions between "smooth" sliding and stick-slip behavior at certain critical velocities and loads [18–20], (d) more than one type of sliding motion [20], (e) inverted stick-slip motion [20], (f) substantial increase of the effective viscosity in liquid sheared films relative to the bulk [21], (g) dramatic slowing down of dynamics compared to the bulk [4, 21] and (h) a dependence of friction on the previous history of the system [22]. These and other observations have motivated various theoretical efforts [1, 3, 5, 8, 9], both numerical and analytical, but many phenomena remain unexplained or still controversial.

8.2 Theoretical Approaches

The theoretical approaches introduced to investigate frictional forces in sheared systems include large-scale molecular dynamics (MD) simulations [5, 8, 23–29], phenomenological rate-and-state models [1, 30–34] which attempt to incorporate the basic physics and are usually convenient to apply, and "minimalistic" models [35–38], which take into account only a few interactions those that are believed to be most relevant. Each approach has its advantages and disadvantages and emphasizes different aspects.

Atomistic MD simulations have a wide range of applicability and have reached a high level of accuracy. They help in the understanding of liquid layering in nano-confinement [25, 29], the relationship between static and kinetic friction [1,27], the nature of transitions between stick-slip and smooth sliding [24], slippage at solid-liquid interfaces [24,39], shear thinning [40], and the friction of rough surfaces [41]. But MD simulations are currently limited to time-scales not greater than tens of nanoseconds and length scales of tens of nanometers, which are too short for analysing the much slower relaxation processes occurring at shearing interfaces or confined liquid films [5].

An important question is how to reduce the large scale and many parameter MD simulations to a simpler description with only a few equations of motion that are not sensitive to all the specific details of the simulations, yet are sufficient for modeling the processes under consideration. Various phenomenological rate-state models [1,30–34] provide such a reduced description by assuming that the friction depends on the shear rate and a small number

of 'state variables' that describe the properties of the interface. Different interpretations of the microscopic properties that these state variables describe have been proposed, such as the amount of dilation (molecular volume) at the interface [34] or the degree of crystallinity of the sheared film [1,31–33]. Most approaches leave the nature of the state variables unspecified and assume that they depend on the most recent velocity. The coefficients of simple dynamical equations are fitted to experiment and can then be used to describe a wide range of observed frictional behavior, including the transition between stick-slip (regular or chaotic) and smooth sliding friction. However, it remains unclear how to relate the parameters introduced in these state-and-rate models to measurable molecular or system properties.

An important step towards developing an understanding of friction came with 'minimalistic' models that focus on a small number of the most relevant degrees of freedom of confined molecules needed to describe a particular level of complexity [3,35–38,42]. The dynamical response of AFM has been successfully modeled when the embedded system is represented by a single particle which corresponds also to many non-interacting particles (see Fig. 8.1). In this case the motion of the driven plate and the embedded particles have been described by the following deterministic equations [35,36]

$$M\ddot{X} + \eta(\dot{X} - \dot{x}) - K(X - Vt) + \frac{\partial U(x - X)}{\partial X} = 0 \tag{8.1}$$

$$m\ddot{x} + \eta\dot{x} + \eta(\dot{x} - \dot{X}) + \frac{\partial U(x)}{\partial x} + \frac{\partial U(x - X)}{\partial x} = 0 \tag{8.2}$$

Here m and M are the masses of the particle and the plate, and x and X are the coordinates of the particle and the driven plate. The plate is pulled by a spring with a force constant K connected to a stage that moves with a velocity V. The interactions between the particle and each of the plates are described by a periodic potential $U(x)$ with the period b. The second term in Eq. (8.1) and the second and the third terms in Eq. (8.2) describe the dissipative forces between the particle and the plates which are proportional to their relative velocities. These terms account for dissipation which arises from interaction with phonons and/or other excitations.

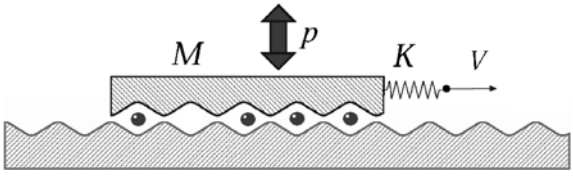


Fig. 8.1. Schematic presentation of the "minimalistic model" that reduces the system to its bare essentials: representing an embedded system by non-interacting particles embedded between two surfaces

This model has been also generalized to include the effects of thermal fluctuations, coupling between lateral and normal motion of the plates and the embedded particles and incommensurability of confining plates [37, 38, 43–45]. It has been shown that when the embedded system is represented by a single particle most of the experimental observations mentioned above are qualitatively recovered. Moreover, these simple single particle models enabled predictions to be made that were later verified experimentally. These include: (a) chaotic behavior of sheared system characterized by intermittent force fluctuations, (b) new phases of motion (two types of sliding and inverted stick-slip motion), (c) dependence of observed frictional properties on mechanical characteristics of the apparatus, and (d) mechanical control of friction via external manipulations.

The reason for the success of the "minimalistic" approach is that it naturally leads to two characteristic states of the embedded system when sheared in the presence of thermal noise: '*trapped*' and '*sliding*' states [45]. These are the ingredients that lead to stick-slip and the transition to sliding and therefore are the essential requirements for successful models, including rate-and-state models and MD simulations. The minimalistic models have emphasized the non-linear nature of frictional dynamics and opened a way to the application of the arsenal of approaches of nonlinear dynamics to characterize and tune frictional response.

8.3 Control of Friction

The ability to control and manipulate frictional forces is important for a variety of applications. From a practical point of view one wishes to be able to control frictional forces so that the overall friction is reduced (or enhanced), the chaotic regime is eliminated, and instead, smooth sliding is achieved. Such control can be of high technological importance for micromechanical devices and computer disk drives, where the early stages of motion and stopping often exhibit undesired stick-slip or damage [10]. In contrast, chaotic stick-slip behaviour might be desirable, e. g., in string instruments. Controlling frictional forces has been traditionally approached by chemical means, usually by supplementing base lubricants with friction modifier additives [28, 46–48]. However, standard lubrication techniques used for large objects are expected to be less effective in the micro- and nano-world. Novel methods for control and manipulation are therefore needed.

8.3.1 Mechanical Control

A new approach to 'tuning' frictional response, which has recently attracted interest, is the mechanical control of a system, via externally imposed vibrations of small amplitude and energy [26, 43, 49–54]. In this case, the idea is not to change considerably the physical properties of the interfaces, but to

either reduce the frictional force or to eliminate stick-slip motion through the stabilization of desirable modes of motion, which are unstable in the absence of control. The goal of this approach is twofold: (a) to achieve smooth sliding at low driving velocities, which otherwise correspond to the stick-slip regime; (b) to decrease the frictional force.

Below we discuss two different methods of mechanical control, via: (i) normal, and (ii) lateral vibrations, both of which are applied externally to the system.

Normal vibrations

Two different methods of control via normal vibrations have been discussed: the first one uses a feedback control [49–51], and the second one relies on a "brute-force" modification of the system dynamics without a feedback [26, 43, 53, 54]. We start from the feedback mechanism of control. The analysis of the mechanism has been done on a one-dimensional model [49], which includes two rigid plates with embedded non-interacting particles between them. The effect of normal load is introduced through the dependence of the amplitude of the particle-substrate interaction U_0 on the load P_n

$$U_0(P_\mathrm{n}) = U_0(1 + \chi(P_\mathrm{n} - P_\mathrm{n}^0)) \tag{8.3}$$

Here U_0 is the value of the potential for some nominal value of the normal load P_n^0,and χ is a dimensional constant. The normal load has been used as the control parameter to modify friction. Eq. 8.3 assumes small load variations around P_n^0,which, as shown in Ref. [49], are sufficient to achieve control. The control method is characterized by two independent steps: (a) reaching the vicinity of an unstable sliding mode of motion, and (b) stabilizing it. The control has been realized by small variations of the normal load, which has been externally adjusted employing a proportional feedback mechanism.

The aim of this approach is to stabilize a smooth sliding state for low driving velocities, (velocities smaller than a critical value $V < V_\mathrm{c}$), where one expects chaotic stick-slip motion. Within the model discussed here sliding states correspond to periodic orbits of the system with two periods: (a) period $T = b/V$, which corresponds to a motion of the particles being trapped by one of the plate; and (b) period $T = 2b/V$, which corresponds to the particles moving with the drift velocity $V/2$. In the chaotic stick-slip region both orbits still exist, but are unstable. The approach is therefore to drive the system into a sliding state by stabilizing these unstable periodic orbits. This makes it possible to extend the smooth sliding to lower velocities. The control of unstable periodic orbits in dynamical systems was previously proposed and experimentally applied to a wide variety of physical systems including mechanical systems, lasers, semiconductor circuits, chemical reactions, biological systems etc. In Ref. [49] we proposed for the first time such a control for "cleaning" chaotic stick-slip motion.

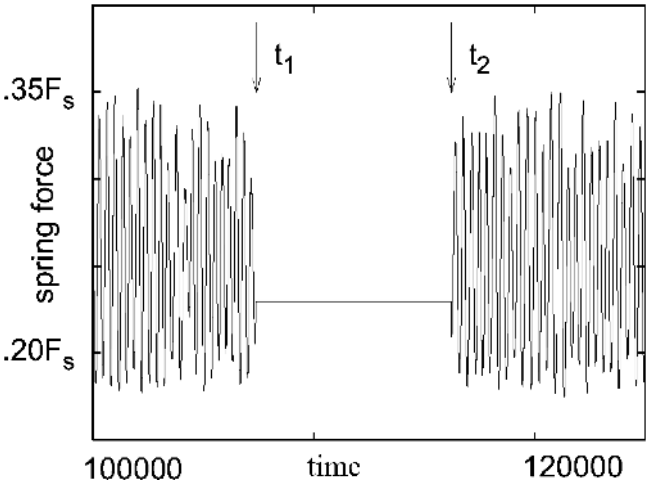


Fig. 8.2. Eliminating the chaotic stick-slip motion under mechanical control [49]. The friction force is given in units of static friction

Figure 8.2 demonstrates the effect of the mechanical control on the time dependence of the spring force. The results correspond to the control of the trapped-sliding state [49], in which the particles cling to one of the plates and move either with velocity V=0 (sticking at the bottom plate) or velocity V (sticking at the top plate). The control is switched on at time t_1 and is shut down at time t_2. We clearly see that as a result the chaotic motion of the top plate is really replaced by smooth sliding.

In Ref. [49] the possibility to control friction has been discussed in model systems described by differential equations.In realistic systems, what is usually available are time series of dynamical variables, rather than governing equations. In this case the time-delay embedding method [55] can be applied in order to transform a scalar time series into a trajectory in phase space. This procedure allows to find the desired unstable periodic orbits and to calculate variations of the parameters required to control friction.

The "brute-force" method of control based on a harmonic modulation of the normal load, $P_{\rm n}(t)$, has been studied within various approaches which include the generalized Tomlinson model [43], one-dimensional rate-state models [26, 53] and grand-canonical MD simulations [26]. All the calculations demonstrated that oscillations of the normal load could lead to a transition from a high-friction stick-slip dynamics to a low-friction sliding state. This effect can be controlled through the selection of the oscillation frequency. The mechanism behind this phenomenon in lubricated junctions has been clarified by MD simulations [26] which show that oscillations of the normal load frustrate ordering in the lubricated film, maintaining it in a nonequilibrium sliding state with low friction. The theoretical predictions on friction control have been supported by recent experiments [52–54] which indicate that nor-

mal vibrations generally stabilize the system against stick-slip oscillations, at least for a modulation frequency much larger than the stick-slip one.

Lateral vibrations

In order to demonstrate an effect of lateral vibrations on friction we consider below [56] an AFM configuration in which a nanoscale tip is driven along a substrate, which oscillates in the lateral direction (see Fig. 8.3). The motion of the tip in the lateral and normal directions is governed by the coupled Langevin equations:

$$M\ddot{x}(t) = -\eta_x(z)(\dot{x}(t) - \dot{x}_0(t)) - \partial U(x - x_0, z)/\partial x + F_x + f_x \,, \tag{8.4}$$

and

$$M\ddot{z}(t) = -\eta_z(z)\dot{z}(t) - \partial U(x - x_0, z)/\partial z + F_z + f_z \,, \tag{8.5}$$

where

$$U(x, z) = U_0 \left[1 + \sigma \sin\left(\frac{2\pi}{b}(x - x_0)\right)\right] \exp(1 - z/\lambda) \,, \tag{8.6}$$

$$\eta_{x,z}(z) = \eta^0_{x,z} \exp(1 - z/\lambda) \,. \tag{8.7}$$

Here M and x, z are the mass and the lateral and normal coordinates of the tip, $U(x, z)$ is the potential experienced by the tip due to the interaction with the substrate, b is its periodicity in the lateral direction, and σ characterizes the amplitude of corrugation in the lateral (x) direction. The parameters η_x and η_z are responsible for the dissipation of the tip kinetic energy due to the motion in the x and z directions, respectively. Here we take into account the dependence of U and $\eta_{x,z}$ on the tip-substrate separation [56]. As an example, we assume an exponential decrease of U and $\eta_{x,z}$ with a rate λ^{-1} as z increases. The tip is held at the surface by a normal load $F_z = K_z(z_0 - z(t))$ applied by a linear spring of spring constant K_z, and it is laterally pulled, $F_x = K_x(Vt - x(t))$, by a spring of spring constant K_x connected to a stage which moves with a constant velocity V. The effect of lateral vibrations of

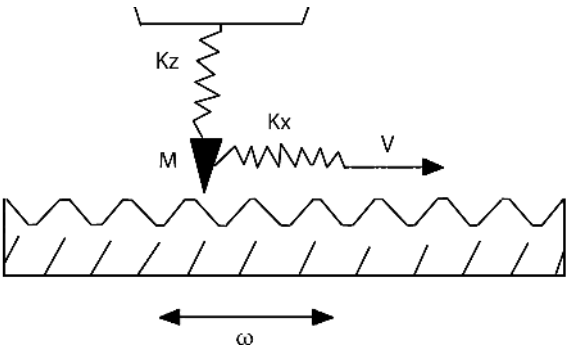


Fig. 8.3. Schematic sketch of a system under an externally applied lateral vibrations in order to control friction

the substrate is included through a time dependence of its position, $x_0 = A_0 \sin(2\pi\omega t)$, where A_0 and ω are the amplitude and the frequency of the oscillations. The random forces, $f_{x,z}$, represent thermal noise satisfying the fluctuation-dissipation relation, $< f_i(t) f_j(0) >= 2\eta_i k_\mathrm{b} T \delta(t) \delta_{i,j}$, where $i, j = x, z$.

It is convenient to introduce the dimensionless coordinates and time $X = x/b$, $Z = z/b$, $\tau = t\omega_0$, where $\omega_0 = (1/b)\sqrt{U_0\sigma/M}$ is the frequency of the small oscillations of the tip in the periodic potential. The dynamical behavior of the system is determined by the following dimensionless parameters: $A = A_0/b$, $\Omega = \omega/\omega_0$, $k_\mathrm{B}T/U_0$, $\eta_{x.z}/M\omega_0$, λ/b, $K_x b^2/\left(4\pi^2 U_0 \sigma\right)$, $K_z\lambda^2/U_0$ and $\tilde{v} = V/\left(\omega_0 b\right)$.

Equation (8.4) shows that the substrate vibrations cause a time-periodic (ac) force acting on the tip, $F_\mathrm{ac} = M(2\pi\omega)^2 A_0 \sin(2\pi\omega t)$. This force presents the effect of inertia. Its amplitude depends on both the amplitude and frequency of vibrations. Recent studies of surface diffusion under ac forcing [57, 58] demonstrated that the diffusivity D may be strongly enhanced and even exceed the free (Brownian) diffusivity, $D_\mathrm{free} = k_\mathrm{B}T/\eta_x$, for an optimal matching of the driving frequency, ω, and the amplitude A_0. Similar effect has been found in our calculations. The mechanism of this phenomenon can be understood analyzing the tip trajectories. At resonance frequencies, $\Omega = \Omega_*$, which correspond to the maximum of the diffusion coefficient, the tip approaches the top of the surface potential at the end of half cycle of the plate vibrations, where the driving force, $F_\mathrm{ac} = 0$. Then, even a weak thermal noise splits the ensemble of tips into two parts that relax to the neighboring minima of the surface potential, and the resonance enhancement of diffusion is observed.

Our calculations suggest an AFM configuration for the observation of the vibration-induced enhancement of the diffusion. In this configuration the tip experiences the influence of two potentials: the periodic surface potential and the harmonic potential, $K_x(x - x_\mathrm{sup})^2/2$, due to the elastic coupling to the support of the microscope of coordinate x_{sup}, which remains fixed. Our simulations in Fig. 8.4 demonstrate that the experimentally measurable root mean square displacement (rmsd) of the tip, $\Delta L\,(\Omega)$, exhibits a resonance enhancement for the frequency Ω_* corresponding to the maximum of the diffusion coefficient. The results for $\Delta L\,(\Omega)$ can be fitted by the Ornstein–Uhlenbeck equation for the rmsd due to diffusion in the harmonic potential [59]

$$\Delta L_\mathrm{OU} = \sqrt{D_\mathrm{free}\eta_x/K_x}\,, \tag{8.8}$$

when the free diffusion coefficient, D_free, is substituted by the Ω-dependent diffusion coefficient $D\,(\Omega)$, which corresponds to the enhanced diffusion induced by the lateral vibrations.

Under the conditions which are typical for AFM measurements [60], $M = 8.7 \cdot 10^{-12}$kg, $U_0 = 0.25\,\mathrm{eV}$ and $b = 0.4\,\mathrm{nm}$, we arrive at the resonance frequency $\omega_* = \omega_0\Omega_* = 7 \cdot 10^4\,\mathrm{Hz}$. This value lies within the frequency interval exploited by the shear modulation technique [61] and agrees qualitatively

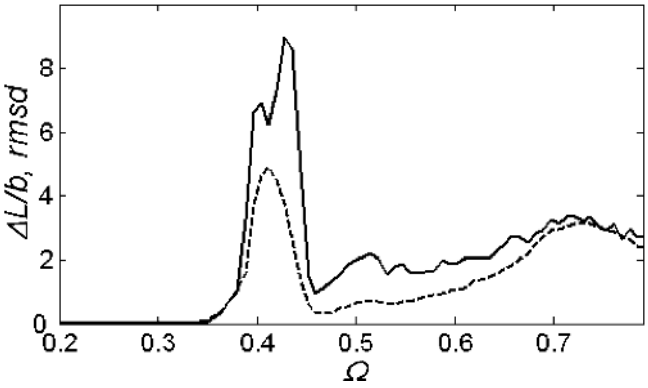


Fig. 8.4. The effect of lateral vibrations on the frequency dependence of the root mean squared displacement (rmsd) of the tip [56]. *Solid curve* – numerical simulations, *dashed curve* - calculation according to the equation $\Delta L(\Omega) = \sqrt{D(\Omega)\eta_x/K_x}$. Parameter values: $A = 1$, $\lambda/b = 1$, $\sigma = 1$, $\eta_{x,z}/M\omega_0 = 3.2$, $k_{\mathrm{B}}T/U_0 = 0.01$, $K_x b^2/\left((2\pi)^2 U_0\sigma\right) = 3.2 \cdot 10^{-4}$, $K_z >> K_x$

with the value of the frequency for which the resonance reduction of friction under the oscillatory drive has been observed [60]. The experiment suggested here can be considered as a diffusion "spectroscopy" of surfaces. Measuring the "spectrum" of diffusion, $D(\Omega)$, one can determine the parameters of the surface potential.

We have found that the substrate vibrations can cause also a significant reduction of friction in the stick-slip regime of motion. Figures 8.5 and 8.6 show the Ω-dependence of the time-averaged friction force, $\langle F_x \rangle$, and the

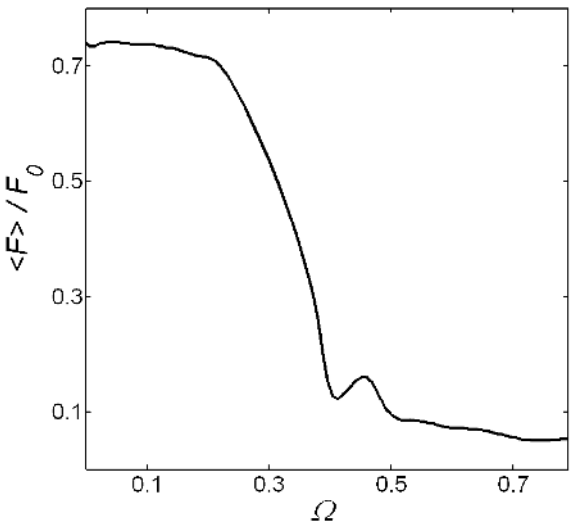


Fig. 8.5. The relative frictional force versus frequency of the in-plane vibrations of the substrate calculated for an AFM configuration [56]. The friction force is given in units of static friction, F_0 . Parameter values: $V/V_0 = 0.16$, where $V_0 = \omega_0 b$, and other parameters as in Fig. 8.4

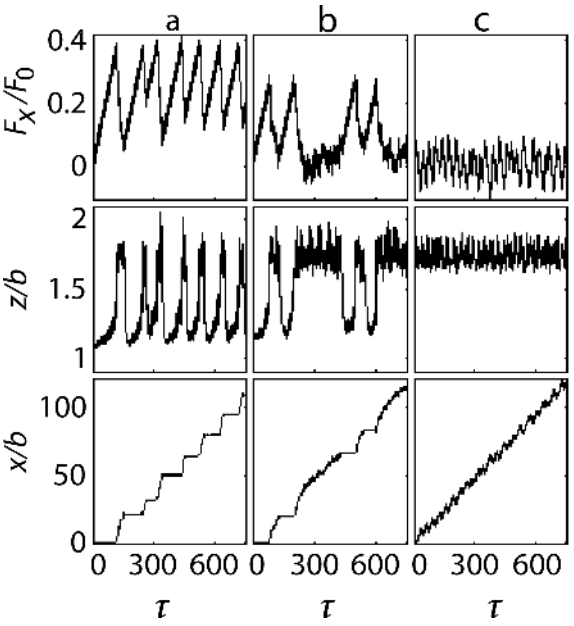


Fig. 8.6. Time dependencies of the lateral tip displacement (top panel) and the tip-surface separation (bottom panel) calculated for four vibrational frequencies: **a** $\Omega = 0.26$, **b** $\Omega = 0.32$, **c** $\Omega = 0.39$, **d** $\Omega = 0.48$. Parameter values: $A = 1$, $K_z\lambda^2/U_0 = 0.63$, and other parameters as in Fig. 8.4

instantaneous friction force, $F_x = K_x\,(Vt - x)$, tip displacement and tip-surface separation, respectively. One can see that for low frequencies the lateral vibrations do not affect the frictional response. Both the spring force and displacement traces show the patterns which are typical of the stick-slip behavior, and the average force is independent of Ω. In the vicinity of the threshold frequency $\Omega_{\rm th}$, for which the enhanced diffusion begins to emerge, we find a drastic decrease of the kinetic friction. Figures 8.5 and 8.6 demonstrate that the lateral vibrations not only reduce the friction force but they also transform the stick-slip motion to a "smooth" sliding. However, the application of lateral vibrations does not allow to eliminate completely the force fluctuations. Even under the optimal conditions the variance of the friction force remains of the order of K_xA. The main feature in Fig. 8.5 is a reduction of friction for all frequencies above the threshold one $\Omega_{\rm th}$. In contrast to the enhancement of diffusion, the reduction of friction does not exhibit pronounced resonance features.

In AFM experiments the tip is held near the surface by a normal load applied through the spring with the spring constant K_z. As a result the tip driven in a lateral direction performs oscillations also in the normal direction [43]. The tip-surface separation, which is initially z_0, at equilibrium, starts growing before a slippage occurs and stabilizes at a larger distance, z_h, as long as the tip motion continues (see Fig. 8.6). The amplitude of the normal oscillations depends on the surface potential and the stiffness of the normal spring. Including the normal motion of the tip does not change qualitatively the effect of lateral vibrations on friction. As in the one-dimensional case,

the main feature of the frequency dependence of frictional force is a sharp decrease of $\langle F \rangle$ and a transition from the stick-slip motion to sliding at the threshold frequency, $\Omega_{\rm th}$. The threshold frequency decreases with a decrease of the stiffness of the normal spring, K_z. In the majority of cases, a decrease of K_z leads also to a reduction of the time-averaged frictional force, $\langle F \rangle$. This can be easily understood, because the decrease of the stiffness, K_z, results in the reduction of the actual amplitude of the surface potential experienced by the tip. It should be noted (see Fig. 8.6) that the vibration induced reduction of friction is accompanied by a dilatancy transition; namely the disappearance of the state with a small tip-substrate separation, z_0. Both the reduction of friction and dilatancy transition originate from the excitation of the large scale tip oscillations by the substrate vibrations. As a result both effects arise at the same threshold frequency $\Omega_{\rm th}$.

8.4 Chemical Control

As we mentioned earlier, controlling frictional forces has been traditionally approached by chemical means, usually supplementing base lubricants by friction modifier additives. Each additive in such molecular mixtures has a different role: some decrease the static friction and eliminate undesirable stick-slip motion, some influence the temperature dependence of viscosity, others inhibit corrosion [28,46,48]. While the behavior of single component lubricants in nanoscale confinements has been extensively studied both experimentally and theoretically [1,2,4,8,9], investigations of the behavior of molecular mixtures under similar conditions are at their first stages [28, 47, 48]. Questions on frictional forces of mixtures, regimes of motion, stability, microscopic behaviors of the additives and their interactions with the base lubricants and substrates are still open.

Recently we have suggested a new approach which might help decide how to tailor molecular mixtures so that they provide desirable frictional properties [62, 63]. The idea has been to modify frictional behaviors through doping a base solvent by molecules that induce dynamic phase transitions (or phase transitions under shear) in the embedded system. We have considered a mixed monolayer embedded between the plates of the SFA which consists of two types of molecules A and B, a "base solvent" and an "additive", respectively. Molecules A and B are chosen in such a way that they tend to occupy different sites on the surfaces of the plates, for instance at the top of a substrate atom and at the center of the substrate lattices. The A–B intermolecule interaction has been described by a potential which provides an attraction between different types of molecules and a repulsion between identical molecules.

In Fig. 8.7 we present an example of the dependence of the time-averaged spring force on the number fraction of additives. The total number of embedded molecules was kept constant. The calculations have been done for three

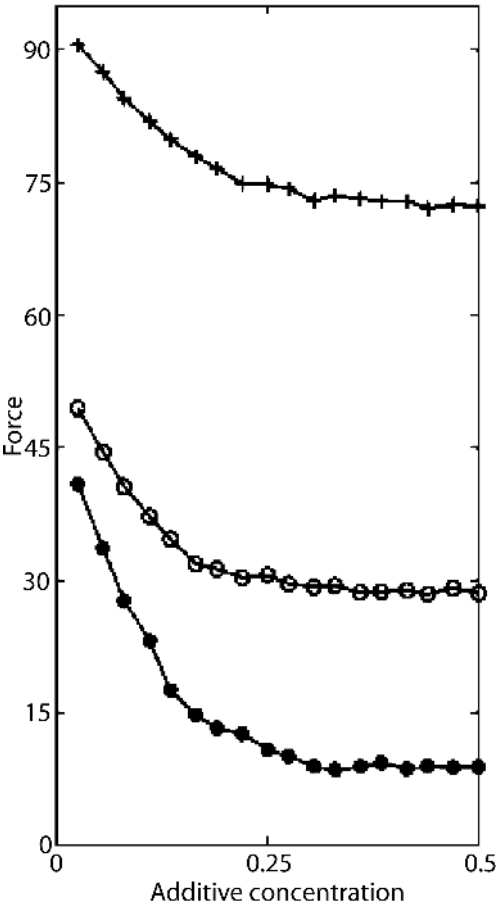


Fig. 8.7. Dependence of the average spring force on the concentration of additives for driving velocities, corresponding to stick-slip (*bold symbols*), intermittent type (*circles*), and sliding (*plus signs*) motions. The force is presented in units of the static friction for the based lubricant [62]

values of driving velocities which correspond to periodic and chaotic stick-slip behaviors typical for low driving velocities, and to steady sliding typical to higher velocities. Figure 8.7 demonstrates a dramatic decrease of friction with the concentration of additives. Indeed there is more than fourfold lowering in friction when fraction of additives was changed from 0 to 0.5. We observe the decrease of friction for all regimes of motion, however the strongest effect is found at low driving velocities where the stick-slip motion occurs.

In order to clarify the mechanism of modifying friction by adding an additive we have performed a detailed study of the geometrical structure of the mixed embedded molecular layer under shear [62, 63], and have established the relationship between geometry and frictional response. Figure 8.8 shows the time series of the spring force and of the ensemble-averaged distance between molecules A and B, calculated for an equal number of A and B molecules. In addition, Fig. 8.9 presents the snapshots of the embedded system observed during stick (a) and sliding (b) states of motion. The instants

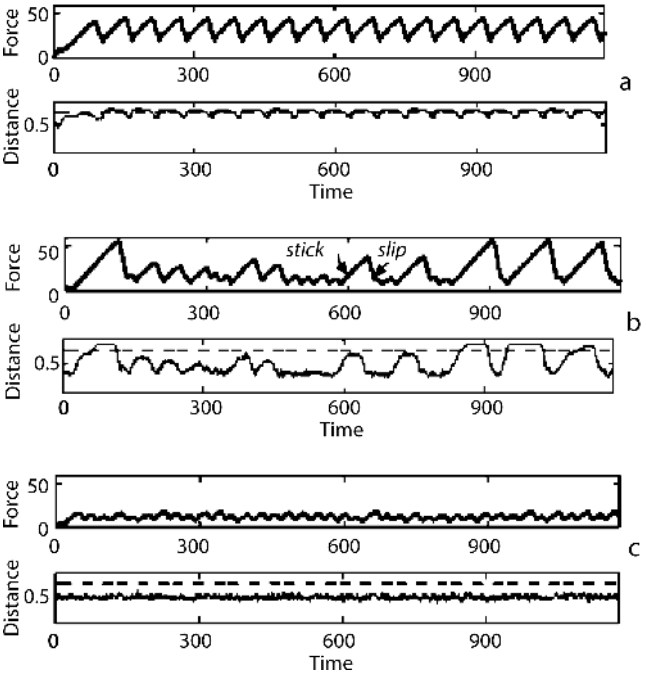


Fig. 8.8. Time dependence of the spring force for **a** weak, **b** intermediate, and **c** strong attractions between A and B molecules. Bottom panels show the corresponding time dependence of the ensemble averaged distances between neighboring A and B molecules [62]

corresponding to the snapshots are marked on the time series of the force by arrows: see Fig. 8.8. The snapshots are complemented by the two-dimensional Fourier transforms of the instantaneous correlation function for the sheared monolayer, which are also shown in the insets to Fig. 8.9.

Our results clearly demonstrate that the embedded monolayer has mostly tetragonal symmetry in the stick state, while a new *hexagonal* symmetry arises during sliding. Here the different molecules group into A–B pairs and form a lattice with well-defined hexagonal symmetry, ignoring essentially the symmetry of underlying potential (see Fig. 8.9). The effect of pair formation is seen in Fig. 8.8, which shows a significant decrease of the A–B distances during sliding compared to the value, which is typical of the tetragonal lattice. Thus, due to the attraction between the base and additive, the molecules "pull" each other out of the minima of the corresponding potentials, thereby effectively decreasing the potential barriers to sliding. This leads to a significant reduction of the friction force in the sliding state and to an increase of the time intervals during which the system spends in motion.

Mixing the embedded layer with additives does not only reduce friction, but it also makes it possible to control the regimes of motion. Namely, tuning

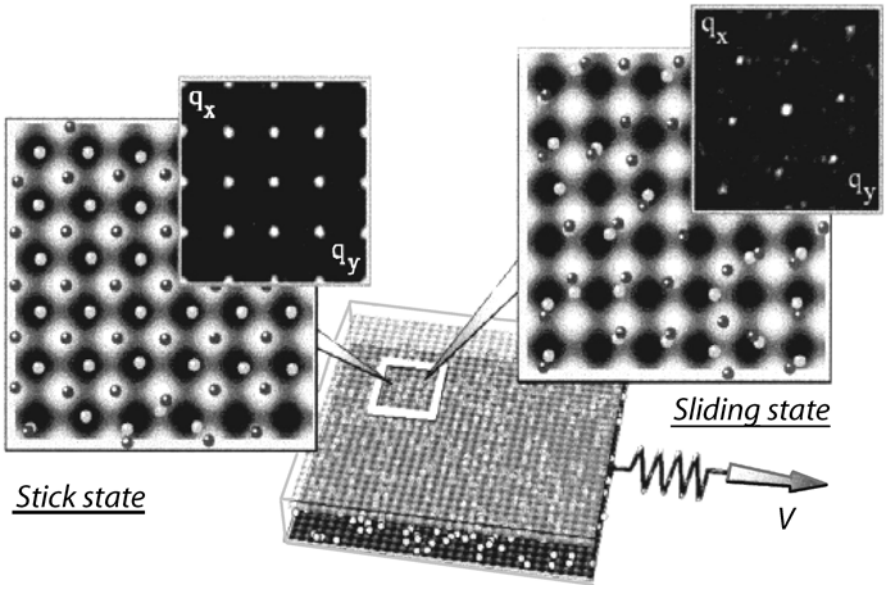


Fig. 8.9. Snapshots of the sheared monolayer showing a tetragonal lattice **a** for the locked states and a hexagonal one **b** for the sliding states. Insets present the corresponding correlation functions [62]

the concentration of additives and/or the attraction between the additive and the base solvent allows to eliminate stick-slip motion and to achieve sliding at low driving velocities.

Acknowledgement. This research was supported by the Israel Science Foundation (grant No. 1116/05) and by the Deutsche Forschungsgemeinschaft (HA 1517/26-1).

References

1. B.N.J. Persson, Sliding Friction, Physical Properties and Applications; Springer Verlag, Berlin, 2000
2. B. Bhushan, J.N. Israelachvili and U. Landman, Nature, 374, 607 (1995).
3. M.G. Rozman, M. Urbakh, J. Klafter and F.-J. Elmer; J. Phys. Chem. B, 102, 7024 (1998).
4. S. Granick, Physics Today, 52, 26 (1999).
5. O. Robbins and M.H. Muser, in Modern Tribology Handbook, ed. B. Bhushan (CRC Press, Boca Raton, 2001).
6. M. Scherge and S.N. Gorb, Biological Micro- And Nanotribology: Nature's Solutions, Springer Verlag, Berlin, 1998.
7. E. Gnecco, R. Bennewitz, T. Gyalog and E. Meyer, J. Phys. Cond. Matter 13, R619 (2001).

8. M.H. Muser, M. Urbakh and M.O. Robbins, Adv. Chem. Phys. 126, 1888 (2003).
9. M. Urbakh, J. Klafter, D. Gourdon and J. Israelachvili, Nature 430, 525 (2004).
10. Micro/Nanotribology and Its Applications, Series E: Applied Sciences, NATO Advanced Sciences Institutes; Edited by B. Bhushan, Kluwer Academic Publishers, Dortrecht, 1997.
11. Tribology and Mechanics of Magnetic Storage Devices, edited by B. Bushan (Springer Verlag, New York, 1996).
12. G. Binnig, C.F. Quate, and Ch. Gerber,. Phys. Rev. Lett. 56, 930 (1996).
13. J. Israelachvili, and G.E. Adams, J. Chem. Soc. Faraday Trans. I 74, 975 (1978).
14. J. Klein, and E. Kumacheva, Science 269, 816 (1995).
15. C. Drummond, N.A. Alcantar, and J.N. Israelachvili, Phys. Rev. E 66, 011705 (2002).
16. A.L. Demirel, and S. Granick, Phys. Rev. Lett. 77, 4330 (1996).
17. C. Drummond, and J.N. Israelachvili, Macromolecules 33, 4910 (2000).
18. C. Drummond, and J.N. Israelachvili, Phys. Rev. E 63, 041506 (2001).
19. D. Gourdon, and J. Israelachvili, Phys. Rev. E 68, 021602 (2003).
20. C. Drummond, J. Israelachvili, and P. Richetti, Phys. Rev. E 67, 066110 (2003).
21. H.W. Hu, G.A. Carson, and S. Granick, Phys. Rev. Lett. 66, 2758 (1991).
22. H., Yoshisawa, Y.-L. Chen and J. Israelachvili,. J. Phys. Chem. 97, 4128 (1993).
23. U. Landman, W.D. Luedtke and E.M. Ringer, Molecular dynamics simulations of adhesive contact formation and friction. In Fundamentals of Friction: Macroscopic and Microscopic Processes; I.L. Singer and H.M. Pollock Eds.; Kluwer, Dordrecht, 1992, pp. 463–510.
24. P.A. Thompson, and M.O. Robbins, Science 250, 792 (1990).
25. J.P. Gao, W.D. Luedtke, and U. Landman, Phys. Rev. Lett. 79, 705 (1997).
26. J.P. Gao, W.D. Luedtke, and U. Landman, J. Phys. Chem. B 102, 5033 (1998).
27. G. He, M.H. Muser and M.O. Robbins, Science, 284, 1650 (1999).
28. N.J. Mosey, M.H. Muser and T.K. Woo, Science, 307 1612 (2005).
29. B.N.J. Persson, J. Chem. Phys. 113, 5477 (2000).
30. A. Ruina, J. Geophys. Res. 88, 10359 (1983).
31. J.M. Carlson, and A.A. Batista, Phys. Rev. E 53, 4153 (1996).
32. M. Urbakh, L. Daikhin, and J. Klafter, Phys. Rev. E. 51 2137 (1995).
33. I. S., Aranson, L.S. Tsimring, and V.M. Vinokur, Phys. Rev. B 65, 125402 (2002).
34. A. Lemaître, Phys. Rev. Lett. 89, 195503 (2002).
35. M. G Rozman, M. Urbakh, and J. Klafter, Phys. Rev. Lett. 77, 683 (1996).
36. M. G Rozman, M. Urbakh, and J. Klafter, Phys. Rev. E 54, 6485 (1996).
37. M.H. Muser, L. Wenning, and M.O. Robbins, Phys. Rev. Lett. 86, 1295 (2001).
38. M.H. Muser, Phys.Rev. Lett. 89, Art. No. 224301 (2002).
39. J.-L. Barrat, and L. Boquet, Faraday Discuss. 112, 1 (1999).
40. P.A. Thompson, M.O. Robbins, and G.S. Grest, Israel J. of Chem. 35, 93 (1995).
41. J.P. Gao, W.D. Luedtke, D. Gourdon, M. Ruths, J. Israelachvili, and U. Landman, J. Phys. Chem. B. 108, 3410 (2004).
42. M.G. Rozman, M. Urbakh and J. Klafter, Europhys. Lett. 39, 183(1997).
43. V. Zaloj, M. Urbakh and J. Klafter, Phys. Rev. Lett. 82, 4823 (1999).
44. M. Porto, V. Zaloj, M. Urbakh. and J. Klafter Tribology Letters, 9, 45 (2000).
45. A.E. Filippov, J. Klafter and M. Urbakh. Phys. Rev. Lett. 87, No 275506 (2001).

46. A.G. Papay, Lubr. Eng. 47, 271(1991).
47. M. Ruths, H. Ohtani, M.L. Greenfield and S. Granick, Tribol. Lett. 6, 207 (1999).
48. M.L. Greenfield and H. Ohtani, Tribol. Lett. 7, 137(1999).
49. M.G. Rozman, M. Urbakh and J. Klafter, Phys. Rev. E 57, 7340 (1998).
50. F.J. Elmer, Phys. Rev. E, 57, R4903 (1998).
51. Y. Braiman, J. Barhen, and V. Protopopescu, Phys. Rev. Lett. 90, Art. No. 094301 (2003).
52. M. Heuberger, C. Drummond, and J.N. Israelachvili, J. Phys. Chem. B 102, 5038 (1998).
53. A. Cochard. L. Bureau and T. Baumberger, Trans. ASME, 70, 220 (2003).
54. A. Socoliuc, E. Gnecco, S. Maier, O. Pfeiffer, A. Baratoff, R. Bennewitz, E. Meyer, Science, 313, 207 (2006).
55. H.D.I. Abarbanel, R. Brown, J.L. Sidorowich, and L.S. Tsimring, Rev. Mod. Phys. 65, 1331 (1993).
56. Z. Tshiprut, A.E. Filippov, and M. Urbakh, Phys. Rev. Lett. 95, 016101 (2005).
57. Hu. Gang, A. Daffertshofer, and H. Haken, Phys. Rev.Lett. 76, 4874 (1996).
58. M. Schreier, P. Reimann, P. Hanggi, and E. Pollak, Europhys. Lett. 44, 416 (1998).
59. H. Risken, The Fokker-Planck Equation (Springer, Berlin, 1996).
60. E. Riedo et al., Phys. Rev. Lett. 91, 084502 (2003).
61. S. Ge et al., Phys. Rev. Lett. 85, 2340 (2000).
62. O. Dudko, A.E. Filippov, J. Klafter. and M. Urbakh. Phys. Rev. B 66, 094114 (2002).
63. O. Dudko, A.E. Filippov, J. Klafter. and M. Urbakh, Tribology Letters, 12, 217 (2002).

9 Experimental Observations of Superlubricity and Thermolubricity

Martin Dienwiebel[1,2] and Joost W. M. Frenken[1]

[1] Kamerlingh Onnes Laboratory, Leiden University, P.O. Box 9504, 2300 RA Leiden, The Netherlands frenken@physics.leidenuniv.nl
[2] IAVF Antriebstechnik AG, Im Schlehert 32, D-76187 Karlsruhe, Germany martin.dienwiebel@iavf.de

9.1 Introduction

In this chapter we discuss manifestations of two effects, which we shall refer to as *superlubricity* and *thermolubricity*. Superlubricity is the phenomenon in which two surfaces slide over each other in dry contact without the atomic-scale instabilities that are thought to be the main source for energy dissipation. Superlubricity can reduce friction forces by orders of magnitude. Thermolubricity is the effect that thermal excitations significantly assist the contact between two bodies in overcoming the energy barriers against sliding, resulting in a reduction of the friction forces of contacts that are not superlubric. We shall argue that together, the two effects may lead to near-frictionless sliding over a wide range of conditions. Although we demonstrate these special effects for nanoscale contacts, we propose that they play a key role in the well-known lubricating properties of some layered materials, such as graphite and molybdenum disulfide, and speculate they may hold a promise for further, low-friction applications.

9.1.1 The Transition to Frictionless Sliding in the One-Dimensional Case

In friction force microscopy (FFM) experiments at the atomic scale the lateral force signals often show a sawtooth-like modulation with the periodicity of the lattice of the substrate over which the tip is being moved. Observations of this type have been made on many different materials, such as graphite [1], mica [2], MoS_2 [3], copper [4], diamond [5, 6], and alkali-halides (NaF, NaCl, KF, KCl, KBr) [7–9]. The general characteristics of these observations are described well by a simple model that has been formulated first by Prandtl [10] and by Tomlinson [11]. Applied to the nanoscale geometry of a friction force microscope the model describes the motion of a point-like tip which is coupled by a spring to a moving support. The tip is in contact with a rigid solid, which is treated as a periodic potential energy surface. If this potential energy landscape has only a single Fourier component, with period a and amplitude

V_0, we can write the lateral force on the tip as

$$\frac{2\pi}{a} V_0 \sin\left(\frac{2\pi}{a} x_\mathrm{t}\right) = k(x_\mathrm{m} - x_\mathrm{t}) \tag{9.1}$$

where k is the stiffness of the spring and x_t and x_m denote the positions of the tip and the support.

The relative strength of the interaction potential with respect to the stiffness of the spring is often expressed in the form of a dimensionless parameter $\gamma \equiv 4\pi^2 V_0/ka^2$. When γ exceeds unity, multiple solutions exist to Eq. (9.1). The tip remains stuck in a metastable equilibrium position until the spring force is large enough to force the tip to rapidly slip to the next equilibrium position, which may again be merely metastable. These two elements, the sticking and the slipping, represent the stick-slip motion, commonly observed in FFM experiments. A tacit assumption in this description is that the excess potential energy that is released during the slip event is irretrievably lost to other degrees of freedom of the system, such as phonons in the substrate and the tip. As a consequence, stick-slip motion implies energy dissipation. By contrast with this scenario for energy dissipation, Eq. (9.1) predicts continuous sliding of the tip over the counter surface for $\gamma < 1$, i. e. when the surface potential is sufficiently weak and the spring is sufficiently stiff. In this case, the lateral force oscillates between negative (against the direction of attempted motion) and positive (forward) values and the average lateral force, i. e. the *friction* force, is zero. The transition from zero dissipation to non-zero friction at $\gamma = 1$, due to the breaking of analyticity, is known as an Aubry transition [12]. The Aubry transition has been investigated primarily in the context of the one-dimensional Frenkel–Kontorova model [12, 13], in which two one-dimensional lattices are sheared over each other. In this model, static friction and the Aubry transition depend strongly on the ratio p/q of the lattice constants of the top and bottom solid.

9.1.2 Superlubricity

The term *superlubricity* has been introduced by Hirano and Shinjo [14]. Originally, it was defined as the extension of the Aubry transition to a two-dimensional geometry and describes the effect that friction can vanish almost completely when two crystalline surfaces slide over each other in dry contact without wear and plastic deformation. This phenomenon was first demonstrated in a quasistatic calculation for rigid crystals with fcc, bcc and hcp symmetry and for various surface orientations [15]. In the two-dimensional case it was found that the frictionless or superlubric regime can be reached for a much wider range of values of γ and they noted that superlubricity should appear for any combination of flat and clean metals when the interaction potential is weak. Hirano and Shinjo concluded that a way to tune the interaction potential experimentally, is to change the commensurability

between two surfaces. Of the theoretical work that has been performed after these first calculations we mention that by Sørensen et al. [16], who used molecular dynamics simulations to investigate friction at $T = 0\,\mathrm{K}$ between flat copper asperities, e. g. 19×19 atoms large, and a copper surface. When the asperity and the surface were both (111) oriented, the sliding did not involve wear. For an aligned contact, regular stick-slip motion was observed with high friction, whereas the friction force vanished when the contact was twisted 16.1° out of registry.

In this chapter, we will use the word "superlubricity", even though it suggests an analogy between the *structural lubricity* at an incommensurate interface and the phenomena of superconductivity and superfluidity [17]. Lately, the term superlubricity has been used by several authors to also indicate other situations with extraordinarily low friction forces, not involving a lattice mismatch effect. Here, we will stick to the original meaning of the word and concentrate on experiments that probe the effect of commensurability on friction.

9.1.3 In Search for Superlubricity

In an early experiment, Hirano *et al.* [18] have employed a tribometer to measure the orientation dependence of friction between mica sheets. They found a friction force of $(8 \cdot 10^{-4}\,\mathrm{N})$ when the orientations of the mica sheets matched. The friction force was reduced by as much as a factor 4 when the crystallographica directions of the mica sheets were misoriented relative to each other. Since in the incommensurate case friction was still relatively strong, either the superlubricity effect was incomplete or other mechanisms were responsible for additional channels of energy dissipation.

As was shown by Ko and Gellman [19], one such additional type of dissipation can be the internal friction that arises when the contact pressure is high enough to cause plastic deformation. These authors measured the friction force as function of the misfit angle between two Ni(100) crystal surfaces using an ultrahigh vacuum (UHV) tribometer and found a lower friction coefficient for 45° and 135° misfit angles than for 0° and other orientations. Although at first sight, this observation seems consistent with superlubricity, the orientational variations were still observed after adsorption of as much as 20 monolayers of ethanol or sulfur on the nickel surfaces prior to contact formation, which made Ko and Gellman conclude that the low friction in certain directions was caused by easy shearing along the preferred slip planes in the bulk. This explanation is consistent with the computer-simulation result obtained by Sørensen et al. [16] for shearing contacts between clean copper surfaces, which revealed that the shear occurred predominantly along the (111) planes, even for (001) oriented surfaces.

In another macroscopic experiment Martin et al. found a remarkably low friction coefficient between clean MoS_2 surfaces after a short sliding distance

using a UHV tribometer [20]. After the experiment MoS_2 flakes were collected and examined with a transmission electron microscope (TEM). The TEM images showed that the flakes were rotated with respect to each other. The authors concluded that the low friction coefficient was due to the incommensurability between the flakes, thus due to superlubricity.

In 1997 Hirano et al. [21] have performed a scanning tunnelling microscope (STM) experiment and claimed the observation of superlubricity in UHV between a tungsten tip and a Si(001) surface. The tungsten tip was first imaged using field electron microscopy, after which it was advanced towards the Si surface until a tunnelling current could be measured. Since an STM is usually not capable of detecting forces, the bending of the tip was monitored optically and translated into a lateral force. When the major crystallographic axes of the two surfaces were aligned, a bending of the tungsten wire over an estimated 100 nm was measured, which was absent when the orientation was rotated over 45°.

9.2 Atomic-Scale Observation of Superlubricity

9.2.1 Commensurability-Dependent Superlubricity between Finite Graphite Surfaces

In this chapter we concentrate on nanotribological experiments, conducted with a dedicated friction force sensor, the Tribolever [22]. This sensor is part of an unconventional friction force microscope [23] that allows quantitative tracking of the forces on the scanning tip in three directions, with a high resolution in the lateral forces, down to 15 pN. The instrument can rotate the sample to change the relative orientation between the tip and sample lattices. Initial measurements with this instrument have been performed on low-grade, highly oriented pyrolytic graphite (HOPG) [24, 25].

Although in some cases, these measurements showed traditional friction loops, with stick-slip character and measurable energy dissipation, in many friction loops the average friction force was very low and the tip was sliding over the graphite surface without stick-slip motion. Also it was found that the variation of the friction force with the normal force was rather weak. These observations strongly suggested that the sliding had been taking place between two graphite surfaces, one being the HOPG substrate, and the other being a small piece of graphite, i. e. a graphite flake, that was attached to the tungsten tip. The difference between the high- and low-friction force loops could then be attributed to the difference in commensurability between the flake and the substrate, high friction corresponding to a fully commensurate contact and low friction to an incommensurate contact.

In order to obtain further support for our interpretation of these initial observations, we have repeated the experiment much more carefully with

a high-quality HOPG sample that had an average grain size of several millimeters. In the experiment we rotated this sample in small steps with respect to the tip. For each orientation, we performed a complete set of friction force spectroscopy measurements for a range of normal forces between +25 nN and pull-off (−22 nN) and a range of sliding directions. We recorded the lateral forces in the X- and the Y-direction of the sensor, from which we reconstructed the average friction force in the sliding direction (for details see [25]).

Figure 9.1 shows lateral force maps and force loops measured in the X-direction for different rotational orientations. A typical force loop is shown in Fig. 9.1d, which was measured at a normal force of 18 nN. The lateral force in Fig. 9.1d displays clearly resolved atomic-scale stick-slip sliding and the average friction force parallel to the sliding direction is 203.3 ± 20 pN. Figure 9.1b,e and c,f show FFM measurements obtained with the graphite substrate rotated +12° and −22° with respect to Fig. 9.1a,d around an axis normal to the surface and parallel to the tip. The rotation by 12° has caused the average friction force to reduce by more than one order of magnitude, to 15 ± 15 pN. Rotating 22° away from the first measurement in the opposite direction also has caused a reduction, to 8^{+16}_{-8} pN, which is equal to zero friction within the detection limit of our instrument. This variation of the friction force with the rotation angle Φ was completely reversible. Notice that the ultra-low lateral forces in Fig. 9.1e–f still exhibit regular variations with the periodicity of the graphite substrate.

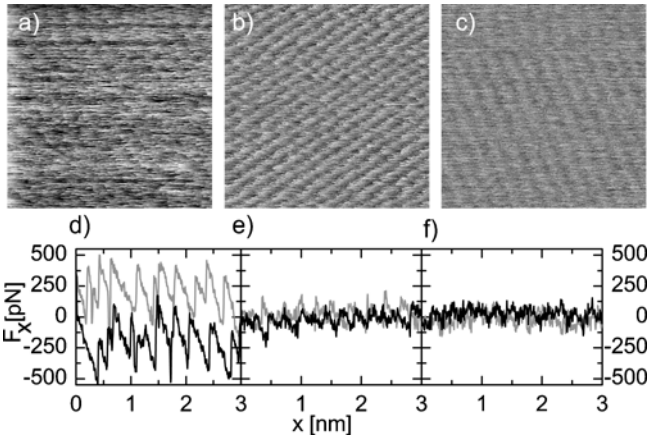


Fig. 9.1. Lateral force images (forward direction) and friction loops measured between a tungsten tip and a graphite substrate. The displayed signals correspond to the X-direction of the Tribolever sensor and rotation angles Φ of the graphite sample of 60° (**a**,**d**), 72° (**b**,**e**) and 38° (**c**,**f**). The normal force between the tip and the substrate amounted to $F_N = 18$ nN in (**a**,**d** and **c**,**f**) and $F_N = 30.1$ nN in **b**,**e**. The grey scales in the force images cover force ranges of **a** 590 pN, **b** 270 pN, and **c** 265 pN. The image size is 3 nm × 3 nm. After [25]

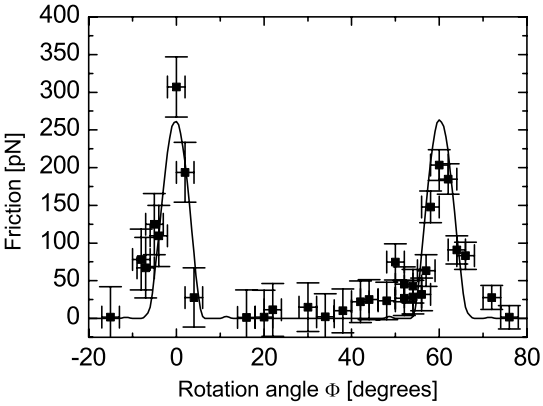


Fig. 9.2. Average friction force between a tungsten tip and a graphite substrate, plotted versus rotation angle Φ of the graphite sample with respect to an axis normal to the sample surface. Two narrow peaks of high friction are observed at 0° and 61°, respectively. Between these peaks a wide angular range with ultra-low friction, close to the detection limit of the instrument, is found. The first peak has a maximum friction force of 306 ± 40 pN, and the second peak has a maximum of 203 ± 20 pN. The curve through the data points shows results from a calculation according to the Tomlinson model for a symmetric 96-atom graphite flake sliding over the graphite surface (see text). After [25]

Figure 9.2 displays the average friction forces measured over a 100° range of substrate rotation angles. We recognize two narrow angular regions with high friction, separated by a wide angular interval with nearly zero friction. The distance between the two friction peaks is $61 \pm 2°$, which corresponds well with the 60° symmetry of individual atomic layers in the graphite lattice. After every 60° rotation, the lattices of the substrate and the graphite flake align and the friction is high. For intermediate angles, the lattices are incommensurate and the friction force is close to zero.

The peak width in Fig. 9.2 can be used to estimate the flake diameter. For finite-size contacts, the cancellation of lateral forces, which causes superlubricity, can be considered complete when the mismatch between the two lattices adds up to one lattice spacing over the diameter of the contact. The mismatch condition provides us with the estimate that $\tan(\Delta\Phi) = 1/D$, where $\Delta\Phi$ is the full width at half maximum of the friction peak, and D is the flake diameter, expressed in lattice spacings. From the widths of the two peaks in Fig. 9.2, of $5.4 \pm 1.0°$ for the first peak and $6.5 \pm 0.8°$ for the second, we estimate that the flake diameter is between 7 and 12 lattice spacings. The contact size was determined more precisely by Verhoeven et al. [26]. He modelled the flake as a rigid, finite lattice, with the hexagonal symmetry of a single layer of graphite. Because the relative positions of the atoms in the N-atom flake $(x_i, y_i, 0)$ with respect to the position $(x_\mathrm{t}, y_\mathrm{t}, z_\mathrm{t})$ of the center

of mass (CM) of the flake were fixed, the flake-surface interaction potential could simply be obtained by the summation over N atomic contributions:

$$V_{\text{int}}(x_{\text{t}} + x_i, y_{\text{t}} + y_i, z_{\text{t}}) = -V_0(z_{\text{t}})[2\cos(b_1(x_{\text{t}} + x_i))\cos(b_2(y_{\text{t}} + y_i)) + \cos(2b_2(y_{\text{t}} + y_i))] + V_1(z_{\text{t}})\,, \tag{9.2}$$

with $b_1 = 2\pi/(0.246\,\text{nm})$ and $b_2 = 2\pi/(0.426\,\text{nm})$. The height-dependent corrugation amplitude per flake atom is given by $V_0(z)$, while $V_1(z)$ indicates the overall, i. e. N-atom position-averaged z-dependence of the interaction, expressed per atom. The amplitude of the summed potential depended strongly on the orientation angle Φ of the flake lattice with respect to the substrate lattice. The flake was coupled to a support by springs in the x- and y-directions (see Fig. 9.3), with which it was dragged through this N-atom interaction potential.

Symmetric flakes of various sizes were considered in the calculation. Each flake was a piece of graphene sheet and had a shape with 60° rotational symmetry. As expected, the friction force was maximal if the misfit angle Φ was zero (or a multiple of 60°). For these orientations, the friction force increased linearly with the number of atoms N in the flake. In order to compare all different flake sizes for the same fixed total interaction between the flake and the surface, the potential amplitude per atom V_0 was lowered with increasing flake size such that the calculated friction force with the flake and substrate in registry was the same for all flakes, namely 265 pN at

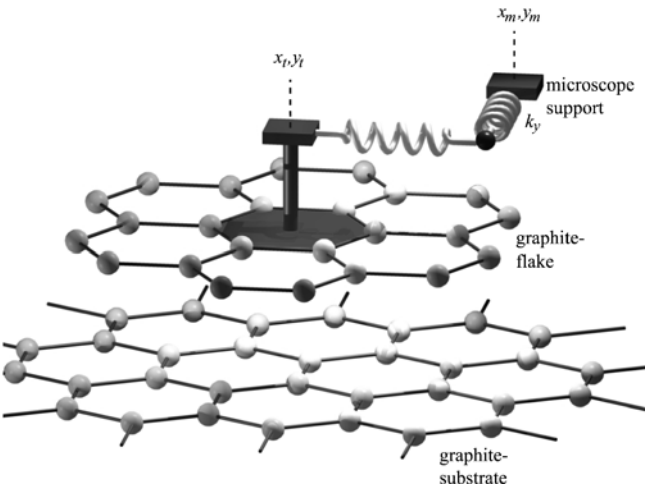


Fig. 9.3. Illustration of the modified Tomlinson model used in our calculations. A rigid graphite flake consisting of N carbon atoms (here $N = 24$) is connected by an x-spring and a y-spring to the support of the microscope. The support is moved in the x-direction. The substrate is modelled as a rigid, infinite, single layer of graphite. From [26]

0° pulling direction, between the values measured experimentally at misfit angles of 0° and 60°.

The effective interaction potential energy surface (PES) for the flake as a whole $V_{\mathrm{int}}^{\mathrm{flake}}$ is shown in Fig. 9.4 a for matching lattices ($\Phi = 0°$) for a flake size of $N = 96$. The small grey areas overlayed on the PES are the flake positions recorded in the 'forward' scan direction, during the 3 nm × 3 nm scan, parallel to the x-axis ($\Theta = 0°$). In Fig. 9.4a, the flake is only found in limited regions, slightly displaced to the upper right with respect to the minima of the PES. Also shown are flake pathways for three separate scan lines. During the scanning process the flake moves continuously through the grey 'sticking' regions, while force is built up in the spring. From the end of such a region it jumps (slips) to the beginning of the next sticking region. When the 96-atom flake is misaligned by 7°, the calculated lateral forces become small and for most trajectories the average lateral force, i. e. the friction force, vanishes completely within the precision of the calculation.

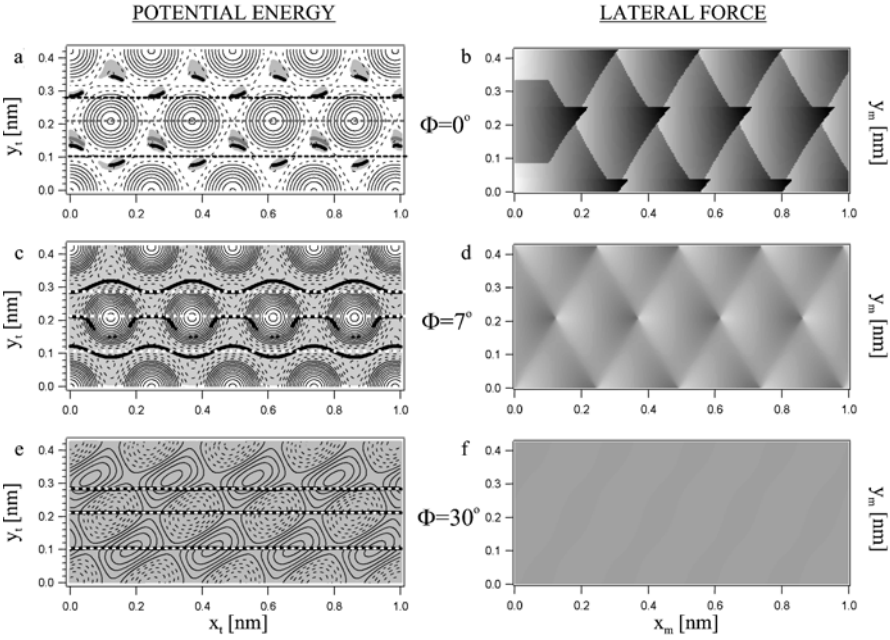


Fig. 9.4. Total potential energy surfaces and lateral force images (1.0 nm × 0.426 nm), calculated in the forward x-direction for a symmetric, 96-atom graphite flake sliding over a graphite substrate, for misfit angles $\Phi = 0°$ (**a**,**b**), $\Phi = 7°$ (**c**,**d**) and $\Phi = 30°$ (**e**,**f**). The *grey scale* in the lateral force images corresponds to the range $[-1.04, 0.63]$ nN. For this range, **b** has maximal contrast. The *grey areas* in the potential energy contour plots denote positions that were visited by the flake. The *black lines* denote pathways of the flake during single scan lines of the support. From [26]

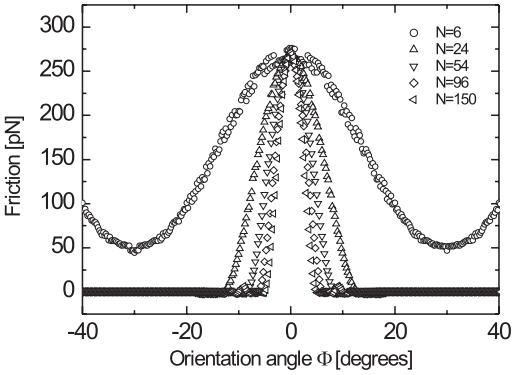


Fig. 9.5. Friction force as a function of the orientation angle for different symmetric graphite flakes ranging in size from 6 to 150 atoms, sliding over a graphite substrate. The potential amplitude per atom has been chosen such that all flakes share the same maximum friction value of 265 pN at $\Phi = 0°$. From [26]

The corrugation of the PES has decreased with respect to the situation at $\Phi = 0°$ and the regions addressed by the flake have merged, indicating that the flake moves continuously through most of the PES.

If the misalignment between the 96-atom flake and the substrate is further increased to 30°, the corrugation of the PES becomes so low that the pathway of the flake through the PES is identical to that of the support, within the precision of the calculation. The flake-graphite contact is now completely superlubric.

Figure 9.5 displays the computed friction force as a function of the misfit angle Φ, for five symmetric flakes with different sizes. We find an angular region with high friction around 0°, repeating every 60° due to the rotational symmetry of the flakes. At intermediate angles, near-zero friction is calculated, except for the 6-atom flake, for which the friction drops to 52 pN. These numerical calculations confirm the simple geometrical estimate, mentioned above, the best fit being produced by a flake with a size of $N = 96$ atoms. Further calculations showed that the shapes of the peaks in Fig. 9.5 also depended on the shapes of the flakes, the best fit to the experimental data being obtained for a symmetric flake, as shown by the curve in Fig. 9.2.

9.2.2 The Role of the Normal Force

As is clear from the description in Sect. 9.1.1, sliding without instabilities is possible also in the case of a single-atom contact or a contact between two commensurate lattices, provided that γ is smaller than 1. The required reduction of V_0 can be achieved by making the normal force sufficiently small, with which the tip is pressed against the substrate. In case of an attractive tip-substrate interaction, this may even require a negative external force, i. e. pulling the tip. This approach has been demonstrated first by Takano and Fujihira [27] and more recently by Socoliuc et al. [28]. In the latter FFM experiment, a silicon tip was scanned over a NaCl surface along the (100) direction in UHV. The observed force loops showed excellent agree-

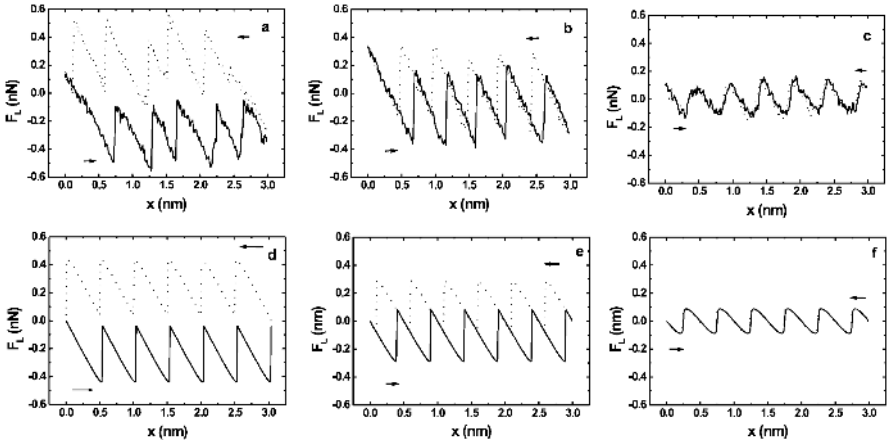


Fig. 9.6. Friction force loops of a Si tip sliding along the (100) direction of NaCl at **a** $F_N = 4.7\,nN$, **b** $F_N = 3.3\,nN$, and **c** $F_N = -0,47\,nN$. Prandtl–Tomlinson calculation with **d** $\gamma = 5$, **e** $\gamma = 3$, and **f** $\gamma = 1$. Reprinted from [28] with permission by A. Socoliuc and E. Gnecco

ment with the transition to frictionless sliding that is predicted by the one-dimensional Prandtl–Tomlinson model. As γ approached unity at a normal force of $F_N = -0.47\,nN$, the area enclosed in the friction loop and, thus, the energy dissipated in a cycle reduced to zero and the tip was observed to slide over the NaCl surface without stick-slip motion. At that point the lateral force still showed slightly distorted sinusoidal variations with the periodicity of the surface lattice, but the average force was zero (Fig. 9.6c).

Following the original definition by Hirano and Shinjo [14], we should not refer to this form of near-frictionless sliding at low (or even negative) normal forces as "superlubricity", since it does not involve the cancellation between lateral forces on individual atoms in the contact resulting from a mismatch between the two contacting surface lattices. By contrast, in the case of superlubricity, the normal force can be made surprisingly high [25]. We expect that the natural limit in the normal force, or rather contact pressure, will be that a new energy dissipation channel is introduced when the contact pressure is made high enough to induce noticeable lateral elastic deformations in the two contacting surfaces. In the case of such deformations, we can no longer describe the situation as that of two translating, rigid bodies [17]. In addition, the contact will then carry a rapidly shifting deformation pattern that one may view as a lattice of dislocation lines. Moving this pattern involves the combination of "breaking" atoms out of registry on one side of these lines and the "popping" of atoms back into registry on the other side. As these changes in atomic positions are no longer rigidly connected, much of the energy released on one side will not be re-invested on the other side and will be lost in the form of heat (phonons).

9.3 The Role of Temperature

Our discussion in the previous sections has been presented in terms of the classical mechanics of the sliding system, combined with the assumption of instantaneous loss of the excess energy during each slip event. These elements form the basis of the Tomlinson model that was used to fit the data in Fig. 9.2. One of the obvious simplifications in the Tomlinson model is the complete absence of effects due to the spontaneous thermal excitations that are present at finite temperatures. Such effects have been anticipated already by Prandtl in 1928 [10]. Here, we summarize how they lower the average friction force and may lead to a strongly modified type of sliding motion.

9.3.1 Weak Thermal Effects

It is easy to see that thermal excitations will assist the tip in overcoming the energy barrier for sliding from one well in the potential energy surface to the next. One might expect this phenomenon to become noticeable only when the amplitude of the potential V_0 (see Eqs. (9.1) and (9.2)) and, thus, the energy barriers would be limited to only a few times the thermal energy $k_{\mathrm{B}}T$. However, one should realize that the tip moves in the *combined* potential of its interaction with the surface and its interaction with the spring that connects it to the moving support. Within the Tomlinson model the tip remains stuck in one well of this combined potential until the spring is sufficiently extended that the energy barrier to the next well vanishes, at which point the system is unstable and the tip necessarily slips into the next well. This means that even when the barrier to the next well starts out at a high value, it decreases continuously to zero while the spring is being stretched. This implies that thermal excitations will *always* play a role, since the "stick" part of the stick-slip cycle always contains a portion during which the barrier is sufficiently low with respect to $k_{\mathrm{B}}T$ that it can be overcome by a thermally activated jump. This so-called pre-critical jump somewhat lowers the maximum lateral force that is exerted by the spring and thus reduces the friction force. The effect depends weakly on the sliding velocity because at lower velocities there is time for more attempts of the system to thermally overcome each barrier, which should make friction reduce more. It is easy to provide an estimate of this reduction in the friction force. Due to the exponential nature of the thermal excitations there is a range of sliding velocities over which the friction force is expected to increase proportionally with the logarithm of the sliding velocity v [29–31] or with $(\log v)^{2/3}$ [32]. Such a weak dependence of friction on sliding velocity has indeed been observed experimentally [31, 33, 34].

9.3.2 Strong Thermal Effects: Thermolubricity

As the measurements in Fig. 9.2 show, the rotation angle Φ of the graphite flake with respect to the graphite substrate determines the amplitude of the

potential V_0 and can tune it anywhere between a high value of γ, at which there is strong friction, and low values $\gamma < 1$ that result in superlubricity. If the spring coefficient of the sensor k is low, as was the case in the measurements of Fig. 9.2, the transition to superlubricity takes place at a comparatively low value of the potential amplitude V_0. For example, at a typical (low) spring coefficient of $k = 1\,\mathrm{N/m}$ and a typical lattice constant of $a = 0.25\,\mathrm{nm}$ the transition to superlubricity is at a potential amplitude of only 10 meV, which is lower than the thermal energy at room temperature of 25 meV. This implies that for a contact that is close to superlubricity the role of thermal excitations will be much more dramatic than that described in the previous section. Rather than to merely facilitate the jumps that were bound to happen anyway (pre-critical jumps), thermal excitations are now sufficient to efficiently promote the system over all barriers, both to the next well(s) and to the previous one(s) [29]. As a result, the tip will conduct a random walk over the surface, its average position following the slow translation of the support. The stochastic nature of this driven diffusion of the tip is characterized by rapid force variations that replace the periodic stick-slip character, typical for higher γ values. The average lateral force, i. e. the observed friction force, is strongly reduced by these thermal jumps. It is this behavior for which we have proposed the term *thermolubricity* [35].

Figure 9.7 shows two selected force loops measured between a graphite substrate and a tungsten tip (dressed with a graphite flake) for two relative orientations, corresponding to two different values of γ. While the force loop for the higher γ still displays recognizable stick-slip behavior, the force

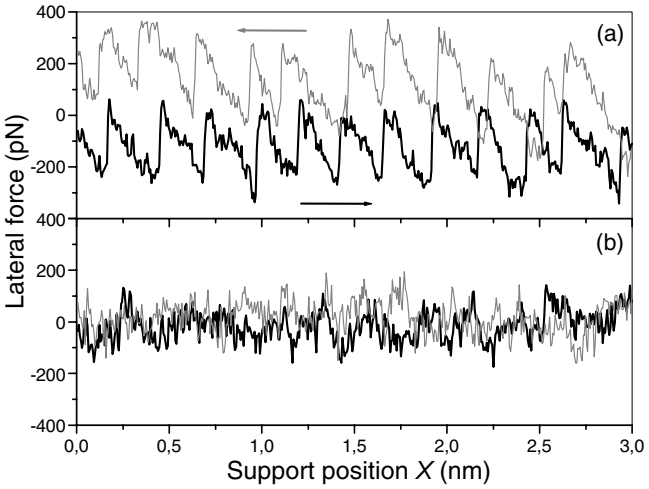


Fig. 9.7. Characteristic lateral force loops measured with a tungsten tip (with a graphite flake) on a graphite substrate at two different relative orientations, corresponding to **a** $\gamma = 5.0$ and **b** $\gamma = 2.5$. After [35]

variations at the lower γ value are almost completely stochastic. Note that the average friction force of the lower loop is close to zero, although γ is still well above unity, i. e. the system is not yet superlubric. These qualitative features agree well with the thermolubricity scenario. A stronger test is a quantitative confrontation of these observations with numerical calculations of thermolubricity. In these calculations we describe the surface potential as a one-dimensional sine function (Eq. (9.1)) to which we add the potential due to the interaction with the spring, $\frac{1}{2}k(x_\mathrm{t} - x_\mathrm{m})^2$. If $\gamma < 1$ this combined potential shows only a single minimum for every position x_m of the support and the sliding is frictionless (superlubricity). When $\gamma > 1$ the combined potential shows several wells. Rather than to calculate individual trajectories or concentrate on average behavior, we describe the process in terms of the probabilities p_i for the tip to reside in each well i at every point in time. These probabilities evolve according to a simple set of continuity equations of the type:

$$v\frac{\mathrm{d}p_i}{\mathrm{d}x_\mathrm{m}} = -(r_i^+ + r_i^-)p_i + r_{i-1}^+ p_{i-1} + r_{i+1}^+ p_{i+1} \,, \tag{9.3}$$

where the role of time is played by the coordinate of the support $x_\mathrm{m} = vt$. The rates r_i^+ and r_i^- of jumps from well i to the right and to the left are calculated according to the Arrhenius law:

$$r_i^\pm = r_0 \exp(-\frac{\Delta E_i^\pm}{k_\mathrm{B}T}) \,. \tag{9.4}$$

Here, $\Delta E_i^\pm$ are the energy barriers from well i to the next well and to the preceding one. The locations and heights of these barriers depend on the support position x_m. The prefactor r_0 is the frequency of attempted jumps, which is treated as a fitting parameter (see below). Equation (9.3) can be solved analytically in the limit of very strong thermolubricity, i. e. low velocities v, low amplitudes of the potential V_0 and/or high temperatures T. For other conditions the probabilities need to be evaluated numerically.

Figure 9.8 demonstrates that for $\gamma > 1$ the experiments presented before, probing the lateral forces between a graphite-decorated tungsten tip and a graphite substrate, clearly exhibit thermolubricity. For each value of the relative strength of the potential γ the experimental data fall significantly below the dashed curve, which is the friction force expected according to the Tomlinson model. The full curves show the results obtained from Eq. (9.3) for the values of T, v, a, and k taken from the experiment and for four different values of the attempt frequency r_0. Like the experiment, the theoretical curves fall below the curve for the Tomlinson model. The second curve, for $r_0 = 1.6\,\mathrm{kHz}$, provides an excellent fit to the experiment. This frequency is in the order of the eigenfrequency of the employed Tribolever sensor. Although this may not seem very surprising, further work will be necessary to resolve why the much higher vibration frequencies of the apex of the tip seem not to dominate the value of r_0.

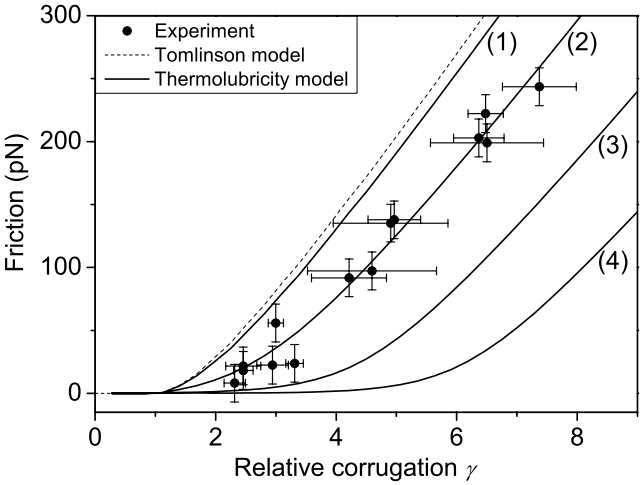


Fig. 9.8. Friction force as a function of relative surface corrugation γ. Experiments for a graphite substrate and a tungsten tip with a graphite flake are compared with numerical solutions of Eq. (9.3) for $v = 30\,\mathrm{nm/s}, a = 0.25\,\mathrm{nm}, k = 1.8\,\mathrm{N/m}$ and $T = 290\,\mathrm{K}$ and for $v/ar_0 = 7.53 \times 10^{-n}$ with (from left to right) $n = 1, 2, 3, 4$. Experiment and calculations all fall well below the dashed curve, which shows the corresponding result from the Tomlinson model, i. e. the friction force in absence of thermal excitations. After [35]

9.4 Other Manifestations of Superlubricity and Thermolubricity

9.4.1 Lubrication by Graphite and other Lamellar Solids

Graphite is a popular solid lubricant that is usually applied in the form of flaky powder. Several other layered materials also show excellent lubricating properties, for example MoS_2 and Ti_3SiC_2. Traditionally, the good lubrication by these materials is ascribed to the weak interaction between adjacent layers [36]. However, the intuitive idea that this would lead to "easy shear" cannot be correct, since it would still require the simultaneous rupture of all bonds in a plane, which involves a tremendously high energy, even for the weakly interacting layers in graphite [37]. Based on the lateral force measurements discussed in this chapter it seems natural to add the extra element of superlubricity and possibly also that of thermolubricity to the low-friction scenario of graphite. As argued above, the "easy shear" cannot take place *within* individual pieces of graphite, but it can occur *between* flakes of graphite since in a lubrication film flakes will be oriented randomly with respect to each other, which introduces the mismatch required for superlubricity and thermolubricity for almost all graphite-graphite contacts. It is known that when metals are lubricated by graphite under ambient conditions, the oxide

layers on the metals tend to fix some of the graphite to the metal, leading to a smooth transfer film [38], which would indeed concentrate most of the shear motion within the film of graphite flakes, rather than between the graphite and the metal surfaces. In a practical sliding geometry, the normal and shear forces will not be carried by a single, macroscopic flake-flake contact but it will be distributed over a large ensemble of simultaneous microcontacts between flakes, most of which will be slippery due to superlubricity or thermolubricity, while only a small fraction will be in temporary registry.

Strong support for the suspicion that it is the slipperiness of misoriented flakes that makes the friction low when macroscopic contacts are lubricated by layered solids comes from transmission electron microscopy observations by Martin et al. on MoS_2 [20]. TEM inspection of wear particles harvested from a MoS_2 lubrication film that had been exposed to sliding friction revealed Moiré patterns characteristic for superimposed flakes that were rotated with respect to each other around their c-axis.

9.4.2 Lubrication by Diamond-like Carbon and related Coatings

We also briefly speculate about the extremely good lubricating properties of diamond-like carbon (DLC) coatings and related, carbon containing materials and propose that it is again the superlubricity and thermolubricity of graphite that might be responsible. During the running-in phase, some of the DLC coating may be transformed into graphite and remain loosely bound to the DLC film. A relatively small amount of graphitized material should be sufficient to decorate all asperities and thus dominate the shear response. The main role of the DLC film would thus be to provide the material (carbon) and the conditions (e. g. through its hardness) necessary to produce small amounts of graphite. An essential element of this scenario is that it is self-terminating. The high friction forces at the beginning of run-in provide local pressures and temperatures that should be high enough to shear off carbon from the DLC film and graphitize it. Once it has been formed, the graphite dramatically reduces friction, so that the local shear stresses on the DLC film are too low to continue wearing off the film and graphitizing it. It has indeed been demonstrated that under sliding conditions, a graphitized tribolayer is formed on top of diamond-like carbon (DLC) coatings, which goes hand in hand with the decrease of the friction coefficient during run-in [39].

9.4.3 Lubrication by Fullerenes and Carbon Nanotubes

Fullerenes show very low friction when they are oriented in an incommensurate fashion. Miura et al. [40] have constructed a molecular ball bearing by placing a monolayer of C_{60} molecules between two graphite sheets of $1\,mm^2$. When they moved the upper graphite flake with an FFM tip, they observed that the forward and backward traces were identical and no energy

dissipation was measurable up to normal loads of more than 100 nN. Surprisingly the lateral force traces still showed sawtooth-like force variations which are typical for instabilities and thus significant friction should be expected. Therefore the Prandtl–Tomlinson model apparently is not able to describe the low friction behavior of this system. Falvo et al. [41, 42] manipulated carbon nanotubes (CNTs) on a graphite surface using the tip of an FFM. They observed that the CNTs changed from sliding to rolling motion, depending on the orientation of the tubes on the substrate. The rolling motion of the CNTs in the case of a commensurate contact was found to require a higher lateral force than the sliding motion of the CNTs in the case of an incommensurate contact. Cumings and Zettl [43] have used a TEM to estimate the friction force between two tubes of a multiwall carbon nanotube (MWNT) in the direction of the long axis. They pulled the core tube out of the outer tube and calculated the friction force from the retraction time, which was below $1.5 \cdot 10^{-5}$ nN/atom.

Nanoparticles form another interesting model system to study the friction between two finite surfaces. By deposition of Sb particles on graphite and MoS_2, Ritter et al. [44] have created incommensurate contacts of various sizes. At a certain size the particles undergo a transition from amorphous to crystalline and at that point an increase in friction was observed. Although their friction was low, the amorphous particles were found not to be completely superlubric in air [45]. For future applications in superlubric meso- or macroscopic mechanical systems it is very important to learn why superlubricity is not seen in this case.

9.5 Concluding Remarks

In this chapter we have reviewed friction force microscopy experiments on extremely low friction and their interpretation in terms of lattice mismatch effects (superlubricity) and effects due to thermal excitations (thermolubricity). The prospect of exploiting these effects in practical applications, for example in nano- and micro-electro-mechanical systems (NEMS and MEMS) is exciting. The examples of excellent lubrication by graphite and other layered materials and by modern, carbon containing coatings such as DLC strongly suggest that these effects can indeed manifest themselves on macroscopic length scales and under realistic loading conditions.

It is evident that many questions have not been addressed in the studies reviewed here on nanoscale contacts. Similarly, many aspects remain to be explored on the way from a single, nano-contact under modest loading pressures to a large ensemble of larger contacts with possibly higher loading. An example is the effect of the load-dependent elastic deformations that may be expected to undermine the slipperiness, as mentioned in Sect. 9.2.2.

Another important issue is whether we can evoke superlubricity- and thermolubricity-based slipperiness also on other materials than layered solids

such as graphite and related materials. Simulations for copper sliding over copper [16] suggest that there is no fundamental reason against this possibility, which is therefore demanding to be examined.

Acknowledgement. The authors are grateful to a large number of people for their valuable contributions to the work reviewed in this chapter. In particular we mention J. A. Heimberg for the design and construction of the friction force microscope, K.B. Jinesh and N. Pradeep for performing part of the experiments and analysis, S. Yu. Krylov for setting up the theory of thermolubricity and G. S. Verhoeven and H. Valk for numerical calculations of superlubricity and thermolubricity.

The work presented in Sects. 9.2.1 and 9.3.2 is part of the research program of the "Stichting voor Fundamenteel Onderzoek der Materie (FOM)" and was made possible by financial support of the "Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO)".

References

1. C. M. Mate, G. M. McClelland, R. Erlandsson, and S. Chiang, Phys. Rev. Lett. **59**, 1942 (1987).
2. S. Fujisawa et al., J. Vac. Sci. Technol. B **12**, 1635 (1994).
3. S. Morita, S. Fujisawa, and Y. Sugawara, Surf. Sci Rep. **23**, 1 (1996).
4. R. Bennewitz et al., Phys. Rev. B **60**, R11301 (1999).
5. G. J. Germann et al., J. Appl. Phys. **73**, 163 (1993).
6. R. J. A. van den Oetelaar and C. F. J. Flipse, Surf. Sci. Lett. **384**, L828 (1997).
7. R. W. Carpick, Q. Dai, D. F. Ogletree, and M. Salmeron, Tribol. Lett. **5**, 91 (1998).
8. L. Howald et al., J. Vac. Sci. Technol. B **12**, 2227 (1994).
9. R. Lüthi et al., J. Vac. Sci. Technol. B **14**, 1280 (1996).
10. L. Prandtl, ZS f. angew. Math. u. Mech. **8**, 85 (1928).
11. G. A. Tomlinson, Phil. Mag. S.7. **7**, 905 (1929).
12. S. Aubry, in *Solitons and Condensed Matter Physics (A.R. Bishop and T. Schneider, eds.)* (Springer-Verlag, Berlin, 1979), Chap. The New Concept by Breaking of Analyticity in a Crystallographic Model, p. 264.
13. S. Aubry, Physica D **7**, 240 (1983).
14. K. Shinjo and M. Hirano, Surf. Sci. **283**, 473 (1993).
15. M. Hirano and K. Shinjo, Phys. Rev. B **41**, 11837 (1990).
16. M. R. Sørensen, K. W. Jacobsen, and P. Stoltze, Phys. Rev. B **53**, 2101 (1996).
17. M. H. Müser, Europhys. Lett. **66**, 97 (2004).
18. M. Hirano, K. Shinjo, R. Kaneko, and Y. Murata, Phys. Rev. Lett. **67**, 2642 (1991).
19. J. S. Ko and A. J. Gellman, Langmuir **16**, 8343 (2000).
20. J. M. Martin, C. Donnet, T. LeMogne, and T. Epicier, Phys. Rev. B **48**, 10583 (1993).
21. M. Hirano, K. Shinjo, R. Kaneko, and Y. Murata, Phys. Rev. Lett. **78**, 1448 (1997).
22. T. Zijlstra et al., Sensors and Actuators A **84**, 18 (2000).
23. M. Dienwiebel et al., Rev. Sci. Instrum. **76**, 043704 (2005).

24. M. Dienwiebel et al., Phys. Rev. Lett. **92**, 126101 (2004).
25. M. Dienwiebel et al., Surf. Sci. **576**, 197 (2005).
26. G. S. Verhoeven, M. Dienwiebel, and J. W. M. Frenken, Phys. Rev. B **70**, 165418 (2004).
27. H. Takano and M. Fujihira, J. Vac. Sci. Technol. B **14**, 1272 (1996).
28. A. Socoliuc, R. Bennewitz, E. Gnecco, and E. Meyer, Phys. Rev. Lett. **92**, 134301 (2004).
29. T. Baumberger, P. Berthoud, and C. Caroli, Phys. Rev. B **60**, 3928 (1999).
30. C. Caroli and P. Nozières, in *Physics of Sliding Friction*, Vol. 311 of *NATO ASI Series E: Applied Sciences*, edited by B. N. J. Persson and E. Tosatti (Kluwer, Dordrecht, 1996), p. 27.
31. E. Gnecco et al., Phys. Rev. Lett. **84**, 1172 (2000).
32. Y. Sang, M. Dubé, and M. Grant, Phys. Rev. Lett. **87**, 174301 (2001).
33. E. Riedo et al., Phys. Rev. Lett. **91**, 084502 (2003).
34. K. B. Jinesh and J. W. M. Frenken, to be published.
35. S. Y. Krylov et al., Phys. Rev. E **71**, 065101(R) (2005).
36. W.Bragg, in *An Introduction to Crystal Analysis* (G.Bell and Sons Ltd., London, 1928), p. 64.
37. G. I. Finch, Proc. Phys. Soc. A **63**, 785 (1950).
38. F. P. Bowden and D. Tabor, *The Friction and Lubrication of Solids* (Clarendon Press, Oxford, 1964), Vol. 2.
39. Y. Liu, A. Erdemir, and E. I. Meletis, Surf. Coat. Technol. **86-87**, 564 (1996).
40. K. Miura, S. Kamiya, and N. Sasaki, Phys. Rev. Lett. **90**, 055509 (2003).
41. M. R. Falvo et al., Nature **397**, 236 (1999).
42. M. R. Falvo, J. Steele, R. M. Taylor II, and R. Superfine, Phys. Rev. B **62**, R10665 (2000).
43. J. Cumings and A. Zettl, Science **289**, 602 (2000).
44. C. Ritter, M. Heyde, B. Stegemann, and K. Rademann, Phys. Rev. B **71**, 085405 (2005).
45. C. Ritter, private communication.

10 Theoretical Aspects of Superlubricity

Martin H. Müser

Department of Applied Mathematics, University of Western Ontario, London, ON, Canada N6A 5A7 mmuser@uwo.ca

10.1 Introduction

Everyday experience tells us that a finite threshold force, namely the static friction force, F_s, has to be overcome whenever we want to initiate lateral motion of one solid body relative to another. In order to maintain the motion, a force equal to or larger than the kinetic friction, F_k, has to be applied, however small the relative velocity. Conversely, when attempting to drag a solid through a fluid medium, there is no threshold to initiate or another one to maintain the sliding motion. Instead, one only needs to counteract friction forces linear in the (final) sliding velocity v_0. It came as a surprise when Hirano and Shinjo suggested that static friction between solids in ultra-high vacuum may essentially disappear as well [1, 2]. While their suggestion of super-low static friction, or superlubricity, contradicts our intuition based on everyday experience, it does not necessarily contradict classical mechanics. If the slider and substrate have homogeneous surfaces and wear and plastic deformation are negligible, then one may expect the same (free) energy at the beginning of the sliding process as at its end, because of translational invariance. Consequently, no work would have to be done on the system implying the possibility of very low friction. Not only the fundamental or theoretical considerations evolving around superlubricity are intriguing, but even more so the possibilities to exploit it technologically. For instance, in small mechanical devices, friction-induced wear and heat often cause the main limits to further miniaturization. These limits could be overcome if superlubric surfaces could be designed. Achieving technological exploitation of superlubricity will certainly benefit from a thorough theoretical understanding.

One of the key ingredient of superlubricity is that all atoms within each solid must move in one correlated Larkin domain. Ideally, the solids move like rigid plates and have no energy-dissipating boundary lubricant immersed between them. Heat could not be produced in this scenario, because there would be no random motion of atoms, and unless the surface corrugation of both solids matched by design or by chance, there could be no interlocking and hence no macroscopic static friction. Of course, solids do not behave like rigid plates but can be deformed. This makes it possible for two solids to interlock at the microscopic scale. It is interesting to note that Coulomb discussed elastic deformation as a potential scenario to lead to solid friction: [3]

"... ou bien il faut supposer que les molécules des surfaces des deux plans en contact contracte, par leur proximeté, une cohérence qu'il faut vaincre pour produire le movement."[1] Today's pursuit for superlubricity still very much evolves around the question of how one can avoid the "coherence" of atoms or asperities in contacts between two surfaces. During the beginning of the last century, works by Prandtl [4] and Tomlinson [5] showed on a fundamental level how the "contraction" of surfaces may not only induce interlocking and hence static friction but also kinetic friction: Whenever individual atoms become mechanically unstable during sliding, they will "pop" into the next available potential energy minimum. In this process, lattice vibrations will be excited, whose energy will be irreversibly lost as heat. The argument applies to collective degrees of freedom as well [6]. The intriguing conclusion from Hirano and Shinjo's work is that the energy-dissipating "pops" do not have to be expected when two atomically smooth surfaces are in contact.

Since Hirano and Shinjo's pioneering works in the early 1990s, a lot of progress has been made on superlubricity by both theoreticians and experimentalists. The chapter by Dienwiebel and Joost in this book will focus on the experimental aspects. This chapter will be concerned with an overview of the theoretical concepts related to superlubricity. We will try to describe these without any mathematical formalism first and then sketch the quantitative analysis. We will also review some key computational studies of superlubric interfaces. One of the important messages resulting from many realistic computer simulations is that quite a few materials, or pairs of materials, may be good candidates to show superlow friction. The main limitation in real life appears to be to produce surfaces that are as flat and as flawless as those in the virtual computer experiments.

The remainder of this chapter is organized as follows: In Sect. 10.2, we will first present some purely analytical concepts of superlubricity. This section will open with an attempt to define superlubricity, because there is no generally accepted definition as yet. The definition is based on measurable quantities and will thus hopefully serve theoreticians and experimentalists alike. The theory section, however, will mainly be concerned with the analysis of two rigid solid bodies in contact and how elastic instabilities (as best described in the Prandtl–Tomlinson model) and long-range elastic deformations (as most easily introduced in the Frenkel–Kontorova model) alter the rigid-body picture, including a short treatment of drag forces in the superlubric regime. Section 10.3 gives an overview of computer simulations relevant to the topic. The section is subdivided according to the type of materials studied by computer experiments. The analyzed systems range from "generic materials" such as Lennard Jonesium and "finite-element materials" with excluded volume interaction to "realistic materials", including bare metals and

[1] ... or one has to assume that the surface molecules of the two opposing planes contract due to their proximity into a coherence, which needs to be overcome to produce motion.

hydrogen-terminated surfaces. One of the advantages of computer simulations is that geometries can be designed at will, allowing one to study effects of surface alignment, roughness at the nanometer scale, contamination, etc. with more rigor than with analytical or even experimental methods. Conclusions will be drawn in Sect. 10.4.

10.2 Theory

The microscopic justification for the possibility of the virtual absence of lateral forces between solids can be supported by the following argument: There are as many bumps (or atoms) in the substrate pushing the slider to the right as there are surface irregularities in the substrate pushing the slider to the left. Hence, statistically speaking, there is the possibility of an almost perfect annihilation of lateral forces. In this section, we will investigate the nature of this process in more detail and explore how elasticity affects this picture. First, elasticity will lead to dissipative forces which behave similarly to the drag forces that a solid experiences when moving through a fluid. Thus, the interface may remain superlubric despite the abilities of the solids to deform. However, once the solids become sufficiently compliant, elasticity will destroy superlubricity and induce finite pinning forces. The concept of drag forces due to elasticity in the solids will be discussed within the framework of linear-response theory. As for the breaking of superlubricity, particular attention will be given to the Prandtl–Tomlinson model and the Frenkel–Kontorova model. We will also be concerned with the role of thermal fluctuations and mechanisms other than elastic instabilities that can lead to instabilities and thus to Coulomb-type friction.

10.2.1 Definition of Superlubricity

Defining superlubricity is not an easy task. There is no formation of a Bose Einstein condensate or of Cooper pairs as is the case in superfluids or superconductors. The mobility of solids does not diverge to infinity when they are superlubric, unlike the mobility of helium atoms in a superfluid or that of electrons in a superconductor. Also, the friction forces between two solids in sliding motion remain finite at finite velocities v_0. As a side comment, one can note that the friction between two incommensurate *supercrystals* would probably disappear completely. Because, in general, there is no true divergence of the mobility of the solids, it will be necessary to use an ad-hoc definition of superlubricity such as the kinetic friction coefficient $\mu_\mathrm{k} = F_\mathrm{k}/L$ must be less than, say, 10^{-3}.

Although μ_k may appear almost constant in many cases for a large range of loads L and sliding velocities v_0, μ_k will ultimately be a function of both L and v_0, e. g., μ_k may vanish at astronomically small values of v_0 due to thermal activation. Therefore, it will be necessary to make a second ad-hoc

assumption on how to choose L and v_0. The (local) load should be high, but not as high as to plastically deform the system, e. g., in the order of 10% of the indentation hardness of the softer material in contact. v_0 should also be sufficiently high so that the system cannot relax stress through creep or thermal relaxation. However, v_0 should certainly be small compared to the velocity of sound c, e. g., less than in the order of $10^{-3}c$.

It is important to keep in mind that the kinetic friction coefficient μ_k and the static friction coefficient μ_s are not necessarily correlated. Kinetic friction is due to hysteresis while static friction is due to energy barriers. In this chapter we will encounter some model systems that have finite energy barriers and hence exhibit finite static friction, although hysteresis effects are negligible leading to an absence of kinetic friction. Given these dissipation-free systems, one may as well define superlubricity as the absence of (mechanical) hysteresis. For example, significant static friction was observed with an atomic force microscope in a recent experiment by Socoliuc et al., [7], yet, kinetic friction was very small. More details of that study and other experiments will be given in the chapter by Dienwiebel and Joost.

10.2.2 Cancellation of Lateral Forces. Symmetry Considerations

When studying lateral forces exerted between two surfaces, symmetry is crucial. For instance, as will be shown in more formal detail further below, two flat, rigid, commensurate solids, i. e., solids that systematically share a common periodicity, have a friction coefficient that is independent of the area of contact, or, more generally speaking, independent of the number N of atoms in direct contact with the substrate. The reason is that the forces exerted on the individual atoms in the slider add up in a systematic fashion in commensurate interfaces. Conversely, for flat, rigid, incommensurate solids there is a systematic annihilation of lateral forces similar to the destructive interference in optics. This leads to friction coefficients that vanish linearly with the area of contact. For disordered surfaces, lateral forces have random direction or random sign. Consequently they add up stochastically so that the absolute lateral force only grows with $\sqrt{N}$ resulting in $\mu_\mathrm{s} \propto 1/\sqrt{N}$, as we assume that the load L increases linearly with system size N. The geometric arguments presented here are generally known among theoreticians, though it is not clear who presented them first and no original literature is known to the author. It may yet be beneficial to outline the concepts in more detail than in previous reviews [8, 9].

Let us start by making a minimalist model of two flat, crystalline solids in contact. It consists of a rigid substrate with lattice constant b and a slider with lattice constant a. A sketch is shown in Fig. 10.1. The surface atoms in the slider are connected to their lattice sites by springs of stiffness k_1 and to their neighbors by springs of stiffness k_2. To explore the effects of symmetry to a zero degree order, we will assume that both types of springs are infinitely stiff. In later sections, we will allow the springs to be flexible.

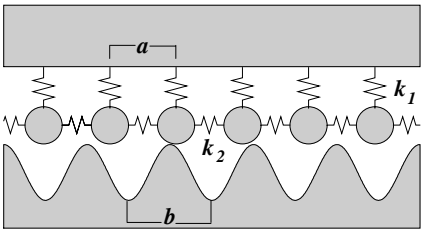


Fig. 10.1. Schematic representation of two solid bodies in contact. The substrate is considered rigid. In the slider, each atom is coupled with a spring of stiffness k_1 to its ideal lattice site and with a spring of stiffness k_2 to its neighbor. Both springs act in lateral direction

To model the effect of corrugation on the motion in lateral direction, it is often assumed that the potential energy in a system such as the one shown in Fig. 10.1 only depends on the lateral direction x and that the normal deflection of the atoms can be neglected. For purely periodic systems, in which the normal position of the atoms is assumed to be fixed, the potential energy of the slider V in the substrate potential can be expressed as a Fourier sum of the form [10]

$$V = \sum_{m=-\infty}^{\infty} \tilde{V}(G_m) \sum_{n=0}^{N-1} \mathrm{e}^{iG_m(x_0+na)} \,, \tag{10.1a}$$

which can be simplified to yield

$$= \sum_{m=-\infty}^{\infty} \tilde{V}(G_m)\mathrm{e}^{iG_m x_0} \times \begin{cases} N & \text{if } \dfrac{aG_m}{2\pi} \text{ is an integer} \\ \dfrac{1-\mathrm{e}^{iNG_m a}}{1-\mathrm{e}^{iG_m a}} & \text{otherwise} \end{cases} . \tag{10.1b}$$

Here, G_m are the reciprocal lattice vectors of the substrate with $G_m = 2\pi m/b$, $\tilde{V}(G_m)$ are the Fourier expansion coefficients of the substrate potential, and $x_0 + na$ is the position of atom n in the chain. One must ensure V to be a real function by requiring that $\tilde{V}(-G_m) = \tilde{V}^*(G_m)$. In the most simple case, where the ground harmonic dominates, Eq. (10.1) reduces to

$$V = NV_0 + V_1 \sum_{n=1}^{N} \cos\left\{\frac{2\pi}{b}(x_0 + na) + \varphi_1\right\} , \tag{10.2}$$

where V_1 is real valued and $2\tilde{V}(G_1) = V_1 \exp(i\varphi_1)$. At a given position of the rigid slider, as denoted for example by the value of x_0, the (lateral) force F exerted from the substrate on the slider will be $F = -\mathrm{d}V/\mathrm{d}x_0$ and thus

$$F = \frac{2\pi}{b} V_1 \sum_{n=1}^{N} \sin\left\{\frac{2\pi}{b}\,(x_0 + na) + \varphi_1\right\} + \ldots \,, \tag{10.3}$$

where the periods indicate that the full Fourier series may have more terms than only the first harmonic. We will now explore the effects of symmetry on the way in which F grows with system size.

Surfaces with Identical Lattice Constants

If $a = b$, the sum in Eq. (10.3) is easy to evaluate because all summands are identical, thus

$$F = N\frac{2\pi}{b}V_1 \sin\left(\frac{2\pi}{b}x_0 + \varphi_1\right) , \tag{10.4}$$

where we have neglected the effect of higher harmonics for reasons of simplicity. If we want to initiate sliding, the maximum lateral force has to be overcome, thus the static friction force reads

$$F_\mathrm{s} = N\frac{2\pi}{b}V_1 . \tag{10.5}$$

Therefore, F_s is proportional to the number of atoms in contact. As the argument extends to higher dimensions as long as the identical surfaces are perfectly oriented, one may say that the friction force is proportional to the area of contact. This is not a contradiction to Amontons' laws, which state that static friction is proportional to the load irrespective of the (apparent) area of contact, because in our example, we would (implicitly) double the whole system and thus double the load L when we double N. In other words, the static friction coefficient would be independent of the area of contact for two ideally-oriented, rigid surfaces with identical lattice constants. Note that the *kinetic* friction force remains zero in this purely mechanical model without *internal* degrees of freedom, which would justify an ad-hoc damping term.

Commensurate Surfaces with $a \neq b$

Two surfaces are called commensurate when the ratio a/b is a rational number, i. e., if there are two natural numbers p and q such that

$$\mathcal{L} = pa = qb , \tag{10.6}$$

where $\mathcal{L}$ is the smallest common period of the two surfaces. We have considered $p = q = 1$ above. Now we will be concerned with the case in which $\mathcal{L}$ is greater than both a and b. Consider again Eq. (10.1). Terms related to Fourier components G_m that cannot be represented as an integer multiple of $2\pi/\mathcal{L}$ will not add up systematically as can be learned from Eq. (10.1b). Thus for sufficiently large interfaces, the terms that grow proportionally with N will dominate. These terms require that $a \cdot m$ be equal to an integer multiple of b, e. g., $a \cdot m$ is equal to $\mathcal{L}$, in which case $m = p$. Thus we can write the "corrugation potential" V_corr as

$$V_\mathrm{corr} = N \sum_{m,\,\frac{a}{2\pi}G_m \in Z} \tilde{V}(G_m)\,\mathrm{e}^{iG_m x_0} \tag{10.7a}$$

$$= N v(x_0) , \tag{10.7b}$$

where $v(x_0)$ can be interpreted as an average atomic potential. Typically, the expansion coefficients in surface potentials decay exponentially fast with increasing index [10]. Therefore, one has to expect static friction that is exponentially small in the smallest common period of the two surfaces in contact, although it does grow linearly with system size.

The exponential decay of the V_m's with increasing index m usually justifies to only keep the first term in the expansion, which is related to the smallest common period. Therefore, the first non-constant and non-vanishing term to contribute to $V_{\rm corr}$ is related to the lattice vector $G_{\rm p} = 2\pi\mathcal{L}/ab$ and hence one often may approximate

$$V_{\rm corr} \approx NV_{\rm p} \cos\left(\frac{2\pi p}{b}x_0 + \varphi_{\rm p}\right) . \tag{10.8}$$

$V_{\rm p}$ and $\varphi_{\rm p}$ can be constructed from $\tilde{V}_{\rm p}$ in a similar way as the equivalent terms related to the first harmonic introduced in Eq. (10.2).

As a side comment, I wish to note that two non-adhesive two Lennard Jones solids with $\mathcal{L} = 4a = 5b$ are sufficiently "out of tune" to produce friction coefficients of less than 10^{-3} when pressed against each other (unpublished data by the author). Mathematically speaking, one would not refer to such surfaces as superlubric, however, by means of practical applications it would be justified to classify such friction coefficients as extremely low.

Incommensurate Surfaces

When the ratio a/b cannot be expressed by a rational number, the surfaces are said to be incommensurate. One may argue that the two surfaces have a common period $\mathcal{L}$ that is infinite and given the argument that the corrugation potential vanishes exponentially with $\mathcal{L}$, one would conclude that the corrugation potential per surface atom v/N tends to zero for large N. Alternatively, consider again Eq. (10.1). There is no term in the corrugation potential that would increase with system size. Therefore, Eq. (10.7b) would have to be replaced with

$$V_{\rm corr} = v_{\rm inc}(N, x_0) , \tag{10.9}$$

where $v_{\rm inc}$ can be calculated from Eq. (10.1). While $v_{\rm inc}(N, x_0)$ does not vanish exactly, one can argue that it does not grow systematically with system size N, because the fractions on the r.h.s. of Eq. (10.1b) are oscillating functions of N and only the first few terms contribute in a significant way. Keeping only the first term related to $m = \pm 1$, one can easily see that the maximum lateral force per atom and hence $\mu_{\rm s}$ vanishes with $1/N$ for increasing system sizes or alternatively, the net friction force

$$F_{\rm s} \propto N^0 . \tag{10.10}$$

Flat, Disordered Surfaces

For incommensurate surfaces, the annihilation of lateral forces is rather systematic, similar to the deconstructive interference in optics. In order to include the effect of disorder into our model, several avenues can be pursued. One is to introduce randomness into the substrate potential, another one is to allow for stochastic variations of the interatomic spacing within the slider. Both procedures yield similar results for rigid planes, but the latter approach is more easily discussed in the present context. Let us assume that two adjacent neighbors in the slider have a spacing of $a+\delta a_{n,n+1}$, where $\delta a_{n,n+1}$ is an (independent) random number of mean zero, $\langle \delta a_{n,n+1} \rangle = 0$, and well defined second moment of $\langle \delta a_{n,n+1}^2 \rangle = \sigma_a^2$. The position of particle n can then be given by

$$x_n = x_0 + na + \tilde{a}_n \tag{10.11}$$

with $\tilde{a}_n = \sum_{n'=1}^{n-1} \delta a_{n',n'+1}$. Thus, the positions of two atoms m and n become more and more uncorrelated as the distance between m and n decreases. Specifically, the expectation values of $(x_m - x_n)$ and $(x_m - x_n)^2$ satisfy

$$\langle (x_m - x_n) \rangle = (m-n)\, a \tag{10.12a}$$

$$\langle (x_m - x_n)^2 \rangle = (m-n)^2 a^2 + |m-n|\, \sigma_a^2 \,. \tag{10.12b}$$

We can rewrite Eq. (10.3) as

$$F = \frac{2\pi}{b} V_1 \Im \left(\sum_{n=0}^{N-1} \exp \left\{ i \left[\frac{2\pi}{b} (x_0 + na + \tilde{a}_n) + \varphi_1 \right] \right\} \right) \,, \tag{10.13}$$

where $\Im(\bullet)$ denotes the imaginary part its argument. Higher harmonics were suppressed for simplicity. Due to the randomness in the configuration, we can only be concerned with statistical averages. For a given value of x_0, the expectation value of the force reads:

$$\langle F \rangle = \frac{2\pi}{b} V_1 \Im \left(\sum_{n=0}^{N-1} \exp \left\{ i \left[\frac{2\pi}{b} (x_0 + na) + \varphi_1 \right] \right\} \left\langle \mathrm{e}^{i\, 2\pi\, \tilde{a}_n / b} \right\rangle \right) \,. \tag{10.14}$$

As we are interested in large systems, it will be sufficient to discuss the properties of $\tilde{a}_m$ for the many terms with $m \gg 1$. The standard deviation of $\tilde{a}_m$ is $\sqrt{m}\sigma_a$, which will exceed b for sufficiently large indices m. Therefore, the expectation value of $\exp(i 2\pi \tilde{a}_m / b)$ on the right-hand-side of Eq. (10.14) will be zero, leading to

$$\lim_{N \to \infty} \frac{1}{N} \langle F \rangle = 0 \,. \tag{10.15}$$

As for the second moment of the lateral force F, one can start from Eq. (10.13) and make use of the fact that the statistical properties of the expression would

remain unaltered, if we took the real part instead of the imaginary part on the right hand side of the equation. Thus,

$$\langle F^2 \rangle = \left(\frac{2\pi V_1}{b} \right)^2 \frac{1}{2} \sum_{n,n'} \exp\left\{ i \frac{2\pi}{b} (n - n') a \right\} \left\langle \mathrm{e}^{i2\pi(\tilde{a}_n - \tilde{a}_{n'})/b} \right\rangle . \tag{10.16}$$

Note that the expressions on the r.h.s. of Eq. (10.16) are translationally invariant, so that we can replace the sum over n' with a sum over Δn. Except for "surface" terms, one can therefore write

$$\langle F^2 \rangle = \frac{1}{2} N \left(\frac{2\pi V_1}{b} \right)^2 \sum_{\Delta n} \exp\left\{ i \frac{2\pi a}{b} \Delta n \right\} \left\langle \mathrm{e}^{i2\pi \tilde{a}_{\Delta n}/b} \right\rangle \tag{10.17a}$$

$$= \frac{1}{2} N \left(\frac{2\pi V_1}{b} \right)^2 f\left(\frac{a}{b}, \frac{\sigma_a}{b} \right) , \tag{10.17b}$$

where we can approximate the function $f\left(\frac{a}{b}, \frac{\sigma_a}{b}\right)$ with

$$f\left(\frac{a}{b}, \frac{\sigma_a}{b} \right) = \sum_{\Delta n} \exp\left\{ i \frac{2\pi a}{b} \Delta n \right\} \mathrm{e}^{-\frac{1}{2}(2\pi \Delta n \sigma_a / b)^2} , \tag{10.18}$$

which is exact for a Gaussian distribution of the random numbers $\delta a_{n,n+1}$ (as can be seen by seen by terms of a cumulant expansion).

An interesting aspect of Eq. (10.17b) is that the disorder introduced in Eq. (10.11) leads to an expectation value $\langle F^2 \rangle \propto N$ irrespective of the position x_0 and independent of the precise value of a/b. The function $f\left(\frac{a}{b}, \frac{\sigma_a}{b}\right)$ is merely a prefactor, which is maximal for $a = b$. However, the "average commensurability," $a = b$ with $\sigma_a > 0$, does not induce the $F \propto N$ behavior found for truly commensurate surfaces. Since Eq. (10.17b) is valid for any position x_0, it is also valid at the position where F is maximum, and hence we may conclude that

$$F_\mathrm{s} \propto \sqrt{N} , \tag{10.19}$$

or in other words, $\mu_\mathrm{s} \propto 1/\sqrt{N}$.

Higher Dimensions

The calculations done above generalize to higher dimension if we associate N with the number of atoms in the bottom most layer of the slider. Thus for two three-dimensional solids with a two-dimensional interface, the model of rigid surfaces would predict that

$$\mu_\mathrm{s} \propto \begin{cases} A^0 & \text{commensurate} \\ A^{-1} & \text{incommensurate} \\ A^{-1/2} & \text{amorphous} \end{cases} . \tag{10.20}$$

It is important to reemphasize that the results strictly depend on the assumption that the planes are ideally rigid and atomically smooth. For commensurate surfaces in dimensions larger than one, it is not sufficient for the two surfaces to have identical periods, but the orientation of the two surfaces need to be aligned. In numerical simulations of atomically smooth but disordered surfaces, the $\mu_\mathrm{s} \propto 1/\sqrt{A}$ hypothesis could be confirmed [11].

10.2.3 Role of Instabilities in Simple Models

The way in which elastic deformations can induce finite friction and thus break superlubricity has been casted in a semi-quantitative fashion within the Prandtl–Tomlinson (PT) [4, 5] model and within the Frenkel–Kontorova (FK) [6] model. The PT model describes the effects of elasticity within the Einstein model of solids, i. e., each (surface) atom in the slider is coupled to its ideal lattice site via a harmonic spring of strength k. Moreover surface atoms experience a sinusoidal interaction $V_\mathrm{s}(x) = V_0 \cos(2\pi x/a)$ with the effectively rigid substrate plus a drag force linear in the atom's velocity, which will be motivated later. The equation of motion for the particle reads

$$m\ddot{x} + \gamma\dot{x} = \frac{2\pi V_0}{a} \sin(2\pi x/a) - k\left[x - x_\mathrm{d}(t)\right] \,, \tag{10.21}$$

where $x_\mathrm{d}(t)$ is the position of the atom's ideal lattice site. Above a critical value k^* for the spring constant k, the PT model predicts zero kinetic friction even in the absence of thermal noise. To be specific, if k exceeds the maximum curvature $\kappa = (2\pi/a)^2 V_0$ of the substrate potential, then no instabilities (hysteresis loops) are found and kinetic friction vanishes in the PT model. Thus, if the slider is sufficiently stiff, F_k may disappear and hence the interface can be superlubric. For $k < k^*$, zero-temperature F_k remains finite no matter how slowly we pull the spring, which is akin of Coulomb's law of friction, [3] which states that kinetic friction is (rather) independent of the sliding velocity v_0. The mechanism is described in more detail in Fig. 10.2.

The FK model is similar to the PT model, except that surface atoms are not coupled elastically to their lattice site but instead to their neighbors in a one-dimensional chain. The advantage of the FK model over the PT model that tribologists have seen is that the FK model incorporates the effect of long-range elastic deformation. However, one needs to keep in mind that the FK model, being a one-dimensional model, tremendously overestimates elastic deformations, artificially suppressing superlubricity.[2] The FK model and its generalizations to higher dimensions are yet useful to describe various tribological phenomena conceptually and sometimes even quantitatively. In fact, in one of their first studies, Hirano and Shinjo parametrized a generalization

[2] The PT model is the mean-field variant of the one-dimensional FK model and therefore more realistic in higher dimensions, in particular above the upper critical dimension for elastic manifolds sliding through external potentials.

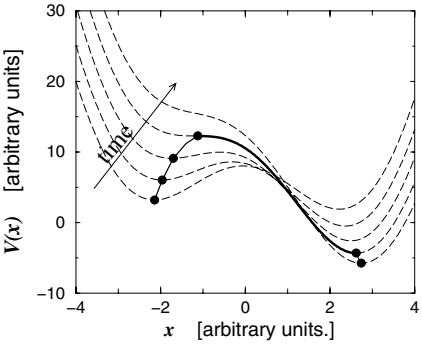


Fig. 10.2. Schematic representation of an instability in the PT model. The time-dependent potential $V(x)$ (*dashed lines*) is the superposition of a sinusoidal substrate potential, which is fixed in space, and a moving parabolic potential, which represents the elastic interaction between a degree of freedom (DOF), e. g., an atom or an AFM tip, and the moving driving device. The DOF, whose position at various instances of time is indicated by the *full circles*, becomes unstable when its mechanical equilibrium position is $x \approx -1.7$, from where it jumps to $x \approx 2.6$. In this process, an energy $\Delta E \approx 16.6$ is dissipated, provided the stage moves slowly and that thermal fluctuations can be neglected. The jumps occurs every time the driving device moves by one lattice constant, which is 2π in this example. Sliding over a long distance will therefore result in an average kinetic friction force of $F_{\mathrm{k}} \approx 16.6/2\pi \approx 2.65$

of the FK model with realistic interatomic potentials for copper and found that incommensurate copper surfaces should often remain superlubric [2].

Of course, there can be other ways than local, elastic instabilities that induce finite friction and break superlubricity. Adhesive jump-to-contact instabilities, plastic deformation, instabilities in confined boundary lubricants, chemical reactions, cold welding, and related processes are all characterized by sudden pops of microscopic degrees of freedom, which lead to energy dissipation [9].

10.2.4 Effect of Temperature

Temperature-induced relaxation can both increase or decrease friction [9]. It can increase friction because the real contact area can increase due to thermally-assisted plastic flow, which increases the effective load and consequently the friction. Moreover, thermally assisted aging can increase the strength of the junctions formed between two solids at rest. In superlubric systems, one should generally hope that plastic flow is negligible. Temperature then helps the instability (such as the one shown in Fig. 10.2) to occur prematurely and therefore temperature can reduce kinetic friction. When pulled sufficiently slowly, the system will be able to be close to thermal equilibrium, e. g., the coordinate x in the PT model will occur approximately with

a probability $\exp(-V(x)/k_{\mathrm{B}}T)$ at a given position of the driving stage and not only close to the (previous) mechanically stable position as indicated in Fig. 10.2. In this case, linear-response theory is applicable and friction must be proportional to velocity.[3] In fact, any *finite* system at *finite* temperature will eventually approach the linear-response regime [12]. Even finite, *commensurate* surfaces, which according to any reasonable definition should not be classified as superlubric, will ultimately enter a regime in which $F_{\mathrm{k}} \propto v$, as demonstrated in numerical simulations of dry contacts [13].

10.2.5 Damping in the Superlubric Regime

When friction is extremely small, it is tempting to assume that instabilities do not play a significant role and that the system is close to thermal equilibrium at all times. As argued in the previous section, friction would then have to be considered to be linear in velocity v_0 at small v_0. The response of the slider to a time-dependent external force can be calculated by using linear-response theory and the fluctuation-dissipation theorem, [12] i. e., the fluctuations of the lateral force at any fixed position of the driving device can be related to the the damping that the slider would exert on a slowly moving driving device. An excellent overview on the topic and tips how to make use of the concepts in computer simulations is given in the book *Molecular Simulation* by Frenkel and Smit [14]. Note that the damping coefficient of a degree of freedom (be it an atom or be it a collective degree of freedom such as an AFM tip) can be position dependent in principle.

The linear-response formalism has so far been employed in the so-called non-contact mode of atomic force microscope (AFM) tips [15,16] (see also the chapter by Kantorovich and Trevethan in this book), however, the formalism is the same for intimate contacts [13]. It appears that the theoretically predicted values for non-contact damping are smaller than the ones measured experimentally. The reason for the discrepancy may be that most calculations assume sliding of ideal crystals. Surface defects, dislocations, and contamination are usually neglected. They will all lead to additional force fluctuations and hence to increased friction between substrate and slider. In essentially every real-life contact, there will be isolated points of contact where instabilities cannot be prevented. These will lead to dissipation that dominates the damping-induced interactions. Single-asperity contacts in sufficiently well-defined microcontacts may be an exception to this rule [7].

Lastly, it is worth mentioning that linear response does not always imply very low friction, for example, when internal deformation in solids is highly viscous – as is the case for rubber. Formalism to describe dissipation for sliding rubber contacts, [17] or generalizations thereof, should form a promising starting point for formal theories of damping forces in superlubric contacts.

[3] For a discussion of the velocity-dependence of F_{k} at more elevated velocities, see the chapter by Evstigneev and Reimann in this book.

10.2.6 Long-Range Elastic Deformations

In real solids, atoms are not coupled elastically to their lattice sites but they interact with the other atoms in the solid. For many purposes, it is sufficient to treat this interaction within one solid as harmonic interactions between adjacent atoms. When one solid is placed on top of another one as shown schematically in Fig. 10.1, instabilities of surface atoms do not necessarily involve the sudden motion of single atoms but they may involve the collective motion of many atoms. In order to ascertain whether such collective instabilities occur, it is then necessary to analyze whether the intrabulk, elastic interactions dominate the interfacial interactions on length scales larger than atomic scales. If the answer is positive, the system can be superlubric.

The equations of motion for an elastic solid that is sliding with respect to a (rigid) substrate have the same structure as those that describe charge density waves (CDWs) [18]. These latter systems have been studied thoroughly [19]. One result is that the CDWs or, to be more precise, the vortices in type-II superconductors can be treated as essentially rigid domains within a characteristic length called the Larkin length. Adjacent Larkin domains are coupled only weakly so that (roughly speaking) its center of mass can have more than one (meta)stable position. This multistability would automatically break superlubricity in a way similar to that shown in Fig. 10.2.

In order to apply the concept of Larkin domains to tribological phenomena, one can proceed as follows: [20] Assume that a block of linear dimension $\mathcal{L}$ is rigid. Then assess how such a rigid block would couple elastically to its neighboring block (so that the solid's elastic constants remain independent of the coarse-grain length $\mathcal{L}$) and how the block manages to interlock with the substrate, see also Fig. 10.3. To get an order of magnitude estimate for the relevant variables at the atomic scale, $\mathcal{L} = \mathcal{L}_{\text{atomic}}$, it is reasonable to assume that the elastic coupling k between adjacent atoms is in the order of the bulk modulus B, say 40 GPa for a soft solid, times a lattice spacing, say 2 Å. Thus, $k(\mathcal{L}_{\text{atomic}}) \approx 8\,\text{N/m}$ would be a reasonable value for relatively soft solids. For hydrogen-terminated diamond, the estimate for $k(\mathcal{L}_{\text{atomic}})$ would be more than a factor 10 larger. A similar estimate can be done for the maximum local curvature of the substrate potential κ, which was introduced in the PT model following Eq. (10.21): The atoms of opposed surfaces interact via weak physical bonds, provided the surfaces are chemically passivated. Solids that are tied together by physical bonds have bulk moduli in the order of 4 GPa and nearest neighbor separation are in the order of 3 Å, resulting in an estimate of $\kappa \approx 1.2\,\text{N/m}$. Thus, at the atomic scale, $k > \kappa$, so that instabilities cannot be expected to occur. This is a necessary conditions for superlubricity.

Once the values for k and κ are estimated at the atomic scale, $\mathcal{L}_{\text{atomic}}$, it is possible to also estimate those values at any length scale $\mathcal{L}$. For instance, if we replace a linear chain of N beads with separation a by $N/2$ beads that are separated by $2a$, then we need to reduce the stiffness of the coarse-grained

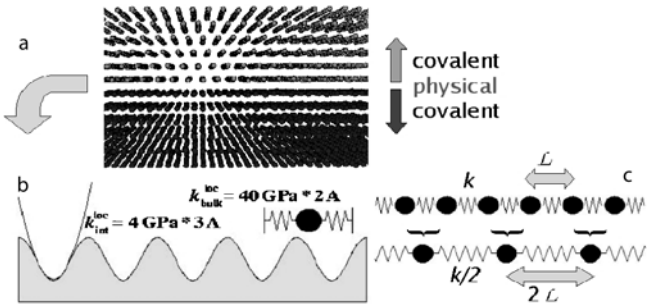


Fig. 10.3. Schematic view of the interaction between chemically passivated solids. Part **a** shows a snapshot of a simulation. Part **b** represents the coupling of a surface atom to its neighbors (reflected by the harmonic springs) and to the substrate (reflected by the sinusoidal line). The parabola indicates the maximum curvature of the atom-substrate potential, which corresponds to the curvature κ of the substrate potential in the PT model. Part **c** describes the scaling procedure for a one-dimensional elastic chain. From Ref. [20]

springs by a factor of 2, as illustrated in Fig. 10.3. Thus, the springs become softer upon coarse-graining in one-dimensional systems, just like the effective capacitance of capacitors connected in series is reduced. In two dimensions, we would not only connect springs (capacitors) in series but also in parallel. The net effect is that the coarse-grained spring (capacitor) has the same stiffness (capacitance) as the original one. Each dimension added in parallel makes the stiffness of the springs harder. If D is the dimensionality of the solid, one finds

$$k(\mathcal{L}) = \left(\frac{\mathcal{L}}{\mathcal{L}_{\rm atomic}}\right)^{D-2} k(\mathcal{L}_{\rm atomic})\,. \tag{10.22}$$

It is a bit more difficult to find a similar scaling law for the curvature of the slider-substrate potential. However, it is reasonable to assume that (apart from a prefactor) κ scales similarly as the static friction force $F_{\rm s}(\mathcal{L})$ of a rigid domain of linear scale $\mathcal{L}$. In scaling studies, it is necessary to keep the intrinsic thermodynamic such as the (average) normal pressure $p_\perp$ constant so that we can say $\kappa(\mathcal{L}) \propto F_{\rm s}(\mathcal{L}) = \mu_{\rm s}(\mathcal{L}) \cdot (\mathcal{L}^{D_\int} p_\perp)$, where $D_\int$ is the dimension of the interface and thus $(\mathcal{L}^{D_\int} p_\perp)$ the load carried by an "area" A of linear dimension $\mathcal{L}$. The scaling of the friction coefficient, however, strongly depends on the order at the interface. Using Eq. (10.20) and the $\kappa \propto \mu L$ relationship for rigid or correlated domains, one finds that

$$\kappa(\mathcal{L}) \propto \begin{cases} \mathcal{L}^{D_\int} & \text{commensurate} \\ \mathcal{L}^0 & \text{incommensurate} \\ \mathcal{L}^{D_\int/2} & \text{amorphous} \end{cases}. \tag{10.23}$$

Whenever the ratio of $k(\mathcal{L})/\kappa(\mathcal{L})$ increases systematically with $\mathcal{L}$, one should expect the intra-bulk elasticity to dominate and thus to have the possibility of

superlubricity. In the technically relevant case of disordered, two-dimensional interfaces and three-dimensional solids, both $k(\mathcal{L})$ and $\kappa(\mathcal{L})$ increase linearly with $\mathcal{L}$. Allowing for some elastic deformation within the Larkin domains effectively yield logarithmic correction to the scaling laws, which would result in finite friction [21, 22]. However, as the Larkin domains would be relatively large and no local instabilities should be expected, these corrections can probably be seen as irrelevant, in particular for systems such as hydrogen-terminated diamond with large values of k/κ at the atomic scale.

So far we have neglected the highly non-uniform distribution of normal loads, which are the consequence of realistic surface topologies. Real contacts have roughness on many different length scales and the distribution of normal pressures sometimes allows for large normal values, which would increase the values for κ at the intimate point of contacts. It is plausible to expect elastic instabilities in such point of high pressure. However, elastic instabilities in bulk systems require pressures that are typically higher than the yield strength of the material, e. g., incommensurate interfaces between two atomically smooth gold surfaces only show friction in molecular dynamics simulations when $p_\perp$ is raised above 4 GPa [23]. (The periodic boundary conditions in lateral direction allow the gold solids to sustain these unusually large pressures.)

10.3 Simulations

Analytical calculations, such as the ones presented in the last section, can generally not be used to make quantitative predictions for specific systems. Even the most simple models without any chemical detail elude analytical tractability. For example, there are no closed-form solutions for the dynamical properties of the Frenkel–Kontorova (FK) model, which consists of a one-dimensional, elastic chain embedded into a simple, sinusoidal potential. Only some aspects of the continuum variant of the FK model, namely the sine-Gordon (SG) model, can be solved analytically. Due to its continuum character, there is no finite kinetic friction in incommensurate SG models. Discreteness corrections can be applied, however, the calculations are tedious [24]. Given the fact that most analytical models are one-dimensional and the paramount importance of dimensionality emphasized in the last section, it appears to be a sensitive choice to resort to computer simulations of sliding solids.

The main stream technique for these simulations is molecular dynamics (MD) [8]. Atomic configurations are set up in a computer experiment, interactions between atoms are assumed (or calculated with quantum chemical first-principle techniques) and Newton's equations of motion are solved for each individual atom. The boundary walls are typically coupled to a driving device and boundary conditions are chosen such that a tribological experiment is mimicked in the best possible way. Sliding produces heat, which then requires

the use of (artificial) thermostats that remove the heat from a sliding contact. Refs [8,9] give an overview of the results of atomistic simulations, a more technical review on how to set up simulations will soon be available too [25].

10.3.1 Generic Models

A large part of the literature is concerned with so-called generic models, where the interaction between atoms are only simple two-body potentials. Results of these simulations can generally not be applied to specific systems. However, they allow one to study the effect of dimensionality, commensurability, contamination, surface geometry, etc. on tribological properties. Simulations of generic models can capture many tribological effects qualitatively or in some cases even semi-quantitatively. One popular approach of generic models is the study of "Lennard Jonesium", which is a (virtual) material in which atoms interact through the Lennard Jones (LJ) potential $V(r) = 4\epsilon[(\sigma/r)^{12} - (\sigma/r)^6]$, where σ and ϵ are the LJ length and LJ energy, respectively. σ can be roughly associated with the sum of the radii of the two interacting atoms and ϵ with the cohesive energy of a dimer or nearest neighbors in a solid.

In the context of superlubricity, one important question to address is whether solids generically have the tendency to show finite static and kinetic friction when brought into contact or whether one should expect superlubricity only for very specific materials. In order to evaluate at what interfacial strength perfect [111] surfaces of face-centered-cubic (fcc) solids would become unstable instantaneously if the models are more realistic than simple bead spring models, Müser placed two LJ solids on top of each other (see also Fig. 10.3a) [26]. All interactions between identical atoms, i. e., those atoms that originate from the same solid, were chosen such that $\epsilon = \sigma = 1$. The LJ parameters for pairs of atoms originating from opposed solids were chosen $\sigma_1 = \sigma$ and ϵ_1 was varied. In order to obtain *instantaneous* instabilities and thus pinning, the value ϵ_1 had to exceed that of ϵ by a factor of eight. These instabilities, however, were not elastic in nature, but they involved large rearrangement of the atoms which could be interpreted as cold welding. If the simulations could be run for much longer times, the mixing instabilities could of course be observed for any value of $\epsilon_1 > \epsilon$. Other simulations also support the idea that instabilities in solids without directed bonds are typically not elastic in nature, e. g., the elastic instabilities found by Lançon at normal pressures of 4 GPa in simulations of incommensurate solids of gold, [23] would imply plastic flow in laboratory experiments, see for instance Fig. 10.4(b) where the periodic boundary conditions in lateral direction prevent significant (further) plastic flow.

As long as the pressure conditions are not extreme and plastic deformation is absent or relatively small, all-atoms simulations of curved tips show that many of the analytical arguments advanced in Sect. 10.2.2 for flat contacts also apply to curved tips, [27], for instance that friction is largest for commensurate and smallest for incommensurate contacts and intermediate

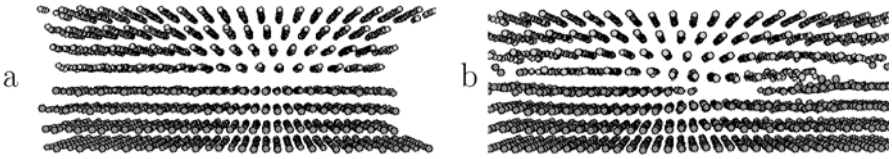


Fig. 10.4. a Low-load configuration of chemically-passivated solids with terraces. Both solids are fcc Lennard Jonesium with [111] surfaces. The surfaces are misoriented by 90°. Static and kinetic friction are too small to be defined, because thermal fluctuations at ambient conditions are strong enough to depin the contact. **b** Snapshot of a configuration that results from configuration **a** as a consequence of high normal pressure. Despite the onset of plastic deformation, the static friction coefficient for this contact remains smaller than 0.01. From Ref. [26]

for disordered surfaces. One of the conclusions to be drawn from those simulations is that large radii of curvature lead to a smaller ratio of lateral and normal pressure. This means that layered materials may be one possibility to achieve ultra-small friction.

Recent approaches have incorporated roughness on more length scales and allowed to include the long-range elastic deformation in the solids by means of multi-scale methods [28]. While superlubricity is broken in these systems (the load is carried at isolated locations only, which induces local pinning), it needs to be emphasized that (a) exceedingly small "terraces" were allowed in that study, sometimes as small as one atom, and (b) despite of the existence of these small terraces friction coefficients were typically still well below 0.1. When the solids were allowed to deform plastically as well, the small terraces were rubbed away resulting in smaller friction coefficients. This is a typical observation in MD simulations of unlubricated (non-metallic) surfaces that are driven at moderate or small loads. See also Fig. 10.4, which shows a contact with moderate plastic flow.

Many tribological aspects cannot be modeled by Lennard Jonesium, for instance whenever directed bonds are important. These play an obviously crucial role in molecular solids, layered materials and other chemically complex solids such as the low-friction material Teflon. Also metals are poorly described by LJ potentials, for instance the well-known necking during contact formation between bare metals is a consequence of the particular many-body form of metals. Generally, metals have a larger propensity to annihilate free surface than Lennard Jonesium, which favors cold welding. Lennard Jonesium cannot reproduce these effects and thus more realistic studies are required to understand friction (or the absence thereof) in these materials.

10.3.2 Layered Materials

Given the analysis in the previous sections, one would expect that layered solids are natural candidates for superlubric materials. First, the elastic cou-

pling within a sheet, i. e., k in the PT model, is very large, while the interaction with the opposed surface, i. e., κ in the PT model, is small. Moreover, long terraces are formed and as argued in the last section, large local radii of curvature lead to small friction in unlubricated and non-yielding contacts. Thus, almost all simulations of one layered material placed on another layered material show small friction.

An interesting aspect of many layered materials, in particular graphite, is that they have the ability to form nanotubes. These nanotubes allow one to study friction in a single-asperity contact for well-defined geometries as opposed to AFM tips, whose detailed atomic structure has remained elusive. Experimentally [29] and in simulations, [30,31] registry or commensurability is found to play the role that one would expect from classical mechanics and the concept of geometric interlocking. The static friction is large if the nanotubes orient with the graphite substrate, while misaligned nanotubes show much reduced friction.

In many cases nanotubes are multiwalled. The friction between an inner tube and an outer tube could be measured experimentally [32]. Many times, inner tube and outer tube are incommensurate due to the different ways in which graphite sheets can fold to form a tube. When the tubes are incommensurate, non-extensive shear stresses are found suggestive of the dominant roles of surface effects in those systems.

Theoretical simulations of double-walled nanotubes often show extremely small friction indeed, which lead to the suggestion to use them as mechanical nano-oscillators [33]. The main force inducing lateral motion between inner and outer tube is related to surface energy. The tubes try to minimize free surface, which can act in sliding direction or against sliding direction. In the absence of instabilities, only little energy is dissipated as heat within one oscillation cycle. It may be surprising that this last statement even holds for commensurate nanotubes, which have large instantaneous/static friction forces [34]. However, commensurate nanotubes also show large instantaneous "anti-friction." The net damping calculated for commensurate and incommensurate nanotubes turns out to be of similar magnitude – at least for the idealized geometries utilized in the simulations [34]. Thus, commensurability does not automatically lead to instabilities and significant kinetic friction. It may be worth pointing out that kinetic friction between slightly contaminated commensurate surfaces has been predicted to be even smaller than that between incommensurate forces [35].

Most simulations idealize the nanotubes initial geometry and underestimate the friction force. However, if chemical details of the fractured end are incorporated into the simulations, experimentally measured and calculated values for kinetic friction agree reasonably well [36]. It is probably safe to assume that the friction between nanotubes is dominated by defects, in particular those that involve chemical bonds between inner tube and outer tube. As the covalent bonds between carbon atoms are orders of magnitude

stronger than the van der Vaals interactions, these chemical bonds (and their rupture) between atoms in the inner and atoms in the outer tube will most significantly contribute to the dissipation in real nanotubes.

In practice, one disadvantage of layered materials certainly is that the sheets are easily rubbed off. This process leads to the generation of debris and consequently friction increases. This effect may become particularly severe when chemical point defects are present. Indeed, some of the first experimental evidence for superlubricity between layered molybdenum disulfide (MoS_2) was seen to be transient until wear particles probably increased the friction to coefficient to a value above 10^{-3} [37].

10.3.3 Metal on Metal Contacts

While the first realistic calculations suggestive of superlubricity [1, 2] were based on incommensurate contacts, it is probably safe to assume that bare metals are not good candidates for superlubric materials. One of the reason is the "vulnerability" of metal contacts in particular at their boundaries [38]. Upon sliding, dislocations and ultimately wear is easily generated from the contact boundaries, as discussed for ideally blunt copper tips sliding on ideal copper (111) surfaces. Thermal aging of the tips and hence producing more realistic contact geometries results in even more wear [38].

One can yet ask the question, how much friction one should expect in idealized situations, i. e., if it was possible to suppress the boundary effects. A particularly interesting study addressed the role of roughness in the very last layer. Qi et al. [39] studied atomically smooth Ni(100)/Ni(100) interfaces. Their idealized geometries display the same superlubric behavior as the idealized copper interfaces studied by Hirano and Shinjo [1] However, roughening the top layer with a mere 0.8 Å rms variation, changes the behavior completely, with friction coefficients increasing by several orders of magnitude. The calculated values for the surfaces with the 0.8 Å additional roughness matches the available experimental data extremely well [40].

In another study, the friction between chemically passivated aluminum surfaces (Al_2O_3 termination) was calculated [41]. An interesting result of that study is that incommensurate surfaces show small but yet non-negligible friction at moderate normal loads, although no wear occurred. As opposed to the commensurate case, the friction force evolved relatively smoothly as a function of time for incommensurate Al_2O_3 contacts, which would only be consistent with a picture in which the instabilities are localized. The "wearless" instabilities must thus be a consequence of the directed bonds, which are usually not incorporated in a purely theoretical description of contacts. More studies would be needed to identify better candidates for superlubric terminated metals than aluminum, however, it may well be that this quest will not be successful.

10.3.4 Hydrogen-Terminated Surfaces

One of the most promising candidates for superlubric materials is hydrogen-terminated diamond-like carbon (DLC) [42]. Unfortunately, there have not yet been many simulations on hydrogen-terminated DLC, although results will soon be published [43]. Interesting new results are expected in the near future, partly due to the availability of force fields that allow one to model chemical reactions. An example for such a force field is the reactive empirical bond order potentials, [44] which was used to study the compression and sliding induced polymerization reactions within alkyne chains [45]. Chemical reactions may impede superlubricity, as the formation of bonds corresponds to instabilities, which can lead to Coulomb type friction if the reactions are reversible [46, 47].

A large fraction of computational studies of chemically-passivated, "realistic" surfaces has been concerned with self-assembled monolayers (SAM), which are much softer than chemically bonded solids, and thus inappropriate candidates for superlubricity. It has yet to be noted that SAMs have typically low friction and probably good resistance to wear, which make them interesting materials for tribological applications. Many of the results obtained in the literature can be interpreted with the concept of geometric interlocking. For instance, when n-alkane monolayers are grafted to a diamond surface with a period of two lattice constants ($2a$) to form a SAM, and this SAM is slid against a perfectly oriented diamond surface, then the friction force has a period of $2a$, with positive and negative contributions [48]. In these simulations, the maximum instantaneous lateral force, which can be associated with the static friction force F_s reaches 4 nN at normal loads of approximately 80 nN. The average, kinetic friction force F_k was not stated explicitly, but can be estimated from the graphs as being at least a factor of 10 smaller than F_s, thus the system is almost superlubric. It would be interesting to know the friction between SAMs attached to diamond and hydrogen-terminated diamond for incommensurate interfaces. In simulations of generic models of two surfaces separated by one monolayer, it was found that static friction is much larger for commensurate than for incommensurate systems, while the opposite is true for kinetic friction [11].

The crucial role of orientation was also found in another study of friction involving monolayers. Commensurate fluorine-terminated alkanethiol SAMs exhibited a tremendous sensitivity of the (differential) kinetic friction coefficient $\tilde{\mu}_\mathrm{k} = \mathrm{d}F_\mathrm{k}/\mathrm{d}L$ on line defects in one of the two SAMs [49]. When the layers were truly commensurate, $\tilde{\mu}_\mathrm{k}$ turned out much larger than in those cases where line defects significantly reduced commensurability and thus $\tilde{\mu}_\mathrm{k}$. Also the packing density in SAMs has the effect on friction that one would expect from the theoretical analysis, i.e, less densely packed and hence softer systems will become more easily unstable than dense, hard systems. For instance, friction between an amorphous carbon tip and a SAM decreased with increasing packing. Two systems had been compared [50]. When the poly-

mers in the SAM were identical so that all of them contained $N = 14$ carbon atoms, the structures were dense, while they were less dense when N was either 12 or 16 (with equal probability). Kinetic friction in the latter system was twice as large than in the dense system.

10.4 Conclusions

Theoretical considerations and atomistic simulations clearly indicate that there should be the possibility of identifying many materials showing superlubricity. The ingredients favoring superlubricity are smoothness of the surfaces, in particular, smoothness of the last layer, and chemical passivation of the last layer. Studies investigating the relevance of roughness on larger length scales are still in their infancy, however, first results indicate that the local roughness is paramount. Thus, layered solids such as graphite and molybdenum disulfide appear as natural candidates for superlubric materials. However, the sheets in layered solids are easily rubbed off, which results in debris and ultimately increased friction.

The currently best candidate for superlubricity remain systems similar to diamond-like carbon, in which the dangling bonds are saturated. Currently, the surfaces seem to be predominantly passivated with hydrogen. From a theoretical point of view, it might be possible to make the surfaces smoother by terminating the dangling bonds of carbon with atoms that are larger than hydrogen. Thus, if feasible, theory would suggest a termination of the dangling bonds in diamond-like carbon with fluorine rather than with hydrogen. Potential alternatives are silicon saturated with either fluorine or perhaps chlorine.

Lastly, it is important to stress that superlubricity is not necessarily a well-defined term. For instance, commensurate systems typically have high static friction, but they do not necessarily show high kinetic friction, as was discussed in this chapter in the context of (idealized) commensurate carbon nanotubes. One may well classify these structures as superlubric, because they dissipate little energy upon sliding. However, other systems have small kinetic friction *and* small static friction. The smallness in static friction is typically due to the incompatibility of geometries of the two opposed surfaces. For this scenario, the term structural lubricity was suggested. The terms structural lubricity and superlubricity can best be illustrated in the context of the experiments by Socoliuc et al. [7] and Dienwiebel et al. [51] In Socoliuc's experiment, the instantaneous lateral forces between the AFM tip and the substrate remained (relatively) large. This clearly indicates that the contact was not structurally lubric. However, due to the stiff cantilever, no instabilities occurred, and thus the contact can be identified as superlubric. In Dienwiebel's experiment, a graphite flake was brought out of registry. This resulted not only in a small kinetic friction but even in a small instan-

taneous force. Therefore, the contact was not only superlubric (absence of instabilities) but even structurally lubric (absence of geometric interlocking).

References

1. M. Hirano and K. Shinjo, Phys. Rev. B **41**, 11837 (1990).
2. K. Shinjo and M. Hirano, Surf. Sci. **283**, 473 (1993).
3. C. A. Coulomb, *Théorie des Machines Simples*, in Memoirs de Mathematique et de Physique de l'Academie Royale, 161-342 (1785).
4. L. Prandtl, Z. Angew. Math. Mech. **8**, 85 (1928).
5. G. A. Tomlinson, Philos. Mag. Series **7**, 905 (1929).
6. Y. I. Frenkel and T. Kontorova, Zh. Eksp. Teor. Fiz. **8**, 1340 (1938).
7. A. Socoliuc, R. Bennewitz, E. Gnecco, and E. Meyer, Phys. Rev. Lett. **92**, 134301 (2004).
8. M. O. Robbins and M. H. Müser, in *Modern Tribology Handbook*, ed. B. Bhushan (CRC Press, Boca Raton, 2001).
9. M. H. Müser, M. Urbakh, and M. O. Robbins, Adv. Chem. Phys. **126**, 187 (2003).
10. W. Steele, Surf. Sci. **36**, 317 (1973).
11. M. H. Müser, L. Wenning, and M. O. Robbins, Phys. Rev. Lett. **86**, 1295 (2001).
12. R. Kubo, Rep. Prog. Phys. **29**, 255 (1966).
13. M. H. Müser and M. O. Robbins, Phys. Rev. B **61**, 2335 (2000).
14. D. Frenkel and B. Smit, *Understanding Molecular Simulation* (Elsevier, San Diego, 2002).
15. M. Gauthier and M. Tsukada, Phys. Rev. B **60** 11716 (1999).
16. L. N. Kantorovich, Phys. Rev. B **64**, 245409 (2001).
17. B. N. J. Persson, J. Chem. Phys. **115**, 3840 (2001).
18. B. N. J. Persson and E. Tosatti, Solid State Commun. **109**, 739 (1999).
19. G. Blatter, M. V. Feigelman, V. B. Geshkenbein, A. I. Larkin, and V. M. Vinokur, Rev. Mod. Phys. **66**, 1125 (1994).
20. M. H. Müser, Europhys. Lett. **66**, 97 (2004).
21. J. B. Sokoloff, Phys. Rev. Lett. **86**, 3312 (2001).
22. J. B. Sokoloff, Phys. Rev. B **65**, 115415 (2002).
23. F. Lançon, Europhys. Lett. **57**, 74 (2002).
24. O. M. Braun and Y. S. Kivshar, Phys. Rep. **306**, 1 (1998).
25. M. H. Müser and N. J. Mosey, Rev. Comp. Chem. (in press).
26. M. H. Müser, Tribol. Lett. **10**, 15 (2001).
27. B. Q. Luan and M. O. Robbins, Nature **435**, 929 (2005).
28. B. Luan, S. Hyun, M. O. Robbins, and N. Bernstein, Mater. Res. Soc. Symp. Proc. **841**, R7.4 (2005).
29. M. R. Falvo et al., Nature **397**, 236 (1999).
30. A. Buldum and J. P. Lu, Phys. Rev. Lett. **83**, 5050 (1999).
31. J. D. Schall and D. W. Brenner, Mol. Simulat. **25**, 73 (2000).
32. A. N. Kolmogorov and V.H. Crespi, Phys. Rev. Lett. **85**, 4727 (2000).
33. J. L. Rivera, C. McCabe, and P. T. Cummings, Nano Lett. **3**, 1001 (2003).
34. P. Tangney, S. G. Louie, and M. L. Cohen, Phys. Rev. Lett. **93**, 065503 (2004).

35. M. H. Müser, Phys. Rev. Lett. **89**, 224301 (2002).
36. Z. Xia and W. A. Curtin, Phys. Rev. B **69**, 233408 (2004).
37. J. M. Martin, C. Donnet, and Th. Le Mogne, Phys. Rev. B 48, 10583 (1993).
38. M. R. Sørensen, K. W. Jacobsen, and P. Stoltze Phys. Rev. B 53, 2101-2113 (1996).
39. Y. Qi, Y.-T. Cheng, T. Cagin and W. A. Goddard III, Phys. Rev. B **66**, 085420 (2002).
40. J. S. Ko and A. J. Gellman, Langmuir **16**, 8343 (2000).
41. Q. Zhang, Y. Qi, L. G. Hector, Jr., T. Cagin, and W. A. Goddard, III Phys. Rev. B 72, 045406 (2005).
42. A. Erdemir, Surf. Coat. Techn. **146**, 292 (2001).
43. J. A. Harrison (private ccommunication).
44. D. W. Brenner et al., J. Phys. Condens. Matt. **14**, 783 (2002).
45. G. M. Chateauneuf, P. T. Mikulski, G. T. Gao, and J. A. Harrison, J. Phys. Chem B **108**, 16626 (2004).
46. N. J. Mosey, T. K. Woo, and M. H. Müser, Phys. Rev. B **72**, 054124 (2005).
47. N. J. Mosey, M. H. Müser, and T. K. Woo Science **307**, 1612 (2005).
48. P. T. Mikulski and J. A. Harrison, Tribol. Lett. **10**, 29 (2001).
49. B. Park et al., Langmuir **20**, 10007 (2004).
50. P. T. Mikulski, G. T. Gao, G. M. Chateauneuf, and J. A. Harrison, J. Chem. Phys. **122**, 024701 (2005).
51. M. Dienwiebel et al., Phys. Rev. Lett. **922**, 126101 (2004).

11 First-Principles Atomic-Scale Study of Superlow Friction

S. Ciraci[1], S. Dag[2], O. Gulseren[1], and T. Yildirim[3]

[1] Department of Physics, Bilkent University, Ankara 06800, Turkey
ciraci@fen.bilkent.edu.tr
[2] Center of Nanophase Materials Science (CNMS) and Computer Science and Mathematics Division, Oak Ridge National Laboratory, 37830, Oak Ridge, TN, USA
[3] NIST Center for Neutron Research, Gaithersburg, MA, USA

11.1 Introduction

Friction is the resistance to the relative motion of two sliding or rolling objects imposed by nonconservative forces [1–3] and, in general, it occurs along with dissipation of mechanical energy and wear. These forces are generated from short- and long-range interactions between the sliding surfaces [4, 5]. The interaction potential can be either attractive or repulsive depending on the distance between surfaces and also on their relative lateral positions. The moving objects are either in direct contact through asperities, or gaseous, liquid or solid lubricants may be introduced between them to reduce the friction.

The dry sliding friction of two surfaces that are in direct contact through their asperities involves many interesting and complex phenomena, such as adhesion, wetting, atom exchange, the breaking and formation of bonds, as well as elastic and plastic deformation. During the relative motion, phonons are generated and electron–hole pairs are created at the expense of damped mechanical energy. Photons may even be emitted. The nonequilibrium phonon distribution generated locally is dissipated by phonon-phonon and electron–phonon coupling. Simulations of dry sliding friction between a metal asperity and an incommensurate metal surface have revealed unusual atomic processes [6]. For example, the lateral force exhibits a quasiperiodic variation with the displacement of an asperity; each period consists of two different stick–slip processes involving structural transitions. It has also been found that the perpendicular elastic deformation of the substrate that is induced by the sliding object is crucial to the energy damping associated with friction [7]. In certain conditions, due to the elastic deformation of the substrate, the corrugation of the surface potential energy can be inverted under high loading forces. This situation gives rise to the occurrence of a second state (*bistability*) in the stick–slip motion and anisotropy in the hysteresis curve [7]. It is also very well-known that the stiffer the sliding surfaces, the smaller the friction coefficient [7].

The dry sliding friction between atomically flat, commensurate or incommensurate sliding surfaces is perhaps the simplest but most fundamental type of friction in tribology. The relative motion of two commensurate surfaces can take place through repeating stick–slip stages, which can help us to visualize the energy damping under weak elastic deformation. However, the situation is rather complex if the sliding surfaces are incommensurate and undergo an elastic or plastic deformation involving atom exchange and wear. An atomic-scale analysis of the interaction between sliding surfaces is required to understand the nature of nonconservative lateral forces and the various mechanisms of energy damping. In fact, studies based on the Tomlinson's model [8] or calculations made using the Frenkel–Kontorova model [9] have revealed valuable information about the atomic processes involved with friction. Furthermore, the inventions of the atomic force microscope [10] and the friction force microscope [11, 12] have had a significant impact on the science of friction and opened up a new field called nanotribology. Nowadays, various atomic processes can be easily observed, and lateral forces in the range of a fraction of nanoNewton ($1\,\mathrm{nN} = 10^{-9}\,\mathrm{N} = 0.62415\,\mathrm{eV/\AA}$) can be measured with precision using these microscopes. As the precision of friction force measurement have increased and various atomic-scale processes have been resolved, atomic-scale simulations involving several atoms have also been performed using realistic empirical potentials [13–18]. Moreover, first-principles studies treating relatively small systems based on density functional theory (DFT) [19] have appeared [20]. First-principles studies have also led to the development of empirical potentials. Theoretical studies, on the other hand, have started to investigate microscopic aspects of energy transfer and energy damping processes [21–24].

Because of the heat generated by the dissipation of mechanical energy and material losses resulting from wear, the objects in relative motion become flawed after some operational time and are eventually destroyed. Since friction results in much resource loss, lowering the friction coefficient has been the principal goal in various fields of science and technology. Lubricants have been used to lower friction coefficients and to eliminate the wear in the machining and transportation industries. Over the last decade, progress made in materials science and surface coating technologies has led to a steady drop in the friction coefficient.

11.2 Superlow Friction

11.2.1 General Theoretical Arguments

Whether *superlubricity* – which is somewhat analogous to superconductivity (i. e., a state of matter leading to zero electrical resistivity) or superfluidity (i. e., a liquid state with zero viscosity) – can be achieved has been questioned. The absence of energy damping in mesoscopic objects has been

pointed out previously [25]. This question can be clarified by examining the energy damping agents involved in friction. These are long- and short-range interactions between two surfaces and various elementary excitations, such as phonons, electron–hole creation, charge density waves, and photon emission. High-energy excitations such as surface plasmons do not contribute to the energy damping process. Bond-breaking or rebonding, atom exchange between surfaces and local surface reconstruction can damp mechanical energy and or mediate excitations. Normally acoustic phonons with small excitation energies can easily be excited and hence contribute to the energy damping process. Experiments performed using a noncontact AFM [26] have shown that the vibration of the tip over the sample gives rise to energy dissipation even if its minimum spacing from the surface is greater than the range of the short-range forces involved [27–30]. This argument eliminates the possibility that an absolutely zero kinetic friction coefficient ($\mu_{\mathrm{k}} = 0$) can ever occur. Apparently, superfluidity with $\mu_{\mathrm{k}} = 0$ cannot be achieved, but a superlow friction coefficient is a target one can reach.

The interaction energy, $E_{\mathrm{i}}(\boldsymbol{\rho}, z)$ between two flat surfaces is a function of their spacing z and their relative lateral positions $\boldsymbol{\rho} = x\boldsymbol{i}+y\boldsymbol{j}$. Usually, the interaction energy is small and attractive (i. e., $E_{\mathrm{i}} < 0$) for large $z(z < 0)$, but decreases (becomes more attractive) as z decreases. After passing through a minimum, it then starts to increase and eventually becomes repulsive (i. e., $E_{\mathrm{i}} > 0$). The attractive interaction energy is specified as an adhesion between the two surfaces and involves the formation of bonds between the surfaces, which may give rise to a high friction coefficient during the sliding motion. Under loading forces, E_{i} can increase and become repulsive, in which case elastic and or at least local plastic deformations may occur. Substances (solid lubricants, inert gas atoms, etc.) may be placed between the surfaces in order to weaken E_{i}. Under ultrahigh vacuum conditions, friction coefficients as low as $\mu = 0.01$ have been observed for MoS_2 and diamond-like carbon (DLC) coatings [31–33]. Even if the lubrication of surfaces coated with such low friction coefficient materials appears to be desirable, the low friction coefficient can increase under different ambient and operational conditions. The coating of surfaces with special materials that result in repulsive interactions for a wide range of the loading force F_{N} is desirable. The loading force will then be balanced by the repulsive force derived from the interaction energy, $F_z(\boldsymbol{\rho}, z) = -\partial E_{\mathrm{i}}(\boldsymbol{\rho}, z)/\partial z$, and the atoms on one surface will be prevented from merging into the other surface because a large gap is maintained between the surfaces. In this way, bond-breaking, rebonding and severe deformations can be eliminated. The flights of trains over superconductive rails is reminiscent of the sliding of one coated surface over another when there is a repulsive interaction between them.

In order to reduce the energy damping during the relative motion and hence to lower μ_{k}, one must also take the force constants (which determine the vibrational frequencies of the atoms) into account. It is well-known that

the stiffer the sliding surface, the smaller the friction constant. The principal energy-damping agents are phonons, and phonons can be excited by any elastic deformation. Therefore, such elastic deformations are not favored. In this respect, coating materials comprising short and stiff surface bonds are desirable for superlow friction.

11.2.2 Recent Experimental Progress

In an effort to lower the friction coefficient, Erdemir et al. [34–36] reported superlow friction and wear between diamond-like carbon (DLC)-coated surfaces using a hydrogen-rich plasma. They achieved kinetic friction coefficients μ_k as low as 0.001 and wear rates of 10^{-9} to 10^{-10} mm^3/Nm in an inert gas environment under 10 N load and sliding velocities of 0.2 – 0.5 m/s. It has been shown that the magnitude and time-variation of μ_k are close correlated with the hydrogen content of the source gas. This work by Erdemir and his coworkers was a breakthrough in research into superlow friction and prolonged durability of moving parts in various mechanical applications ranging from the automotive industry to nanotechnology.

11.3 Theoretical Method

The structures of sliding surfaces contain several types of defects (such as asperities of different shapes and sizes, vacancies, impurities, domains, etc.). A realistic simulation of dry sliding friction must include all of these defects. Hence, atomistic models of sample surfaces require a large number of atoms. In this respect, the classical molecular dynamics (CMD) method that uses empirical potentials to represent atomic-scale interactions is convenient for simulating friction processes. Recently, various processes have been simulated and new structures have been predicted by using empirical potentials developed for certain systems. Since numerical calculations using these empirical potentials are not time-consuming, large systems comprising several thousands of atoms have been treated. However, the main drawback of CMD simulations appears when a completely new system is treated, particularly when the characteristics of the surface atoms (such as their effective charges and bonds) deviate dramatically from those of the bulk structure. Under these circumstances, whether the empirical potential can be parameterized using the bulk properties becomes questionable. On the other hand, first-principles calculations can provide reliable results for the optimized atomic structure, mechanical, electronic and magnetic properties and phonon density of states of a given system, if it involves a small number (200–300) of atoms. Various mechanisms behind the energy dissipation and estimations of the friction coefficient with upper and lower limits can be elucidated. In this respect, first-principles calculations are superior to classical methods if the system

can be represented by 200–300 atoms. Besides, first-principles methods are complementary to CDM in that they reveal the correct charge and bond structure and hence aid the development of reliable empirical potentials.

11.3.1 Details of First-Principles Calculations

In this study, the atomic processes and forces involved in sliding friction were investigated by carrying out calculations from first principles within DFT. These calculations were proven to yield accurate predictions for many metal and insulator surfaces. Here we present the crucial parameters for the first-principles calculations.

The sliding friction is treated either via a supercell method using periodic boundary conditions or by finite-size surfaces using a local basis set. In the supercell method, where the wavefunctions are expressed in momentum space,

$$\Psi_{n,\boldsymbol{k}}(\boldsymbol{r}) = \sum_{\boldsymbol{G}} a_{n,\boldsymbol{k}+\boldsymbol{G}} \exp[-i(\boldsymbol{k}+\boldsymbol{G})\boldsymbol{r}] \,. \tag{11.1}$$

The magnitude of the largest wave vector sets the cutoff energy, $\hbar^2|\boldsymbol{k}+\boldsymbol{G}|^2/2m$, and hence the number of plane waves used in the expansion. The ionic potentials are represented by ultrasoft pseudopotentials ([37]; numerical calculations were performed by using the VASP package [38]) and so the cutoff energy is taken to be 300 eV. The exchange correlation potential is represented via the generalized gradient approximation [39]. The Brillouin zone corresponding to the supercell is sampled within the Monkhorst-Pack special $\boldsymbol{k}$-point scheme [40].

The sliding surfaces are represented by two infinite slabs made from atomic layers of the coating materials. The atoms in the slabs fall into two different categories, which are treated differently. The first category of atoms, those at the back surfaces of both slabs, are kept fixed in their ideal configurations, $x_\mathrm{i}, y_\mathrm{i}, z_\mathrm{i}$. The layers of fixed atoms represent the sample or coating layers far away from the sliding surface. They are not affected from the friction process. By displacing all of the fixed atoms of one slab relative to the fixed atoms of the other slab, one can achieve a lateral displacement of two slabs and induce a loading force. The atoms at the surface region of the slabs facing each other form the second category, and are relaxed, unlike the atoms from the first category which are fixed at given $x_\mathrm{i}, y_\mathrm{i}, z_\mathrm{i}$ positions. In this way, the processes involved in the relative sliding of the two slabs, including atomic displacements, elastic and plastic deformations, etc., can be modeled from first principles. The atomic positions are optimized by a conjugate gradient method. The lateral components F_x, F_y and the perpendicular component F_z of the net force induced between the two slabs are calculated. We did not include long-range van der Waals forces since they are negligible compared to the perpendicular force F_z induced under high loading force F_N.

11.4 Atomic-Scale Study of Superlow Friction Between Hydrogenated Diamond Surfaces

11.4.1 Atomistic Model

In this section we will present our study of superlow friction between two hydrogenated diamond(001)-(2×1) surfaces performed using a first-principles plane wave method [41]. Hydrogenated DLC (H:DLC) coatings have complex, amorphous structures showing various irregularities. The sliding surfaces cannot be commensurate and they contain irregularly distributed asperities and perhaps voids. We believe that determining the structure of the DLC is itself an important goal, and this issue was addressed previously [42]. However, even if the structure of the H:DLC realized in superlow friction [34–36] as well as the physical and chemical processes associated with friction are stochastic in nature, the local bond orders and the C–H bond topology are expected to be similar to various hydrogenated diamond surfaces. Therefore, the interaction between the H:DLC surfaces and the nature of the interaction between these surfaces can be understood by the present model. Clearly, our study does not promise to provide a realistic simulation of the experiment yielding superlow friction [34]. Our objective in this atomic-scale study is to better understand the physical mechanisms involved in the superlow friction observed between hydrogenated DLC-coated surfaces [34]. We hope that the components of the superlow friction revealed in our study will be useful in the development of new coating materials that are stable under the operating conditions desired. In particular, our objective is to develop coating materials that are stable under ambient conditions and to oxidation.

When assessing how simple we can make our model, two features are of particular importance. These are the full relaxation of surface atoms at any instant of the sliding process, and the accurate calculation of the variations in the lateral force components under the constant loading force F_N. Diamond(001)-(2×1) surfaces are represented by two slabs facing each other at a specific distance. Each slab consists of six layers of carbon atoms. Carbon atoms at the back surface of each slab are saturated with hydrogen atoms. The atomic structure of each individual slab is first optimized and then the carbon atoms at the sixth layer (at the back surface of the slab) and saturating H atoms (i. e., those atoms from the first category) are kept at their equilibrium positions. We believe that such a configuration mimics the semi-infinite slab (or a thick coating).

Figure 11.1a illustrates two diamond (001)-(2×1) slabs with H-saturated, fixed back surfaces. The other surfaces of the slabs face each other and are free when the distance d between them is large. The structural parameters of the bare surface, which reconstructs to form dimer bonds, are successfully reproduced. Contour plots calculated for the self-consistent surface charge density are presented in Fig. 11.1b. The dimer and back bonds are clearly seen to have covalent character.

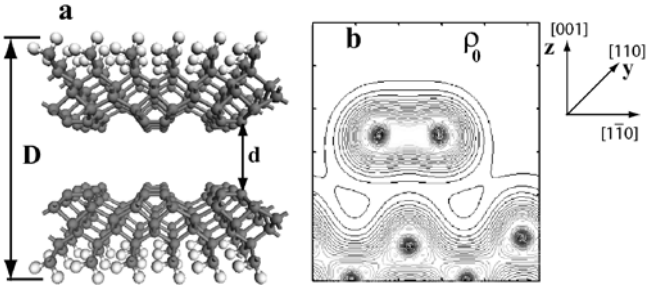


Fig. 11.1. **a** Two diamond(001)-(2×1) slabs used to model the sliding of two diamond(001) surfaces over each other. Carbon atoms at the back surfaces of the slabs are saturated with H atoms. The positions of these carbon atoms and those of the saturating H atoms are fixed at the configuration corresponding to that obtained from the optimization of individual (free) slabs. The distance between the back surfaces of the slabs is D, and that between the two sliding surfaces facing each other is d. The crystal directions are identified by Cartesian axes shown in the *inset*. C and H atoms are shown by *filled* and *empty spheres*, respectively. **b** Charge density contour plots for the bare diamond(001)-(2×1) slab for a vertical plane passing through the dimer bond. (Reproduced from [41])

We first calculated the normal force F_z which originates from the short-range interaction between the surfaces of the slabs. To this end, we kept the distance D between the back surfaces of the slabs at each preset value and calculated the total energy of whole system, $E_{\mathrm{T}}(D, \rho)$, and the total force on one of the slabs. Here the total energy and total force are obtained after optimizing the positions of atoms in the second category. We note that, since the two slabs are pressed against each other by fixing D, the calculated forces on the atoms at the back surface balance the external (loading) forces which maintain D at a preset value. Therefore, the total calculated vertical force on one of the slabs is equal to the vertical interaction force F_z. By definition, the loading force $F_{\mathrm{N}} = -F_z$. The variation of F_z is plotted in Fig. 11.2 with respect to the separation between slab surfaces before relaxation, d_o, as well as the actual separation, d, after the relaxation. The interaction is weak and repulsive for $d > 2.75\,\text{Å}$, but F_z becomes attractive as D decreases and then jumps to contact, attaining a value of approximately $-6\,\text{eV/Å}$. Strong bonds form between the sliding surfaces of two diamond(001)-(2×1) slabs near equilibrium separation corresponding to $F_z \simeq 0$. Once a normal force is applied in order to press the slabs against each other, atoms from different surfaces become close to each other at $d \sim 1.5\,\text{Å}$ and subsequently F_z becomes repulsive. Under these circumstances, since the sliding motion can involve local deformations, bond-breaking and rebonding, the dynamical friction coefficient μ_{k} as well as the wear rate are expected to be high. In fact, the dynamical friction coefficient has been measured to be equal to 0.65 for sliding DLC-coated surfaces which are free of hydrogen.

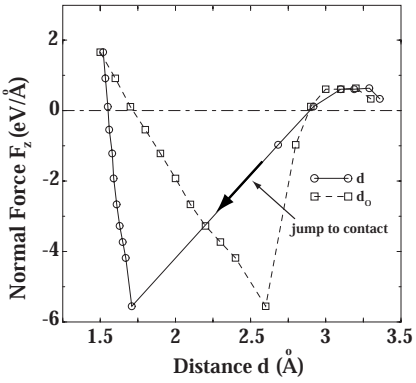


Fig. 11.2. Calculated normal force F_z generated when two diamond(001) slabs are pressed towards each other by decreasing D and hence d. d_o and d correspond to the distance between two sliding diamond(001)-(2×1) surfaces before and after relaxation, respectively.(Reproduced from [41])

11.4.2 Force Variations in the Sliding Friction of Two Hydrogenated Diamond Surfaces

First, we will examine the variation in the normal force when the sliding diamond surfaces are hydrogenated. Dangling bonds of carbon atoms on the two slab surfaces facing each other are saturated with H atoms to form a monohydride phase, i. e., H:diamond(001)-(2 × 1). Upon the saturation of the surface dangling bonds, the dangling bond surface states disappear and a wide energy gap opens between the valence and conduction bands of the slab. The surface charge density differs dramatically from that of the clean diamond(001)-(2 × 1). In Fig. 11.3 we show the atomic configurations of the H:diamond(001)-(2×1) surfaces and a contour plot of the surface charge density.

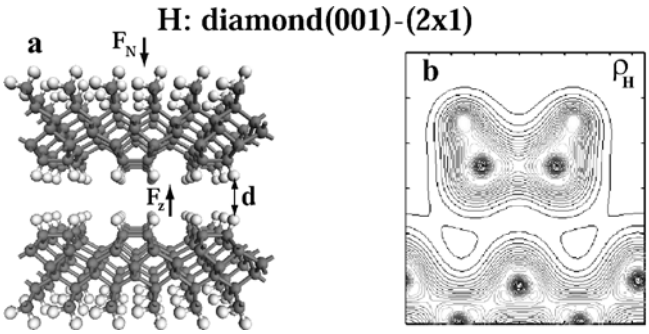


Fig. 11.3. **a** Atomic configurations of two diamond(001)-(2×1) slabs where the dangling bonds on the surfaces facing each other are saturated with hydrogen atoms to form a monohydride phase H:diamond(001)-(2×1). d is the spacing between these surfaces, F_z the normal force, F_N the loading force. **b** Contour plot of total charge density of H:diamond(001)-(2×1) surface for a vertical plane containing the surface dimer bond. (Reproduced from [41])

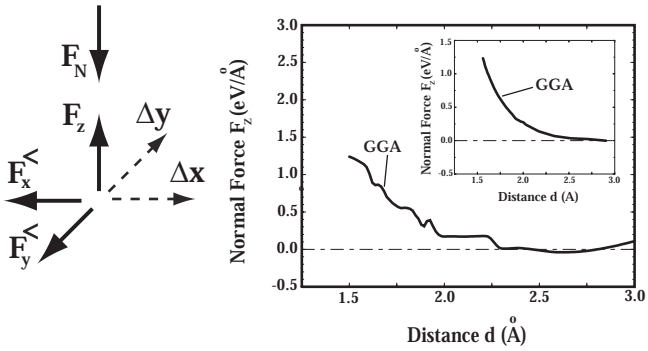


Fig. 11.4. *Left*: Directions of the loading force F_{N}, F_z, lateral force components $F_{x,y}$, and lateral displacements, Δx and Δy. The lateral force component, which acts in the opposite direction to the displacement, is indicated by the superscript "$<$" symbol. *Right*: Variation in the calculated normal force F_z between the two surfaces of H:diamond(001)-(2×1) as a function of their actual, relaxed separation d. F_z is generated when two diamond(001) slabs are pressed towards each other by decreasing D. The *inset* shows the variation in the same force between hydrogenated Si(001)-(2×1) surfaces. (Reproduced from [41])

Moreover, Mulliken analysis indicates that 0.25 electrons are transferred from the H atom to the C atom that is bound to it. This situation correlates with the fact that the C atom is more electronegative than the H atom. As a result, the H atom is positively charged. The depletion of electrons on H atoms induces a repulsive interaction and hence a repulsive F_z, even for $d < 2.5$ Å, between the H:diamond(001)-(2×1) surfaces. This is the most essential feature for obtaining superlow friction coefficients from H:DLC-coated sliding surfaces. The variation of F_z with spacing d is shown in Fig. 11.4. This repulsive force F_z keeps the sliding surfaces wide apart at a distance d and balances the loading force F_{N}. As a result, the sliding surfaces are prevented from approaching each other too closely. In this way, C–H bond deformation is suppressed to a large extent. It is interesting to note that, like carbon, silicon is also a Group IV element and has a diamond structure. A strong attractive interaction is generated between clean Si(001) slabs. However, similar to diamond(001) slabs, the attractive interaction turns repulsive upon the hydrogenation of the Si(001) surfaces, which generates a strong repulsive normal force. It appears that H:Si(001)-(2×1) displays features similar to those of H:diamond(001)-(2×1). We next examine whether this feature, namely the repulsive normal force between surfaces, can lead to a superlow friction coefficient.

11.4.3 Sliding Friction of Hydrogenated Diamond(001) Slabs

Having examined the perpendicular variation of F_z, we now address the following questions. (i) Does the repulsive interaction continue to keep the sur-

faces wide apart if one of the diamond slabs is laterally displaced relative to the other one? (ii) What is the range of F_{N} where the repulsive interaction between the surfaces persists without any serious deformation? (iii) Can one obtain an upper limit for the friction coefficient? To answer all of these questions, we carried out a series of first-principles calculations for the interaction energy E_{i}, normal force F_z, and lateral force F_{L} corresponding to different loading forces (and hence D) and displacements (Δx, Δy) of the upper slab. In these calculations, all of the atoms were relaxed except for the C and H atoms at the back surfaces of both slabs. The latter atoms are kept fixed in their ideal configurations after their planes are displaced to different perpendicular and lateral positions by varying D and $(\Delta x, \Delta y)$ in sequential increments. We note that keeping the two back ends of slabs at a specific distance D but relaxing the other atoms induces a loading force $F_{\mathrm{N}}(D)$, which in turn is balanced by F_z. F_z itself is obtained from the sum of the perpendicular components of the forces calculated for the fixed atoms on one of the slabs, namely $F_z = \sum_i F_{z,i}$, where i is the index of a fixed atom from one slab. Performing the same sum on the other slab yields F_z with the same magnitude but in the opposite direction. Similarly, the lateral forces along the x-axis (or the y-axis) are obtained from the sum $F_{x(y)} = \sum_i F_{i,x(y)}$. Performing ab initio calculations of E_{i}, F_z, $F_{x,y}$ as a function of d (or D) at different relative lateral positions yields a database of Δx and Δy values. In these calculations, the values of D were varied in small steps to yield normal forces over an appropriate range of interest.

Figure 11.5 presents the calculated variations in E_{i} and F_z as a function of d for different lateral displacements, Δx and Δy, of the top slab. Here we note that the interaction energy $E_{\mathrm{i}} = E_{\mathrm{T}} - E_{\mathrm{T},d=\infty}$, where $E_{\mathrm{T},d=\infty}$ is the total energy corresponding to very large d (or twice the total energy of one slab in the absence of the other one). We note that the variations in E_{i} and F_z are not smooth functions due to the discrete changes in D and to the relaxation of the C–H bonds. Note that since F_z is always repulsive and strong, even at significantly large spacings, so the sliding surfaces are kept apart, even for large loading forces. As a result, the C–H bonds from different surfaces neither merge nor interfere with each other. During the course of sliding, the C–H bonds experience neither significant deformation (i. e., bending, stretching or shrinking) nor wear through bonding–rebonding.

If the sliding motion were adiabatic, no energy would be damped during the sliding motion of two commensurate surfaces such as those we treat here. However, this is not the case; various rapid processes generate excitations and give rise to energy damping. However, we will delay a discussion of the microscopic theory of energy damping and dissipation to the next section and instead we now present a global approach to estimating an upper limit for μ_k using the variation in the lateral force obtained from the present calculations. To this end, we consider displacements along the $x-$ and $y-$axes and we derive the variation in lateral force under the given constant loading force using our database, namely $F_{\mathrm{L}=x}(\Delta x, \Delta y = 0, F_{\mathrm{N}})$ and $F_{\mathrm{L}=y}(\Delta x = 0, \Delta y, F_{\mathrm{N}})$.

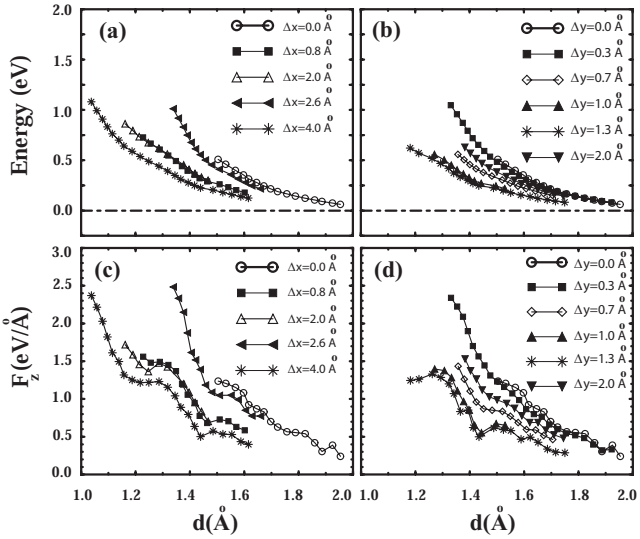


Fig. 11.5. a and **b** Variation in the calculated interaction energy E_{i} as a function of the perpendicular distance d calculated at different lateral displacements Δx and Δy. **c** and **d** The same variation but for the normal force F_z. Energy and force units correspond to per (2×1) unit cell. (Reproduced from [41])

Keeping the loading force F_{N} constant is the most difficult part of our study and requires a large number of numerical calculations corresponding to different Δx, Δy and D values. We considered that the loading force F_{N} =1 and 1.2 eV/Å per cell, which are actually values that are higher than the loading forces used in the experiment [34] and in practical applications. In this respect, our estimation of μ_k is a stringent test. The variations in F_x and F_y are illustrated in Fig. 11.6. For the reasons pointed out earlier, the variation in lateral force is not smooth. Since the lateral force is calculated using coarse displacement steps of D, the elastic deformations of the slabs and the C–H bonds induced by sliding are released suddenly. This gives rise to the stick-slip process described by Tomlinson's model [8]. Of course, there are error bars involved in the calculation of forces. In particular, achieving the constraint of a constant loading force via the limited number of data points in the database generated through ab initio calculations can lead to hysteric variations in the lateral force. Now, as an ad hoc approach to estimating μ_k in an energy-damping medium, we assume that the work done by the lateral force F_{L} (i. e., when it is parallel to the direction of motion as denoted by $F_{\mathrm{L}}^{>}$) is totally lost. Then the average friction force is calculated by $\overline{F}_{\mathrm{f}} = \int F_x^{<} \mathrm{d}x/R$, where R is the period of the motion.

Using the data in Fig. 11.6, we can extract the average friction force, $\overline{F}_{\mathrm{f}} \sim 0.05\,\mathrm{eV/Å}$ for $F_{\mathrm{N}} = 1\,\mathrm{eV/Å}$ and $\overline{F}_{\mathrm{f}} \sim 0.07\,\mathrm{eV/Å}$ for $F_{\mathrm{N}} = 1.2\,\mathrm{eV/Å}$. Then the kinetic friction coefficient is calculated from $\mu_{\mathrm{k}} = \overline{F}_{\mathrm{f}}/F_{\mathrm{N}}$ to be ap-

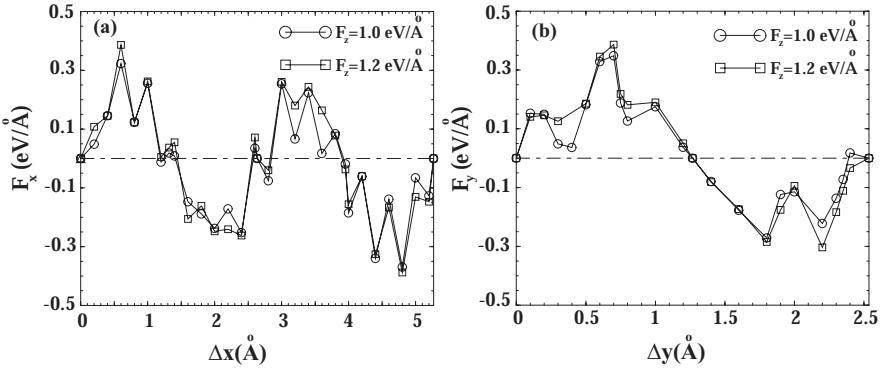


Fig. 11.6. a Variation in the lateral force, $F_{\mathrm{L}=x}$ [in eV/Å per (2×1)cell] as a function of the displacement Δx of the top slab relative to the bottom one. **b** Same as **a** but for the displacement Δy. During the course of sliding, the loading force F_{N} is taken to be approximately constant. (Reproduced from [41])

proximately 0.05 in both cases. A more realistic estimation could be obtained from $\overline{F}_{\mathrm{f}} = \int (F_x^{<} + F_x^{>})\,\mathrm{d}x/R$ if the lateral force variation was calculated precisely. Although the force variations shown in Fig. 11.6 are too crude to obtain precise values, μ_{k} has been calculated for the sake of comparison to be ~ 0.01.

It should be noted that during the sliding of commensurate surfaces the lateral forces acting on each atom or cell are added constructively to yield a high total lateral force. These lateral forces are, however, conservative, and do not give rise to energy damping if the sliding motion is adiabatic. In the case of incommensurate surfaces, the total lateral force is lower due to the cancellations. H:DLC-coated surfaces can be viewed to be incommensurate except that the disorder gives rise to higher energy damping. Consequently, the above estimation of μ_{k}, obtained from hydrogenated diamond surfaces with the assumption that all mechanical energy stored into elastic energy is damped, is an upper limit for H:DLC-coated surfaces; however, it is still too low.

11.4.4 Microscopic Theory of Energy Damping

The mechanisms of energy damping and energy transfer from sliding objects or lubricants have been studied theoretically and experimentally [14, 20–22, 43–45]. In the sliding friction of commensurate surfaces, the elastic energy (which is related to the interaction energy E_{i}) V_{T} exhibits a periodical variation with the displacement; the period is determined by the lateral lattice parameters of the sliding surfaces. In the present case, V_{T} attains its maximum value when the dimer bonds of the two surfaces face each other, but it becomes minimum when the upper slab is displaced by half of the unit cell. During sliding, V_{T} varies between maximum and minimum values. If

the variation is adiabatic there will be no energy damping; mechanical energy is stored into elastic energy mainly through the deformation of C–H bonds, and the first half of the period will be released as kinetic energy during the second half of the period. As for the lateral force with components $F_{x,y} = -\partial V_{\mathrm{T}}(\boldsymbol{r})/\partial x, y$ it will have also the same period as V_{T}. Moreover, it is conservative for fully adiabatic sliding motion. First, it is parallel to the direction of sliding, then it becomes antiparallel so that $\int F_{\mathrm{L}}\,\mathrm{d}\eta = 0$ for full-period displacement along the direction of the vector $\boldsymbol{\eta} = x\boldsymbol{i} + y\boldsymbol{j}$. However, the sliding motion is not adiabatic but instead involves sudden changes which can create various types of excitations, in particular nonequilibrium phonons, at the expense of the mechanical energy of the sliding objects.

In the sliding friction of hydrogen-saturated diamond(001)-(2×) surfaces and also H:DLC, the characteristics of the C–H bonds are crucial to the damping of mechanical energy. The C–H bonds are associated with a salient surface phonon [46] stretch mode of $\hbar\Omega_q \sim 360\,\mathrm{meV}$ and a bending mode of 150 meV. The C–H bonds are short and stiff and cannot be easily excited to large amplitude vibrations in order to dissipate mechanical energy.

Excitation of phonons over the thermal equilibrium distribution with ambient temperature T_o is the prime mechanism in energy damping. It involves two stages. (i) Excitation of nonequilibrium phonon distribution. (ii) Dissipation of excess phonons from the sample. Both processes are sample specific; in other words they depend on the sample materials, the atomic structures of the sliding surfaces and operational conditions. In particular, one needs to know the phonon frequency spectrum (or the density of states $D(\Omega)$). In principle, the density of phonon frequencies can be calculated if the atomic structures and force constants of the sliding objects are known. If the sliding takes places over a number of asperities, the situation becomes even more complex. In what follows we present a concise theory of a phononic energy damping process through a single asperity.

Let us consider a mode frequency Ω_q and denote the occupation numbers of the corresponding phonon for the ambient temperature T_o and for high temperature T as n_q^o and n_q, respectively. Here, n_q denotes Planck's distribution for a given Ω_q and T. We take $T > T_o$. Therefore, the excess phonons for this particular mode are expressed as

$$\Delta n_q = n(\Omega_q, T) - n^o(\Omega_q, T_o)\,. \tag{11.2}$$

Here q is the mode index, including the polarization. The equilibrium state of the sample (or asperity) can be expressed by the occupation number representation as $\Psi(n_1^o, n_2^o, \ldots n_q^o, \ldots n_{3N}^o)$, with $3N$ being the degrees of freedom in the motion of N atoms of the system. Subsequent to a deformation, the occupation number state becomes $\Psi(n_1, n_2, \ldots n_q, \ldots n_{3N})$. Here, the crucial problem is to relate the excess phonon density, Δn_q, to the deformation. In principle, a given deformation state in terms of the displacements of individual atoms, $u_{i,x}$, $u_{i,y}$ and $u_{i,z}$, can be expressed in normal coordinates by

using the appropriate transformation. In other words, this problem reduces to finding the vibrational states of a system (consisting of N atoms connected by springs) when its preset deformation is suddenly released. In order to provide a fundamental understanding of the phononic dissipation, one can follow a simpler approach and relate the deformation u_q to the number of excited phonons by using a semiclassical equation,

$$\delta V_q = \sum_q M \Omega_q^2 u_q^2 / 2 \,. \tag{11.3}$$

Here, δV_q is the mechanical energy used to excite Δn_q excess phonons with frequency Ω_q, namely $\delta V_q = \hbar \Omega_q \Delta n_q$. We note that the amount of energy damped by phonons upon the release of one of the deformed states is the sum of the phonon excitation energies over the mode index q, $V_{\mathrm{T}} = \sum_q \delta V_q$. Knowing $n_q = \Delta n_q + n_q^o$, and using Planck's distribution, one can estimate the local temperature T corresponding to the excited phonons.

Having calculated the excited phonons subsequent to the release of one deformed state, we now discuss the dissipation of excess phonons. The process is closely related to the transfer of energy through molecules and has been treated in several theoretical and experimental studies. The decay of n_q to n_q^o is usually expressed by a rate equation,

$$n_q(t) = n_q^o + n_q(t=0) \exp[-R(\Omega_q)] \,. \tag{11.4}$$

Here $R(\Omega_q)$ is sample-specific and obtained from scaling arguments. Model calculations on a Cu asperity consisting of 14 atoms by Buldum et al. [22] showed that low-energy modes experience the highest excitation probabilities but the lowest decay rates. Therefore, low-energy modes determine the phononic energy damping. Moreover, calculations based on nonequilibrium statistical mechanics and Keldysh Green's function formalism [24] show that the excess phonon distribution dissipates within a picosecond if the couplings to the substrate are strong.

11.4.5 Effect of Oxidation

The most serious issue is that the superlow friction coefficient obtained from H:DLC-coated surfaces cannot be sustained under ambient conditions [34,35]. The oxygen atom could potentially destroy the superlow friction when the H:DLC coating is exposed to the air. In what follows, we clarify the effect of oxygen on the hydrogenated DLC coating leading to superlow friction. To test the effect of oxygen, we placed O atoms at different sites on the H:diamond(001)-(2×1) surface. Upon relaxation, the system attains the minimum energy configuration, whereby O atoms break surface bonds to form new C–O–C or C–O–H and C–O bonds, and hence they become attached to the surface. Fortunately, they attack the C–H bonds to form C–O–H radicals. Charge transferred to O from H and C makes the O atom negatively

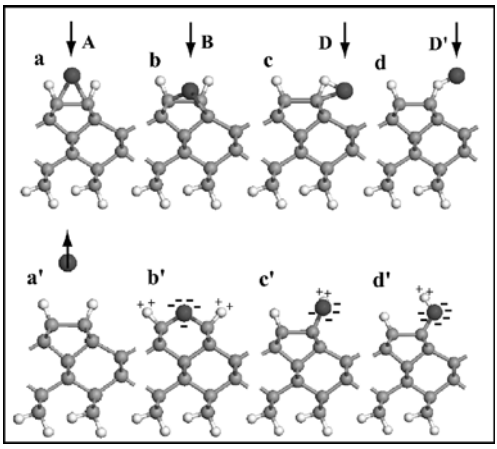


Fig. 11.7. Calculated atomic configurations showing the effect of an oxygen atom on the H:diamond(001)-(2×1) surface. **a**–**d** The oxygen atom is placed at different sites in the surface unit cell before the relaxation of the surface. **a'**–**d'** Atomic structure and bonding after relaxation of the system. The adsorbed oxygen atom is negatively charged. (Reproduced from [41])

charged, as shown in Fig. 11.7. In this way, the interaction between two atoms on different surfaces can be attractive when they carry charges of different polarities. As a result, the steady and strong repulsive interaction between the H:diamond(001)-(2×1) surfaces gradually becomes weaker or turns attractive. Eventually, the superlow friction ends.

11.5 Conclusions

This work reports on an extensive study of the interaction between two bare and hydrogenated diamond(001)-(2×1) surfaces. The interaction between bare surfaces is strongly attractive up to a small distance between surfaces $d \sim 1.5$ Å, at which point it becomes repulsive. Hydrogen atoms donate charge to the carbon atom and become positively charged. This appears to be the most important ingredient of the superlow friction. The repulsive interaction persists at any relative position of the sliding surfaces, and is strong even at large distances from each other, preventing C–H bonds from merging. Strong and stiff C–H bonds and the stiff diamond crystal prevent large amounts of energy from being dissipated. It was found that oxygenation of surfaces under atmospheric conditions destroys the steady repulsive interaction.

Acknowledgement. This work was supported by Scientific and Technological Council of Turkey, TÜBİTAK under Grant No. TBAG-104T537.

References

1. F.P. Bowden and D. Tabor, *Friction and Lubrication* (Methuen, London, 1965).
2. E. Rabinowitz, *Friction and Wear* (Wiley, New York, 1965).

3. B.N.J. Persson, *Sliding Friction: Physical Principles and Applications* (Springer-Verlag, Berlin, 2000).
4. J.N. Israelachvili, *Intermolecular and Surface Forces* (Academic Press, London, 1985).
5. S. Ciraci, A. Baratoff and I.P. Batra, Phys. Rev. B **42**, 7168 (1990); S. Ciraci, E. Tekman, A. Baratoff and I.P. Batra, Phys. Rev. B **46**, 10411 (1992).
6. A. Buldum and S. Ciraci, Phys. Rev. B **55**, 12892 (1997).
7. A. Buldum and S. Ciraci, Phys. Rev. B **55**, 2606 (1997).
8. G.A. Tomlinson, Philos. Mag. **7**, 905 (1929).
9. J. Frenkel and T. Kontorova, Phys. Z. Sowjet, **13**, 1 (1938).
10. G. Binnig, C.F. Quate and Ch. Gerber, Phys. Rev. Lett. **56**, 960 (1986).
11. C. M. Mate, G.M. McClelland, R. Erlandsson and S. Chiang, Phys. Rev. Lett. **59**, 1942 (1987).
12. E. Meyer, R. Overney et al., Phys. Rev. Lett. **69**, 1777 (1992)
13. A. P. Sutton, JB. Pethica, J. Phys.: Condens. Matter **2**, 5317 (1990).
14. M. Cieplak, E. D. Smith and M. O. Robins, Science **265**, 1209 (1994).
15. B. Bhushan, J.N. Israelachivili and U. Landman, Nature, **347**, 607 (1995).
16. M. R. Sorensen, K. W. Jacobsen and P. Stoltze, Phys. Rev. B **53**, 2101 (1996).
17. A. Buldum and S. Ciraci, Phys. Rev. B **57**, 2468 (1998).
18. A. Buldum and S. Ciraci, Phys. Rev. B **60**, 1982 (1999).
19. P. Hohenberg and W. Kohn, Phys. Rev. B **136**, B864 (1964); W. Kohn and L.J. Sham, Phys. Rev. **140**, A1133 (1965).
20. W. Zhong and D. Tomanek, Phys. Rev. Lett **64**, 3054 (1990); D. Tomanek, W. Zhong and H. Thomas, Europhys. Lett. **15**, 887 (1991).
21. A. Buldum, D.M. Leitner and S. Ciraci, Europhys. Lett. **47**, 208 (1999).
22. A. Buldum, D.M. Leitner and S. Ciraci, Phys. Rev. B **59**, 16042 (1999).
23. A. Ozpineci, D. M. Leitner and S. Ciraci, Phys. Rev. B **62**, 10558 (2000).
24. A. Ozpineci and S. Ciraci, Phys. Rev. B **63**, 125415 (2001).
25. J.B. Sokoloff, Phys. Lett. **71**, 3450 (1993).
26. F.J. Giessibl, Science **267**, 68 (1995).
27. M. Bammerlin, R. Lüthi, E. Meyer, A. Baratoff, J. Lu, M. Guggisberg, Ch. Gerber, L. Howald, and H.-J. Güntherodt, Probe Microsc. **1**, 3 (1997).
28. M. Gauthier and M. Tsukada, Phys. Rev. B **60**, 11716 (1999).
29. M. Guggisberg, M. Bammerlin, R. Lüthi, Ch. Loppacher, F. Battiston, J. Lu, A. Baratoff, E. Meyer and H.-J. Güntherodt, Appl. Phys. A **66**, S245 (1998).
30. L.N. Kantorovich, Phys. Rev. B **64**, 245409 (2001).
31. J.M. Martin, C. Donnet Th. Le Mogne, and Th. Epicier, Phys. Rev. B **48**, 10583 (1993).
32. I. Sugimoto and S. Miyake, Appl. Phys. Lett. **56**, 1868 (1990).
33. C. Donnet and A. Grill, Surf. Coat. Technol. **94-95**, 456 (1997).
34. A. Erdemir, O.L. Eryilmaz and G. Fenske, J. Vac. Sci. Technol. A **18**, 1987 (2000).
35. A. Erdemir, Surf. Coat. Technol. **146-147**, 292 (2001).
36. J.A. Heimberg, K.J. Wahl and I.L. Singer and A. Erdemir, Appl. Phys. Lett. **78**, 2449 (2001).
37. D. Vanderbilt, Phys. Rev. B **41**, 7892 (1990).
38. Numerical calculations were performed by using the VASP package: G. Kresse and J. Hafner, Phys. Rev. B **47**, 558 (1993); G. Kress and J. Furthmüller, *ibid* **54**, 11169 (1996).

39. J.P. Perdew, J.A. Chevary, S.H. Vosko, K.A. Jackso, M.R. Pederson, D.J. Singh and C. Fiolhais, Phys. Rev. B **46**, 6671 (1992).
40. H.J. Monkhorst and J.D. Pack, Phys. Rev. B **13**, 5188 (1976).
41. S. Dag and S. Ciraci, Phys. Rev. B **70**, 241401 (2004).
42. G.T. Gao, P.T. Mikulski, and J.A. Harrison, J. Am. Chem. Soc. **124**, 7202 (2002); G.T. Gao, P.T. Mikulski, G.M. Chateauneuf and J.A. Harrison, J. Phys. Chem. B **107**, 11082 (2003).
43. L.S. Levitov, Europhys. Lett. **8**, 449 (1989).
44. B.N.J. Persson, Phys. Rev. B **44**, 3227 (1991); B.N.J. Persson and A.I. Volotkin, J. Chem. Phys. **103**, 8679 (1995).
45. A. Dayo, W. Alnasrallah, and J. Krim, Phys. Rev. Lett. **80**, 1690 (1998).
46. B.D. Thoms and J.E. Butler, Phys. Rev. B **50**, 17450 (2000).

12 NanoMechanics: Elasticity in Nano-Objects

Lina Merchan, Robert Szoszkiewicz, and Elisa Riedo

Schol of Physics, Georgia Institute of Technology, 837 Sate Street, Atlanta, Georgia 30332, USA

12.1 Introduction

Nanotubes [1], oxide nanobelts and rods [2], and semiconductor nanowires [3] are structures with dimensions of the order of a nanometer. These have exciting technological potential in applications that include integrated nanoelectronic and photonic circuits, nano-sensors, interconnects and electromechanical nanodevices. All of these applications require knowledge of, and the ability to control the mechanical behavior of the relevant nano-materials. For example, the assembly of nanowires and nanotubes between electrodes requires a balance of rigidity and strength. In addition, the electronic properties of nanotubes and nanowires are strongly affected by mechanical deformation [4, 5].

The study of the elastic properties of nano-objects (N-O) is very challenging both from an experimental and theoretical point of view. From the experimental side, investigating the mechanical properties of nano-objects is critical because of the lack of reliable methods to quantitatively measure the elasticity and sometimes the friction at the nanoscale. The problems are related to spatial and force resolution, instrument calibration as well as not well defined surface shape and chemistry (and at this scale each atom makes a difference). From the theoretical side, developing a theory of elasticity at the nanoscale is an intriguing theoretical challenge, which lies at the cross-over between the atomic level and the continuum.

Nowadays, new powerful methods have been developed to study the elastic properties of nano-objects, such as atomic force microscopy methods and optical tweezers. The development of these experimental methods has opened a new world of possibilities by allowing us to measure elastic properties at the nanoscale. In particular, biology has seen its horizons broaden with applications ranging from studying the DNA to understanding diseases such as cancer and heart disease.

This chapter is organized in three sections. In the first section we describe the theoretical background necessary to study the elastic properties of nano-objects and to interpret the experimental results. In the second section, we report the state of the art experimental methods to investigate the elastic properties at the nanoscale. Finally in the last section we discuss the most

recent findings related to the elastic properties of inorganic nano-objects and biological samples.

12.2 Theories of Contact

We are going to present a brief summary of some analytical approaches describing the contact between two bodies. There are two extreme cases, the ideally elastic and the ideally plastic, and most surfaces are a combination of both. An ideally elastic surface regains its original shape after being subjected to a deformation, while an ideally plastic surface cannot recover from the deformation.

From now on, we will neglect the plastic behavior of the two bodies and assume that one or both bodies in contact are elastic. In order to relate the indenter and sample deformations to the applied force we have to learn about several analytical approaches to contact theory. We will present a brief compilation of classical continuum mechanics taken in part from a review article [6].

12.2.1 Hertz and Sneddon

Hertz theory dates back to 1882 [7]. This approach assumes that surfaces forces and adhesion can be neglected. When one body is exerting pressure on another one, as in an atomic force microscope (AFM), there are various combinations of behaviors depending on the relative stiffness of the two bodies. We can look at some common cases:

Tip as an Elastic Sphere Versus a Rigid Flat Surface

This theory is commonly used for indentations of a few angstroms or nanometers. In this case the sphere suffers a deformation δ as it is pressed against the surface with a force F and pressure $P(y^*)$. The contact area is a circle of radius a, and the distance between this circle and the center of the sphere is y. Expressions relating the contact radius and the applied load are the following:

$$a = \left(\frac{3RF}{4E^*}\right)^{1/3}, \tag{12.1}$$

$$F = \frac{4}{3}E^* R^(1/2)\delta^{\frac{3}{2}} \tag{12.2}$$

$$P(y^*) = \frac{3F\sqrt{1-y^{*2}}}{2\pi a^2} \tag{12.3}$$

where $y^* = y/a$, and the reduced Young modulus E^* is given by

$$\frac{1}{E^*} = \left(\frac{1-\nu^2}{E} + \frac{1-\nu_i^2}{E_i}\right) \tag{12.4}$$

E and E_i are the Young modulus, and ν and ν_i are the Poisson ratios for surface and indenter respectively. From Eq. (12.1), we can see that the relation between F and δ is a simple analytical formula.

If we want to assume the indenter to be a rigid body and the surface to be a soft elastic medium, we should employ the Sneddon analysis which gives the next case.

Tip as a Rigid Sphere Versus a Soft Elastic Surface

The force and the elastic surface deformation are given by a transcendental equation that can be computed numerically

$$F = \frac{E^*}{2}\left[(a^2+R^2)\ln\left(\frac{R+a}{R-a}\right) - 2aR\right] , \tag{12.5}$$

$$\delta = \frac{1}{2}a\ln\left(\frac{R+a}{R-a}\right) . \tag{12.6}$$

Tip with an Axis-Symmetric Shape Versus a Soft Elastic Surface

For a more general geometry, with symmetry along an axis perpendicular to the surface,

$$F = \frac{3}{2}E^* a\int_0^1 \frac{x^2 f'(x)}{\sqrt{1-x^2}}\,\mathrm{d}x . \tag{12.7}$$

$$\delta = \int_0^1 \frac{f'(x)}{\sqrt{1-x^2}}\,\mathrm{d}x , \tag{12.8}$$

where $f(x)$ describes the indenter's profile. This is very useful if the true geometry of the tip is known.

Tip as a Rigid Cone Versus a Soft Elastic Surface

For deep indentations, where the tip can no longer be modelled as a sphere, pyramidal tips have been modelled as cones [8,9]. This approach is appropriate when studying biological samples much softer than the cantilever tip and when the indentation on the surface is more than 10 nm [10]. The relations between the indentation δ, the loading force F, the diameter of the contact area d are

$$F = \delta^2\frac{\pi}{2}\frac{E}{(1-\nu^2)}\tan(\alpha) , \tag{12.9}$$

$$d = 2F\sqrt{\frac{2\pi}{\tan(\alpha)}\frac{(1-\nu^2)}{E}} , \tag{12.10}$$

$$\delta = \sqrt{\frac{k}{\frac{\pi}{2}\frac{E}{(1-\nu^2)}\tan(\alpha)}}\sqrt{d} , \tag{12.11}$$

where E is the sample Young modulus, ν is the sample Poisson ratio and α is the opening of the cone. In reality, the tip is somewhat blunted, so a blunted cone model has been developed for a wide range of indentations [11,12]. Since the actual geometry of pyramidal tips is not a blunted cone but a blunted pyramidal, Rico et al. [13] developed a contact model for a blunted pyramidal tip. The blunted pyramidal model agrees with the spherical model with a radius equal to the spherical cap, for indentations where the sample only contacts the spherical cap, and at deeper indentations the blunted pyramidal model approaches the ideal pyramidal model. Rico and coworkers found that for soft agarose gels probed with an AFM tip the Young modulus varied with indentation until it reached a constant value at depths larger than $\sim 300\,\text{nm}$. At deep indentations, the Young modulus and the shear modulus measured with the sphere was two times smaller than with the pyramid.

When surface forces, such as friction and adhesion, are negligible Hertz and Sneddon, mentioned previously, can be used. The (DMT) Derjaguin–Müller–Toporov theory expands Hertz by including only long range (van der Waals) forces between the bodies in the contact region [14]. DMT is applicable to AFM experiments where the cantilever tip is small and the adhesion between the surfaces is small. JKR theory, named after Johnson, Kendall and Roberts, takes into account only short range forces inside the contact region and neglects long range forces outside of it [15]. For AFM experiments, this theory can be employed for highly adhesive systems when the tip has a low stiffness and large radius. A very complete contact mechanics theory was developed by Maugis and Pollock and it deals with a wide range of behaviors [16]. By varying a single parameter this theory can go continuously from DMT to JKR and can be applied to compliant, large and adhesive bodies as well as small rigid materials with low adhesion. There is an enormous bibliography on analytical and numerical studies about contact between different geometries ([17,18]).

All these theories assume no plastic deformation and no viscoelastic phenomena. To treat this range of behaviors there are finite element studies for bodies of different geometries in contact ([18–20]). All surfaces are assumed to be ideally flat as well as frictionless. In reality, surfaces are rough at the nanoscale and friction is inherent to all surface contacts. There is a lingering question as to whether Hertz theory and its derivatives should be used at the nanoscale. At what point does the discreteness of matter start to play an important role such that continuum mechanics can no longer be used? At some length scale, the real surface topography cannot and should not be assumed as ideally smooth. The existence of roughness complicates tremendously the calculations of the real area of contact because the contact occurs at several points where the asperities of the surfaces meet. How to incorporate friction to contact mechanics has been studied for a long time. One of the first multi-asperity models was done by Kragelsky [21]. He modelled a rough surface as an assembly of rods of different lengths fixed on one extreme to a rigid base. Later, Greenwood and Williamson, assumed the surface to be covered

by hemispherical asperities all with the same radius [22]. Whitehouse and Archard extended that model by allowing random radii of curvature. Others improved on this model, incorporating the notion of volume conservation when asperities are deformed upon contact. Nayak introduced random process theory to the study of rough surfaces by characterizing them with parameters such as the distribution of asperities heights, the density of asperities, the mean surface gradient and the mean curvature of asperities [23]. Ogilvy predicted the friction force between rough surfaces numerically avoiding the need to define the exact geometry of the asperities by generating a surface with a Gaussian distribution [24]. There is an extensive research on stochastic methods to treat contact among rough surfaces ([25,26]). A review on the historical and modern understanding of friction was written by Gao et al. [27].

This approach has limited applicability to real life because it simplifies the geometry of the asperities, assumes a Gaussian distribution for the height of the asperities and neglects the interactions between adjacent asperities. Since it assumes all asperities to be of the same length scale it oversimplifies the problem of real surface roughness. The statistical parameters have the added problem of depending strongly on resolution and sampling length of the instrument used to measure the roughness so they do not uniquely characterize a surface. If plasticity is considered, several asperities can deform and give rise to larger contact spots and smaller asperities can also come into contact.

Some microscopic measurements of roughness have shown fractal structure in different surfaces from the nanometer to the millimeter scale. Majumdar and Bhushan developed a model of contact between isotropic rough surfaces based on the scale-independent fractal roughness parameters [28]. Using a power-law relation for the size-distribution of the contact spots, they related the real contact area of an elastic deformation to the load for a fractal dimension between one and two. Their study predicted that contact spots under certain size were in plastic contact and these plastically behaved spots would merge into elastic spots under larger loads. Zahouani extended the 2D methods to 3D rough surfaces [29]. Yan and Komvopoulos introduced a two-variable fractal surface description in a 3D elastic-plastic contact mechanics analysis [30]. Instead of using the power law relation to predict the number of contacts, Chung and Lin developed expressions for size distribution functions for the elastic, elastoplastic, and fully plastic deformations [31].

Another approach to include fractals into contact mechanics is that of Persson. He developed a theory of contact mechanics between randomly rough surfaces valid also for self-affine fractal surfaces [32]. A self affine fractal surface has the property that its morphology does not change under different scale changes along different directions. He assumed the bodies to deform elastically under a certain yield stress, and plastically above it. When an elastic medium is pressing on a substrate with roughness at different length scales, it will tend to fill out the small cavities at the top of the large asperities, but will not do the same at the bottom of a large cavity where the load

is smaller. This will happen successively as the magnification increases. He showed that the area of real contact is proportional to the load.

This theory does not include adhesion, but it is quite important in elastically soft objects where it may pull the two surfaces in contact over a large region of the nominal contact area. There seems to be a very interesting relationship between adhesion and roughness, as Tabor and Fuller found in 1975 [33]. For a simple model with roughness at one length scale, they found that a small roughness can completely remove adhesion. Persson and Tosatti studied the adhesion among surfaces with roughness on many length scales and fractal self-affinity [34]. They found that for surfaces with fractal dimension greater than 2.5, the adhesion force may vanish or be largely reduced. They studied the pull-off force when partial and complete contact occurred in the nominal surface area. For a more in depth treatment refer to chapter "A multiscale approach to adhesion and friction: from continuum mechanics to molecular dynamics".

Buzio et al. [35] have approached this topic from the experimental side by studying the roughness of self-affine fractal surfaces with the AFM. They found that by varying the applied load the average frictional force followed a power law behavior in the single asperity regime and a linear behavior in the multi-asperity regime. Another interesting feature they found was that the average indentation depth depended strongly on the fractal parameters of the surface. A detailed overview can be read in chapter "The role of nanoroughness in contact mechanics and friction".

Even considering the roughness at all length scales, surfaces are still being treated as continuous down to atomic dimensions. There are some molecular dynamics simulations studying when continuum mechanics break down. Recently, Luan and Robbins [36] did molecular dynamics simulations to test whether Hertzian theory is still valid at the nanoscale. They found that the atomic discreteness at the bulk does not seem to affect as much as the atomic discreteness at the surface, which gives rise to the roughness. They conclude that continuum mechanics may underestimate the area of AFM contacts by up to 100% at small loads, though the error decreases with increasing load; and radius and the yield stresses may also be underestimated by a factor of 2 or more. On the other hand, the friction and contact stiffness may be overestimated.

Miesbauer et al. [37] did molecular dynamics simulations of the contact between two NaCl nanocrystals. They studied the adhesion and deformations between two cuboids, and the forces and deformations between a hemisphere and a cuboid. After comparing with the Hertzian theory, they concluded that this theory could be employed to describe systems with sizes down to 5 nm.

For more details refer to chapters "A multiscale approach to adhesion and friction: from continuum mechanics to molecular dynamics" and "The Role of nanoroughness in contact Mechanics".

12.3 Experimental Methods

So far, measurements of local elastic properties have been addressed by both static and dynamic methods. The static methods include force-distance curves provided by an atomic force microscope, nanoindentation tests, optical and magnetic tweezers (and wrench), as well as various local realizations of the classic triple point contact tests. The dynamic techniques utilize AFM tips vibrating either at low (a few kHz) or ultrasonic frequencies. Elastic properties of some nanoscopic objects (i. e. nanotubes, nanorods) were also reported based on in situ observations of their vibrational frequencies. Depending on the shape of the investigated objects and type of elastic property studied (tensile, shear or radial) these techniques are reported to work either in compression, bending or tension. Depending on the compliance of the investigated samples a wide choice AFM cantilevers is used, along with nano-indentation systems and recently optical and magnetic tweezers (and wrench). Below we describe these techniques in some detail.

12.3.1 Static Experimental Methods

Force-Distance Curves Provided by an AFM

As already described, a sum of interaction forces F between atomic-size contacts in an elastic limit is approximated by the continuum contact mechanics, which assumes frictionless contact, and works best in the case of small roughness. The outcome is a power law relating F with indentations δ. In proximity of a given indentation point the forces do not change too much, and in an analogy to the Hook's law of elastic deformation, it has become very convenient to introduce the normal contact stiffness k_{n} defined as:

$$k_{\mathrm{n}} = \frac{\partial F}{\partial \delta} \,. \tag{12.12}$$

In the case of the Hertz contact mechanics [38] the value of k_{n} is a measure of local elasticity because:

$$k_{\mathrm{n}} = 2aE^* = (6FRE^{*2})^{1/3} \,. \tag{12.13}$$

Here, a is the contact radius, E^* is the reduced Young modulus of the tip and the sample, F is the load force (equalling as well the sum of all the interaction forces) and R is the reduced tip-sample curvature radius (see Sect. 12.2.1). The force-distance curves are routinely measured by an AFM (see Fig. 12.1), and an extensive review was provided by Cappella and Dietler [6]. The major trouble with these curves is, however, in their calibration. The forces must by calibrated from volts into newtons via a two step procedure. First, the raw force in volts is multiplied by a vertical sensitivity factor (nm/V) representing

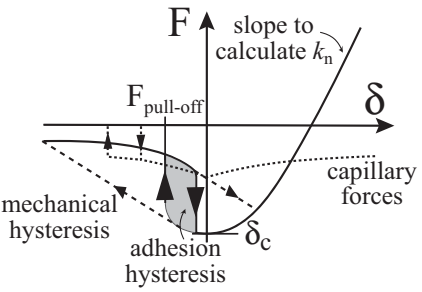


Fig. 12.1. Typical force-distance curve between an AFM tip and a sample surface

the amount of cantilever bending (in nanometers) per volt of a raw force signal. Next, the force is multiplied by a cantilever spring constant. The values of δ are obtained by subtracting the cantilever position change from the initially measured scanner displacement. The errors in calibrations come from many sources: detection of the cantilever's position and exerted upon it forces, coupling between torsion and vertical cantilever's bending, hysteresis and resonances of a piezoscanner, interference between light beams reflected from the lever and the surface, etc. Furthermore, Heim et al. [39] as well as Hutter [40] reported recently on some systematic errors in force-distance curves due to usual cantilever tilt with respect to the surface.

Many instrumental developments have been pursuit in order to improve the force-distance sensitivity, which include: (A) use of segmented scanners with lower hysteresis (i. e. Molecular Imaging), (B) separation and minimization of X, Y, Z scanners to obtain better S/N ratio (i. e. Asylum Research, PSIA), (C) use of low-coherence lasers to minimize any interference patterns on the force-distance curves (many commercial systems). Still however, the comparative force-distance measurements with the same AFM configuration are the most meaningful and trusted.

Despite the calibration issues, state-of-the art force-distance curves were obtained with proper scanner's creep calibration, as i. e. reported by Li et al. [41] in unambiguous detection of confined water layers on glass, mica and HOPG. From the onset on a plastic deformation on the force-distance curves between diamond tips and a Si surface, the local hardness was obtained by Manyes et al. [42] with indentations smaller than one nm. By pushing the electronics to the limits and building scanners with much higher than usual resonance frequencies, Rost et al. [43] shown that force-distance imaging can be obtained with video rates. In order to minimize the errors in force-distance curves due to poorly defined contact there are also attempts to use carbon nanotubes, micrometer-size spheres, and focused electron beam (FEB) or focused ion beam (FIB) in order to create custom cantilever's geometries [44, 45].

AFM-Based Methods for Measuring the Axial Elasticity of Wire-Like Nano-Objects

There are several techniques that have been developed for measuring the axial elasticity of single wire-like nano-objects. The technique demonstrated by Lieber et al. [46] was based on quantifying the deflection of a carbon NT that was affixed at one end and the other end was free to be deflected by an AFM tip. The NT was laid in parallel to a solid substrate, and the elastic modulus of a carbon NT was calculated from the force-deflection curve.

The technique by Wang et al. [47], relied on the electromechanically resonance of a NT/NW by in situ transmission electron microscopy (TEM). The resonance was stimulated by applying an AC voltage across two electrodes, one of which was a carbon NT that was glued to a metal tip affixed on a specimen holder. The resonance frequency together with the geometrical parameters of the NT provided by TEM yielded the elastic modulus.

The technique of Yu et al. used two AFM tips to stretch a carbon NT that was glued at both ends to the two tips, respectively; the stretching force-displacement curve gave the tensile strength and elastic modulus.

Another interesting method was developed by Salvetat et al. in Ref. [48]. The N-Os was deposited on a porous substrate. On such a substrate, the N-Os occasionally lie over the pores. Attractive interactions between the tubes and the substrate clamped the tubes to the substrate. An AFM tip was then used to apply a force and to measure the resulting deflection of the tube/belt.

For all of these techniques, the NT/NWs have to be removed from the substrate used in the growth and are manipulated for the measurements.

Another method to study the axial elasticity of NT/NW vertically aligned on a substrate without destroying and/or removing the NT/NW has been recently demonstrated [49]. The elasticity of vertically aligned ZnO nanowires has been measured by simultaneously acquiring the topography and lateral force image of the aligned nanowires in AFM contact-mode. The measurement is based on quantifying the lateral force required to induce the maximal deflection of the nanowire where the AFM tip was scanning over the surface in contact mode (see Fig. 12.2). For the [0001] ZnO nanowires grown on a sapphire surface with an average diameter of 45 nm, the elastic modulus is measured to be 29 ± 8 GPa. This technique can be applied to any as-grown one-dimensional nanomaterials that are aligned on a solid substrate.

Nanoindentation

Indentation tests are perhaps the most commonly applied means of testing the mechanical properties of materials. In such a test, a hard tip, typically a diamond, is pressed into the sample with a known load. After some time, the load is removed. The area of the residual indentation in the sample is measured and the hardness, H, is defined as the maximum load, $F_{\max}$, divided by the residual indentation area, A_{r}, or $H = F_{\max}/A_{\mathrm{r}}$.

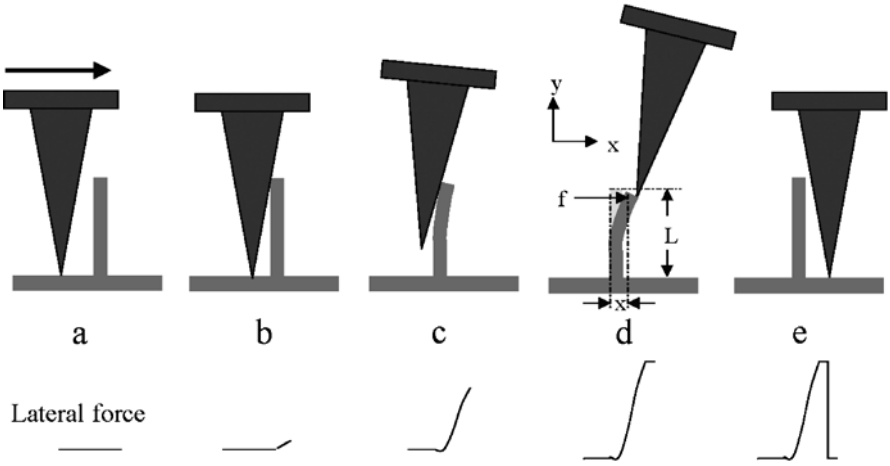


Fig. 12.2. By measuring simultaneously the topography and the lateral force when an AFM tip scans over a forest of vertically aligned nanorods, it is possible to derive the axial modulus of each individual nanorod

The idea of nanoindentation arose from the realization that an indentation test is an excellent way to measure very small volumes of materials. In principle, if a very sharp tip is used, the contact area between the sample and the tip, and thus the volume of material that is tested, can be made arbitrarily small. The only problem is determining the indentation area. It is easy to make an indentation that is so small that it is difficult to see without a powerful microscope.

To solve this problem depth sensing indentation methods were developed. In this method, the load and displacement of the indenter are recorded during the indentation process and these data are analyzed to obtain the contact area, and thereby mechanical properties, without having to see the indentations [50, 51].

By far the most successful and widespread model for nanoindentation data analysis is one in which the unloading data are assumed to arise from a purely elastic contact (Hertzian contact). This approach was developed over 40 years with contributions from a number of groups around the world. The form most often used is that presented by Oliver and Pharr, and is known as the Oliver and Pharr method.

The basic assumptions of this approach are: a) Deformation upon unloading is purely elastic, b) the compliance of the sample and of the indenter tip can be combined as springs in series, and c) the contact can be modelled using an analytical model for contact between a rigid indenter of defined shape with a homogeneous isotropic elastic half space. Following this model the reduced elastic modulus E^*, as defined in Eq. (12.1) can be estimated using

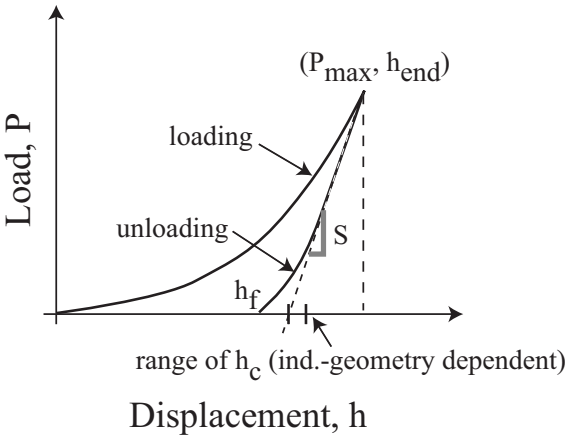


Fig. 12.3. Loading and unloading curves during a nanoindentation experiment

the relationship:

$$E^* = S\frac{\sqrt{\pi}}{2\sqrt{A}} \tag{12.14}$$

where S is the slope of the unloading data at maximum load $F_{\max}$ and A is the contact area (see Fig. 12.3). If the shape of the indenter is known, it is possible to calculate the contact area as a function of the indentation, h, and then calculate the elastic modulus using Eq. (12.14).

Nanoindentation refers to depth-sensing indentation testing in the submicrometer range and has been made possible by the development of 1) machines that can make such tiny indentations while recording load and displacement with very high accuracy and precision, and 2) analysis models by which the load displacement data can be interpreted to obtain hardness, elastic modulus, and other mechanical properties.

All commonly used nanoindenters are "soft" load controlled machines. This means that the load is applied in such a way that small changes in displacement do not change the force significantly. A number of methods for high-resolution force actuation have been developed. One of the earliest and most common means of applying force is the electromagnetic actuation, introduced by Pethica in 1981 [52]. Other kind of force actuation methods are: electrostatic actuation [53, 54] (also used in Hysitron Picoindenter) and actuation through springs [55,56]. Recently, it has been shown that also AFM can be used for nanoindentation measurements [57].

Magnetic and Optical Tweezers

The AFM generated force-distance curves lack the sensitivity for measuring local elastic properties through forces of the order of piconewtons (pN), which are often found in biological systems [58]. Here, magnetic (MT) or op-

tical (OT) tweezers are used, which can currently provide calibrated forces measurements down to the hundreds of femtonewtons [59–71].

Magnetic Tweezers

In typical MT setup magnetic fields manipulate the position of a superparamagnetic bead. MT manipulations consist on stretching and/or twisting the molecules attached to the bead. The bead is usually composed of highly crosslinked polystyrene with some magnetic materials (i. e. Fe_2O_3 and Fe_3O_4) evenly distributed inside of the bead. In a magnetic field flux B the net magnetic dipole moment p_m is generated in the bead, and dipole moments of magnetic impurities align along the field direction in order to minimize the potential $U = -p_\mathrm{m}B$. For large magnetic fields the magnetization M of the bead (magnetic moment per bead's volume) saturates and becomes independent of B. Then, the magnetic force F_m, which acts on p_m is given by:

$$F = -\nabla U = \nabla(p_\mathrm{m}B) \simeq p_\mathrm{m}\nabla(B) \tag{12.15}$$

Therefore, quantitative force measurements with MT rely on well-defined magnetic-field gradients as well as uniform and large B so that the bead magnetization is constant. Various MT setups are encountered in the literature. In one of the first designs, Fig. 12.4, Smith et al. [59] utilized two movable permanent magnets in conjunction with a flow field. By approaching the magnets towards the measurements' cell, the amount of vertical stretching force was varied, while by turning the magnets twists were imposed on a superparamagnetic beads. Similar setup was used by Strick et al. [60]. Massive manipulation of many objects and thus expanded manipulation field was achieved by Assi et al. [61] using stacked permanent magnets. Bausch and co-workers [62] as well as Wang et al. [63] substituted the permanent magnets for magnetic coils containing cylindrical soft-iron cores. Haber and Wirtz [64] used two sets of magnetic coils to achieve better control of MT over

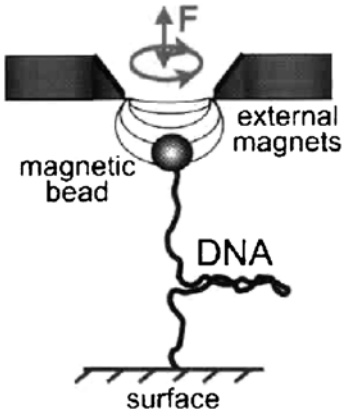


Fig. 12.4. Typical Magnetic Tweezers setup as used by Smith et al. [59]

larger volumes. Magnetic coils offer an advantage over permanent magnets, because the generated fields can be easily controlled. Furthermore, literature pertaining to nuclear magnetic resonance instrumentation provides a wealth of information on how to design magnetic field gradient coils [72]. With respect to the optical tweezers, the MT setup potentially minimize any sample damage due to local heating [73].

Since Eq. (12.15) assumes no edge effects and field dispersion, calibrated measurements via magnetic tweezers are done as follows. The magnetic force is calibrated by measuring the Stokes drag on a magnetic bead in a viscous medium (i. e. glycerol). For a given configuration of magnetic tweezers (i. e. distances of the magnets from the measurements' cell or voltage applied to electromagnets), the unloaded bead velocity υ is measured by observing the bead's trajectory through i. e. a CCD camera. Then the magnetic force F_{M} is obtained as $F_{\mathrm{M}} = 6\pi\eta r\upsilon$. Here, η is the liquid's viscosity and r is the bead's diameter. Once the forces are calibrated, the mean position $\langle x\rangle$ of a loaded bead (i. e. with a molecule of interest attached) is obtained from images provided by the CCD camera. As a result, equivalents of AFM force-distance curves are obtained, but due to sometimes complicated nature of investigated molecules (i. e. DNA) any elasticity information is not straightforward. For calibrating torques, a similar method is applied. Unloaded bead are twisted in a viscous medium due to a torque T, which relates the twisting force F_{tw} and the dipole moment p_{m} with the liquid's rotational viscosity ζ calculated from the Stokes law $\zeta = 8\pi\eta r^3$. We have: $T = p_{\mathrm{m}}F_{tw}sin\theta = -\zeta\frac{\mathrm{d}\theta}{\mathrm{d}t}$, where θ is the twist angle.

Optical Tweezers

OT (called as well an *optical gradient trap*) were pioneered by works of Ashkin et al. [74]. In OT a polymeric beam (to which the molecule of interest is attached) is initially trapped in the center of a focused light beam as shown in Fig. 12.5. While the detailed theory behind OT is complicated the optical force, F_{op}, exerted on a sphere of radius a and dielectric constant ϵ by monochromatic light with wavenumber k in the Rayleigh limit ($a \ll 2\pi k$) is [75, 76]:

$$\begin{aligned} F_{\mathrm{op}} &= F_{\nabla} + F_{\mathrm{s}} \\ &= 2\pi a^3 \frac{(\epsilon_o)^{1/2}}{c}\left(\frac{\epsilon-\epsilon_o}{\epsilon+2\epsilon_o}\right)\nabla \mid S \mid + (8/3)\pi k^4 a^6 \frac{(\epsilon_o)^{1/2}}{c}\left(\frac{\epsilon-\epsilon_o}{\epsilon+2\epsilon_o}\right)^2 S\,. \end{aligned} \tag{12.16}$$

F_{∇} is the gradient force due to gradients in the lights intensity as the bead develops an electric dipole moment in response to the light's electric field. F_{s} is the scattering force due to scattering of light by the particle; ϵ_o is the dielectric constant of a medium, S is the Poyting vector (indicating direction of an optical power transfer). From Eq. (12.16) we can clearly see that when $\epsilon > \epsilon_o$ the gradient force attracts the particle to a focal point of the

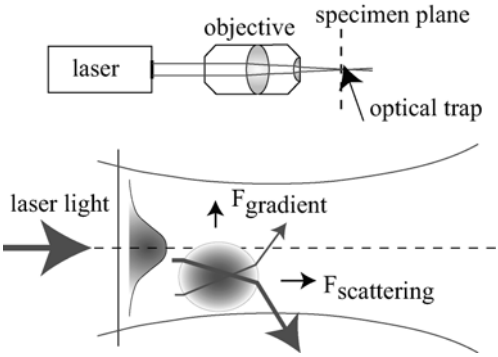


Fig. 12.5. Typical Optical Tweezers setup as used by Block et al.

beam, while the scattering force, drives the particle along the direction of light's propagation. In order to form a full three dimensional trap, the axial intensity gradient must be large enough to overcome the scattering force. Therefore the light beam must be tightly focused to a diffraction limited spot with a high numerical aperture lens. Such configuration is different than in magnetic tweezers, which can work either with a trapped bead, or with beads on which the magnetic force (Eq. (12.15)) is constantly applied.

Light can also induce torques on the OT's beads. These include use of azimuthally phase modulated beams (Laguerre–Gaussian beams) or combinations of beams to introduce an orbital angular momentum component in the light beam [71,77,78]. In particular, the holographic optical tweezers apply these concepts in a very elegant way [79]. Here, a phase of an incident light beam is modulated (like in holograms) by initial diffraction on a grating prepared by liquid crystals domains, which could be oriented in a predefined way (like in LCDs). As a result various superpositions of beams can be created and in particular vortexes of in-plane polarized light with a given radius [80]. Such vortexes can then encompass and rotate the trapped bead. Another examples of optical wrench include use of magnetic fields to apply torque to free or optically trapped magnetic particles [81], or use of linearly or circularly polarized light to orient and apply torque to birefringent particles (i. e. made of calcite or quartz) [82,83].

There are a few methods of calibrating forces exerted by optical tweezers on attached to them particles [84]. The easiest method, however, relies on the equipartition theorem. The Brownian motion of the unloaded bead is observed for a few seconds and then related with the thermal energy to yield: $k_{\mathrm{OT}} = k_{\mathrm{B}}T/(\langle x^2 \rangle)$. Here, k_{OT} is the optical tweezers stiffness, which roughly varies linearly with light intensity [84], k_{B} is the Boltzmann's constant, T is temperature, and $\langle x^2 \rangle$ is the mean value of the square displacement of the bead. Once k_{OT} is calculated, the forces are a product of k_{OT} and a mean displacement of a particle with respect to the beam focus (analogy to the Hook's law). For torque measurements, either similar ways as in the magnetic tweezers setup can be applied (ref. Sect. 12.3.1), or polarizations of incident

and transferred light beams can be compared, and the amount of momentum transferred calculated [83].

Finally, any heating effects are minimized in optical tweezers when photons are not absorbed either by the polymeric bead or the molecules being manipulated, which should be carefully looked after in particular for biological samples.

12.3.2 Dynamic Experimental Methods

Modulated Nanoindentation

The modulated nanoindentation method is a development of the force modulation method [85,86]. Normal modulated nanoindentation consists in indenting an AFM tip in a sample up to a fixed distance while small oscillations are applied to the sample (see Fig. 12.6). Oscillations and indentation are co-linear, normal to the substrate and to the N-O long axis. The amplitude of the oscillations has to be very small, (about 1 Å), in order to remain in the elastic regime. In this amplitude range and experimental geometry, the normal force F_{N} required to vertically move the substrate by a distance D with respect to the cantilever support coincides with the force needed to elastically stretch two springs in series [87, 88]: the cantilever, with normal stiffness $k^{\mathrm{N}}_{\mathrm{lev}}$, and the tip-sample contact, with normal stiffness $k^{\mathrm{N}}_{\mathrm{cont}}$. If D is the total normal displacement of the substrate, i. e. D is equal to cantilever bending plus tip and N-O normal deformation, and F_{N} is the total normal force, this configuration allows the measurement of the total stiffness $k^{\mathrm{N}}_{\mathrm{tot}}$ at each load, defined by the relation:

$$\mathrm{d}F_{\mathrm{N}}/\mathrm{d}D = k^{\mathrm{N}}_{\mathrm{tot}} = (1/k^{\mathrm{N}}_{\mathrm{lev}} + 1/k^{\mathrm{N}}_{\mathrm{cont}})^{-1} \tag{12.17}$$

Since $k^{\mathrm{N}}_{\mathrm{lev}}$ is known, a measurement of $\mathrm{d}F_{\mathrm{N}}/\mathrm{d}D$ at different normal loads leads to the value of $k^{\mathrm{N}}_{\mathrm{cont}}$ as a function of F_{N} [89]. We underline that the

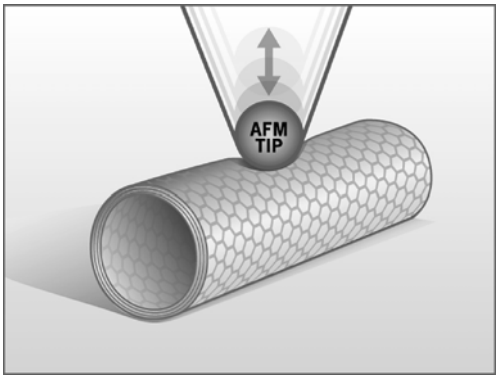


Fig. 12.6. Modulated nanoindentation

stiffness of the substrate usually does not play any role, however each individual case needs to be examined and in the case a three spring model has to be considered. By integrating the equation $\mathrm{d}F_\mathrm{N} = k^\mathrm{N}_\mathrm{cont} \cdot \mathrm{d}z$, where z is the indentation of the tip in the NT [90], it is possible to obtain F_N vs. indentation z from the experimental curves $k^\mathrm{N}_\mathrm{cont}(F)$.

It is the possible to extract the transversal N-O Young modulus from $k^\mathrm{N}_\mathrm{cont}$ vs. F_N measurements by modelling the contact between the AFM tip and the N-O with the Hertz model [5,91,92]. This model might be questioned due to the nanometer dimensions of the contact and the anisotropic geometry and mechanical properties of the N-O.

Under the assumptions of standard elasticity theory, the Hertz model gives the dependence of the indentation distance z versus the normal force F_N between two elastic solids in contact [93]. This dependence is an analytical function of the normal load, the tip and sample Young modulus and some geometrical parameters, e. g. the tip radius. By fitting the experimental z vs. F_N curves with the appropriate Hertz function we obtain the transversal Young modulus of the N-O.

In order to measure the shear elasticity of an individual N-O it is possible to use a method very similar to that one described in the previous paragraph. The position of the sample is modulated laterally in respect to the cantilever holder while the torsional bending of the cantilever, $\mathrm{d}F_\mathrm{L}$, is recorded. In this way we measure the lateral contact stiffness $k^\mathrm{L}_\mathrm{cont}$ as a function of the normal load:

$$\mathrm{d}F_\mathrm{L}/\mathrm{d}D = k^\mathrm{L}_\mathrm{tot} = (1/k^\mathrm{L}_\mathrm{lev} + 1/k^\mathrm{L}_\mathrm{cont})^{-1} \,. \tag{12.18}$$

By using the Hertz model to compute the contact area between the AFM tip and the N-O form the $k^\mathrm{L}_\mathrm{cont}$ vs. F_N curves we can extract the shear modulus of our N-O. If the material is isotropic the shear modulus is proportional to the Young modulus. This can be of great relevance because in shear modulated experiments the lateral stiffness of the cantilever, no more the normal stiffness, is involved during the measurements. Since the lateral stiffness of a cantilever can be two orders of magnitude larger than normal cantilever and sample stiffness, the lateral compression will be now mostly taken by the sample, without limiting the sensitivity of normal force acquisition.

Finally, this method permits to acquire at the same time a map of the topography (with nanoscopic resolution) and a map of the transversal Young and shear modulus of the nano-object. This gives in principle access to the elastic properties of each point of the sample. In Fig. 12.7 we show the measurement of $k^\mathrm{L}_\mathrm{tot}$ while the topography of the sample is acquired simultaneously.

Ultrasonic Methods

Classic acoustic microscopy [94] is a very well developed non-destructive testing method. Its resolution is nevertheless limited almost like in the case of an

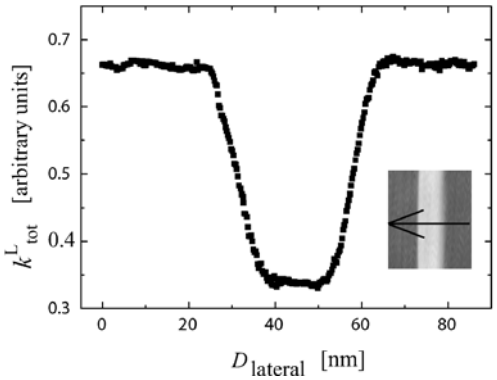


Fig. 12.7. k^{L}_{tot} in the different locations across the nanotube

optical microscopy, meaning up to about a micron. The near field combinations of ultrasound with AFM produced a whole bunch of other techniques sensible to elastic (and/or adhesive) properties of nanocontacts. These include scanning microdeformation microscopy (SMM), acoustic force atomic microscopy (AFAM), scanning local acceleration microscopy (SLAM), heterodyne force microscopy (HFM), and ultrasonic force microscopy (UFM).

SMM

The SMM method was developed by Cretin at al. [95] and is based on the micro-indentations done by an actuated sapphire tip fixed at the cantilever extremity. Another transducer placed underneath the sample measures the resulting deformation field, out of which the quantitative sample's elasticity is calculated. The original resolution was limited by curvature radii of sapphire tips of a few microns, but these can nowadays by crafted with much better precision i. e. through the FEB or FIB techniques.

AFAM

AFAM (Fig. 12.8) was developed by Rabe and Arnold [96]. Here, a small amplitude (angstroms) vibrations are applied to the cantilever in contact with

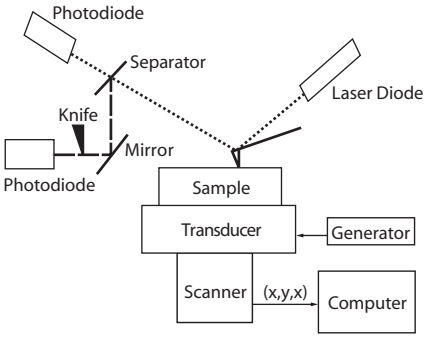


Fig. 12.8. Principles of the AFAM technique

the sample, and at a one of its high frequency (MHz) flexural or torsional contact resonances. The corresponding cantilever vibrations at the excitation frequency are detected (i. e. via a knife-edge detector), while the quastistatic cantilever's position changes provide topography of a sample.

It has been proven by many authors that higher frequency flexural modes of cantilever response are controlled by normal contact stiffness ([97] and references in the chapter VI). Therefore, the AFAM ultrasonic response provides a way to calculate the values of k_{n}, out of which the Young moduli are obtained through Eq. (12.13).

SLAM

The SLAM technique was developed by Burnham et al. [98]. Here, the sample is excited in the kHz range by an amplitude d_{s} via transducer placed underneath. The cantilever's response, z_{c} at excitation frequency is detected by a standard beam-bounce technique. It has been shown [99, 100] that at frequencies higher than the tip-sample contact resonance the ratio of $d_{\mathrm{s}}/z_{\mathrm{c}}$ is a sensitive measure of the contact stiffness [101]. Therefore, subsequent SLAM imaging at different frequencies reveals the changes of contrast between areas of different elasticity and allows local comparisons of the Young moduli.

UFM

The UFM method was invented by Kolosov et Yamanaka [102]. Here (Fig. 12.9), an ultrasonic excitation of a few MHz, which does not coinciding with any of the cantilever's resonance, is applied to a tip-sample contact via an actuator placed underneath the sample. The fast ultrasonic excitation is modulated by a slow frequency (between 100 Hz to a few kHz) at which the cantilever can readily respond, and which does not interfere with an AFM feedback (set to be low). As developed recently by Szoszkiewicz et al. [103–105], the cantilever response at the low frequency is a quantitative measure of local elasticity and adhesion. In the absence of capillary forces between an AFM

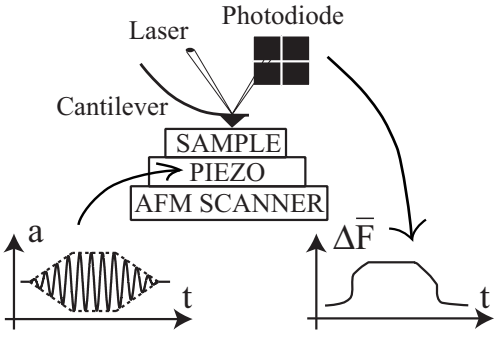


Fig. 12.9. Principles of the UFM technique based on an AFM

tip and a sample (either in dry environment or in liquids) the force jump-up value (Fig. 12.9) is correlated with E^* through the following relation:

$$4(F_{\mathrm{SP}}(a_2 - a_1) + \Delta\overline{F}_2 a_2 - \Delta\overline{F}_1 a_1) \simeq 1.37\pi^{5/3} R^{4/3} w(w/E^*)^{2/3} \quad (12.19)$$

Here F_{SP} is the static load force (setpoint), a_1, a_2 are the ultrasonic amplitudes before and after the force jump-up, $\overline{F}_1$, $\overline{F}_2$ are changes of the mean ultrasonic force before and after the jump-up, w is the tip-sample adhesion energy, and R is the tip-sample reduced curvature radius. With the knowledge of w and R the tip-sample reduced Young modulus can be easily obtained.

HFM

Cuberes et al. developed the HFM technique [106]. The HFM works by simultaneous application of two ultrasonic excitation to the tip-sample contact. The sample is excited at the transducer placed underneath and the cantilever is subjected to another excitation at its base. Both frequencies are in a range of MHz and differ only by a few kHz to produce beats. The cantilever's response in the beat frequency is detected by a conventional beam-bounce technique. The resulting HFM response (or image) depends on elasticity, but quantitative interpretation remains still unclear.

12.4 Mechanical Properties of Inorganic and Biological Nano-Objects

12.4.1 Thin Films and Embedded Structures

Mapping elastic properties of the thin films with lateral resolution of less than hundreds of nanometers presents a challenge to classical nanoindentation methods. Here, the near field combinations of ultrasound with AFM come handy. These methods (SLAM, AFAM, UFM) can achieve lateral resolution of the order of 20 – 100 nm limited only by an AFM tip blunting as a result of its high frequency tapping. In the case of soft samples (polymers) usually only qualitative information is provided by these techniques, owing to the relatively complex relationship between the contact stiffness and the dynamic Youngs modulus [98, 99]. Moreover, in the presence of sharp variations in sample topography, the variations in the tip-sample contact area may lead to some artifacts [107].

SLAM measurements have been performed on a wide range of engineering materials, from Al_2O_3/Al metal-matrix composites to PVC/PB polymer blends [99, 108], and 100 thin silicon oxide coatings on polyamide [100].

The UFM method was applied to probe relative elasticity of defects in highly oriented pyrolitic graphite (HOPG) [109], Ge dots on silicon substrate [110], as well as heterogeneous nanostructures [111]. Also polymeric materials embedded into stiff [112] or more compliant matrices have been

measured. The UFM depth sensitivity for elastic imaging was also discussed to be several hundreds of nanometers [113].

AFAM has measured reduced Young moduli for 50 nm thin films Ni deposited on Si, glass-fiber/polymer composites [114], and thin diamond-like-carbon coatings [115].

12.4.2 Carbon Nanotubes

The exceptional mechanical, electrical and thermal properties [4,48,116–121] of carbon nanotubes (CNTs) have attracted great scientific and technological interest. CNTs have cylindrical symmetry with axial mechanical properties characterized by the strong in plane covalent C–C bond. The strength of this bond gives rise to an extraordinary axial stiffness, as pointed out by several experimental [46–48] and theoretical studies [120–122] finding values for the axial Young modulus of about 1 TPa. In graphite, the C_{11} in plane elastic constant is 1.06 TPa, while the perpendicular elastic constant C_{33} is only 36 GPa [120]. Similarly the radial Young modulus of CNTs is expected to be much smaller than the axial one. Evidence for the softness of CNTs in the radial direction has been reported in experiments under hydrostatic pressure [123], and in measuring NTs deformation due to Van der Waals forces between the tubes and the substrate [124, 125]. Achieving a fundamental understanding of the radial deformability of CNT is important for applying them in nano-electro-mechanical and nano-electronic systems. For example, the radial deformation of CNTs may strongly affect their electrical properties [4, 5, 126–128].

The axial elasticity of multi-walled nanotubes and ropes of nanotubes has been investigated with several methods, as described in Sect. 12.3.1. The quantitative understanding of the radial elasticity of CNTs is less well known [91, 92, 129–131]. To measure the radial elasticity of CNTs one can indent an AFM tip into a NT and measure force vs indentation curves. However, such measurements are very challenging, since in order to stay in the linear elastic regime, one has to measure forces of a few nN vs displacements of a few Å. Examples of plastic deformation in CNTs using this method have been shown in Ref. [92]. One proposed possibility is to vertically vibrate the cantilever in non-contact mode with amplitudes in the range of several hundreds of Å and with the turning point situated a few Å above the sample [91]. But in this way, a considerable fraction of the signal arises from the van der Waals forces, and only a small part comes from the NT elastic properties.

Very recently, by means of the modulated nanoindentation method (see Sect. 12.3.2), Palaci et al. have measured the radial stiffness of multiwalled CNT with external radii ranging from 0.2 to 12 nm and having a constant ratio of external to internal radii of $R_{\text{ext}}/R_{\text{int}} = 2.2 \pm 0.2$ [131]. They have shown that the radial stiffness strongly increases with decreasing external diameter. The radial effective Young modulus E_{rad} is extracted from the experimental results applying the Hertz model. E_{rad} is found to decrease

to an asymptotic value of 30 GPa for larger tube sizes. This last value is, within the experimental error, equal to the Young modulus of graphite along its C-axis, $E_{\text{graphite}} = 36\,\text{GPa}$ [132]. For the NTs studied in this work R_{ext} is proportional to R_{int} and both are proportional to the number of layers since the distance between layers is approximately constant [120]. Thinking to the elastic energy necessary to enroll a plane one could deduce that the radial rigidity and hence E_{rad} of a NT should increase by increasing the number of layers and by decreasing the internal radius. This is confirmed by measurements of radial deformations of NTs due to van der Waals forces between the tube and the substrate [124,125]. In both these studies the radial deformation increases with the radius for single-walled NT and decreases with the number of layers. The experiments of Palaci et al. show that for small $R_{\text{int}}, E_{\text{rad}}$ increases sharply by decreasing R_{int}, thus in this size range the radial rigidity is controlled by the magnitude of R_{int}, whereas the number of layers plays a minor role. This result is in agreement with a previous theoretical study [120] that shows that the elastic properties of a NT with $R_{\text{int}} = 0.34\,\text{nm}$ do not change by increasing the number of layers as long as the interlayer distance is fixed to 0.34 nm, i. e. the distance between planes in graphite. A similar finding is also obtained in the simulations of Ref. [125] where the radial deformation of a single-walled NT is the same of a multi-walled NT when the radius of the first one is equal to R_{int} of the second one. For large R_{int}, E_{rad} is almost constant. This could mean that the effect due to the increase of R_{int} is counter balanced by the increase of the number of layers, up to the point at which NT's properties reach asymptotically those of graphite.

The radial stiffness of multiwalled carbon NTs has been investigated experimentally by M. F. Yu et al. [91]. In their experiments, one NT with an unknown number of layers is compressed, the maximum indentation distance being larger than 40% of the initial diameter. The force vs indentation distance curves are obtained through a model of the tip-NT van der Waals forces. By interpreting these curves with the Hertz model they find, for a NT with a diameter of 8 nm, a radial Young modulus between 0.3 and 4 GPa, which is roughly one order of magnitude lower than the results obtained by Palaci et al. for NTs of similar diameters. This discrepancy can be ascribed to a difference in the number of graphene layers forming the NT, which is plausible since the NT preparation techniques are different.

Buckling has been shown to strongly influence the elastic properties of CNT in the axial direction. Few calculations and experiments have investigated the role of buckling in the radial deformability of nanotubes [133]. It remains unclear the role of tube radius and number of layers in multiwalled C and BN NTs buckling.

The elastic properties of C and BN NTs are influenced by the presence of structural defects [134–136]. Some experimental investigations have shown that the axial Young modulus of C–NT can drop from 810 GPa for defect-free NT down to 30 GPa for NT with structural defects [48, 137, 138]. No exper-

imental data are present for the effect of defects on the radial deformability of NT. In general it is not clear the role of the defects on the mechanical properties of NT. The reason for this lack of knowledge is the experimental difficulty in studying experimentally the mechanical properties of the NT and at the same time their local structure.

Experimental investigation on the role of temperature, adsorbates and environment, such as water, on the elastic properties of NT are also lacking.

Finally, technical difficulties make the experimental determination of the Poisson's ratio and shear moduli a challenging task. For a theoretical study see [139]. So far, there are no quantitative measurements of shear moduli of single or multi-walled NT.

12.4.3 Oxide Nanowires, Nanobelts and Nanorods

Interest in nano-wires, belts and rods continues to grow, in part, because of their potential in nanoelectronics and optoelectronics [133]. To date, a wide variety of nano-wires, belts and rods based devices has been demonstrated (see Fig. 12.10), including photodetectors, photodiodes, and sensors [140]. These materials offer also certain advantages over planar devices.

Recently, Dr. Wang and co-workers have synthesized belt-like ZnO and SnO_2, so-called nanobelts (rectangular cross section) [2]. The belt-like morphology is distinct from that one of semiconductor nanowires. With well-defined geometry and perfect crystallinity, semiconducting oxide belts are likely to be a model material family for understanding mechanical behavior at the nanoscale in the absence of dislocations and defects, excluding point defects. The most common growth direction of the nanobelt is along the *c*-axis showing no piezoelectric property across the thickness. However, by controlling growth kinetics, Dr. Wang and co-workers have shown the

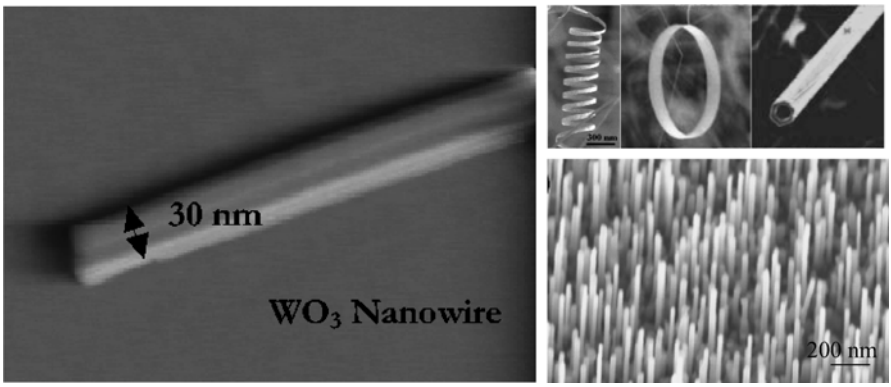


Fig. 12.10. AFM image of a WO_3 wire (*left*). TEM images of a ZnO spring, loop, rod, forest (*right*)

success of growing nanobelts whose surfaces are dominated by the polarized (0001) facets. Owing to the positive and negative ionic charges on the zinc- and oxygen-terminated (0001) surfaces, respectively, a spontaneous polarization is induced across the nanobelt thickness. As a result, right-handed helical nanostructures and nanorings are formed by rolling up single-crystal nanobelts; this phenomenon is attributed to a consequence of minimizing the total energy contributed by spontaneous polarization and elasticity. In an AFM-TEM experiment Gao et al. show that these ZnO nanohelix can be manipulated, and their elastic properties are measured, in particular they find that a nanohelix has a spring constant of about 4 N/m [141]. This suggests possible uses in electromechanically coupled sensors, transducers, and resonators.

The hardness of ZnO nanobelts have been investigated by means of a AFM-based nanoindentation technique [142]. It was found that ZnO nanobelts have a hardness of about 5 GPa, very close to the bulk value. Furthermore, AFM-based methods have been developed to study the piezoelectric effect in ZnO nanobelts [143]. It was found that the piezoelectric coefficient of ZnO nanobelts is frequency dependent and it can reach the value of 25 pm/V, as compared with 10 pm/V for bulk ZnO.

12.4.4 Cancer Cells and Blood Cells

The knowledge of the mechanical properties of tumor cells can give important insights into their motility and their metastasis process. To understand more about the locomotion of cancer cells is important to study the structure and properties of their pseudopodia as well as their role in cell invasion. The elastic properties of cancer cells determine whether they can pass through the microvasculature to form metastases and whether they can withstand the shear-induced deformation due to the blood flow if they do metastasize through the blood vasculature. Just to have an idea, the fluid flow along an artery wall exerts a shear stress on an adherent leukocyte of around 10 to 100 Pa [144]. Even larger forces, of the order of 100 Pa, are exerted by the cortex of locomoting keratocytes [145].

Zhang et al. [135], studied the mechanical properties of hepatocelular carcinoma cells with a micropipette aspiration technique and were able to measure the elastic coefficient. But this technique doesn't permit the characterization of individual molecules but an ensemble average. On the other hand, Lekka et al. [146] using an atomic force microscopy measured the elasticity of normal and cancerous bladder cells. They found that the normal cells have a Young modulus one order of magnitude higher than cancerous ones. Oncogenically transformed cells are different form healthy ones in their cell growth, morphology, cell membrane and organization of the cytoskeleton. In this case, the difference in stiffness seems depend on the structure of the cytoskeleton because a deep indentation is needed to observe this difference.

If the indentation were smaller the difference could be due to the cell membrane. They observe that the lower stiffness of cancer cells may be caused by a partial loss of actin filaments and/or microtubules, and therefore a lower density of the cellular scaffold. Goldmann et al. [147–149] also studied the elasticity of cancer cells, in particular wild-type F9 mouse embryonic carcinoma and vinculin-deficient F9 cell lines. Through indentation with AFM, they produced viscoelastic maps of the two types of cells. For a review on atomic force microscopy on cell biology refer to references [150] and [151].

Red blood cells have been studied extensively by varied methods such as suction with pipettes, flicker spectroscopy, and optical tweezers. This has shown the cytoskeleton of red blood cells to be very different to the cytoskeleton of other cells and much information is now known about its viscoelasticity. Another blood component that has gathered attention are human platelets. They are small cells that play an important role in wound healing and blood clotting. When they are activated with an injury, they go through a shape transformation thanks to reorganization of its cytoskeleton.

Radmacher et al. in 1995 measured the Young modulus of human platelets over large sections of the cell by doing force mapping [10]. They took force displacement curves continuously while scanning over the surface. With this information they were able to construct a force reliefs of the surface by doing a 3D plot of sample position, sample height cantilever deflection. This way by plotting one scanned line at a time you can see how the slope of the contact region of cantilever deflection versus sample height changes and likewise the stiffness of the different regions within the cell. Another interesting feature of this method is that by plotting sample position versus height for a specific value of force over all the area, we obtain an isoforce topograph. This topography is similar to the obtained by the usual scanning where the force is held constant but can show more information about the elasticity of the sample. By plotting isoforce topographs for increasing values of the force, we can see in this case how the cell is deformed under the compression. For the human platelet some areas are completely compressed and some features don't appear when the force is too high. On the other hand, with this method,

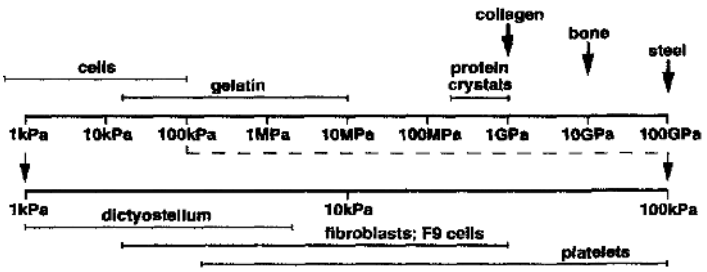


Fig. 12.11. In this figure, we can appreciate the enormous range in Young modula present in biological tissues. Taken from [151]

imaging can be done with smaller forces than with the usual scanning. One interesting point is that the cell is able to recover from the deformation at each point probably due to a fast relaxation.

One big problem encountered when imaging with the fast scan is that the lateral forces might displace the sample or even detach it from the substrate, as mentioned earlier. In this case, since there are no lateral forces, this doesn't present a problem. With this method they were able to see the relative behavior under compression of the different components of the cell, and measured the elastic modula of the pseudonucleus and the filopodia. Although gels of actin and microtubules, components of the cytoskeleton of the cell, had been already measured in different ways, this was one of the first attempts to study the rigidity of the cytoskeleton in-vivo.

In the late 1980's, researchers managed to trap optically living cells and since then there have been several efforts to use optical tweezers in biological research. For a review on single-molecule biomechanics with optical methods refer to Ref. [152]. Viruses, bacteria and eukaryotic cells have been manipulated non-invasively. It also allows for the measurement of forces with a precision of pN. To minimize photo damage, the laser wavelength should be chosen in the infrared. In this way, cells can be directly trapped by the focused laser. Another approach is to indirectly manipulate the cells by using silica or latex micro-beads, which can be done in 2D or 3D. This can be done in two ways, by fast deflection of the laser with acusto-optics deflectors (AOD's) or by splitting the laser into more beams using diffractive optical elements.

Recently, Ferrari et al. [153] used both methods to manipulate Escherichia coli cells by designing two and three dimensional beads configurations to surround the cell. This indirect manipulation could be used, as they say, to stretch and shrink cells to investigate cell reactions to mechanical stimuli. Another variation to this tool are holographic optical tweezers (HOT), which allow for varied three-dimensional traps such as optical vortices which can be moved around at ease. This tool has recently been used to study the pericellular matrix, a cross-linked polymer network attached to the outer membrane of cells. By understanding its mechanical properties, such as elasticity, some light might be shed on the adhesion of cancer cells when undergoing metastasis, for instance.

12.4.5 Arteries

Heart disease is one of the leading causes of deaths worldwide. To treat blocked arteries, the most common procedure uses a catheter-based balloon angioplasty to dilate the arteries. It has a very immediate success rate, but 40 percent of the cases renarrow within 6 months. This has led to great interest in the elasticity of heart arteries. In particular how does the artery wall go from being able to stretch to accommodate the changing flux of blood, to stiffening due to fatty plaque deposits called cardiac calcifications. Lopez

et al., using a modified AFM incorporating the features of a nanoindenter, measured the mechanical properties of femoral artery tissue in- vitro and of diseased calcified arteries [154]. They concluded that calcified deposits where many orders of magnitude stiffer that the healthy tissue and much more stiff than previously known. Knowing that the fatty plaque deposits are less elastic than anticipated, might help devise an improved method to treat cardiac calcifications disease. This same type of modified AFM was used to study extensively the mechanical properties of bones and teeth at the nanoscale [155–157].

12.4.6 DNA

The elastic behavior of DNA has been studied with several experimental methods, such as hydrodynamic drag, micropipettes, and optical and magnetic traps [158–162]. The elastic behavior is different for different force regimes [159, 163–167]. There have been several models to study DNA elasticity, such as the free jointed chain model (FJC) [59] and the worm-like chain (WLC) [168]. The FJC model treats the polymer as a chain of statistically independent segments whose orientations are uncorrelated in the absence of external forces. The WLC model characterizes a polymer with a single parameter, the flexural persistence length A, and assumes the polymer as a line that bends smoothly under the influence of random thermal fluctuations [168]. The value A defines the distance over which the direction of this line persists. The correlation between the orientations of two polymer segments falls off exponentially with decay length A according to the contour length that separates them. For double stranded DNA in buffer, A is approximately 50 nm.

Optical tweezers have been used to study the behavior of DNA upon stretching and twisting. To stretch a single molecule of DNA, a force is exerted on the tether molecule and its extension is measured. A very important result has been the "overstretching transition" at approximately 65 pN, over which the molecule stretches to almost twice its contour length ([159, 160]). What happens with the DNA molecule when it overstretches has been highly debated.

Magnetic tweezers have also been used to study single molecules of DNA ([59, 169, 170]). Manipulation of the external magnetic field can be used to exert forces or torques on the tether molecule to induce stretching or twisting. It can induce superhelicity by applying some torque on the beads and rotating in a certain direction. Zlatanova and Leuba used magnetic tweezers to measure the superhelicity density of a single DNA molecule, by counting the number of rotations of the tether molecules while measuring the force and extension of the molecule at the same time [170]. Research on the area has shown a marked difference between the elastic properties between single and double stranded DNA. This allows the study of enzymes acting on DNA. For

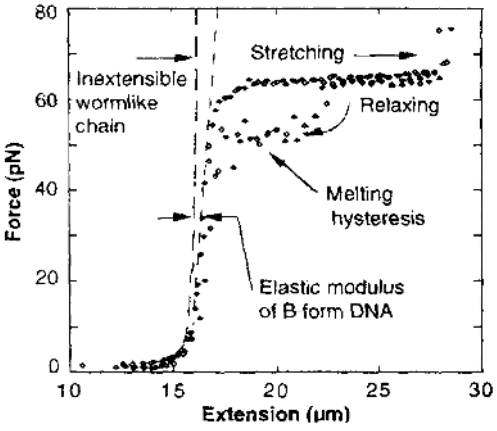


Fig. 12.12. Stretching of λ-phage DNA (NEB) in 50 mM NaCl, 10 mM tris, 1 mM EDTA, pH 8.0. The "inextensible worm like chain" for a persistence length A of 53 nm and contour length of 16.4 μm. Image taken from [160]

example, the replication of DNA can be monitored by following the changes in molecule extension or the rigidity of the tether.

The capabilities of the AFM have also been used to try to elucidate more about the mechanical properties of DNA. There is a very interesting review by Hansma on applications of the AFM to the surface biology of DNA [150]. Lee et al. [171] measured the forces between two strands of double stranded DNA for the first time by attaching complementary DNA oligonucleotides on the tip and the substrate. They needed 0.13 nN to separate the two strands. In similar experiments done at the Guntherdot laboratory the unbinding forces scaled with the number of base pairs ruptured [172].

The overstretching phenomenon has been observed with the AFM, as with the optical tweezers. By pulling on a single double stranded DNA molecule, force-distance curves can be generated as shown in Fig. 12.12. The B-DNA is stretched through the B-S transition at ∼65 to an overstretched form almost 1.7 times its B-DNA contour length ([159], [160]). For larger forces, the DNA base pairs melt. For the AFM experiment, the unstretching and stretching curves overlap which means that the process can be reversible [173]. The force needed for melting DNA depends on several parameters such as temperature, ionic strength of the medium, and the DNA sequence. For this last reason, it has been said that it could be use to discriminate between different sequences. Sequence-dependent differences in rupture forces were measured directly on guanine-cytosine (GC) versus adenine-thymine (AT) sequences by AFM pulling on poly(dG-dC) and poly(dA-dT) [174].

12.4.7 The Challenge: Control of Nano-Object Chemistry, Size, Structure, etc.

A major challenge in nano-mechanics is the possibility to to perform nanomechanical measurements on nano-object of well defined structure, chemical composition, geometrical shape and etc.. For example, the geometrical shape

can be imaged by AFM during the mechanical measurements, however in order to know the structure of the individual nano-objects new strategies need to be developed. One approach is to grow N-Os on a patterned substrate which will permit to move from an instrument to another one and recognize the same nano-object. A second approach could be to perform AFM tip-enhanced Raman spectroscopy measurements coupled with AFM force measurements [175].

References

1. S. Iijima, *Helical microtubules of graphitic carbon*, Nature **354**, 56 (1991).
2. Z.W. Pan, Z.R. Dai, and Z.L. Wang, *Nanobelts of Semiconducting Oxides*, Science **291**, 1947 (2001).
3. X. Duan, Y. Huang, Y. Cui, J.F. Wang, and C.M. Lieber, *Indium phosphide nanowires as building blocks for nanoscale electronic and optoelectronic devices*, Nature **409**, 66 (2001).
4. R. Martel, T. Schmidt, H.R. Shea, T. Hertel, and P. Avouris, *Single- and multi-wall carbon nanotube field-effect transistors*, Appl. Phys. Lett. **73**, 2447 (1998).
5. C. Gómez-Navarro, P. J. d. Pablo, and J. Gómez-Herrero, *Radial electromechanical properties of carbon nanotubes*, Adv. Mater. **16**, 549 (2004).
6. B. Cappella and G. Dietler, *Force-distance curves by atomic force microscopy*, Surf. Sci. Rep. **34**, 1 (1999).
7. H.J. Hertz, *On the contact of elastic solids*, Reine Angew. Math **92**, 156 (1882).
8. M. Radmacher, M. Fritz, and P.K. Hansma, *Imaging soft samples with the atomic-force microscope -gelatin in water and propanol*, Biophys. J. **69**, 264 (1995).
9. H.W. Wu, T. Kuhn, and V.T. Moy, *Mechanical properties of l929 cells measured by atomic force microscopy: Effects of anticytoskeletal drugs and membrane crosslinking*, Scanning **20**, 389 (1998).
10. M. Radmacher, M. Fritz, C.M. Kacher, J.P. Cleveland, and P.K. Hansma, *Measuring the Viscoelastic Properties of Human Platelets with the Atomic Force Microscope*, Biophysical Journal **70**, 556 (1996).
11. B.J. Briscoe, K.S. Sebastian, and M.J. Adams, *The effect of indenter geometry on the elastic response to indentation*, J. Phys. D **27**, 156 (1994).
12. A.B.M. et al., *?*, J. Biochem **34**, 1545 (2001).
13. F. Rico, P. Roca-Cusachs, N. Gavara, R. Farré, M. Rotger, and D. Navajas, *Probing mechanical properties of living cells by atomic force microscopy with blunted pyramidal cantilever tips*, Physical Review E **72**, 021914 (2005).
14. B.V. Derjaguin, V.M. Muller, and Y.P.T. Toporov, *Effect of contact deformations on adhesion of particles*, J. Colloid Interface Sci. **53**, 314 (1975).
15. K.L. Johnson, K. Kendall, and A.D. Roberts, *Surface energy and contact of elastic solids*, Proc. R. Soc. A **324**, 301 (1971).
16. D. Maugis and H.M. Pollock, *Surface forces, deformation and adherence at metal microcontacts*, Acta Metall **32**, 1323 (1984).
17. K. Shull, *Contact mechanics and the adhesion of soft solids*, MATERIALS SCIENCE & ENGINEERING R-REPORTS **36**, 1 (2002).

18. J.R. Barber and M. Ciavarella, *Contact mechanics*, INTERNATIONAL JOURNAL OF SOLIDS AND STRUCTURES **37**, 29 (2000).
19. S. Schmauder, *Computational mechanics*, ANNUAL REVIEW OF MATERIALS RESEARCH **32**, 437 (2002).
20. C. Tsakmakis, *Description of plastic anisotropy effects at large deformations – Part I: restrictions imposed by the second law and the postulate of Il'iushin*, INTERNATIONAL JOURNAL OF PLASTICITY **20**, 167 (2004).
21. I. Kragelsky, M. Dobychin, and V. Kombalov, *Friction and wear calculation methods* (New York: Pergamon Press, ADDRESS, 1982).
22. J. Greenwood and J. Williamson, *Contact of nominally flat surfaces*, Proc.Roy.Soc.Lond. **A295**, 300 (1966).
23. P. Nayak, *Random process model of rough surfaces*, ASME J Lubr Tecnol **93**, 398 (1971).
24. J. Ogilvy, *Numerical simulations of friction between contacting rough surfaces*, J.Phys. D. **24**, 2098 (1991).
25. J. Sugimura, *Stochastic modeling of surface roughness*, JOURNAL OF JAPANESE SOCIETY OF TRIBOLOGISTS **43**, 933 (1998).
26. P. Meakin, *The growth of rough surfaces and interfaces*, PHYSICS REPORTS-REVIEW SECTION OF PHYSICS LETTERS **235**, 189 (1993).
27. J. Gao, W.D. Luedtke, D. Gourdon, M. Ruths, J.N. Israelachvili, and U. Landman, *Frictional forces and Amontons' law: From the molecular to the macroscopic scale*, JOURNAL OF PHYSICAL CHEMISTRY B **108**, 3410 (2004).
28. A. Majumdar and B. Bhushan, *Fractal Model of Elastic-Plastic Contact Between Rough Surfaces*, Journal of Tribology-Transactions of the ASME **113**, 1 (1991).
29. H. Zahouani, R. Vargiolu, and J.L. Loubet, *Fractal models of surface topography and contact mechanics*, Math. Comput.Modell. **28**, 517 (1998).
30. W. Yan and K. Komvopoulos, *Contact Analysis of Elastic-Plastic fractal surfaces*, J. Appl. Phys. **84**, 3617 (1998).
31. J.C. Chung and J.F. Lin, *Fractal Model Developed for Elliptic Elastic-Plastic Asperity Microcontacts of Rough Surfaces*, Transactions of the ASME **126**, 646 (2004).
32. B.N.J. Persson, *Elastoplastic Contact between Randomly Rough Surfaces*, Physical Review Letters **87**, 116101 (2001).
33. K.N.G. Fuller and D. Tabor, *Effect of surface-roughness on adhesion of elastic solids*, Proc.R.Soc.Lond A **345**, 327 (1975).
34. B.N.J. Persson and E. Tosatti, *The effect of surface roughness on the adhesion of elastic solids*, Journal of Chemical Physics **115**, 5597 (2001).
35. R. Buzio, C. Boragno, and U. Valbusa, *Contact mechanics and friction of fractal surfaces probed by atomic force microscopy*, Wear **254**, 917 (2003).
36. B. Luan and M. Robbins, *The breakdown of continuum models for mechanical contacts*, Nature **435**, 929 (2005).
37. O. Miesbauer, M. Gotzinger, and W. Peukert, *Molecular dynamics simulations of the contact between two NaCl nano-crystals: adhesion, jump to contact and indentation*, Nanotechnology **14**, 371 (2003).
38. Which as discussed in our theoretical section is justified for all models at high indentations.
39. L.-O. Heim, M. Kappl, and H.-J. Butt, *Tilt of atomic force microscope cantilevers: effect on spring constant and adhesion measurements*, Langmuir **20**, 2760 (2004).

40. J. Hutter, *Comment on tilt of atomic force microscope cantilevers: effect on spring constant and adhesion measurements*, Langmuir **21**, 2630 (2005).
41. T.-D. Li, J. Gao, R. Szoszkiewicz, U. Landman, and E. Riedo, *Water molecules confined in sub-nanometer gaps*, submitted to Nature (2005).
42. S. Garcia-Manyes, A. Guell, P. Gorostiza, and F. Sanz, *Nanomechanics of silicon surfaces with atomic force microscopy: An insight to the first stages of plastic deformation*, J. Chem. Phys. **123**, 114711 (2005).
43. M. Rost, L. Crama, P. Schakel, E. van Tol, G. van Velzen-Williams, C. Overgauw, H. ter Horst, H. Dekker, B. Okhuijsen, M. Seynen, A. Vijftigschild, P. Han, A. Katan, K. Schoots, R. Schumm, W. van Loo, T.H. Oosterkamp, and J. Frenken, *Scanning probe microscopes go video rate and beyond*, Rev. Sci. Instr. **76**, 053710 (2005).
44. B. Bhushan, *Springer Handbook of Nanotechnology* (Springer-Verlag, Heidelberg, 2004).
45. A. Kueng, C. Kranz, A. Lugstein, E. Bertagnolli, and B. Mizaikoff, *AFM-Tip-Integrated Amperometric Microbiosensors: High-Resolution Imaging of Membrane Transport*, Angewandte Chemie Int. Ed. **44**, 3419 (2005).
46. E.W. Wong, P.E. Sheehan, and C.M. Lieber, *Nanobeam mechanics: elasticity, strength, and toughness of nanorods and nanotubes*, Science **277**, 1971 (1997).
47. P. Poncharal, Z.L. Wang, D. Urgarte, and W.A. de Heer, *Electrostatic deflections and electromechanical resonances of carbon nanotubes*, Science **283**, 1513 (1999).
48. J.P. Salvetat, G.A.D. Briggs, J.M. Bonard, R.W. Bacsa, A.J. Kulik, T. Stockli, N.A. Burnham, and L. Forró, *Elastic and shear moduli of single-walled carbon nanotube rope*, Phys. Rev. Lett. **82**, 944 (1999).
49. J.H. Song, X. Wang, E. Riedo, and Z. Wang, *Elastic Property of Vertically Aligned Nanowires*, Nano Letters **5**, 1954 (2005).
50. W. Oliver and G.M. Pharr, *An improved technique for determining hardness and elastic-modulus using load and displacement sensing indentation experiments*, Journal of Materials Research **7**, 1564 (1992).
51. W. Oliver, *Alternative technique for analyzing instrumented indentation data*, Journal of Materials Research **16**, 3202 (2001).
52. J. Pethica, R. Hutchings, and W. Oliver, *Hardness measurement at penetration depths as small as 20-nm*, Philosophical Magazine A **48**, 593 (1983).
53. E.T. Lilleodden, W. Bonin, J. Nelson, J.T. Wyrobek, and W.W. Gerberich, *In-situ imaging of Mu-N load indents into gas*, J. of Mat. Res, **10**, 2162 (1995).
54. C. Schuh, J. Mason, and A. Lund, *Quantitative Insight into Dislocation Nucleation from High-temperature Nanoindentation Experiments*, Nature Materials **4**, 617 (2005).
55. N. Burnham and R. Colton, , J. Vac. Sci. Tech. A **7**, 2906 (1989).
56. T. Bell, J. Field, and M. Swain, *Elastic plastic characterization of thin-films with spherical indentation*, Thin Solid Films **220**, 289 (1992).
57. C.A. Clifford and M. Seah, *Quantification issues in the identification of nanoscale regions of homopolymers using modulus measurement via AFM nanoindentation*, Appl. Surf. Sci. **252**, 1915 (2005).
58. T. Strick, J.-F. Allemand, V. Croquette, and D. Bensimon, *Twisting and stretching single DNA molecules*, Prog. in Biophys. & Mol. Biol. **74**, 115 (2000).

59. S.B. Smith, L. Finzi, and C. Bustamante, *Direct Mechanical Measurements of the Elasticity of Single DNA Molecules by Using Magnetic Beads*, Science **258**, 1122 (1992).
60. T. Strick, J.-F. Allemand, D. Bensimon, and V. Croquette, *Behavior of supercoiled DNA*, Biophys. J. **74**, 2016 (1998).
61. F. Assi, R. Jenks, J. Yang, C. Love, and M. Prentiss, *Massively parallel adhesion and reactivity measurements using simple and inexpensive magnetic tweezers*, J. Appl. Phys. **92**, 5584 (2002).
62. A. Bausch, F. Ziemann, A. Boulbitch, K. Jacobson, and E. Sackmann, *Local measurements of viscoelastic parameters of adherent cell surfaces by magnetic bead microrheometry*, Biophys. J. **75**, 2038 (1995).
63. N. Wang, J. Butler, and D. Ingber, *Mechanotransduction across the cell-surface and through the cytoskeleton*, Science **260**, 1124 (1993).
64. C. Haber and D. Wirtz, *Magnetic tweezers for DNA micromanipulation*, Rev. Sci. Instr. **71**, 4561 (2000).
65. K. Svoboda and S. Block, *Optical trapping of metallic Rayleigh particles*, Opt. Lett. **19**, 930 (1994).
66. P. Ke and M. Gu, *Characterization of trapping force on metallic Mie particles*, Appl. Opt. **38**, 160 (1999).
67. L. Chislain, N. Switz, and W. Webb, *Measurements of small forces using and optical trap*, Rev. Sci. Instr. **65**, 2762 (1994).
68. R. Litvinov, H. Shuman, J. Bennett, and J. Weisel, *Binding strength and activation state of single fibrinogen-integrin pairs on living cells*, Proc. Natl. Acad. Sci. **99**, 7426 (2002).
69. F. Gittes and C. Schmidt, *Signals and noise in micromechanical measurements*, Methods Cell. Biol. **55**, 129 (1998).
70. A. Pralle, M. Prummer, E. Florin, E. Stelzer, and J. Horber, *Three-dimensional high-resolution particle tracking for optical tweezers by forward scattered light*, Microsc. Res. Tech. **44**, 378 (1999).
71. D. Grier, *A revolution in optical manipulation*, Nature **424**, 810 (2003).
72. V. Bangert and P. Mansfield, *Magnetic-field gradient coils for NMR imaging*, J. Phys. E **15**, 235 (1982).
73. Y. Liu, D. Cheng, G. Sonek, M. Berns, C. Chapman, and B. Tromberg, *Evidence for localized cell heating induced by infrared optical tweezers*, Biophys. J. **68**, 2137 (1995).
74. A. Ashkin, J. Dziedzic, J. Bjorkholm, and S. Chu, *Observation of a single-beam gradient force optical trap for dielectric particles*, Opt. Lett. **11**, 288 (1986).
75. Y. Harada and T. Asakura, *Radiation forces on a dielectric sphere in the Rayleigh scattering regime*, Opt. Commun. **124**, 529 (1996).
76. E. Dufresne and D. Grier, *Optical tweezer arrays and optical substrates created with diffractive optics*, Rev. Sci. Instr. **69**, 1974 (1998).
77. L. Paterson, M. MacDonald, J. Arlt, W. Sibbett, P. Bryant, and K. Dholakia, *Controlled rotation of optically trapped microscopic particles*, Science **292**, 912 (2001).
78. V. Bingelyte, J. Leach, J. Courtial, and M. Padgett, *Optically controlled three-dimensional rotation of microscopic objects*, Appl. Phys. Lett. **82**, 829 (2003).
79. J. Curtis, B. Koss, and D. Grier, *Dynamic holographic optical tweezers*, Opt. Commun. **207**, 169 (2002).
80. J. Curtis and D. Grier, *Structure of optical vortices*, Phys. Rev. Lett. **90**, 133901 (2003).

81. L. Sacconi, G. Romano, R. Ballerini, M. Capitanio, M.D. Pas, M. Giuntini, D. Dunlap, L. Finzi, and F. Pavone, *Three-dimensional magneto-optic trap for micro-object manipulation*, Opt. Lett. **26**, 1359 (2001).
82. M. Friese, T. Nieminen, N. Heckenberg, and H. Rubinsztein-Dunlop, *Optical alignment and spinning of laser-trapped microscopic particles*, Nature **394**, 348 (1998).
83. A.L. Porta and M. Wang, *Optical torque wrench: Angular trapping, rotation, and torque detection of quartz microparticles*, Phys. Rev. Lett. **92**, 190801 (2004).
84. J. Joykutty, V. Mathur, V. Venkataraman, and V. Natarajan, *Direct measurement of the oscillation frequency in an optical-tweezers trap by parametric excitation*, Phys. Rev. Lett. **95**, 193902 (2005).
85. P. Maivald, H.J. Butt, S.A.C. Gould, C.B. Prater, B. Drake, J.A. Gurley, V.B. Elings, and P.K. Hansma, *Using force modulation to image surface elasticities with the atomic force microscope*, Nanotechnology **2**, 103 (1991).
86. E. Meyer, R. Overney, K. Dransfeld, and T. Gyalog, *Friction and Rheology on the Nanometer Scale* (World Scientific, Singapore, 2002).
87. R.W. Carpick, D.F. Ogletree, and M. Salmeron, *Lateral stiffness: A new nanomechanical measurement for the determination of shear strengths with friction force microscopy*, Appl. Phys. Lett. **70**, 1548 (1997).
88. M.A. Lantz, S. J. O'Shea, M.E. Welland, and K.L. Johnson, *Simultaneous force and conduction measurements in atomic force microscopy*, Phys. Rev. B **55 (56)**, 10776 (15345) (1997).
89. A lock-in amplifier is used to vertically modulate the sample position, dD, and measure dF. We have chosen $dD = 1.3$ Å and we have worked at a frequency of 1.8967 kHz.
90. Tip deformations can be neglected due to the large difference in stiffness, one exception being presumably the smallest tube.
91. M.F. Yu, T. Kowaleweski, and R.S. Ruoff, *Investigation of the radial deformability of individual carbon nanotubes under controlled indentation force*, Phys. Rev. Lett. **85**, 1456 (2000).
92. W. Shen, B. Jiang, B.S. Han, and S.-s. Xie, *Investigation of the radial compression of carbon nanotubes with a scanning probe microscope*, Phys. Rev. Lett. **84**, 3634 (2000).
93. A.P. Boresi and O.M. Sidebottom, *Advanced Mechanics of Materials* (John Wiley & Sons, 5th Ed., ADDRESS, 1993).
94. G. Briggs, *Acoustic microscopy* (Oxford University Press, Oxford, 1992).
95. B. Cretin and F. Stahl, *Scanning microdeformation microscopy*, Appl. Phys. Lett. **62**, 829 (1993).
96. U. Rabe and W. Arnold, *Acoustic microscopy by atomic force microscopy*, Appl. Phys. Lett. **64**, 1493 (1994).
97. E. Dupas, Ph.D. thesis, Ecole Polytechnique Federale de Lausanne, 2000.
98. N. Burnham, A. Kulik, G. Gremaud, P. Gallo, and F. Oulevey, *Scanning local-acceleration microscopy*, J. Vac. Sci. Techn. B **14**, 794 (1996).
99. F. Oulevey, Ph.D. thesis, Ecole Polytechnique Federale de Lausanne, 1999.
100. G. Rochat, Y. Leterrier, C. Plummer, J. Manson, R. Szoszkiewicz, and A. Kulik, *Effect of substrate crystalline morphology on the adhesion of plasma enhanced chemical vapor deposited thin silicon oxide coatings on polyamide*, J. Appl. Phys. **95**, 5429 (2004).

101. We neglect any spurious cantilever's resonances or resonances of the AFM hardware and electronics.
102. O. Kolosov and K. Yamanaka, *Nonlinear detection of ultrasonic vibrations in an atomic force microscope*, Jpn. J. Appl. Phys. **32**, 22 (1993).
103. R. Szoszkiewicz, B. Bhushan, B.D. Huey, A. Kulik, and G. Gremaud, *Correlations between Adhesion Hysteresis and Friction at Molecular Scales*, J. Chem. Phys. **122**.
104. R. Szoszkiewicz, A. Kulik, and G. Gremaud, *Quantitative measure of nanoscale adhesion hysteresis by Ultrasonic Force Microscopy*, J. Chem. Phys. **122**, 134706 (2005).
105. R. Szoszkiewicz, B. Bhushan, B.D. Huey, A. Kulik, and G. Gremaud, *Adhesion hysteresis and friction at nanometer and micrometer lengths*, accepted in J. Appl. Phys. (2006).
106. T. Cuberes, G. Briggs, and O. Kolosov, *AFM-modes for non-linear detection of ultrasonic vibration* (Oxford University Press, Oxford, 1998).
107. F. Dinelli, M. Castell, D. Ritchie, N. Mason, G. Briggs, and O. Kolosov, *Mapping surface elastic properties of stiff and compliant materials on the nanoscale using ultrasonic force microscopy*, Phil. Mag. A **80**, 2299 (2000).
108. F. Dinelli, N. Burnham, A. Kulik, P. Gallo, G. Gremaud, and W. Benoit, *Mechanical properties studied at the nanoscale using Scanning Local-Acceleration Microscopy (SLAM)*, J. Phys IV **6**, 731 (1996).
109. K. Yamanaka, *UFM observation of lattice defects in highly oriented pyrolytic graphite*, Thin Solid Films **273**, 116 (1996).
110. O. Kolosov, M.R. Castell, C.D. Marsh, and G.A.D. Briggs, *Imaging the elastic nanostructure of Ge islands by ultrasonic force microscopy*, Phys. Rev. Lett. **81**, 1046 (1998).
111. F. Dinelli, H.E. Assender, and N. Takeda, *Elastic mapping of heterogeneous nanostructures with ultrasonic force microscopy (UFM)*, Surf. Interf. Anal. **27**, 562 (1999).
112. K. Porfyrakis, O. Kolosov, and H. Assender, *AFM and UFM surface characterization of rubber-toughened poly(methyl methacrylate) samples*, J. Appl. Pol. Sci. **82**, 2790 (2001).
113. H. Geisler, M. Hoehn, M. Rambach, M. Meyer, E. Zschech, M.M.A. Romanov, M. Bobeth, W. Pompe, and R. Geer, *Elastic mapping of sub-surface defects by ultrasonic force microscopy: limits of depth sensitivity*, Proc. of Conf. on Micr. Semicond. Mat. 2001 **169**, 527 (2001).
114. D. Hurley, M. Kopycinska-Muller, A. Kos, and R. Geiss, *Quantitative elastic property measurements at the nanoscale with atomic force acoustic microscopy*, Adv. Eng. Mat. **7**, 713 (2005).
115. S. Amelio, A. Goldade, U. Rabe, V. Scherer, B. Bhushan, and W. Arnold, *Measurements of elastic properties of ultra-thin diamond-like carbon coatings using atomic force acoustic microscopy*, Thin Solid Films **392**, 75 (2001).
116. P. Avouris, J. Appenzeller, R. Martel, and S.J. Wind, *Carbon nanotube electronics*, Proc. IEEE **91**, 1772 (2003).
117. J. Hone, M.C. Llaguno, M.J. Biercuk, A.T. Johnson, B. Batlogg, Z. Benes, and J.E. Fischer, *Thermal properties of carbon nanotubes and nanotube-based materials*, Appl. Phys. A **74**, 339 (2002).
118. E.T. Thostenson, Z. Ren, and T.W. Chou, *Advances in the science and technology of carbon nanotubes and their composites: a review*, Compos. Sci. Technol. **61**, 1899 (2001).

119. L. Roschier, R. Tarkiainen, M. Ahlskog, M. Paalanen, and P. Hakonen, *Manufacture of single electron transistors using AFM manipulation on multiwalled carbon nanotubes*, Microelectron. Eng. **61-62**, 687 (2002).
120. J.P. Lu, *Elastic properties of carbon nanotubes and nanoropes*, Phys. Rev. Lett. **79**, 1297 (1997).
121. V.N. Popov and V.E.V. Doren, *Elastic properties of single-walled carbon nanotubes*, Phys. Rev. B **61**, 3078 (2000).
122. Y. Xia, M.W. Zhao, Y.C. Ma, M.J. Ying, X.D. Liu, P.J. Liu, and L.M. Mei, *Tensile strength of single-walled carbon nanotubes with defects under hydrostatic pressure*, Phys. Rev. B **65**, 155415 (2002).
123. J.A. Elliot, J.K.W. Sandler, A.H. Windle, R.J. Young, and M.S.P. Shaffer, *Collapse of single-wall carbon nanotubes is diameter dependent*, Phys. Rev. Lett. **92**, 095501 (2004).
124. M.H. Park, J.W. Jang, C.E. Lee, and C.J. Lee, *Interwall support in double-walled carbon nanotubes studied by scanning tunneling microscopy*, Appl. Phys. Lett. **86**, 023110 (2005).
125. T. Hertel, R.E. Walkup, and P. Avouris, *Deformation of carbon nanotubes by surface can der Waals forces*, Phys. Rev. B **58**, 13870 (1998).
126. E.D. Minot, Y. Yaish, V. Sazonova, J.-Y. Park, M. Brink, and P.L. McEuen, *Tuning carbon nanotube band gaps with strain*, Phys. Rev. Lett. **90**, 156401 (2003).
127. S. Dag, O. Gulseren, S. Ciraci, and T. Yildirim, *Electronic structure of the contact between carbon nanotube and metal electrodes*, Appl. Phys. Lett. **83**, 3180 (2003).
128. P. Avouris, T. Hertel, R. Martel, T. Schmidt, H.R. Shea, and R.E. Walkup, *Carbon nanotubes: nanomechanics, manipulation, and electronic devices*, Appl. Surf. Sci. **141**, 201 (1999).
129. V. Lordi and N. Yao, *Radial compression and controlled cutting of carbon nanotubes*, J. Chem. Phys. **109**, 2509 (1998).
130. L. Shen and J. Li, *Transversely isotropic elastic properties of single-walled carbon nanotubes*, Phys. Rev. B **69**, 045414 (2004).
131. I. Palaci, S. Fedrigo, H. Brune, C. Klinke, M. Chen, , and E. Riedo, *Radial Elasticity of Multiwalled Carbon Nanotubes*, Phys. Rev. Lett. **94**, 175502 (2005).
132. B.T. Kelly, *Physics of Graphite* (PUBLISHER, ADDRESS, 1981).
133. Z.L. Wang, *Nanobelts, nanowires, and nanodiskettes of semiconducting oxides - From materials to nanodevices*, Adv. Mater. **15**, 432 (2003).
134. M. Buongiorno-Nardelli, J.-L. Fattebert, D. Orlikowski, C. Roland, Q. Zhao, and J. Bernholc, *Mechanical properties, defects and electronic behavior of carbon naotubes*, Carbon **38**, 1703 (2000).
135. G. Zhang, M. Long, Z.-Z. Wu, and W.-Q. Yu, *Mechanical properties of hepatocellular carcinoma cells*, World Journal of Gastroenterology **8**, 243 (2002).
136. H.F. Bettinger, T. Dumitrica, G.E. Scuseria, and B.I. Yakobson, *Mechanically induced defects and stregth of BN nanotubes*, Phys. Rev. B **65**, 041406 (2002).
137. J.P. Salvetat, J.M. Bonard, N.H. Thomson, A.J. Kulik, L. Forró, W. Benoit, and L. Zuppiroli, *Mechanical properties of carbon nanotubes*, Appl. Phys. A **69**, 255 (1999).
138. J.P. Salvetat, A.J. Kulik, J.M. Bonard, and et al., *Elastic Modulus of Ordered and Disordered Multiwalled Carbon Nanotubes.*, Adv. Mater. **11**, 161 (1999).

139. L. Shen and J. Li, *Transversely isotropic elastic properties of multiwalled carbon nanotubes*, Phys. Rev. B **71**, 035412 (2005).
140. L. Vayssieres, *Growth of arrayed nanorods and nanowires of ZnO from aqueous solutions*, Adv. Mater. **15**, 464 (2003).
141. P.X. Gao, Y. Ding, W. Mai, W.L. Hughes, C. Lao, and Z.L. Wang, *Conversion of Zinc Oxide Nanobelts into Superlattice- Structured Nanohelices*, Science **309**, 1007 (2005).
142. S.X. Mao, M. Zhao, and Z.L. Wang, *Nanoscale mechanical behavior of individual semiconducting nanobelts*, Appl. Phys. Lett. **83**, 993 (2003).
143. M.H. Zhao, Z.-L. Wang, and S.X. Mao, *Piezoelectric Characterization of Individual Zinc Oxide Nanobel Probed by Peizoresponse Force Microscope*, Nanoletters **4**, 587 (2004).
144. E. Evans, A. Leung, and D. Zhelev, *Synchrony of cell spreading and contraction force as phagocytes engulf large pathogens*, J.Cell. Biol **122**, 12951300 (1993).
145. T. Oliver, J. Lee, and K. Jacobson, *?*, Semin. Cell Biol **5**, 139 (1993).
146. M. Lekka, P. Laidler, D. Gil, J. Lekki, Z. Stachura, and A.Z. Hrynkiewicz, *Elasticity of normal and cancerous human bladder cells studied by scanning force microscopy*, Eur. Biophys J **28**, 312 (1999).
147. W.H. Goldmann and R.M. Ezzell, *Viscoelasticity in wild-type and vinculin-deficient (5.51) mouse F9 embryonic carcinoma cells examined by atomic force microscopy and rheology*, Experimental Cell Research **226**, 234 (1996).
148. W.H. Goldmann, R. Galneder, M. Ludwig, W. Xu, E.D. Adamson, N. Wang, and R.M. Ezzell, *Differences in elasticity of vinculin-deficient F(cells measured by magnetometry and atomic force microscopy*, Experimental Cell Research **239**, 235 (1998).
149. W.H. Goldmann, R. Galneder, M. Ludwig, A. Kromm, and R.M. Ezzell, *Differences in F9 and 5.51 cell elasticity determined by cell poking and atomic force microscopy*, FEBS Letters **424**, 139 (1998).
150. H.G. Hansma, *Surface Biology of DNA by Atomic Force Microscopy*, Annu. Rev. Phys. Chem **52**, 71 (2001).
151. J.L. Alonso, , and W.H. Goldmann, *Feeling the forces: atomic force microscopy in cell biology*, Life Sciences **72**, 2553 (2003).
152. A.D. Mehta, M. Rief, J.A. Spudich, D.A. Smith, and R.M. Simmons, *Single-Molecule Biomechanics with Optical Methods*, Science **283**, 1689 (1999).
153. E. Ferrari, V. Emiliani, D. Cojoc, V. Garbin, M. Zahid, C. Durieux, M. Coppey-Moisan, and E.D. Fabrizio, *Biological samples micro-manipulation by means of optical tweezers*, Microelectronic Engineering **78-79**, 575 (2005).
154. F. Lopez, A. Lundkvist, M. Balooch, D. Haupt, J. Kinney, S. Oesterle, P. Fitzgerald, and P. Yock, *Plaque extrusion during balloon angioplasty: New evidence from x-ray microtomography*, JOURNAL OF THE AMERICAN COLLEGE OF CARDIOLOGY **29**, 7491 (1997).
155. S. Habelitz, S.J. Marshall, G.W. Marshall, and M. Balooch, *Mechanical properties of human dental enamel on the nanometre scale*, ARCHIVES OF ORAL BIOLOGY **46**, 173 (2001).
156. S. Habelitz, G.W. Marshall, M. Balooch, and S.J. Marshall, *Nanoindentation and storage of teeth*, JOURNAL OF BIOMECHANICS **35**, 995 (2002).
157. J. Kinney, M. Balooch, S. Marshall, G.W. Marshall, and T. Weihs, *Hardness and Young's modulus of human peritubular and intertubular dentine*, ARCHIVES OF ORAL BIOLOGY **41**, 9 (1996).

158. T.T. Perkins, D.E. Smith, R.G. Larson, and S. Chu, *Stretching of a single tethered polymer in a uniform-flow*, Science **268**, 83 (1995).
159. P. Cluzel, A. Lebrun, C. Heller, R. Lavery, J.-L. Viory, D. Chatenay, and F. Caron, *DNA: An Extensible Molecule*, Science **271**, 792 (1996).
160. S.B. Smith, Y. Cui, and C. Bustamante, *Overstretching B-DNA: The Elastic Response of Individual Double-Stranded and Single-Stranded DNA Molecules*, Science **271**, 795 (1996).
161. R.M. Simmons, J.T. Finer, S. Chu, and J. Spudich, *Quantitative measurements of force and displacement using an optical trap*, Biophys. J. **70**, 1813 (1996).
162. M.D. Wang, H. Yin, R. Landick, J. Gelles, and S.M. Block, *Stretching DNA with optical tweezers*, Biophys. J. **72**, 1335 (1997).
163. S.B. Smith, Y. Cui, A.C. Hausrath, and C. Bustamante, *?*, Biophys. J. **68**, A250 (1995).
164. P. Cizeau and J.-L. Viovy, *Modeling extreme extension of DNA*, Biopolymers **42**, 383 (1997).
165. A. Ahsan, J. Rudnick, and R. Bruinsma, *Elasticity theory of the B-DNA to S-DNA transition*, Biophys. J. **74**, 132 (1998).
166. M. Grandbois, M. Beyer, M. Rief, H. Clausen-Schaumann, and H.E. Gaub, *How strong is a covalent bond?*, Science **283**, 1727 (1999).
167. D. Bensimon, A.J. Simon, V. Croquette, and A. Bensimon, *Stretching DNA with a receding meniscus - experiments and models*, Phys. Rev. Lett. **74**, 4754 (1995).
168. O. Kratky and G. Porod, *X-ray investigation of dissolved chain molecules*, Rec.Trav.Chim.Pays.Bas **68**, 1106 (1949).
169. C. Bustamante, Z. Bryant, and S.B. Smith, *Ten years of tension: single-molecule DNA mechanics*, Nature **421**, 423 (2003).
170. J. Zlatanova and S.H. Leuba, *Magnetic tweezers: a sensitive tool to study DNA and chromatin at the single-molecule level*, Biochem. Cell Biol **81**, 151 (2003).
171. G. Lee, L. Chrisey, and R. Colton, *Direct measurement of the forces between complementary strands of DNA*, Science **266**, 771 (1994).
172. T. Strunz, K. Oroszlan, R. Schafer, and H.J. Guntherodt, *Dynamic force spectroscopy of single DNA molecules*, Proc. Natl. Acad. Sci. USA **96**, 11277 (1999).
173. H. Clausen-Schaumann, M. Rief, C. Tolksdorf, and H.E. Gaub, *Mechanical stability of single DNA molecules*, Biophys. J. **78**, 1997 (2000).
174. M. Rief, H. Clausen-Schaumann, and H.E. Gaub, *Sequence-dependent mechanics of single DNA molecules*, Nat. Struct. Biol. **6**, 346 (1999).
175. N. Anderson, A. Hartschuh, S. Cronin, and L. Novotny, *Nanoscale vibrational analysis of single-walled carbon nanotubes*, J. Am. Chem. Soc. **127**, 2533.

13 Mechanical Properties of Metallic Nanojunctions

G. Rubio-Bollinger[1,2], J.J. Riquelme[1], N. Agraït[1,2], and S. Vieira[1,2]

[1] Departamento de Física de la Materia Condensada C-III, Universidad Autónoma de Madrid, E-28049 Madrid, Spain.
[2] Instituto Universitario de Ciencia de Materiales "Nicolás Cabrera", Universidad Autónoma de Madrid, E-28049 Madrid, Spain.

13.1 Introduction

The mechanical properties macroscopic metals cannot be scaled down to nanocontacts between metals as discreetness of matter, detailed atomic configuration and quantum effects become relevant when the size of the contact between two macroscopic bodies is of atomic dimensions. Investigation of elastic properties, wear and fracture of single nanometer scale structures has in the last decades become experimentally accessible through the use of local probe devices, such as the scanning tunneling microscope (STM) and related techniques. Theoretically, molecular dynamics (MD) simulations and Density Functional Theory (DFT) calculations have provided deeper insight into the behavior of matter at the nanoscale. In the case of nanocontacts between metals the study of electron transport has proved to be a very convenient tool for their characterization, providing deeper knowledge of the mechanical properties. At the very smallest scale of one-atom contacts quantum effects become dominant, and the interplay between mechanical and electron transport properties is revealed. The term nanojunctions is used to embrace metallic nanocontacts and atomic-size tunneling junctions. The study of the mechanical properties of tunneling junctions between single asperities or tips, a configuration which takes place both before and after the formation or rupture of nanocontacts, provides deeper insight into the imaging mechanisms and tribological processes in local probe imaging of surfaces at the nanometer scale.

13.2 Electron Transport through Metallic Nanocontacts

Measurement of electric transport in atomic-sized contacts between metals provides rich information related to the size, shape, state of strain, elastic deformation, wear and fracture of metallic nanocontacts [1]. The information which is extracted from electrical conductance measurements complements direct experiments on the mechanical properties of metallic contacts. In particular, the minimal cross-section of a nanocontact can be estimated, under some circumstances, from electrical transport properties.

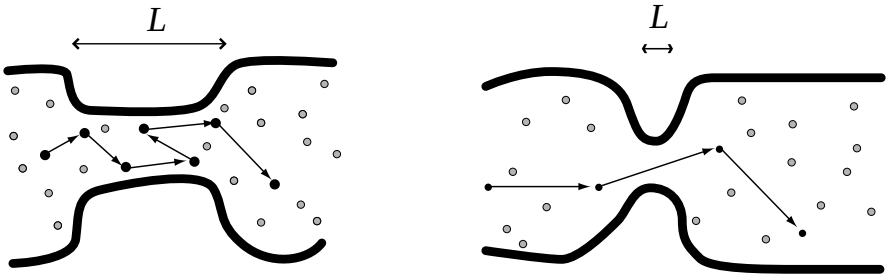


Fig. 13.1. Schematic illustration of electron transport in a diffusive (*left*) and ballistic (*right*) conductor

Electron transport in metallic macroscopic conductors is characterized by Ohm's law, which states that the conductance G of a sample is proportional to the cross-section area S and inversely proportional to its length L: $G = S/(L\rho)$, where ρ is the resistivity of the material.

However, in metallic point-contacts one can identify different electron transport regimes depending on the relative size of various length scales related to the electron scattering mechanisms. An important length scale is the elastic mean free path ℓ, which is roughly the distance between elastic collisions with static impurities in the metal. Depending on the relative values of the mean free path ℓ and the size of the point-contact L, electron transport is said to be in the diffusive regime if $L \gg \ell$ or in the ballistic regime if $L \ll \ell$ (see Fig. 13.1).

Electron motion in the diffusive regime can be viewed as a random walk of step size ℓ. The conductance in the diffusive regime was already addressed by Maxwell [2]. An analytical solution can be obtained if the point-contact is modelled as a constriction of hyperbolic geometry. The conductance is calculated using Ohm's law for the solution of the potential distribution in such geometry. The resulting conductance G_{M} has a very simple form if the opening angle of the hyperboloid is large: $G_{\mathrm{M}} \approx 2a/\rho$, where a is the point-contact minimal cross-section radius and ρ is the resistivity of the metal.

Metallic nanocontacts, whose typical size L can be below 1 nm, usually fall in the ballistic conductance regime $L \ll \ell$ because for pure metallic materials the mean free path can be of the order of tens on nanometers (especially at low temperatures). In this regime electrons pass through the contact ballistically. There is a large potential gradient across the contact and the electrons accelerate within a short distance. A semiclassical analysis of the conductance of conductance of point-contacts in the ballistic regime was considered by Sharvin [3,4], who remarked the similarity with the effusion of gas molecules through a small orifice in the Knudsen regime. The simplest model of a point contact is an orifice of radius a between two electron reservoirs and the so-called Sharvin's conductance G_{S} is given by:

$$G_{\mathrm{S}} = \frac{2e^2}{h}\left(\frac{k_{\mathrm{F}}a}{2}\right)^2 \tag{13.1}$$

where h is Planck's constant, e is the electron charge and k_F is the Fermi wave vector. This formula can also be expressed in terms of the conductance quantum $G_0 = 2e^2/h$. Note that the Sharvin's conductance is material dependent only through k_F, that is, independent of the resistivity and therefore purity and temperature, contrary to the conductance in the diffusive regime. Thus, the conductance of a metallic nanocontact in the ballistic regime provides a reliable estimate of its minimal cross-section area using Sharvin's formula. Note that for gold the Sharvin's formula predicts a conductance close to G_0 per atom in the minimal cross-section of the constriction. In terms of the inverse of the conductance, the resistance is close to 13 kΩ per atom.

13.3 Experimental Tools

A variety of experimental tools have been used to investigate mechanical properties at the nanoscale. We fill focus in this section on devices which are especially well suited for the study of metallic nanocontacts.

13.3.1 The Scanning Tunneling Microscope Supplemented with a Force Sensor

The scanning tunneling microscope (STM) is a versatile tool that allows studying the topography and electronic properties of metallic surfaces with atomic resolution, and is also ideal to study metallic nanocontacts. In order to simultaneously measure mechanical and electrical properties the microscope has to be supplemented with a force sensor. Conventional Atomic Force Microscopes (AFM) with conductive tips are usually not well suited for these experiments because the cantilever elastic constant is too low (≤ 1 N/m) to be able to fabricate and manipulate metallic nanocontacts, whose effective elastic constant is above 10 N/m. Therefore, modified versions of combined AFM/STM techniques are required [5], using bending beams with elastic constants above 100 N/m. One of such devices is shown in the inset of Fig. 13.7.

13.3.2 The Mechanically Controllable Break-Junction Technique

The principle of the mechanically controllable break-junction technique (MCBJ) is shown in Fig. 13.2 [7, 8]. The metal to be studied is in the form of a thin wire (diameter 0.01–0.25 mm), which is fixed at two closely spaced spots on a flexible insulated substrate. A notch is cut in the wire between the two fixing points, producing a weakened region in the wire. The substrate is mounted in a three-point bending configuration, where either the central or the counter supports can be displaced resulting in controlled bending of the substrate. This bending causes the top surface of the substrate to expand, which results in the elongation of the weakest part of the wire at the notch

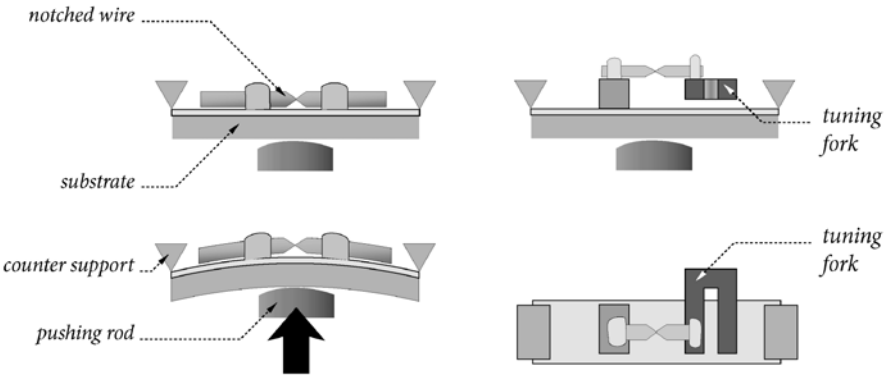


Fig. 13.2. Schematic top and side view of the mounting of a MCBJ (*left*) and a MCBJ supplemented with a tuning fork resonator which acts a a force gradient detector (*right*)

until rupture. In this way two clean fracture surfaces are exposed. These fractured surfaces can be brought again into contact by relaxing the bending force on the substrate. A combination of micrometric screws and piezoelectric transducers are commonly used for fine control of the opening at the notch.

One main advantage of the MCBJ technique is that the freshly exposed surfaces are free of contamination, and can be kept clean if the device is kept in UHV or cryogenic vacuum. A second advantage is the mechanical stability of the two electrodes with respect to each other, which results from the short mechanical loop between electrodes, that is, the distance between the two fixed points of the wire at both sides of the notch, which can be as small as 0.1 mm.

The MCBJ technique has been widely used during the last decade to study electron transport through metallic nanocontacts, and has been recently supplemented with a tuning fork force sensor in order to simultaneously study their mechanical properties [8,9]. The notched wire is not directly glued onto the flexible substrate but instead one of its ends is glued to one prong of a miniature quartz tuning fork (TF) whose base is attached to the flexible substrate, so that the prongs of the fork are freely standing parallel to the substrate (see Fig. 13.2). The other end of the wire is attached to support levelled with the prongs of the TF.

The force gradient or stiffness k of the interaction between the electrodes is proportional, with a factor $\alpha = \Delta f/k$ which depends on the mechanical properties of the TF, to the shift of the resonance frequency Δf of the TF excited with vanishing amplitude. This resonant frequency is tracked by implementing a phase locked loop oscillator [8,10–12].

13.4 Mechanical Properties of Metallic Nanocontacts

This section is devoted to the fabrication procedure, mechanical processes and shape characterization of nanocontacts between gold electrodes. The minimum cross-section of these constrictions ranges from tens of atoms down to single-atom contacts.

13.4.1 Fabrication of Metallic Nanocontacts

Single atomic-sized contacts between metals can be produced by a modified use of a scanning tunneling microscope (STM). In its normal operation mode the tip is scanned over the sample surface without making contact. The tip-sample separation is kept constant by controlling the current that flows between tip and sample, due to the tunneling effect, when a fixed bias voltage is applied between them. Typical currents are of the order of nanoamperes, corresponding to a tunneling junction resistance in the G Ω range, and the tip-sample distance is a few Ångstrom.

However, the STM was soon used to modify the sample surface on a nanometer scale. In the experiment by Gimzewski and Möller [13] the surface was gently touched with the tip and the transition from the tunneling regime to metallic contact was observed as an abrupt jump in the current. This jump in current was due to a change in resistance of about 13 kΩ. According to Sharvin's formula a conductance of $1G_0$ corresponds to a contact diameter of 0.25 nm, suggesting that a contact as small as one-atom can be formed.

In an STM experiment on metallic nanocontacts the bias voltage between tip and sample is kept fixed and the current through the junction is recorded as the tip-sample distance is varied, performing an indentation cycle of approach and subsequent retraction (Fig. 13.3). As the tip is initially approached to the sample, the electronic current depends exponentially on the tip displacement due to the tunneling effect. Further approach results

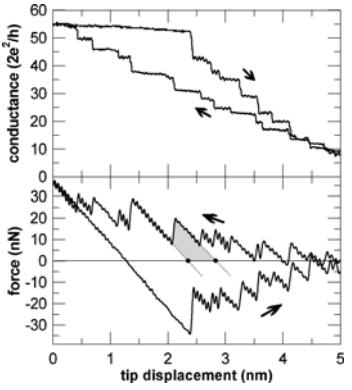


Fig. 13.3. Simultaneous measurement of conductance and force during an indentation cycle. *Arrows* indicate compression and subsequent elongation of the nanocontact. Reprinted figure with permision from [16]. Copyright (1997) by the American Physical Society

in metallic contact, which is signaled by and abrupt increase of the current (the conductance G for this first contact is close to $2e^2/h$). As indentation continues the contact cross-section increases. For ductile metals subsequent retraction of the tip usually results in the formation of a connective neck which gradually gets thinner until it finally breaks.

The whole indentation cycle (without breaking the contact) can be followed measuring the electrical conductance (Fig. 13.3). As the contact size increases its conductance also increases but the curve is not smooth and shows a step like behavior. In this experiment the tensile force was simultaneously recorded using an STM supplemented with a force sensor (see Sect. 13.3). The force curve shows a sequence of linear stages separated by sudden relaxations. The linear stages, which show as nearly constant conductance values (conductance plateaus) are related to the elastic deformation of the nanocontact. At a given tensile force the nanocontact abruptly yields, the atomic configuration changes irreversibly, and there is a sudden force relaxation. This behavior was predicted in early MD simulations by Landman et al. [15]. This yielding reflects as a sharp jump in the contact conductance, related to the sudden change in the minimal cross-section due to the atomic rearrangement. The relaxation in the tensile force is the result of a change in the length of the constriction. This length can be determined from the distance between successive relaxed configurations in the nanocontact separated by a yielding event (*black dots* in Fig. 13.3). In gold nanocontacts this plastic relaxation length ranges from 0.2 to 1 nm, which implies that only a few atomic layers participate in the plastic deformation process.

13.4.2 Elasticity and Fracture of Metallic Nanocontacts

The simultaneous measurement of conductance and tensile force during the indentation cycle described above provides information about the mechanical behavior differences between macroscopic specimens and metallic nanocontacts. When a macroscopic solid is subjected to a load it undergoes a change in shape. For small loads this deformation is elastic and the specimen recovers its original shape as the load is removed. On an atomic scale, elastic strain consists in small changes of the inter-atomic spacing, that is, in the stretching of inter-atomic bonds. Hence the modulus of elasticity E is a measure of the resistance of inter-atomic bonds to deform. The compliance of metallic nanocontacts cam be obtained from the slope of the elastic stages of the force curve during an indentation cycle. The effective elastic constant for gold nanocontacts is consistent with continuum mechanics models.

For most metallic materials elastic deformation is only possible for strains smaller than 0.005 (0.5%). As the material is deformed beyond this point, permanent, nonrecoverable, or plastic deformation takes place. In macroscopic metal specimens this transition from elastic to plastic behavior, or yielding, occurs gradually, and it is difficult to fix the lower limiting stress below which there is no plastic deformation. Commonly the yield strength σ_y is defined as

the stress necessary to produce a plastic strain of 0.002 under uniaxial stress. The yield strength of a metal is very sensitive to prior deformation, to the presence of impurities and to heat treatment, in contrast to the modulus of elasticity which is insensitive to these factors.

On an atomic scale, plastic deformation corresponds to the breaking of bonds between neighboring atoms and subsequent reforming bonds with the new neighbors. Atoms change positions and upon removal of stress they do no return to their original positions resulting in a permanent change of the shape of the specimen. The simplest model of plastic deformation of a perfect crystal considers the sliding of two compact planes with respect to each other. The maximum shear stress $\tau_{\max}$ required for this process was calculated by Frenkel [14] to be $\tau_{\max} \approx G/30$. Where G is the shear modulus of the material. This value of the shear modulus is much larger than those observed in macroscopic metal specimens. This is due to the presence of dislocations, which can glide at low stress values. High values of the shear stress are only observed in dislocation-free specimens like whiskers. In addition, also high values of the shear stress are expected in metallic nanocontacts, where dislocations are unstable and are quickly expelled from nm-volume regions. The maximum pressure that the nanocontact can sustain before relaxation can be computed from the measured force and the contact area. This are is obtained from the measured conductance using Sharvin (see Eq. (13.1)). The apparent pressure ranges from 3 to 6 GPa in gold nanocontacts both for the compressive and tensile branches of the indentation cycle. This apparent pressure is more than 20 times larger than the maximum pressure a macroscopic contact can sustain and is of the same order of magnitude as the theoretical value in the absence of dislocations and is consistent with the theoretical maximum shear stress value for gold.

The energy dissipated in each force relaxation, that is, the energy necessary to produce a configurational change, can be directly obtained from the force cycle. In Fig. 13.3, the energy to pass from one configuration to the next is given by the grayed area. The value of this energy is of the order of 0.1 eV per atom in the minimal cross-section of the contact. If we compare with the heat of fusion (0.13 eV/atom) we find that configurational changes take place only at the zone around the narrowest part of the nanocontact.

13.4.3 The Shape of Metallic Nanocontacts

The STM has been used to estimate nanocontact geometry studied the local modifications produced by touching a Pt-Ir tip a Ag substrate in UHV, by imaging the surface after tip indentation [13]. They showed that for clean metal-metal contacts, after a gentle indentation cycle of the tip into the substrate, the topography image showed a protrusion of nanometer size on the sample surface, indicating that during the indentation a small neck was formed and subsequently stretched and broken.

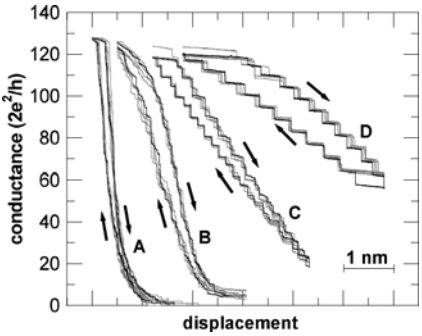


Fig. 13.4. Conductance curves $G(z)$ for four different indentation cycles. The different slope of the $G(z)$ curves are due to different constriction shapes. Reprinted figure with permision from [16]. Copyright (1997) by the American Physical Society

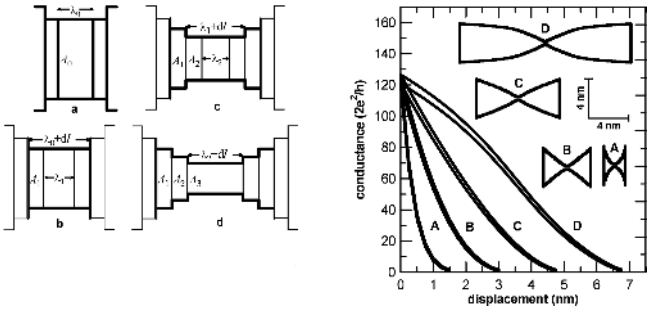


Fig. 13.5. *Left*: slab model of plastic deformation of a constriction. *Right*: Calculated conductance curves, using the slab model, for constriction shapes shown in Fig. 13.4. Reprinted figures with permision from [16]. Copyright (1997) by the American Physical Society

The shape of constrictions formed by such plastic deformation can be estimated from the conductance vs displacement curves $G(z)$ during an indentation cycle (Fig. 13.4) [16]. A slab model (see left panel in Fig. 13.5) of the point contact is considered where the minimal cross-section slab determines the conductance through Sharvin's formula. When the constriction is submitted to a tensile force, stresses are highest in the narrowest part, with cross-sectional area A_i, and we may assume that plastic deformation takes place in that narrowest slab, in a zone of depth λ, leaving the rest of the neck unmodified. The cross-sectional area of the new slab A_{i+1} is given by volume conservation

$$A_{i+1} = \frac{A_i \lambda_i}{\lambda_i + \Delta l} \tag{13.2}$$

where $\lambda_i + \Delta l$ is the length of the new slab. Only the narrowest slab is modified, and the shape of the constriction after a number of plastic deformation processes results in a sequence of values of A_i and λ_i. The plastic deformation length λ_i can be obtained from the experimental $G(z)$ curve noting that for the limit $\Delta l \to 0$, $\lambda = -(\mathrm{d} \ln A / \mathrm{d} l)^{-1}$, where A is the cross-section of the narrowest portion of the constriction.

Figure 13.4 shows four different sets of indentation cycles for gold. The markedly different slopes of the conductance curves can be used to obtain the constriction shape (Fig. 13.5, *right panel*) using the described slab model. Steeper $G(z)$ curves correspond to constrictions with larger opening angle and involve shorter plastic deformation lengths. For constrictions with this shape only one atomic layer is involved in plastic deformation at yielding events.

13.5 Suspended Chains of Single Gold Atoms

It was recently discovered that suspended chains up to 7 atoms in length could be, under some circumstances, extracted from gold nanocontacts. The fabrication procedure and some of their mechanical properties will be described in this section.

13.5.1 Fabrication of Chains

The breaking process of metallic nanocontacts by controlled separation of the electrodes takes place in a discrete sequence of elastic deformation and abrupt yield stages resulting in a non continuous reduction of the minimal cross-section of the contact. It has been shown that for some metals this process takes place down to the smallest nanocontact, a single atom contact between the electrodes [5]. Further separation of the electrodes usually results in breakage of the metallic contact and an abrupt jump into the tunneling regime. The experiments show that a one-atom contact accommodate a maximum elastic deformation below 0.25 nm. It was discovered, however, that under certain circumstances, gold nanocontacts exhibit a different behavior. Figure 13.6 shows a trace of the conductance during the breaking sequence of a gold nanocontact. The conductance decreases step-like until a one-atom contact is formed, which can be identified by a conductance close to G_0. Strikingly, this one atom-contact can be further stretched by a distance larger than 1 nm, without the conductance deviating appreciably from

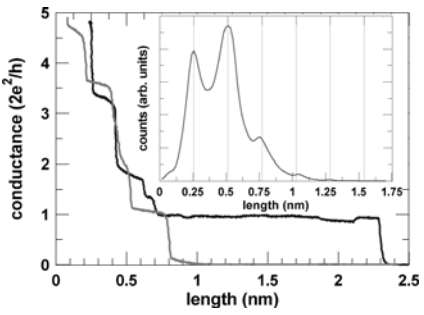


Fig. 13.6. Evolution of conductance during the extraction of an atomic chain out of a gold nanocontact (*black*). After rupture of the chain the electrodes have to travel back to reestablish metallic contact (*gray*). Inset: histogram of chain length obtained in 10,000 nanocontact breaking experiments. Reprinted figure with permision from [19]. Copyright (2002) by the American Physical Society

G_0, showing up in Fig. 13.6 as a conductance plateau. This behavior is due to the formation of a suspended chain formed by several gold atoms as the electrodes are further separated by a distance of up to 1.5 nm [17, 18]. Once the chain starts being pulled the conductance never exceeds G_0, confirming that the chain acts as one-dimensional quantized nanowire. When the chain finally breaks, the electrodes have to travel back a return distance to re-establish metallic contact. This return distance is almost equal to the length of the last plateau itself, suggesting that after the chain breaks, its constituent atoms collapse onto the electrodes on either side.

The conductance traces for successive nanocontact ruptures do not reproduce in detail, as they depend on the exact atomic positions in the contact and not every contact rupture results in the formation of a chain of atoms [17]. The probability of formation of such a structure can be quantified by constructing a histogram of last plateau length, that is, the elongation distance of one-atom contacts, characterized by a conductance close to G_0. Instead of a smooth distribution, a series of equidistant peaks is found. The probability of pulling a chain of length L decreases rapidly for large L. The peaked structure of the histogram shows that atomic chains tend to be elongated by integer multiples of 0.25 nm, which is close to the nearest neighbor spacing of gold atoms in the crystal.

Despite the low probability of formation of chains, once an atomic chain is pulled, the retraction of the electrode can be stopped and the chain remains very stable at liquid helium temperatures (4.2 K): some of the longest chains obtained in the experiments have been held stable for as long as 1 hour. This makes atomic chains suitable for investigation of one-dimensional electron transport and for studies of wear and fracture on low-coordinated metallic nanostructures.

13.5.2 Mechanical Processes During Formation of Atomic Chains

The mechanical processes involved during the formation and rupture of atomic chains of gold have been studied both experimental and theoretically. Figure 13.7 shows a simultaneous measurement of conductance and tensile force while pulling an atomic chain out of the electrodes. The force curve shows a sawtoothlike signal decreasing in amplitude in a sequence of elastic deformation stages separated by sudden force relaxations related to atomic rearrangements in the nanocontact. Once a one atom-contact is obtained, characterized by a conductance close to G_0, the structure is further stretched a distance of about 1 nm while the conductance remains close to G_0 and the force shows large irreversible relaxations between linear stages. In stages with a linearly growing tensile force the chain is elastically stretched, while at the force jumps abrupt atomic rearrangements occur.

The breaking of a nanocontact involves breaking many individual atomic bonds. The question of whether or not an atomic chain is formed while pulling a nanocontact is related to the relative strength of different bonds in specific

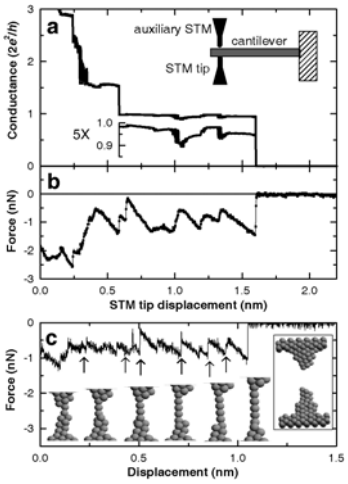


Fig. 13.7. Simultaneous recording of conductance and tensile force during formation of a chain of gold atoms. Bottom: molecular dynamics simulation of this process including snapshots at different stages of elongation. Inset: combined STM/AFM device used in the experiment. The auxiliary STM is used to follow the bending of the cantilever. Reprinted figure with permision from [6]. Copyright (2001) by the American Physical Society

atomic configurations. As a simple model we can consider a chain connected to the electrodes on each side. If the system is stretched beyond a critical force the weakest of the bonds will break. Molecular dynamics simulations show that it is possible that, at a force smaller than the force needed to break the chain, one of the atoms in the electrodes can slip into the chain. This will usually not be the case because pulling an atom out of a surface requires to break more bonds than to break the chain.

However, both the detailed configuration of atoms at the chain to electrode interface and the relative strength of the bonds play a major role. Consider for instance a realistic atomic configuration in which the contact has a pyramidal shape. The atom sitting at the top will have three bonds with the underlying atoms and one with the chain. By pulling, this atom can be moved and incorporated into the chain breaking only one bond to the underlying atoms. In addition, it is well known for metallic cohesion, that the bond strength increases as the coordination number is reduced, a fact which favors the probability of extraction of suspended atomic chains from the electrodes. Molecular dynamics calculations show that generally larger force jumps correspond to incorporation of an atom into the bridging atomic chain. As long as the force required to rearrange the atoms in the electrodes is smaller that the breaking force, the chain will grow in length. Note that mechanical relaxations in Fig. 13.7 take place at force values smaller than the final breaking force.

Nanoelastic properties of the chain have been studied with molecular dynamics simulations, which show that during the elastic elongation of the chain, most of the deformation is accumulated in the electrodes, not in the chain itself (Fig. 13.8). This is a combination of two effects. First, bonds in the chain are stronger due to its low coordinated situation and second, atoms in the electrodes are sitting in arrangements where the breaking of bonds

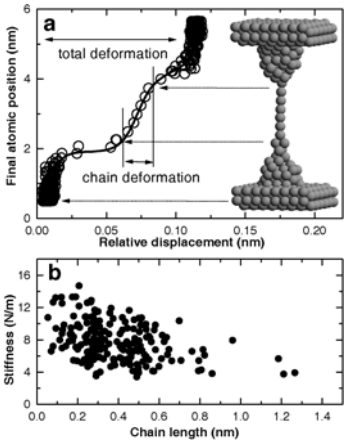


Fig. 13.8. Molecular dynamics simulation showing where the elastic deformation takes place during elongation of an atomic chain (*top*). *Bottom*: Measured chain stiffness as a function of its length. Reprinted figure with permision from [6]. Copyright (2001) by the American Physical Society

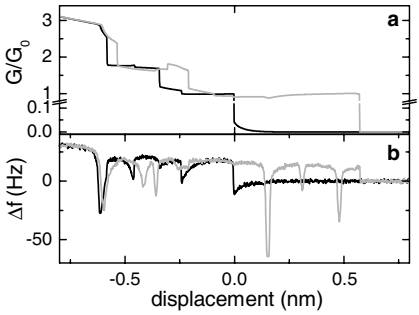


Fig. 13.9. Simultaneous measurement of conductance and force gradient during elongation of a chain of atoms. Reprinted figure with permision from [8]. Copyright (2004) by the American Physical Society

can proceed not through a direct radial stretching but rather through a more concerted motion of the atoms, giving longer paths and hence smaller forces. The consequence is the peculiar feature of the nanoscopic system that thinner is actually stronger.

The MCBJ technique supplemented with a force gradient sensor [8, 9] (described in Sect. 13.3) has also been used to detect bond weakening at extreme tensile forces on atomic chains. Figure 13.9 shows the elastic constant during extraction of the chain.

13.6 Metallic Adhesion in Atomic-Sized Tunneling Junctions

The mechanical properties of small tunneling junctions, as those during STM imaging have also been investigated [20, 21]. The forces between the tip and the surface play an important role in the process of spontaneous jump to contact when the separation in small enough. We focus here on recent experiments on gold nanocontact in which the interaction between tips is studied by measuring simultaneously the force gradient and the tunneling current

as the inter-electrode distance is varied. Increased stability and sensitivity can be obtained by using the MCBJ technique supplemented with a force sensor [8,9] (see Sect. 13.3). In addition, with this technique, freshly exposed fractured wires at cryogenic temperatures (4.2 K) are free of contamination.

Different atomic configurations of the tip apex are measured by making a large contact and breaking it again. Figure 13.10 shows the current and junction stiffness while approaching the tips from the tunneling regime until a one-atom contact is established. As the gap is reduced there is a spontaneous jump to contact during which both the conductance and the force gradient (proportional to the resonant frequency shift) have an abrupt jump. A similar behavior is observed upon retraction (jump out of contact). From the contact portion of the curve we can directly obtain the elastic constant of the one-atom contact, which in this case is 5.8 N/m. The tunneling portion of the force gradient curve during approach is negative, which indicates an attractive force between the tips. Note that the decay length of this force is below 0.5 Å, indicating a short range metallic interaction. This metallic adhesion is due to the overlap of the electronic wave functions. Various theoretical models [22,23] for metallic adhesion predict a decay length similar to the tunneling current decay length.

The mechanical properties of metallic nanojunctions are of fundamental interest, not only theoretically but also from an applied point of view since contact in macroscopic bodies typically occurs at numerous asperities of small size, whose mechanical properties determine those of the contact. This explains the interest of investigating small size contacts for many technologically important problems like adhesion, friction, wear, lubrication, fracture and machining. We have shown how the experiments in which mechanical and electrical measurements are combined are essential for understanding the physics of these systems due to the intimate relation which exists between the mechanical and electrical properties of metallic nanocontacts.

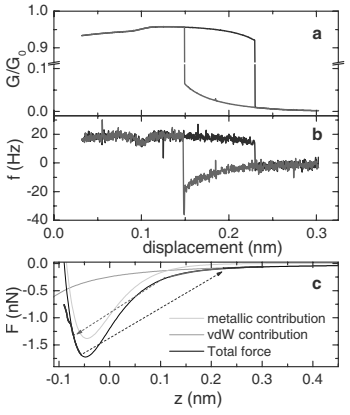


Fig. 13.10. Simultaneous measurement of conductance and force gradient during the formation of a one-atom contact after a jump to contact event from the tunneling regime. Reprinted figure with permision from [8]. Copyright (2004) by the American Physical Society

Acknowledgement. The authors would like to acknowledge financial support from Spanish CAM Contract No. GR-MAT-0111 and MEC under Contract No. MAT-2004-03069, PR2005-053, and Contract No. MAT2002-11982E through the SONS program of the ESF, which is also supported by the EC, Sixth Framework Program.

References

1. For a review on Quantum properties of atomic-sized conductors see N. Agraït, A. Levy-Yeyati, and J.M. van Ruitenbeek: Phys. Rep. **377**, 81 (2003).
2. J. C. Maxwell: *A Treatise on Electricity and Magnetism*, Vol. 1 (Dover Publ., inc., New York 1954).
3. Yu. V. Sharvin: Sov. Phys.–JETP **21**, 655 (1965) (Zh. Eksp. Teor. Fiz. **48**, 984 (1965)).
4. J. A. Torres, J. I. Pascual, and J. J. Saenz: Phys. Rev. B **49**, 16581 (1994).
5. G. Rubio, N. Agraït, and S. Vieira: Phys. Rev. Lett. **76**, 2302 (1996).
6. G. Rubio-Bollinger et al.: Phys. Rev. Lett. **87**, 026101 (2001).
7. C. J. Muller, J. M. van Ruitenbeek, and L. J. de Jongh: Physica C **191**, 485 (1992).
8. G. Rubio-Bollinger, P. Joyez, and N. Agraït: Phys. Rev. Lett. **93**, 116803 (2004).
9. A. M. C. Valkering et al.: Rev. Sci. Instrum. **76**, 103903 (2005).
10. K. Karrai, and R. D. Grober: Appl. Phys. Lett. **66**, 1842 (1995).
11. J. Rychen et al.: Appl. Surf. Sci. **157**, 290 (2000).
12. For a review on this AFM technique see F. J. Giessibl: Rev. Mod. Phys. **75**, 949 (2003).
13. J. K. Gimzewski, and R. Möller: Phys. Rew. B **36**, 1284 (1987).
14. J. Frenkel: Z. Physik **37**, 572 (1926).
15. U. Landman et al.: Science **248**, 454 (1990).
16. C. Untiedt et al.: Phys. Rev. B **56**, 2154 (1997).
17. A. I. Yanson et al.: Nature **395**, 783 (1998).
18. H. Onishi, Y. Kondo, and K. Takayanagi: Nature **395**, 780 (1998).
19. C. Untiedt et al.: Phys. Rev. B **66**, 085418 (2002).
20. A. Stalder, and U. Dürig: Appl. Phys. Lett. **68**, 637 (1996).
21. O. Pfeiffer et al.: Nanotechnology **15**, 101 (2004).
22. J. H. Rose et al.: Phys. Rev. B **28**, 1835 (1983).
23. C. J. Chen: J. Phys. Condens. Matter **3**, 1227 (1991).

14 Contact Mechanics, Friction and Adhesion with Application to Quasicrystals

B.N.J. Persson, G. Carbone, V.N. Samoilov, I.M. Sivebaek,
U. Tartaglino, A.I. Volokitin, and C. Yang

IFF, FZ-Jülich, 52425 Jülich, Germany

14.1 Introduction

The first sample of a quasicrystal was produced in 1982 [1]. Intensive studies of this class of metallic materials have been conducted since that time. Quasicrystals display a unique combination of physical properties, namely low heat conductivity, relatively high hardness, and (under atmospheric condition) low friction coefficient and low surface energy. These properties make them become promising candidates as coatings for, e. g., cookware (see Fig. 14.1), surgical tools and electrical shavers, automotive parts, and for air-space applications.

In this article we present results related to sliding friction, contact mechanics and adhesion. Most of the theory results are very general, and can be applied not only to quasicrystals but also to other materials. In Sect. 14.2 we study how sliding friction depends on the elastic modulus of the solids. In Sect. 14.3 we discuss sliding friction and adhesion for quasicrystals. Sect. 14.4 presents a general discussion about surface roughness, and in Sect. 14.5 and 14.6 we consider contact mechanics and adhesion. Sect. 14.7 contains the summary and an outlook.

Fig. 14.1. A stainless steel pan coated by a quasicrystal material. The coating was made using electron beam vapor deposition in vacuum

14.2 Sliding Friction – Role of Elasticity

Sliding friction for clean solid surfaces, or surfaces separated by a $\sim 1\,\mathrm{nm}$ (or less) thick contamination film (boundary lubrication), usually originates from elastic instabilities occurring at the interface [2]. Elastic instabilities occur if the elastic modulus of the solids is low enough or if the lateral corrugation of the interaction potential at the interface is high enough. This is best illustrated by a one dimensional model, see Fig. 14.2. Here a particle is connected to a spring, and the free end of the spring is pulled with some (small) velocity v_0. If the spring constant is small enough or the potential well U_0 high enough, the particle will perform stick-slip (non-uniform) motion, where during slip the particle moves with a velocity $v(t)$ which is much higher than (and unrelated to) the driving velocity v_0. This will result in a large friction force (spring force averaged over time). On the other hand, if the spring is very stiff or the barrier very small, no stick-slip occurs and the velocity of the particle will be of the order of v_0, and proportional to v_0. In this case the friction force vanishes, at least when $v_0 \to 0$. In reality, the particle may represent some small group of atoms (block atoms and/or contamination atoms) at the interface, and the spring may represent some effective elastic properties which determine the force necessary to displace the group of atoms relative to the center of mass of the solid walls.

It is important to note that the elastic stiffness of solids depends on the length scale over which they are studied. Thus a solid elastic bar of length L will elongate by a distance proportional to L when exposed to some (fixed) forces F and $-F$ at its two ends. However, since hard solids also tend to have small contact areas (with small average diameter L) when squeezed together, this reduces the chances that elastic instabilities will occur at the interface during sliding. Thus, it is clear that hard materials, such as quasicrystals, may exhibit very low friction, in particular since the surfaces will always be incommensurate, thus lowering the barrier U_0.

As illustrations of the discussion above, let us present molecular dynamics simulations for an elastic block sliding on a rigid substrate when the wall atoms are (nearly) incommensurate with the substrate atoms. In Fig. 14.3

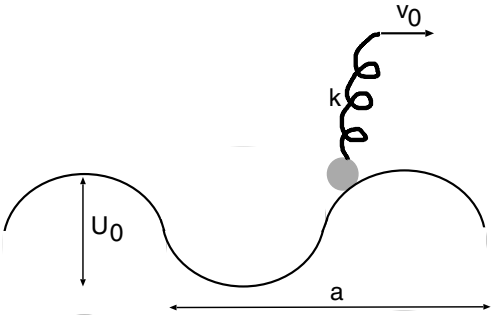


Fig. 14.2. A particle pulled by a spring (with the velocity v_0) in a periodical potential. If the spring k is weak enough or the barrier U_0 high enough ($ka^2 \ll U_0$), the particle will perform stick-slip motion. On the other hand, if $ka^2 \gg U_0$ no stick-slip occurs, and the friction force is very small (it will vanish as $v_0 \to 0$)

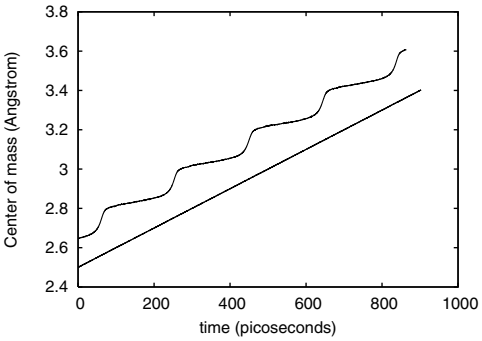


Fig. 14.3. Simulation results for an elastic block sliding on a rigid substrate. The atoms of the bottom surface of the block and of the top surface of the substrate form square lattices which are (nearly) incommensurate. The upper surface of the block is moving with the constant speed $v = 0.1\,\mathrm{m/s}$. *Straight line*: Young modulus $E = 10\,\mathrm{GPa}$, pressure $1\,\mathrm{GPa}$. *Curved line* (stick and slip): Young modulus $E = 1\,\mathrm{GPa}$, pressure $0.1\,\mathrm{GPa}$. For the softer elastic solid stick-slip occurs at the interface while steady motion occurs for the stiffer block

we show the center-of-mass coordinate of the bottom layer of block atoms as a function of time. Both the sliding layer and the substrate have square lattice structure, but with different lattice spacing to have (nearly) incommensurability (ratio 1.625 close to the golden mean). The upper surface of the block is moving with the constant speed $v = 0.1\,\mathrm{m/s}$. When the elastic stiffness of the block is small, stick-slip occurs (curved line), and the friction coefficient is nonzero. For a stiffer block (straight line), the stick and slip behaviour disappears and the friction coefficient gets negligibly small (below the noise level of the simulations).

Recently, a detailed study was performed of the friction between a Si tip and thin hard coatings [3]. As expected, it was observed that the friction coefficient decreases with increasing elastic modulus of the coating. An extreme case is the friction of diamond against diamond where the friction (when the diamond surfaces are passivated by hydrogen) is extremely small (of the order of 0.01).

14.3 Application to Quasicrystals

Quasicrystals differ radically from traditional crystalline materials because they have rotational symmetry which is incompatible with periodicity (translational symmetry). Due to the lack of translational symmetry, the plastic deformation properties of quasicrystals fundamentally differ from those of crystals. The plastic yield stress of most metal crystals is relatively low due to small barriers for motion of dislocations. This is not the case in quasicrystals

because of the absence of long-range translational symmetry. Consequently, the plastic yield stress is much higher for quasicrystals than that for most metallic crystals. Thus, in spite of the fact that quasicrystals only contain metal atoms, they form relatively hard and brittle-like materials. We believe that this is the main reason for the low sliding friction [4] and wear usually observed for quasicrystal materials.

In one set of experiments [5], the adhesion and sliding friction were studied as a sharp tip coated with W_2C in contact with a single grain 10-fold decagonal $Al_{72.4}Ni_{10.4}Co_{17.2}$ quasicrystals. The coated tip had the radius of curvature $\sim$100 nm. For the clean surfaces in ultrahigh vacuum the work of adhesion was found to be $\approx 0.1\,\mathrm{eV/\AA^2}$, but this value is probably an overestimate of the change in the surface energy $\Delta\gamma = \gamma_1 + \gamma_2 - \gamma_{12}$ since some plastic deformation of the tip-sample contact takes place during rupture of the contact. If the quasicrystal surface is exposed to clean O_2 gas, a very thin oxide layer (one or at most two monolayers) is formed on the surface, and the work of adhesion drops to about $\approx 0.03\,\mathrm{eV/\AA^2}$. When the surface is air-oxidized the work of adhesion is only $\approx 3\,\mathrm{meV/\AA^2}$. Similarly, the friction coefficient drops from ≈ 0.4 for the clean surface to ≈ 0.2 for the surface exposed to O_2 and to ≈ 0.1 for the air-oxidized surface.

It has been reported that the oxide formed in air on the quasicrystal surface has a thickness of the order of 26 Å in dry air and 62 Å in humid air. This is much thicker than the in situ grown oxide ($\approx$ 6 Å). Thus, the higher friction and work of adhesion on the very thin oxide formed in vacuum could be explained by the more fragile nature of the film that can be partly destroyed by the tip resulting in (weak) cold-welded regions [5]. In addition, the air exposed surface is likely to have a nanometer thick contamination layer consisting of organic molecules, water and other contamination molecules. This layer will also reduce the sliding friction although it may be at least partly removed after repeated sliding over the same surface area.

In another experiment two *macroscopic* $Al_{70}Pd_{21}Mn_9$ quasicrystals were brought into contact [6]. The crystal surfaces were polished to a mirror with 0.25 μm diamond pasta. The surface roughness amplitude was not measured but should be of the order of several 10 nm. In this case, even after lateral sliding, no adhesive force could be detected during pull-off. This may seem as a paradox taking into account the relatively large pull-off force measured in Ref. [5] when a tip was removed from a quasicrystal. However, the result is easy to understand based on the theoretical results presented in Sect. 14.6. Thus, when two macroscopic solid blocks of hard materials with randomly rough surfaces are brought into contact, the actual contact will only occur in very small, randomly distributed, asperity contact areas. For hard materials with low ductility, such as quasicrystals, a root-mean-square roughness of a few 10 nm (as in the present case) is enough to completely remove the (macroscopic) adhesion between the solids for the following reason. Since the asperities have different sizes they will have different amount of elastic deformation, and will act like elastic springs of different sizes. Thus during pull-off

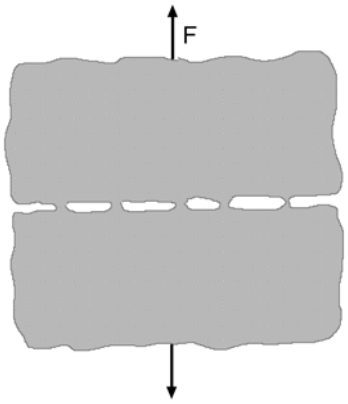


Fig. 14.4. When two ductile metals, e. g., Au or Al, are separated after being in contact, metallic bridges will occur in many asperity contact areas giving rise to a nonzero pull-off force. For (plastically) harder and more brittle metal because of elastic deformation of the asperities, the asperity contact regions will break one after another during pull-off and no adhesion (or pull-off force) will be observed

the different asperity contact areas will break at different times giving rise to a negligible adhesion even though breaking a single asperity contact region requires a non-negligible force as observed in the tip-substrate experiments reported on in Ref. [5]. We point out that a similar effect is observed in silicon wafer bonding (see Sect. 14.4).

For clean surfaces of more ductile metals such as Cu, Au or Al, strong adhesion is usually observed. This is the case even for oxide coated surfaces if sliding occurs before pull-off, as the sliding will break up the oxide coating and result in the formation of cold welded contact areas. During pull-off, because of the high ductility of Cu, Au or Al (and most other metals), "long" metallic bridges may be formed between the solids so that instead of having junctions popping one after another during pull-off, a large number of adhesive junctions may simultaneously impede the surface separation during pull-off (see Fig. 14.4), leading to a large pull-off force.

In Ref. [7] sliding friction measurement was performed both for clean surfaces (in ultra high vacuum) and for O_2 exposed surfaces and for surfaces oxidized in the air. For clean surfaces the friction coefficient was of order ≈ 0.6 which dropped to ≈ 0.4 when exposed to O_2. The friction coefficient of air exposed surfaces was only ≈ 0.1.

14.4 Surface Roughness

Surface roughness has a huge influence on many important physical phenomena such as contact mechanics, sealing, adhesion and friction. Thus, for example, experiments have shown that already a substrate roughness with a root-mean-square (*rms*) roughness of order $\sim 1\,\mu$m can completely remove the adhesion between a rubber ball and a substrate, while nanoscale roughness will remove the adhesion between most hard solids, e. g., metals and minerals; this is the reason why adhesion is usually not observed in most

Fig. 14.5. a Micrometer sized cantilever beam. **b** If the beam is too long or too thin the minimum free energy state corresponds to the beam partly bound to the substrate. Surface roughness lowers the binding energy (per unit area) and hence stabilizes the non-bonded state in **a**

macroscopic phenomena. Similarly, rubber friction on most surfaces of practical interest, e. g., road surfaces, is mainly due to the pulsating forces which act on the rubber surface as it slides over the substrate asperities.

Let us illustrate the importance of surface roughness with three modern applications. At present there is a strong effort to produce small mechanical devises, e. g., micromotors. The largest problem in the development of such devices is the adhesion and, during sliding, the friction and wear between the contacting surfaces [8]. As an example, in Fig. 14.5 we show the simplest possible micro device, namely a micrometer cantilever beam. (Suspended micromachined structures such as plates and beams are commonly used in manufacturing of pressure and accelerator sensors.) If the beam is too long or too thin the free beam state in (a) will be unstable, and the bound state in (b) will correspond to the minimum free energy state [9]. Roughly speaking, the state (b) is stable if the binding energy to the substrate is higher than the elastic energy stored in the bent beam. The binding energy to the substrate can be strongly reduced by introducing (or increasing) the surface roughness on the substrate (see Sect. 14.6.1). In addition, if the surfaces are covered by appropriate monolayer films the surfaces can be made hydrophobic thus eliminating the possibility of formation of (water) capillary bridges.

A second application is the formation of hydrophobic coatings on surfaces by creating the appropriate type of surface roughness [10]. This involves copying Nature where many plant surfaces are found to be highly hydrophobic (Fig. 14.6) as a result of the formation of special types of surface roughness (Fig. 14.7). The surface roughness allows air to be trapped between the liquid and the substrate, while the liquid is suspended on the tips of the asperities.

Fig. 14.6. A water droplet on a superhydrophobic surface: The droplet touches the leaf only in a few points and forms a ball. It completely rolls off at the slightest declination. From [10]

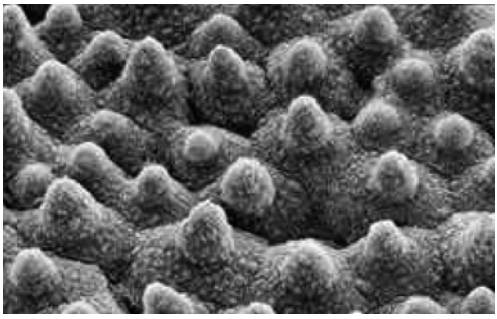


Fig. 14.7. A leaf surface with roughness on several length scales optimized (via natural selection) for hydrophobicity and self-cleaning. Through the combination of micro- (cells) and nanostructure (wax crystals) the water contact angle θ_0 is maximized. From [10]

Since the area of real liquid-substrate contact is highly reduced, the contact angle of the drop is determined almost solely by the surface tension of the liquid, leading to a very large contact angle. New commercial products based on this "Lotus effect", such as self-cleaning paints and glass windows, have been produced.

Finally, we discuss the effect of surface roughness on direct wafer bonding [11]. Wafer bonding at room temperature is due to relatively weak interatomic attraction forces, e. g., the van der Waals interaction or hydrogen bonding, giving (for perfectly flat surfaces) an interfacial binding energy of order $6\,\mathrm{meV/\AA^2}$. The wafer surface roughness is the most critical parameter determining the strength of the wafer bonding. In particular, when the surface roughness exceeds a critical value, the wafers will not bind at all, in agreement with the theory presented in Sect. 14.6.1. Primary grade polished silicon wafer surfaces have *rms* roughness of order $\sim 0.1\,\mathrm{nm}$ when measured over a $10^{10}\,\mu\mathrm{m}$ surface area, and such surfaces bind spontaneously. However, when the surface roughness amplitude is of order 1 nm the surfaces either bind (slowly) when squeezed together at high enough pressure, or they do not bind at all depending on the detailed nature of the surface roughness power spectra.

Surfaces with "ideal" roughness, e. g., prepared by fracture or by some growth process, have been studied intensively for many years [12–15]. However, much less information has been presented for more common surfaces of engineering interest. In what follows we discuss the nature of the power spectra of some surfaces of practical importance. As illustrations we discuss contact mechanics and adhesion.

14.4.1 Surface Roughness Power Spectra: Definition and General Properties

The influence of roughness on the adhesion and friction properties described above is mainly determined by the surface roughness power spectra $C(q)$ defined by [16]

$$C(q) = \frac{1}{(2\pi)^2} \int \mathrm{d}^2x \langle h(\boldsymbol{x}) h(\boldsymbol{0}) \rangle \, \mathrm{e}^{-i\boldsymbol{q}\cdot\boldsymbol{x}} \tag{14.1}$$

Here $h(\boldsymbol{x})$ is the substrate height measured from the average plane defined so that $\langle h\rangle = 0$. The $\langle \ldots \rangle$ stands for ensemble averaging, or averaging over the surface area (see below). We have assumed that the statistical properties of the substrate are translational invariant and isotropic so that $C(q)$ only depend on the magnitude $q = |\boldsymbol{q}|$ of the wave vector $\boldsymbol{q}$. Note that from (14.1) follows

$$\langle h(\boldsymbol{x})h(\boldsymbol{0})\rangle = \int \mathrm{d}^2q\; C(q)\,\mathrm{e}^{i\boldsymbol{q}\cdot\boldsymbol{x}}$$

so that the root-mean-square roughness amplitude $\sigma = \langle h^2\rangle^{1/2}$ is determined by

$$\langle h^2\rangle = \int \mathrm{d}^2qC(q) = 2\pi\int_0^\infty \mathrm{d}q\; qC(q) \tag{14.2}$$

In reality, there will always be an upper, q_1, and a lower, q_0, limit to the q-integral in (14.2). Thus, the largest possible wave vector will be of order $2\pi/a$, where a is some lattice constant, and the smallest possible wave vector is of order $2\pi/L$ where L is the linear size of the surface. In general, one may define a root-mean-square roughness amplitude which depends on the range of roughness included in the integration in (14.2):

$$\langle h^2\rangle(q_0, q_1) = 2\pi\int_{q_0}^{q_1} \mathrm{d}q\; qC(q) \tag{14.3}$$

For a randomly rough surface, when $h(\boldsymbol{x})$ are Gaussian random variables, the statistical properties of the surface are completely defined by the power spectra $C(q)$. In this case the height probability distribution

$$P_\mathrm{h} = \langle\delta[h - h(\boldsymbol{x})]\rangle$$

will be a Gaussian

$$P_\mathrm{h} = \frac{1}{(2\pi)^{1/2}\sigma}\,\mathrm{e}^{-h^2/2\sigma^2}$$

The height distribution of many natural surfaces, e. g., surfaces prepared by fracture, or surfaces prepared by blasting with small particles (e. g., sand blasting or ion sputtering) are usually nearly Gaussian. On the other hand, rough surfaces, e. g., a surface prepared by fracture, which have been (slightly) polished have a non-symmetric height distribution (i. e., no symmetry as $h \to -h$) since the asperity tops have been more polished than the bottom of the valleys, and such surfaces (which are of great practical importance – see below) have non-Gaussian height distribution. For such surfaces it is interesting to study the *top*, C_T, and the *bottom*, C_B, power spectra's defined by

$$C_\mathrm{T}(q) = \frac{1}{(2\pi)^2}\int \mathrm{d}^2x\langle h_\mathrm{T}(\boldsymbol{x})h_\mathrm{T}(\boldsymbol{0})\rangle\,\mathrm{e}^{-i\boldsymbol{q}\cdot\boldsymbol{x}} \tag{14.4a}$$

$$C_\mathrm{B}(q) = \frac{1}{(2\pi)^2}\int \mathrm{d}^2x\langle h_\mathrm{B}(\boldsymbol{x})h_\mathrm{B}(\boldsymbol{0})\rangle\,\mathrm{e}^{-i\boldsymbol{q}\cdot\boldsymbol{x}} \tag{14.4b}$$

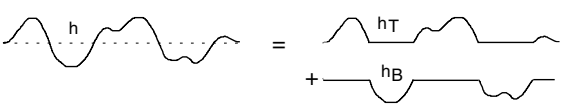


Fig. 14.8. The surface profile $h(x)$ is decomposed into a top $h_{\rm T}(x)$ and a bottom $h_{\rm B}(x)$ profile

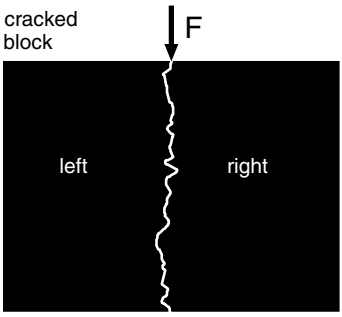


Fig. 14.9. Rough surfaces prepared by crack propagation have surface roughness with statistical properties which must be invariant under the replacement of $h \to -h$. This follows from the fact that what is a valley on one of the crack surfaces (*say the left*) is an asperity with respect to the other crack surface (*right*). Thus the top and bottom power spectra must obey $C_{\rm T}(q) = C_{\rm B}(q)$

where $h_{\rm T}(\boldsymbol{x}) = h(\boldsymbol{x})$ for $h > 0$ and zero otherwise, while $h_{\rm B}(\boldsymbol{x}) = h(\boldsymbol{x})$ for $h < 0$ and zero otherwise, see Fig. 14.8. It is easy to show that $C \approx C_{\rm T} + C_{\rm B}$. It is also clear by symmetry that for a surface prepared by fracture, $C_{\rm T}(q) = C_{\rm B}(q)$, since what is top on one of the cracked block surfaces is the bottom on the other (opposite) crack surface, and vice versa, see Fig. 14.9. However, if the cracked surface is (slightly) polished then, since the polishing will be stronger at the top of the asperities than at the bottom of the valleys [the contact pressure with the polishing object (e. g., sand paper) is highest at the asperity top], $C_{\rm B} > C_{\rm T}$. If $n_{\rm T}$ and $n_{\rm B}$ are the fraction of the nominal surface area (i. e., the surface area projected on the xy-plane) where $h > 0$ and $h < 0$, respectively, with $n_{\rm T} + n_{\rm B} = 1$, then we also define $C^*_{\rm T}(q) = C_{\rm T}/n_{\rm T}$ and $C^*_{\rm B} = C_{\rm B}/n_{\rm B}$. In general, $n_{\rm T} \approx n_{\rm B} \approx 0.5$ and for surfaces prepared by fracture $n_{\rm T} = n_{\rm B} = 0.5$. Roughly speaking, $C^*_{\rm T}$ would be the power spectra which would result if the surface profile in the large valleys (for $h < 0$) is replaced by a surface profile with similar short-wavelength roughness as occurs on the large asperities (for $h > 0$). A similar statement holds for $C^*_{\rm B}$.

Many surfaces tend to be nearly self-affine fractal. A self-affine fractal surface has the property that if part of the surface is magnified, with a magnification which in general is appropriately different in the perpendicular direction to the surface as compared to the lateral directions, then the surface "looks the same", i. e., the statistical properties of the surface are invariant under the scale transformation. For a self-affine surface the power spectrum has the power-law behavior

$$C(q) \sim q^{-2(H+1)} \, ,$$

where the Hurst exponent H is related to the fractal dimension $D_{\rm f}$ of the surface via $H = 3 - D_{\rm f}$. Of course, for real surfaces this relation only holds in some finite wave vector region $q_0 < q < q_1$, and in a typical case $C(q)$ has the form shown in Fig. 14.10. Note that in many cases there is a roll-off

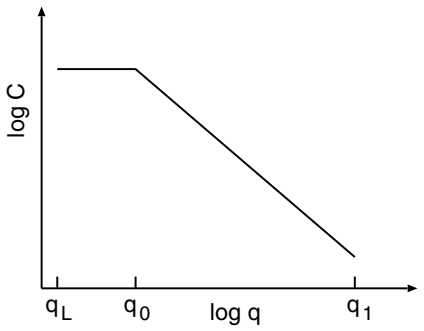


Fig. 14.10. Surface roughness power spectra of a surface which is self affine fractal for $q_1 > q > q_0$. The long-distance roll-off wave vector q_0 and the short distance cut-off wave vector q_1 depend on the system under consideration. The slope of the $\log C - \log q$ relation for $q > q_0$ determines the fractal exponent of the surface. The lateral size L of the surface (or of the studied surface region) determines the smallest possible wave vector $q_L = 2\pi/L$

wavelength q_0 below which $C(q)$ is approximately constant. We will discuss this point further below.

Finally, note that while the root-mean-square roughness usually is dominated by the longest wavelength surface roughness components, higher order moments of the power spectra such as the average slope or the average surface curvature are dominated by the shortest wavelength components. For example, assuming a self affine fractal surface, Eq. (14.3) gives

$$\langle h^2 \rangle(q_0, q_1) \sim \int_{q_0}^{q_1} \mathrm{d}q\, q^{-2H-1} \sim q_0^{-2H} - q_1^{-2H} \approx q_0^{-2H}$$

if $q_1/q_0 \gg 1$. However, the average slope and the average curvature have additional factors of q^2 and q^4, respectively, in the integrand of the q-integral, and these quantities are therefore dominated by the large q (i. e., short wavelength) surface roughness components.

14.4.2 Surface Roughness Power Spectra: Experimental Results

In this section we present power spectra for different surfaces of practical importance. The power spectra have been calculated using Eqs. (14.1) and (14.4), where the height profile $h(\boldsymbol{x})$ has been measured using either an optical method or the Atomic Force Microscope.

Surfaces Produced by Crack Propagation

Figure 14.11 shows the power spectra $C(q)$ for three freshly cleaved stone surfaces, namely a granite and two basalt stone surfaces. Here, and in what follows, we show the power spectra on a log-log scale. Note that the granite and basalt surfaces, in spite of the rather different mineral microstructure (see below), give identical power spectra within the accuracy of the measurement. It has been stated (see, e. g., Ref. [17]) that surfaces produced by crack propagation have self affine fractal structure with the universal fractal

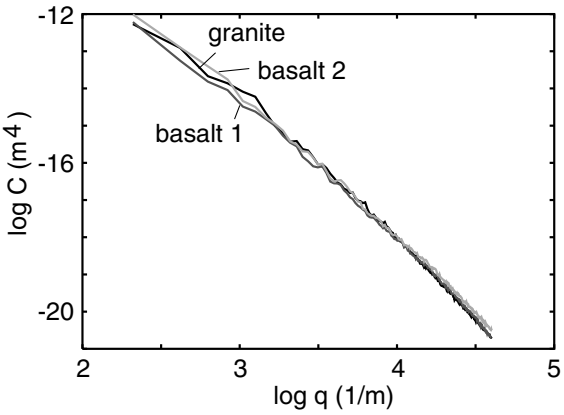


Fig. 14.11. The surface roughness power spectra for two freshly cleaved basalt surfaces and a fresh granite surface

dimension $D_\mathrm{f} \approx 2.2$. However, our measured $\log C - \log q$ relations are not perfectly straight lines, i. e., the surfaces in the studied length-scale range cannot be accurately described as self affine fractal, and the average slope of the curves in Fig. 14.11 correspond to the fractal dimension $D_\mathrm{f} \approx 2$ rather than 2.2.

Note the similarity of the power spectra for the basalt and granite surfaces in Fig. 14.11. Granite and basalt both result from magma and have a similar composition, consisting mainly of minerals from the silicate group. However, granite results from magma which is trapped deep in the crust, and takes very long time to cool down enough to crystallize into solid rock. As a result granite is coarse-textured rock in which individual mineral grains are easily visible. Basalt, on the other hand, results from fast cooling of magma from, e. g., volcanic eruptions, and is therefore fine grained, and it is nearly impossible to see the individual minerals without magnification. In spite of these differences, the surface roughness power spectra of freshly cleaved surfaces are nearly identical. This may indicate some kind of universal power spectrum for surfaces resulting from cleaving of mineral stones of different types.

Note that there is no roll-off region for surfaces produced by fracture (crack propagation), and the surfaces remains fractal-like up to the longest length scale studied, determined by the lateral size L of the surfaces (or of the regions experimentally studied), i. e., with reference to Fig. 14.10, $q_0 = q_L$. One consequence of this is that the *rms*-roughness amplitude is determined mainly by the $\lambda \sim L$ wavelength fluctuations of the surface height, and will therefore depend on the size L of the surface, and the height distribution P_h obtained for any given realization of the rough surface will not be Gaussian, but will exhibit random fluctuations as compared to other realizations (see Fig. 14.12, which illustrates this point for the three stone surfaces discussed above). However, the ensemble averaged height distribution (not shown) should be Gaussian or nearly Gaussian. Thus, when there is no roll-off region in the measured power spectra, averaging over the surface area is

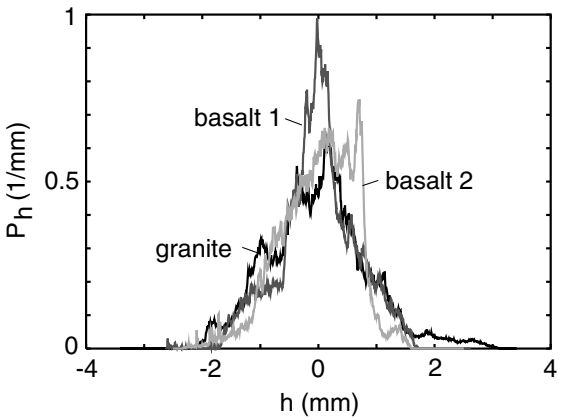


Fig. 14.12. The height distribution $P_{\rm h}$ for two freshly cleaved (cobble stone) basalt surfaces and a fresh granite surface. Note the random non-Gaussian nature of the height profiles

not identical to ensemble averaging. However, when there is a roll-off wave vector $q_0 = 2\pi/\lambda_0$, and if the surface is studied over a region with the lateral size $L \gg \lambda_0$, ensemble averaging and averaging over the surface area $L \times L$ will give identical results for $P_{\rm h}$, and the *rms*-roughness amplitude will be independent of L for $L \gg \lambda_0$.

Polished Crack Surfaces

In the past, cobble stones, made from granite or basalt, were frequently used for road surface pavements. However, these surfaces do not exhibit good frictional properties against rubber. In particular, with increasing time the cobble stone surfaces become polished by the road–tire interaction, which results in a reduced rubber-road friction, even during dry driving conditions. Figure 14.13 illustrates this polishing effect. It shows the power spectrum of a strongly used (basalt) cobble stone, and of two freshly cleaved surfaces

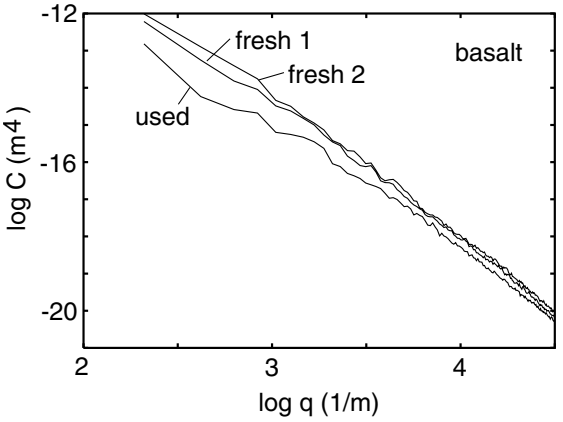


Fig. 14.13. The surface roughness power spectra $C(q)$ for two freshly cleaved cobble stone (basalt) surfaces, and for a used surface

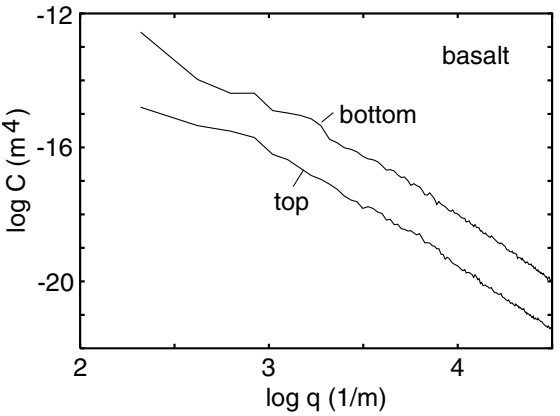


Fig. 14.14. The top C_{T}^{*} and the bottom C_{B}^{*} surface roughness power spectra $C(q)$ for a used cobble stone (basalt) surface

(from Fig. 14.11), from the same cobble stone. At long wavelength the power spectrum of the strongly used surface is nearly one decade smaller than that of the freshly prepared surfaces. The effect of the polishing is even better illustrated by calculating the top and bottom power spectra, C_{T}^{*} and C_{B}^{*}, as shown in Fig. 14.14. The top power spectrum is a factor ~ 30 times smaller than the bottom power spectrum for all wave vectors studied. This arises from the higher polishing of the road asperities than of the valleys (the tire-road contact pressure is highest at the road asperities, resulting in the strongest polishing of the asperity tops during breaking on the road). It is important to take this polishing effect into consideration when designing road pavements.

Surfaces with Long-Distance Roll-Off

As pointed out above, surfaces prepared by fracture have no natural long-distance cut-off and the *rms* roughness amplitude increases continuously (without limit) as the probed surface area increases. This is similar to Brownian motion where the mean square displacement increases without limit (as $\sim t^{1/2}$) as the time t increases. However, most surfaces of engineering interest have a long distance cut-off or roll-off wavelength λ_0 corresponding to a wave vector $q_0 = 2\pi/\lambda_0$, as shown in Fig. 14.10. For example, if a flat surface is sand blasted for some time the resulting rough surface will have a long distance roll-off length, which increases with the time of sand blasting. Similarly, if atoms or particles are deposited on an initially flat surface the resulting rough surface will have a roll-off wavelength which increases with the deposition time, as has been studied in detail in recent growth models. Another way to produce a surface with a long-distance roll-off wavelength is to prepare the solid from small particles. A nominally flat surface of such a solid has still roughness on length scales shorter than the diameter of the particles, which therefore may act as a long distance roll-off wavelength. We illustrate this here with a solid produced by squeezing together corundum

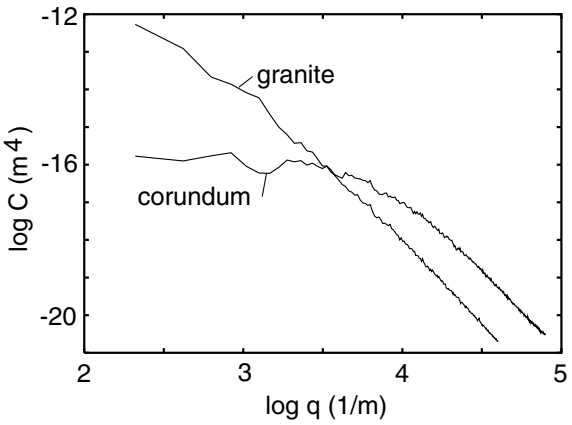


Fig. 14.15. The surface roughness power spectra for a fresh granite surface and a fresh particle-made corundum surface

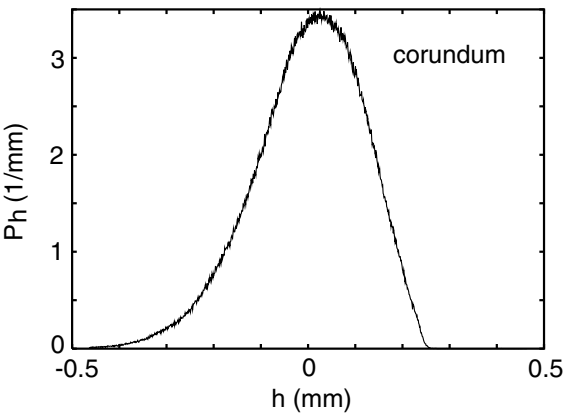


Fig. 14.16. The height distribution $P_{\rm h}$ as a function of the height h for a particle-made corundum surface

particles at high temperature and pressure (Fig. 14.15), and for a sand paper surface (Fig. 14.17). For both surfaces the height distribution $P_{\rm h}$ is smooth and nearly Gaussian (see Figs. 14.16 and 14.18), since averaging over a surface area with lateral size $L \gg \lambda_0$ is equivalent to ensemble averaging.

The sand paper surface in Fig. 14.17 was studied using the AFM at two different resolutions over square areas $20 \times 20\,\mu$m and $100 \times 100\,\mu$m as indicated by the two different lines in Fig. 14.17. The height distribution $P_{\rm h}$ (and hence also the *rms*-roughness amplitude) calculated from these two different measurements over different surface areas, see Fig. 14.18, are nearly identical, as indeed expected when L is larger than the roll-off length λ_0.

14.5 Contact Mechanics

Practically all macroscopic bodies have surfaces with roughness on many different length scales. When two bodies with nominally flat surfaces are

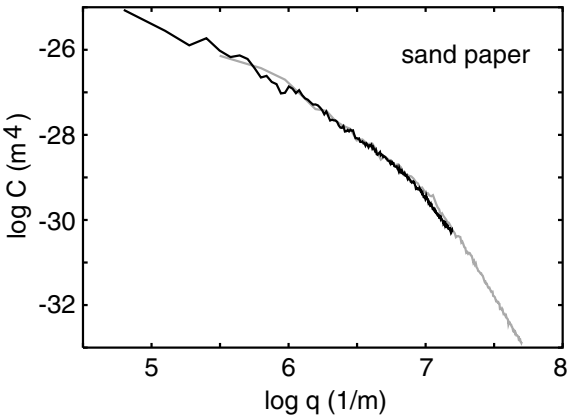


Fig. 14.17. The surface roughness power spectra $C(q)$ for a sand paper surface. The two curves are based on the height profiles measured with an AFM at two different spatial resolution over 20×20 and $100 \times 100\,\mu\text{m}$ square areas

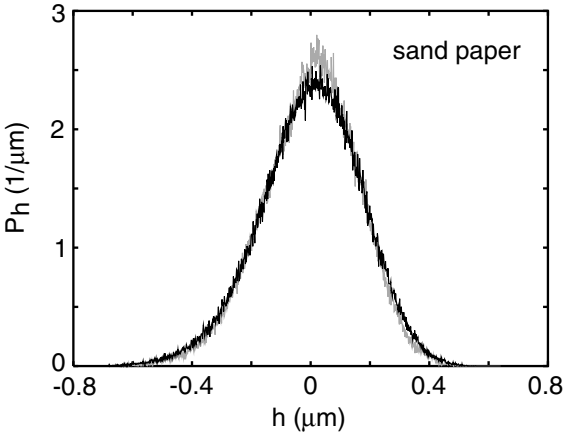


Fig. 14.18. The surface roughness height probability distribution P_h for a sand paper surface. The two curves are based on the height profiles measured with an AFM at two different spatial resolution over 20×20 and $100 \times 100\,\mu\text{m}$ square areas

brought in contact, real (atomic) contact will only occur in small randomly distributed areas, and the area of real contact is usually an extremely small fraction of the nominal contact area. We can visualize the contact regions as small areas where asperities from one solid are squeezed against asperities of the other solid; depending on the conditions the asperities may deform elastically or plastically.

How large is the area of *real* contact between a solid block and the substrate? This fundamental question has extremely important practical implications. For example, it determines the contact resistivity and the heat transfer between the solids. It is also of direct importance for wear and sliding friction [18], e. g., the rubber friction between a tire and a road surface, and it has a major influence on the adhesive force between two solid blocks in direct contact.

Contact mechanics has a long history. The first study was presented by Hertz [19]. He gave the solution for the frictionless normal contact of two

elastic bodies of quadratic profile. He found that the area of real contact ΔA varies nonlinearly with the load or squeezing force: $\Delta A \propto F_{\rm N}^{2/3}$. In 1957 Archard [20] applied the Hertz solution to the contact between rough surfaces and showed that for a simple fractal-like model, where small spherical bumps (or asperities) were distributed on top of larger spherical bumps and so on, the area of real contact varies *nearly linearly* with $F_{\rm N}$. A similar conclusion was reached by Greenwood and Williamson [21–23] who again assumed asperities with spherical summit (of identical radius) with a Gaussian distribution of heights, as sketched in Fig. 14.19b. A more general contact mechanics theory has been developed by Bush et al. [24] They approximated the summit by paraboloids and applied the classical Hertzian solution for their deformation. The height distribution was described by a random process, and they found that at low squeezing force $F_{\rm N}$ the area of real contact increases linearly with $F_{\rm N}$.

Figure 14.20 shows the contact between two solids at increasing magnification ζ. At low magnification ($\zeta = 1$) it looks as if complete contact occurs between the solids at many *macro asperity* contact regions, but when the magnification is increased smaller length scale roughness is detected, and it is observed that only partial contact occurs at the asperities. In fact, if there would be no short distance cut-off the true contact area would vanish. In

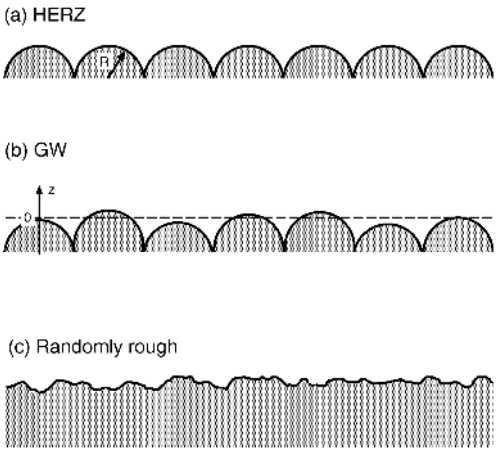


Fig. 14.19. Three models of "rough" surfaces. In case **a** all the "asperities" are equally high and have identical radius of curvature. In this case, according to the Hertz contact theory, the area of real contact ΔA between a solid with a flat surface and the shown surface depends non-linearly on the squeezing force (or load) $F_{\rm N}$ according to $\Delta A \sim F_{\rm N}^{2/3}$. If the asperities have a random distribution of heights as in **b** then, for small $F_{\rm N}$, ΔA is *nearly* proportional to the squeezing force. If the surface roughness is random with "asperities" of different heights and radius of curvature as in **c**, the area of real contact for small $F_{\rm N}$ is *exactly* proportional to the squeezing force

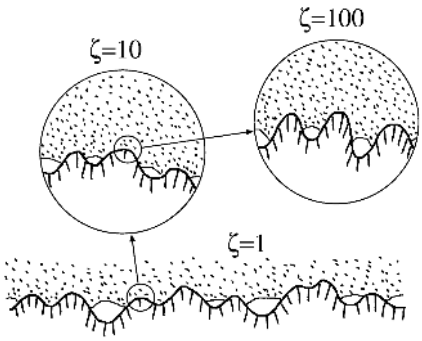


Fig. 14.20. A rubber block (*dotted area*) in adhesive contact with a hard rough substrate (*dashed area*). The substrate has roughness on many different length scales and the rubber makes partial contact with the substrate on all length scales. When a contact area is studied at low magnification ($\zeta = 1$) it appears as if complete contact occurs in the macro asperity contact regions, but when the magnification is increased it is observed that in reality only partial contact occurs

reality, however, a short distance cut-off will always exist since the shortest possible length is an atomic distance. In many cases the local pressure at asperity contact regions at high magnification will become so high that the material yields plastically before reaching the atomic dimension. In these cases the size of the real contact area will be determined mainly by the yield stress of the solid.

From contact mechanics (see, e. g., Ref. [23]) it is known that in the frictionless contact of elastic solids with rough surfaces, the contact stresses depend only upon the shape of the gap between them before loading. Thus, without loss of generality, the actual system may then be replaced by a flat elastic surface [elastic modulus E and Poisson ratio ν, related to the original quantities via $(1-\nu^2)/E = (1-\nu_1^2)/E_1 + (1-\nu_2^2)/E_2$] in contact with a rigid body having a surface roughness profile which results in the same undeformed gap between the surfaces.

One of us (Persson)) has recently developed a theory of contact mechanics [25], valid for randomly rough (e. g., self affine fractal) surfaces. In the context of rubber friction, which motivated this theory, mainly elastic deformation occurs. However, the theory can also be applied when both elastic and plastic deformations occur in the contact areas. This case is, of course, relevant to almost all materials other than rubber.

The basic idea behind the new contact theory is that it is very important not to *a priori* exclude any roughness length scale from the analysis. Thus, if $A(\lambda)$ is the (apparent) area of contact on the length scale λ [26] (see Fig. 14.21), then we study the function $P(\zeta) = A(\lambda)/A(L)$ which is the relative fraction of the surface area where contact occurs on the length scale $\lambda = L/\zeta$ (where $\zeta \geq 1$), with $P(1) = 1$. Here $A(L) = A_0$ denotes the macroscopic contact area [L is the diameter of the macroscopic contact area so that $A_0 \approx L^2$].

Consider the system at the length scale $\lambda = L/\zeta$, where L is the diameter of the nominal contact area. We define $q_L = 2\pi/L$ and write $q = q_L\zeta$. Let $P(\sigma, \zeta)$ denote the stress distribution in the contact areas under the

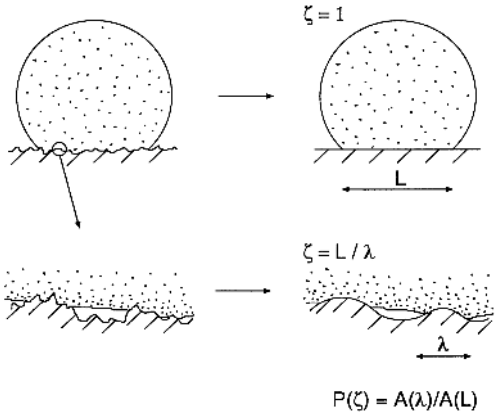


Fig. 14.21. An elastic ball squeezed against a hard, rough, substrate. *Left*: the system at two different magnifications. *Right*: The area of contact $A(\lambda)$ on the length scale λ is defined as the area of real contact when the surface roughness on shorter length scales than λ has been removed

magnification ζ. The function $P(\sigma, \zeta)$ satisfies the differential equation (see Ref. [25]):

$$\frac{\partial P}{\partial \zeta} = f(\zeta) \frac{\partial^2 P}{\partial \sigma^2} \,, \tag{14.5}$$

where $f(\zeta) = G'(\zeta)\sigma_0^2$ with

$$G(\zeta) = \frac{\pi}{4} \left(\frac{E^*}{\sigma_0} \right)^2 \int_{q_L}^{\zeta q_L} \mathrm{d}q \; q^3 C(q) \,, \tag{14.6}$$

where $E^* = E/(1-\nu^2)$.

Equation (14.5) is a diffusion type of equation, where time is replaced by the magnification ζ, and the spatial coordinate with the stress σ (and where the "diffusion constant" depends on ζ). Hence, when we study $P(\sigma, \zeta)$ on shorter and shorter length scales (corresponding to increasing ζ), the $P(\sigma, \zeta)$ function will become broader and broader in σ-space. We can take into account that detachment actually will occur when the local stress reaches $\sigma = 0$ (we assume no adhesion) via the boundary condition [27]:

$$P(0, \zeta) = 0 \,. \tag{14.7}$$

In order to solve the equation (14.5) we also need an "initial" condition. This is determined by the pressure distribution at the lowest magnification $\zeta = 1$. If we assume a constant pressure σ_0 in the nominal contact area, then $P(\sigma, 1) = \delta(\sigma - \sigma_0)$.

We assume that only elastic deformation occurs (i. e., the yield stress $\sigma_Y \to \infty$). In this case

$$P(\zeta) = \int_0^\infty \mathrm{d}\sigma P(\sigma, \zeta)$$

When adhesion is taken into account, tensile stresses can occur at the interface between the two solids, and the boundary condition (14.7) is no longer

valid [28], see Sect. 14.6.1. It is straightforward to solve (14.5) with the boundary conditions $P(0,\zeta) = 0$ and $P(\infty,\zeta) = 0$ to get

$$P(\zeta) = \frac{2}{\pi}\int_0^\infty \mathrm{d}x \frac{\sin x}{x}\,\mathrm{e}^{-x^2 G(\zeta)} = \mathrm{erf}\left(\frac{1}{2\sqrt{G}}\right) . \tag{14.8}$$

Note that for small load σ_0, $G \gg 1$ and in this case (14.8) reduces to $P(\zeta) \approx P_1(\zeta)$ where

$$P_1(\zeta) = [\pi G(\zeta)]^{-1/2} . \tag{14.9}$$

Since $G \sim 1/\sigma_0^2$ it follows that the area of real contact is *proportional* to the load for small load. Using (14.8) and (14.9) we can write in a general case

$$P(\zeta) = \mathrm{erf}\left(\frac{\sqrt{\pi}}{2} P_1(\zeta)\right) . \tag{14.10}$$

The physical meaning of Eq. (14.5) is as follows: When the system is studied at the lowest magnification $\zeta = 1$ no surface roughness can be observed and the block makes (apparent) contact with the substrate everywhere in the nominal contact area. In this case, if we neglect friction at the interface, the stress at the interface will everywhere equal the applied stress σ_0, see Fig. 14.22a, so that $P(\sigma, 1) = \delta(\sigma - \sigma_0)$. When we increase the magnification we observe surface roughness with wavelength down to $\lambda = L/\zeta$. In this case one may observe some non-contact regions as shown in Fig. 14.22b. Since the

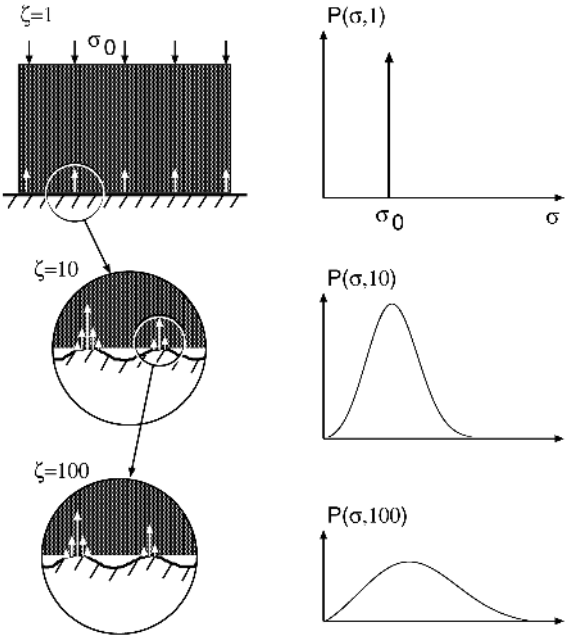


Fig. 14.22. The stress distribution $P(\sigma,\zeta)$ in the contact region between a (rigid) block and an elastic substrate at increasing magnification ζ. At the lowest (engineering) magnification $\zeta = 1$ the substrate surface looks smooth and the block makes (apparent) contact with the substrate in the whole nominal contact area. As the magnification increases, we observe that the area of (apparent) contact decreases, while the stress distribution becomes wider and wider

stress must go continuously to zero at the edges of the boundary between the contact and non-contact regions, it follows that the stress distribution $P(\sigma, \zeta)$ will have a tail extending the whole way down to the zero stress as indicated in Fig. 14.22b (right). There will also be a tail toward larger stresses $\sigma > \sigma_0$ because the average stress must be equal to σ_0. Thus with increasing magnification, the stress distribution will broaden without limit as indicated in Fig. 14.22 (right).

The theory presented above predicts that the area of contact increases linearly with the load for small load. In the standard theory of Greenwood and Williamson [22] this result holds only approximately and a comparison of the prediction of their theory with the present theory is therefore difficult. Bush et al. [24] have developed a more general and accurate contact theory. They assumed that the rough surface consists of a mean plane with hills and valleys randomly distributed on it. The summits of these hills are approximated by paraboloids, the distribution of heights and principal curvatures of which is obtained from the random process theory. This is to be compared with the GW assumption that the caps of the asperities are spherical each having the same mean radius of curvature. As a result of the more random nature of the surface, Bush et al. found that at small load the area of contact depends linearly on the load accordingly to

$$\frac{A}{A_0} = \kappa \frac{F_{\rm N}}{E^*} \left(\int {\rm d}^2 q \; q^2 C(q) \right)^{-1/2} \tag{14.11}$$

where $\kappa = (2\pi)^{1/2}$. This result is very similar to the prediction of the present theory where, for small load, from (14.6) and (14.9), A/A_0 is again given by Eq. (14.11) but now with $\kappa = (8/\pi)^{1/2}$. Thus our contact area is a factor of $2/\pi$ smaller than predicted by the theory of Bush et al. Both the theory of Greenwood and Williamson and of Bush et al., assume that the asperity contact regions are independent. However, as discussed in Ref. [27], for real surfaces (which always have surface roughness on many different length scales) this will never be the case even at a very low nominal contact pressure. We have argued [27] that this may be the origin of the $2/\pi$-difference between our theory (which assumes roughness on many different length scales) and the result of Bush et al.

The predictions of the theories of Bush et al. [24] and Persson [25] have been compared to numerical calculations (see Ref. [27,29,30]). Borri-Brunetto et al. [31] have studied the contact between self affine fractal surfaces using an essentially exact numerical method. They found that the contact area is proportional to the squeezing force for small squeezing forces. Furthermore, it was found that the slope $\alpha(\zeta)$ of the line $A = \alpha(\zeta)F$ decreased with increasing magnification ζ. This is also predicted by the analytical theory [Eq. (14.11)]. In fact, it was found a good agreement between the theory and the computer simulations for the change in the slope with magnification and its dependence on the fractal dimension $D_{\rm f}$.

Hyun et al. have performed a finite-element analysis of contact between elastic self-affine surfaces. The simulations are done for a rough elastic surface contacting a perfectly rigid flat surface. The elastic solid is discretized into blocks and the surface nodes form a square grid. The contact algorithm identifies all nodes on the top surface that attempt to penetrate the flat bottom surface. The total contact area A was obtained by multiplying the number of penetrating nodes by the area of each square associated with each node. As long as the squeezing force is so small that the contact area is below 10% of the nominal contact area, i. e., $A/A_0 < 0.1$, the area of real contact is found to be proportional to the squeezing force in accordance with Eq. (14.11). In Fig. 14.23 we present the results for the factor κ in (14.11) as a function of Hurst exponent H for self affine fractal surfaces. The two horizontal lines gives the predictions of the theories of Bush et al. (solid line) and Persson (dashed line). The agreement with the analytical predictions is quite good considering the ambiguities in discretization of the surface. The algorithm only considers nodal heights and assumes that contact of a node implies contact over the entire corresponding square. This procedure would be accurate if the spacing between nodes where much smaller than the typical size of asperity contacts. However, the majority of the contact area consists of clusters containing only one or a few nodes. Since the number of large clusters grows as $H \to 1$, this may explain why the numerical results approach Persson's prediction in this limit.

Hyun et al. also studied the distribution of connected contact regions and the contact morphology. In addition, the interfacial stress distribution was studied and it was found that the stress distribution remained non-zero as the stress $\sigma \to 0$. This violates the boundary condition (14.7) that $P(\sigma, \zeta) = 0$ for $\sigma = 0$. However, it has been shown analytically [27] that for "smooth" surface roughness this latter condition must be satisfied, and we believe that

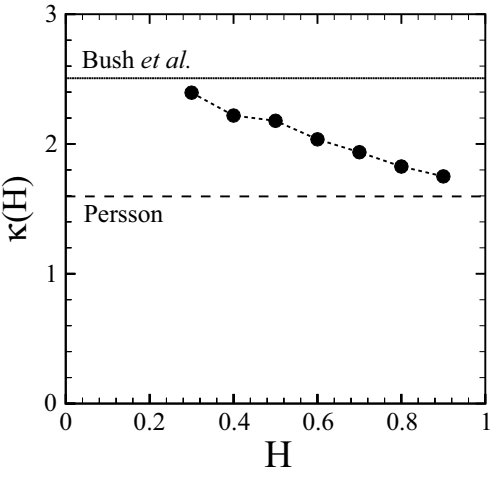


Fig. 14.23. The factor κ as a function of Hurst exponent H for self affine fractal surfaces. The two horizontal lines gives the predictions of the theories of Bush et al. (*solid line*) and Persson (*dashed line*). From Ref. [29]

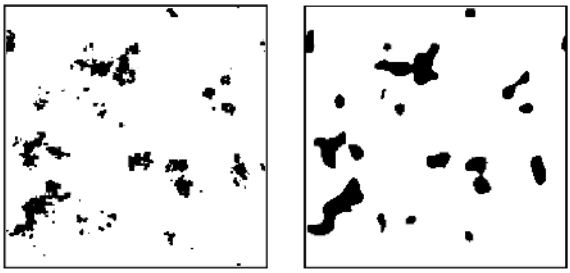


Fig. 14.24. The contact area between an elastic solid block and a randomly rough hard substrate at high (atomic) magnification (*left*), and at a lower magnification (*right*)

the violation of this boundary condition in the numerical simulations reflects the way the solid was discretized and the way the contact area is defined in the numerical procedure.

Yang et al. [30] have studied contact mechanics using Molecular Dynamics. They also found that the contact area varies linearly with the load for small load, and that the contact area at low magnification is larger than at high magnification (see Fig. 14.24), as predicted by the theory [Eq. (14.11)]. The detailed comparison of the simulation results with the theory will be presented elsewhere [30].

Elastic contact theory and numerical simulations show that in the region where the contact area is proportional to the squeezing force, the stress distribution at the interface is independent of the squeezing force. In addition, for an infinite system the distribution of sizes of the contact regions does not depend on the squeezing force (for small squeezing forces). Thus, when the squeezing force increases, new contact regions are formed in such a way that the distribution of contact regions and the pressure distribution remains unchanged. This is the physical origin of Coulombs friction law which states that the friction force is proportional to the normal (or squeezing) force [18], and which usually holds accurately as long as the block-substrate adhesional interaction can be neglected [2].

14.6 Adhesion

In this section we discuss adhesion between rough surfaces. We point out that even when the force to separate two solids vanishes, there may still be a finite contact area (at zero load) between two solids as a result of the adhesional interaction between the solids. We also study the adhesion between a thin elastic film and a randomly rough, rigid substrate.

14.6.1 Adhesion Between Rough Surfaces

A theory of adhesion between an elastic solid and a hard randomly rough substrate must take into account that partial contact may occur between

the solids on all length scales. For the case where the substrate surface is self affine fractal theory shows that when the fractal dimension is close to 2, complete contact typically occurs in the macro asperity contact areas (the contact regions observed when the system is studied at a magnification corresponding to the roll-off wavelength $\lambda_0 = 2\pi/q_0$ of the surface power spectra, see Fig. 14.10), while when the fractal dimension is larger than 2.5, the area of (apparent) contact decreases continuously when the magnification is increased. An important result is that even when the surface roughness is so high that no adhesion can be detected in a pull-off experiment, the area of real contact (when adhesion is included) may still be several times larger than when the adhesion is neglected. Since it is the area of real contact which determines the sliding friction force, *the adhesion interaction may strongly affect the friction force even when no adhesion can be detected in a pull-off experiment.*

The influence of surface roughness on the adhesion between rubber (or any other elastic solid) and a hard substrates has been studied in a classic paper by Fuller and Tabor [32] (see also [33–39]). They found that already a relative small surface roughness can completely remove the adhesion. In order to understand the experimental data they developed a very simple model based on the assumption of surface roughness on a single length scale. In this model the rough surface is modeled by asperities all of the same radius of curvature and with heights following a Gaussian distribution. The overall contact force was obtained by applying the contact theory of Johnson, Kendall and Roberts [40] to each individual asperity. The theory predicts that the pull-off force, expressed as a fraction of the maximum value, depends upon a single parameter, which may be regarded as representing the statistically averaged competition between the compressive forces exerted by the higher asperities trying to prize the surfaces apart and the adhesive forces between the lower asperities trying to hold the surfaces together. This picture of adhesion developed by Tabor and Fuller would be correct *if* the surfaces had roughness on a single length scale as assumed in their study. However, when roughness occurs on many different length scales, a qualitatively new picture emerges [28], where, e. g., the adhesion force may even vanish (or at least be strongly reduced), if the rough surface can be described as a self affine fractal with fractal dimension $D_\mathrm{f} > 2.5$. Even for surfaces with roughness on a single length scale, the formalism used by Fuller and Tabor is only valid at "high" surface roughness, where the area of real contact (and the adhesion force) is very small. The theory presented below is particularly accurate for "small" surface roughness, where the area of real contact equals the nominal contact area.

Qualitative Discussion

Let us estimate the energy necessary in order to deform a rubber block so that the rubber fills up a substrate cavity of height h and width λ. The elastic

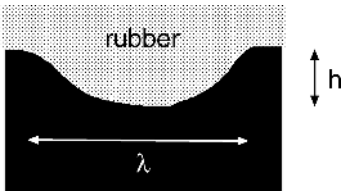


Fig. 14.25. A rubber surface is "pulled" into a cavity in a hard solid by the rubber-substrate adhesional interaction. The elastic energy stored in the deformation field is of order $E\lambda h^2$

energy stored in the deformation field in the rubber is given by

$$U_{\rm el} \approx \frac{1}{2} \int {\rm d}^3x \; \sigma\epsilon$$

where the stress $\sigma \approx E\epsilon$, where E is the elastic modulus. The deformation field is mainly localized to a volume $\sim \lambda^3$ (see Fig. 14.25) where the strain $\epsilon \approx h/\lambda$. Thus we get $U_{\rm el} \approx \lambda^3 E(h/\lambda)^2 = E\lambda h^2$.

Let us now consider the role of the rubber-substrate adhesion interaction. As shown above, when the rubber deforms and fills out a surface cavity of the substrate, an elastic energy $U_{\rm el} \approx E\lambda h^2$ will be stored in the rubber. Now, if this elastic energy is smaller than the gain in adhesion energy $U_{\rm ad} \approx \Delta\gamma\lambda^2$, where $\Delta\gamma = \gamma_1 + \gamma_2 - \gamma_{12}$ is the change of surface free energy (per unit area) upon contact due to the rubber-substrate interaction (which usually is mainly of the van der Waals type), then (even in the absence of an external load $F_{\rm N}$) the rubber will deform *spontaneously* to fill out the substrate cavities. The condition $U_{\rm el} = U_{\rm ad}$ gives $h/\lambda \approx (\Delta\gamma/E\lambda)^{1/2}$. For example, for very rough surfaces with $h/\lambda \approx 1$, and with parameters typical for rubber $E = 1\,{\rm MPa}$ and $\Delta\gamma = 3\,{\rm meV/\AA^2}$, the adhesion interaction will be able to deform the rubber and completely fill out the cavities if $\lambda < 0.1\,\mu{\rm m}$. For very smooth surfaces $h/\lambda \sim 0.01$ or smaller, so that the rubber will be able to follow the surface roughness profile up to the length scale $\lambda \sim 1\,{\rm mm}$ or longer.

The argument given above shows that for elastic solids with surface roughness on a *single length scale* λ, the competition between adhesion and elastic deformation is characterized by the parameter $\theta = Eh^2/\lambda\delta \approx U_{\rm el}/U_{\rm ad}$, where h is the amplitude of the surface roughness and $\delta = 4(1-\nu^2)\Delta\gamma/E$ the so called *adhesion length*, ν being the Poisson ratio of the rubber. The parameter θ is the ratio between the elastic energy and the surface energy stored at the interface, assuming that complete contact occurs. When $\theta \gg 1$ only partial contact occurs, where the elastic solids make contact only close to the top of the highest asperities, while complete contact occurs when $\theta \ll 1$.

Pull-Off Force

Consider a rubber ball (radius R_0) in adhesive contact with a perfectly smooth and hard substrate. The elastic deformation of the rubber can be determined by minimizing the total energy which is the sum of the (positive) elastic energy stored in the deformation field in the rubber ball, and the

(negative) binding energy between the ball and the substrate at the contact interface. The energy minimization gives the pull-off force [40, 41]

$$F_c = (3\pi/2)R_0\Delta\gamma\,. \tag{14.12}$$

Consider now the same problems as above, but assume that the substrate surface has roughness described by the function $z = h(\boldsymbol{x})$. We assume that the surface roughness power spectra has a roll-off wavelength $\lambda_0 = 2\pi/q_0$ (see Fig. 14.10) which is smaller than the diameter of the nominal contact area between the two solids. In this case we can still use the result (14.16), but with $\Delta\gamma$ replaced by $\gamma_{\rm eff}$. The effective interfacial energy $\gamma_{\rm eff}$ is the change in the interfacial free energy when the elastic solid is brought in contact with the rough substrate. $\gamma_{\rm eff}(\zeta)$ depends on the magnification ζ, and the interfacial energy which enters in the rubber ball pull-off experiment is the macroscopic interfacial energy, i. e., $\gamma_{\rm eff}(\zeta)$ for $\zeta = 1$. If A_0 is the nominal contact area and A_1 the true atomic contact area, then

$$A_0\gamma_{\rm eff}(1) = A_1\Delta\gamma - U_{\rm el} \tag{14.13}$$

where $U_{\rm el}$ is the elastic energy stored at the interface as a result of the elastic deformations necessary in order to bring the solids in atomic contact in the area A_1.

Stress Probability Distribution

The theory in Ref. [28] is based on the contact mechanics formalism described in Sect. 14.4.1. Thus, we focus on the stress probability distribution function $P(\sigma, \zeta)$ which satisfies Eq. (14.5):

$$\frac{\partial P}{\partial \zeta} = f(\zeta)\frac{\partial^2 P}{\partial \sigma^2}$$

We assume that detachment occurs when the local stress on the length scale L/ζ reaches $-\sigma_a(\zeta)$. Thus, the following boundary condition is valid in the present case

$$P(-\sigma_a(\zeta), \zeta) = 0$$

This boundary condition replaces the condition $P(0, \zeta) = 0$ valid in the absence of adhesion (see Sect. 14.4.1).

Let us consider the system on the characteristic length scale $\lambda = L/\zeta$. The quantity $\sigma_a(\zeta)$ is the stress necessary to induce a detached area of width λ. This stress can be obtained from the theory of cracks, where for a penny-shaped crack of diameter λ

$$\sigma_a = \left[\frac{\pi\gamma_{\rm eff}(\zeta)E}{(1-\nu^2)\lambda}\right]^{1/2} = \left[\frac{\gamma_{\rm eff}(\zeta)Eq}{2(1-\nu^2)}\right]^{1/2} \tag{14.14}$$

where $q = 2\pi/\lambda = \zeta q_L$. In Ref. [28] we have derived two equations for $\gamma_{\rm eff}(\zeta)$ and $P(\zeta)$ which determine how these quantities depend on the magnification ζ; those equations are the basis for the numerical results presented below.

Numerical Results

Figure 14.26 shows (a) the effective interfacial energy $\gamma_{\rm eff}(\zeta)(\zeta = 1)$ and (b) the normalized area of real contact, $P(\zeta_1) = A(\zeta_1)/A_0$, as a function of $q_0 h_0, h_0$ being the surface r.m.s. roughness and q_0 the roll-off wave vector. Results are shown for different adhesion lengths $\delta = 4(1-\nu^2)\Delta\gamma/E$: $q_0\delta = 0.1$, 0.2, 0.4 and 0.8. We will refer to $\gamma_{\rm eff}(1)$ at the magnification $\zeta = 1$ as the *macroscopic* interfacial free energy which can be deduced from, e. g., the pull off force for a ball according to Eq. (14.12). Note that for $q_0\delta = 0.4$ and 0.8 the macroscopic interfacial energy first increases with increasing amplitude h_0 of the surface roughness, and then decreases. The increase in $\gamma_{\rm eff}$ arises from the increase in the surface area. As shown in Fig. 14.26b, for small h_0 the two solids are in complete contact, and, as expected, the complete contact remains to higher h_0 as $\delta \sim \Delta\gamma/E$ increases. Note also that the contact area is nonzero even when $\gamma_{\rm eff}(1)$ is virtually zero: the fact that $\gamma_{\rm eff}(1)$ (nearly) vanish does not imply that the contact area vanish (even in the absence of an external load), but imply that the (positive) elastic energy stored at

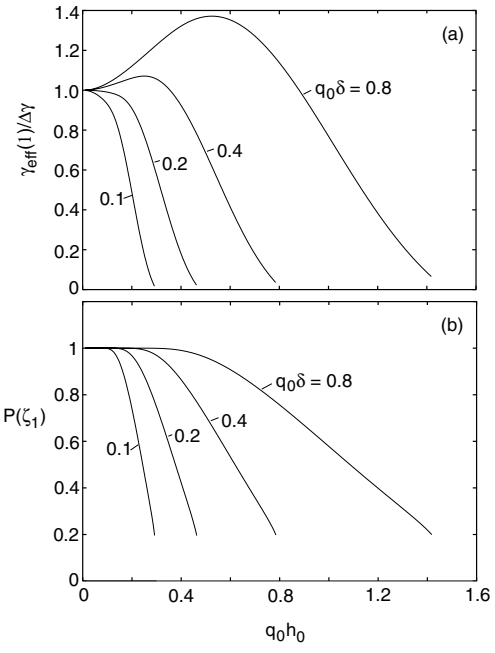


Fig. 14.26. a The macroscopic interfacial energy as a function of the dimensionless surface roughness amplitude $q_0 h_0$. **b** The normalized area of real contact, $P(\zeta_1) = A(\zeta_1)/A_0$, as a function of $q_0 h_0$. The curves correspond to different adhesion energies: $q_0\delta = 0.1$, 0.2, 0.4 and 0.8 as indicated. For $H = 0.8$ and $q_1/q_0 = \zeta_1 = 100$

the interface just balance the (negative) adhesion energy from the area of real contact. *The stored elastic energy at the interface is given back when removing the block, and when* $\gamma_{\rm eff}(1) \approx 0$ *it is just large enough to break the block-substrate bonding.*

Plate Adhesion

In this section we discuss the adhesion of a thin elastic plate to a randomly rough hard substrate. This topic is important for many applications such as thin films used as protective coatings [42], for the manufacturing of multi-layered wafer structures [43], or in bio-films for orthopedic implants [44]. The problem under consideration is also of great importance for understanding the adhesion of flies, bugs, and lizards to a rough substrate (see Fig. 14.27), [45, 46] or the adhesive behavior of recently biologically-inspired adhesive films [47].

Here we consider in detail the case of a thin plate in partial contact with a hard substrate with a self-affine fractal rough surface. Figure 14.28 (thick lines) shows (a) the macroscopic interfacial energy $\gamma_{eff}(1)$, i. e. the effective interfacial energy calculated at the magnification $\zeta = 1$, and (b) the normalized area of real contact $P(\zeta_1)$ at the maximum magnification $\zeta = \zeta_1$, as a function of the dimensionless roughness amplitude $q_0 h_0$. We show results for three different values of $q_0\delta$. The results are for $H = 0.8$, i. e. $D_f = 2.2$, and for a dimensionless thickness of the plate equal to $q_0 d = 0.63$. Note that the macroscopic interfacial energy initially increases with the amplitude h_0 of the rough profile up to a maximum value, and after decreases with h_0. This is caused by the increase of the real contact area produced by the fine structure of the rough profile. Figure 14.26b shows, indeed, that at small h_0 the plate adheres in full contact to the substrate, so that an increase of the surface roughness produces a corresponding increases of the area of contact and, hence, of the surface energy. However this is no more true at large h_0, because of the reduction of the area of real contact. Figure 14.26 also shows that, as expected, the roughness-induced increment of the macroscopic interfacial energy grows by increasing the adhesion length

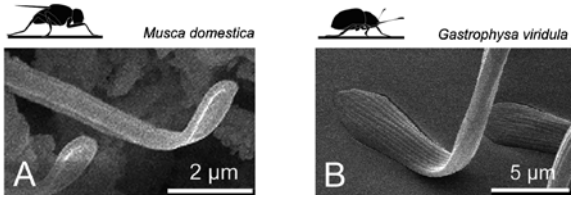


Fig. 14.27. Insect attachment systems consist of fibers or hair which terminates with leaf-like plates which can easily deform (without storing a lot of elastic energy) to bind strongly even to very rough substrates

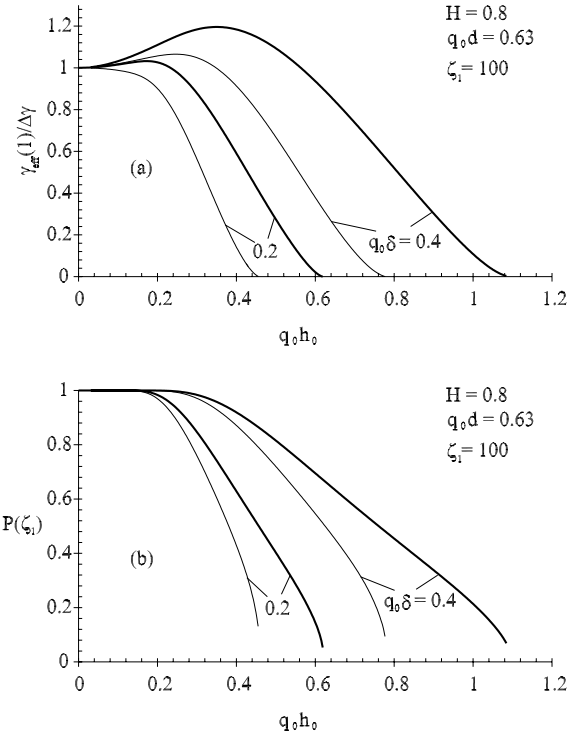


Fig. 14.28. a The normalized macroscopic interfacial energy and **b** the normalized area of real contact, as a function of the dimensionless surface roughness amplitude $q_0 h_0$. *Thick lines* are for the plate case and *thin lines* are for the semi-infinite solid case. Results are for $H = 0.8$ and $q_0 d = 0.63$, and for two different values of $q_0 \delta$

$\delta \sim \Delta\gamma/E$, and that the full contact condition remains to higher amplitude h_0 as δ increases.

In Fig. 14.28 we compare the results obtained for the plate case (thick lines) with those of the semi-infinite solid (thin lines). As expected, because of the higher compliance of the plate, both the macroscopic interfacial energy $\gamma_{\text{eff}}(1)$ and the normalized area of real contact $P(\zeta_1)$ are larger than for the semi-infinite solid case.

To summarize, at small magnification (long length scales) the plate, because of its higher compliance, is able to adhere in apparent full contact to the long wavelength corrugation of the underlying surface. That is, at length scales longer than the plate thickness, the gain in the adhesion energy upon the contact with the substrate overcomes the repulsive elastic energy produced by the elastic deformations, and the plate is able to fill out the large cavities of the rigid substrate. This produces a larger area of contact and an enhanced capability to adhere to a rough surface in comparison to the semi-infinite elastic solid case. However, at large enough magnification (small length scales) the plate behaves as a semi-infinite solid, and, depending on the roughness statistical properties, the area of true atomic contact may be much smaller than the nominal contact area.

Experimental Manifestations

Unfortunately, the surface roughness power spectrum has not been measured for any surface for which adhesion has been studied in detail. Instead only the roughness amplitude (center line average) and the radius of curvature of the largest surface asperities was determined. Nevertheless, the experimental data of Fuller, Tabor, Briggs, Briscoe and Roberts [32, 35, 36] are in good qualitative agreement with our theoretical results. In Fig. 14.29 we show the macroscopic interfacial energy for "hard" and "soft" rubber in contact with Perspex, as a function the substrate (Perspex) roughness amplitude as obtained by Briggs and Briscoe [35]. It is not possible to compare these results quantitatively with the theory developed above since the power spectrum $C(q)$ was not measured for the Perspex substrate. Even if the surfaces would be self affine fractal as assumed above, not only the surface roughness amplitude will change from one surface to another, but so will the long distance cut off length λ_0 and hence also the ratio $\zeta_1 = q_1/q_0$. In the experiments reported on in Ref. [35] the Perspex surfaces where roughened by blasting with fine particles. The roughness could be varied through the choice of the particles and the air pressure.

One practical problem is that most rubber materials have a wide distribution of relaxation times, extending to extremely long times. This effect is well known in the context of rubber friction (see Sect. 14.6.1.), where measurements of the complex elastic modulus show an extremely wide distribution of relaxation times, resulting in large sliding friction even at very low sliding velocities, $v < 10^{-8}$ m/s.

The effect of the stored elastic energy on adhesion has recently been studied using a polyvinylsiloxane rubber block squeezed against a smooth glass surface for a fixed time period before measuring the pull-off force [48]. The square-symbols in Fig. 14.30 show the pull-off force as a function of the

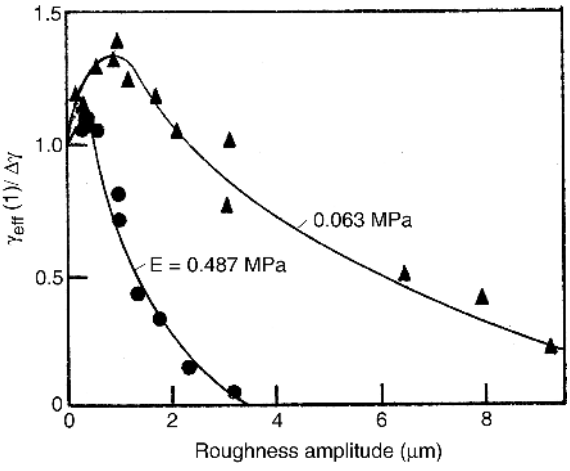


Fig. 14.29. The macroscopic interfacial energy (obtained from the pull-off force) for a smooth rubber surface (*ball*) in contact with Perspex surface as a function of the roughness (*center line average*) of the Persplex. Results are shown for a "soft" rubber ($E = 0.063$ MPa) and a "hard" rubber ($E = 0.487$ MPa). From [35]

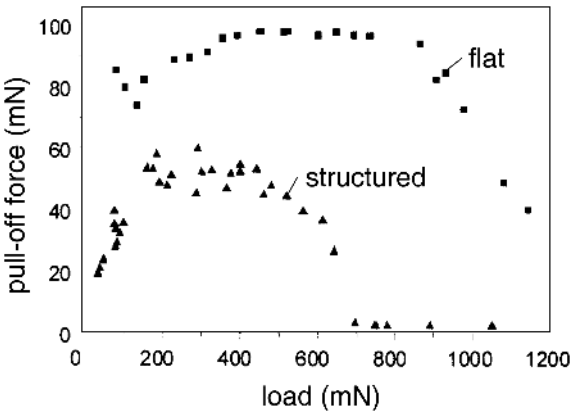


Fig. 14.30. The pull-off force as a function of the squeeze force or load. For silicon rubber in contact with a smooth glass surface. From Ref. [48]

squeezing force. For squeezing forces $F_N > 850\,\text{mN}$ the pull off force decreases. This may be explained by a drastic increase of the elastic energy stored in the rubber because of the strong deformation of the rubber (which remains even when the load is removed as a result of the rubber-glass friction at the interface), see Fig. 14.31(top). This energy, freed during the process of unloading, will help to break the adhesive bonds at the interface. This effect is even stronger when the surface is structured. Thus, the triangles in the figure shows the pull-off force when the rubber surface is covered by a regular array of rubber cylindrical asperities. In this case the pull-off force drops to nearly zero for $F_N > 700\,\text{mN}$. Visual inspection shows that in this case the cylindrical asperities at high load bend and make contact with the glass on one side of the cylinder surface, see Fig. 14.31(bottom). This again stores a lot of elastic energy at the interface which is given back during pull-off, reducing the pull-off force to nearly zero.

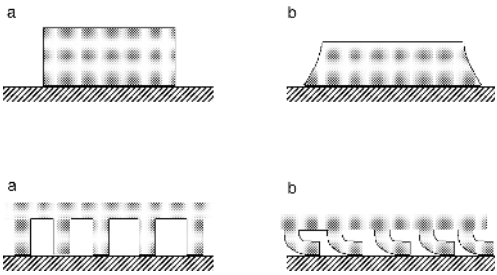


Fig. 14.31. Elastic deformation of a rubber block with a smooth surface (*top*) and a structured surface (*bottom*). **a** shows the initial state before applying a squeezing force, and **b** the new state (without load) after applying (and then removing) a very large squeezing force. In state **b** a lot of elastic energy is stored in the rubber which is "given back" during pull-off resulting in a nearly vanishing pull-off force

The Role of Plastic Yielding on Adhesion

When the local stress in the asperity contact regions between two solids becomes high enough, at least one of the solids yields plastically. This will tend to increase the effective adhesion (or pull-off force) for the following three reasons. First, the area of real contact between the solids will increase as compared to the case where the deformations are purely elastic. Secondly, the amount of stored elastic energy in the contact regions (which is given back during pull-off) will be reduced because of the lowered elastic deformations. Finally, for many materials plastic yielding will strengthen the junctions [49]. For example, most metals are protected by thin oxide layers, and as long as these are intact the main interaction between the surfaces in the contact areas may be of the van der Waals and electrostatic origin. However, when plastic yielding occurs it may break up the oxide films resulting in direct metal-metal contact and the formation of "cold-welded" junctions. When this occurs, because of the high ductility of many metals, during pull-off "long" metallic bridges may be formed between the solids so that instead of having junctions popping one after another during pull-off, a large number of adhesive junctions may simultaneously impede the surface separation during pull-off, leading to a large pull-off force. However, experiment have shown [7] that just squeezing before pull-off will in general only result in very few cold welded junctions, while squeezing *and* sliding will break up the oxide film, resulting in the formation of many more cold welded contact regions, and will hence result in a much larger pull-off force.

14.6.2 The Adhesion Paradox

The biggest "mystery" related to adhesion is not why it is sometimes observed but rather why it is usually not observed. Thus, even the weakest force in Nature of relevance in condensed matter physics, namely the van der Waals force, is relatively strong on a macroscopic scale. For example, even a contact area of order $1\,\mathrm{cm}^2$ can sustain the weight of a car (i. e., a force of order $10^4\,\mathrm{N}$) [see Fig. 14.32a] also when only the van der Waals interaction operates at the interface. [Here we have assumed that the bond breaking occurs uniformly over the contact area as illustrated in Fig. 14.32b.] However, this is never observed in practice and this fact is referred to as the *adhesion paradox*.

There are several reasons why adhesion is usually not observed between macroscopic bodies. For example, on a macroscopic scale the bond-breaking usually does not occur uniformly as in Fig. 14.32b, but occurs by crack propagation, see Fig. 14.32c. The local stress at the crack tip is much higher than the average stress acting in the contact area, and this drastically reduces the pull-off force. Another reason, already addressed in Sect. 14.6.1, is the influence of surface roughness. Thus, for elastically hard surfaces the true (atomic) contact between the solids at the interface is usually much smaller than the nominal contact area. In addition, the elastic energy stored in the

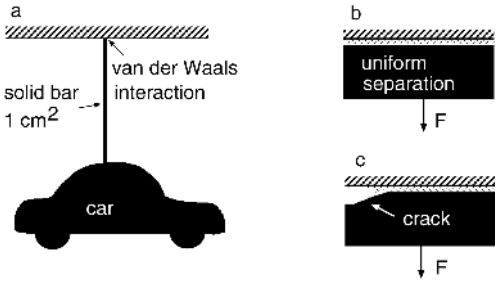


Fig. 14.32. Even the weakest force in Nature of relevance in condensed matters physics, namely the van der Waals force, is relative strong on a macroscopic scale. Thus, for example, if the bond breaking occur uniformly over the contact area as in **b**, already a contact area of order $1\,\mathrm{cm}^2$ can sustain the weight of a car (i. e., a force of order 10^4 N) [see **a**]. However, on a macroscopic scale the bond-breaking does not usually occur uniformly over the contact area, but by crack propagation, see **c**, which drastically reduce the pull-off force. In addition, interfacial surface roughness drastically reduces the pull-off force

solids in the vicinity of the contact regions is given back during pull-off and helps to break the interfacial bonds between the solids (see Sect. 14.6.1).

It is interesting to note that for very small solid objects, typically of order $100\,\mu$m or smaller, the bond breaking may occur uniformly over the contact area (no crack propagation) so that adhesion between smooth surfaces of small objects, e. g., in micromechanical applications (MEMS), may be much stronger than for macroscopic bodies, and this fact must be taken into account when designing MEMS [50, 51].

14.6.3 The Role of Liquids on Adhesion Between Rough Solid Surfaces

As explained in Sect. 14.6.1, surface roughness reduces the adhesion between clean surfaces. First, it lowers the area of real contact. Since the adhesion interaction comes almost entirely from the area where the solids make atomic contact, it is clear that the surface roughness may drastically reduce the adhesion. Secondly, elastic deformation energy is stored in the vicinity of the asperity contact regions. During pull-off the elastic energy is "given back" to the system, usually resulting in a drastic reduction in the effective adhesion and the pull-off force.

Most surfaces have at least nano-scale roughness, and hard solids in the normal atmosphere have at least a monolayer of liquid-like "contamination" molecules, e. g., water and hydrocarbons. Small amount of (wetting) lubricant or contamination liquids between rough solid walls may drastically enhance the adhesion. Thus, for surfaces with nanoscale roughness, a monolayer of a wetting liquid may result in the formation of a large number of nano-bridges between the solids, which increases the pull-off force. This effect is well

known experimentally. For example, the adhesion force which can be detected between gauge blocks (steel blocks with very smooth surfaces) is due to the formation of many very small capillary bridges made of water or organic contamination. For thicker lubrication or contamination films the effective adhesion will be more long-ranged but the pull-off force may be smaller. The thickness of the lubricant or contamination layer for which the pull-off force is maximal will in general depend on the nature of the surface roughness, but is likely to be of order the root-mean-square roughness amplitude. In fact, it is an interesting and important problem to find out at exactly what liquid thickness the pull-off force is maximal.

Some insects such as flies or crickets inject a thin layer of a wetting liquid in the contact region between the insect attachment surfaces and the (rough) substrate. The optimum amount of injected liquid will depend on the nature of the substrate roughness, and it is likely that the insect can regulate the amount of injected liquid by a feedback system involving the insect nerve system.

Here we consider the adhesion between two solid elastic walls with nanoscale roughness, lubricated by octane [38, 39, 52]. We consider two types of substrates (bottom surface) – flat and nano-corrugated (corrugation amplitude 1 nm and wavelength of the corrugation in x and y direction, 4 nm) – and varied the lubricant coverage from $\sim 1/8$ to ~ 4 monolayers of octane. The upper surface (the block) is assumed to be atomically smooth but with

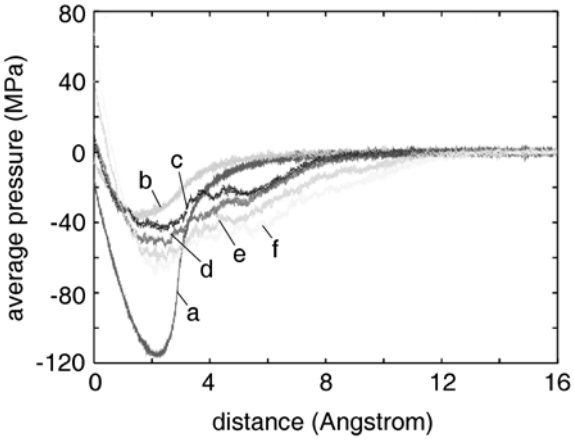


Fig. 14.33. The variation of the average pressure during retraction developed as the block moves a distance of 16 Å away from the substrate. Octane C_8H_{18} was used as lubricant. Pull-off (retraction) velocity was $v_z = 1$ m/s. (a) For the flat substrate without lubricant. (b) For the corrugated substrate without lubricant. Curves (c)–(f) show results for the corrugated substrate with about $1/8, 1/4, 1/2$ and 1 monolayer of octane in the contact region, respectively. For clarity, the curve for the flat substrate (a) is displaced to the right, by 2 Å

a uniform cylinder curvature with a radius of curvature $R \approx 100\,\text{nm}$ (see Fig. 14.35 below). The results presented here have been obtained using standard molecular dynamics calculations [38].

Figure 14.33 shows the variation of the average pressure during retraction as the block moves a distance of 16 Å away from the substrate. The pull-off (retraction) velocity was $v_z = 1\,\text{m/s}$. We have varied the lubricant coverage from 0 to 1 monolayer in the contact region. The pull-off force is maximal when the adsorbate coverage is of the order of one monolayer [curve (f)]. However, the pull-off force is still smaller than for a *flat* substrate without lubricant [curve (a)]. As a function of the octane coverage (for the corrugated

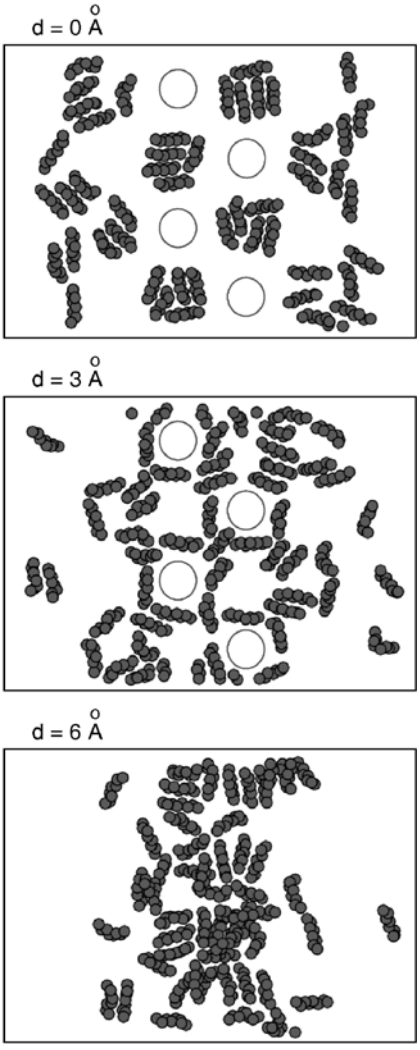


Fig. 14.34. Snapshot pictures (for three different block positions $d = 0, 3$ and 6 Å) of the lubricant layer during retraction. We only show the lubricant molecules in the central part of the contact area between the block and the substrate surfaces (top view, surfaces parallel to the plane of the image). Octane C_8H_{18} was used as lubricant. Pull-off (retraction) velocity was $v_z = 1\,\text{m/s}$. For the corrugated substrate with about 1/4 monolayer of octane in the contact region. The circles indicate the position of several asperity tops of the corrugated substrate surface

substrate) the pull-off force first increases as the coverage increases from zero to ~ 1 monolayer, and then decreases as the coverage is increased beyond monolayer coverage (not shown).

At low octane coverage, the octane molecules located in the substrate corrugation wells during squeezing, are pulled out of the wells during pull-off, forming a network of nano capillary bridges around the substrate nanoasperities, thus increasing the adhesion between two surfaces, see Figs. 14.34 and 14.35. For greater lubricant coverages a single capillary bridge is formed.

Let us discuss the nature of the adhesion for the corrugated substrate, with about 1/4 monolayer of octane in the contact region. Fig. 14.34 shows snapshot pictures of the lubricant layer during retraction, as the block moves away from the substrate for three different block positions $d = 0, 3$ and 6 Å. Only the central part of the contact between the block and the substrate is shown, top view, after removing the block and substrate atoms. In the beginning ($d = 0$ Å) octane molecules are located in the substrate corrugation wells, or cavities with direct metal–metal contact between the block and the top of the substrate nano asperities (see Fig. 14.35). During retraction ($d = 3$ Å) the octane molecules are pulled out of the wells forming an almost symmetric network of nano-bridges around the asperity tops, increasing the

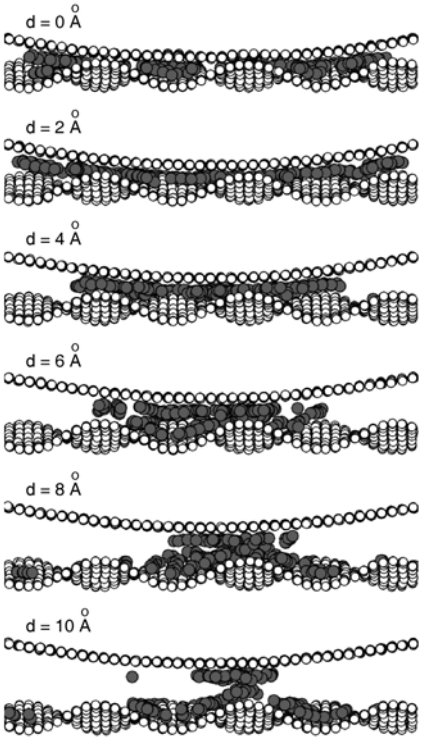


Fig. 14.35. Snapshot pictures (for six different block positions) during retraction. The snapshot pictures show the side view of the central 108 Å × 50 Å section (in the xy-plane) of the contact area. Octane C_8H_{18} was used as lubricant. Pull-off (retraction) velocity was $v_z = 1$ m/s. For the corrugated substrate with about 1/4 monolayer of octane in the contact region

adhesion between the two surfaces. This configuration corresponds to the maximal adhesion force, see curve (d) in Fig. 14.33. Thus maximal adhesion is achieved via the formation of many small capillary nano-bridges, involving just a few molecules for each bridge (see Fig. 14.35). Further retraction ($d = 6$ Å) results in the collapse of the nano-bridges and the formation of a single "large" capillary bridge in the center of the contact region.

14.7 Summary and Outlook

We have discussed the origin of friction and adhesion between hard solids such as quasicrystals. However, most of the results presented above are valid for all types of solids. We have emphasized the fundamental role of surface roughness on many contact mechanics problems, in particular for friction and adhesion between solids bodies.

Surface roughness has a huge influence on many common phenomena. It is the main reason for why macroscopic bodies usually do not adhere to each other with any measurable strength. For example, if the floor and the sole of the shoes would be atomically smooth and clean, it would be impossible to walk on the floor! The (nearly) absence of adhesion in most situations is crucial for the function of many man-made constructions.

The surface to volume ratio of solid objects increases as the lateral size of the object decreases. The role of surface roughness becomes therefore more important as the size of objects decreases. The present drive toward the miniaturization of mechanical devices, e. g., MEMS, requires a better understanding of the role of surface roughness on, e. g., contact mechanics and adhesion.

Surface roughness is also of great importance for the function of many biological systems. Thus, flies, bugs, crickets and lizards have developed very soft layers on their attachment organs which allow them to attach and move on both very smooth and rough vertical solid walls, e. g. stone walls or leafs. Another example is non-wetting coatings on plant surfaces based on surface roughness on many different length scales (the so called Lotus effect) [10].

The roughness of surfaces can today be studied strait forwardly using standard equipments based on optical methods and stylus methods, e. g., the atomic force microscope (AFM). These methods cover the whole length scale from atomic dimension to macroscopic distances. Thus, the AFM can probe the surface profile from ~ 1 nm to 100 μm and optical methods from ~ 1 μm to kilometers. For randomly rough surfaces, the most important quantity which can be deduced from the measured height profile is the surface roughness power spectra. We have shown in this paper how the power spectra determines the contact mechanics and adhesion for solid objects in direct contact. It also govern rubber friction on rough substrates, e. g., tires on a road surfaces, and influence other phenomena of technological importance, e. g., the roughness induced leaking of sealings.

Thus, studies of surface roughness is important not only for understanding many natural and biological phenomena, but also for many technological processes. The present drive toward miniaturization and the design of optimal systems by transfer of ideas from studies on biological systems to materials science (bionics) [10,53], is likely to accelerate the interest and efforts to study and predict the influence of surface roughness on many phenomena.

References

1. D. Shechtman, I. Blech, D. Gratias and J.W. Cahn, Phys. Rev. Lett. **53**, 1951 (1984).
2. B.N.J. Persson, O. Albohr, F. Mancosu, V. Peveri, V.N. Samoilov and I.M. Sivebaek, Wear **254**, 835 (2003).
3. E. Riedo and H. Brune, Applied Physics Letters **83**, 1986 (2003).
4. J.M. Dubois, S.S. Kang, J. von Stebut, J. Mater. Sci. Lett. **10**, 537 (1991); J.M. Dubois, P. Brunet, W. Costin ans A. Merstallinger, J. Non-Cryst. Solids **334-335**, 475 (2004).
5. J.Y. Park, D.F. Ogletree, M. Salmeron, C.J. Jenks and P.A. Thiel, Tribology Letters **17**, 629 (2004).
6. J.S. Ko, A.J. Gellman, T.A. Lograsso, C.J. Jenks and P.A. Thiel, Surface Science **423**, 243 (1999).
7. C.F. McFadden and A.J. Gellman, Tribology Lett. **1**, 201 (1995).
8. Y.-P. Zhao, L.S. Wang and T.X. Yu, J. Adhesion Sci. Technol. **17**, 519 (2003).
9. C.H. Mastrangelo, Trib. Lett. **3**, 223 (1997).
10. See *http://www.botanik.uni-bonn.de/system/bionik_flash.html* for information involving surface roughness in relation to hydrophobicity and surface self cleaning in biological systems.
11. C. Gui, M. Elwenspoek, N. Tas and J.G.E. Gardeniers, J. Appl. Physics **85**, 7448 (1999).
12. See, e. g., A.-L. Barabasi and H.E. Stanley, *Fractal Concepts in Surface Growth*, Cambridge University Press, Cambridge (1995).
13. See, e. g., J. Krug, Advances in Physics **46**, 139 (1997).
14. J. Feder, *Fractals* (Plenum Press, New York, 1988); M.V. Berry and Z.V. Lewis, Proceedings of the Royal Society of London, A **370**, 459 (1980).
15. J. Krim and G. Palasantzas, Int. J. of Modern Phys. B**9**, 599 (1995).
16. P.R. Nayak, ASME J. Lubrication Technology **93**, 398 (1971).
17. E. Bouchaud, J. Phys.: Condens. Matter **9**, 4319 (1997).
18. B.N.J. Persson, *Sliding Friction: Physical Principles and Applications*, Sec. Ed. (Springer, Heidelberg, 2000).
19. H. Hertz, J. Reine Angew. Math. **92** 156 (1882).
20. J.F. Archard, Proc. Roy. Soc. A**243** 190 (1957).
21. J.A. Greenwood, in *Fundamentals of Friction, Macroscopic and Microscopic Processes*, Ed. by I.L. Singer and H.M. Pollack (Kluver, Dordrecht, 1992).
22. J.A. Greenwood and J.B.P. Williamson, Proc. Roy. Soc. A**295**, 300 (1966).
23. K.L. Johnson, *Contact Mechanics* (Cambridge Univ. Press, Cambridge 1985).
24. A.W. Bush, R.D. Gibson and T.R. Thomas, Wear **35**, 87 (1975); A.W. Bush, R.D. Gibson and G.P. Keogh, Mech. Res. Commun. **3**, 169 (1976).

25. B.N.J. Persson, Phys. Rev. Lett. **87**, 1161 (2001); B.N.J. Persson, J. Chem. Phys. **115**, 3840 (2001).
26. We define the apparent contact area $A(\lambda)$ on the length scale λ to be the area of real contact if the surface would be smooth on all length scales shorter than λ. That is, considering the Fourier transform of the surface profile, all the components whose wave vector is larger that $2\pi/\lambda$ have to be set to 0, and the contact area with this new surface is by definition $A(\lambda)$.
27. B.N.J. Persson, F. Bucher and B. Chiaia, Phys. Rev. B**65**, 184106 (2002).
28. B.N.J. Persson, Eur. Phys. J. E**8**, 385 (2002); Phys. Rev. Lett. **89** 245502 (2002).
29. S. Hyun, L. Pei, J.-F. Molinari, and M.O. Robbins, Phys. Rev. E (in press).
30. C. Yang, U. Tartaglino and B.N.J. Persson, Eur. Phys. J. E 19, 47 (2006).
31. M. Borri-Brunetto, B. Chiaia and M. Ciavarella, Comput. Methods Appl. Mech. Engrg. **190**, 6053 (2001).
32. K.N.G. Fuller and D. Tabor, Proc. R. Soc. London A**345**, 327 (1975).
33. K. Kendall, *Molecular Adhesion and it Applications*, (Kluwer, New York, 2001); D. Maugis, *Contact, Adhesion and Rupture of Elastic Solids*, (Springer, Berlin, 1999).
34. K. Kendall, J. Phys. D **4**, 1186 (1971); **6**, 1782 (1973); **8**, 115 (1975). See also the beautiful review article of K. Kendall, Contemp. Phys. **21**, 277 (1980).
35. G.A.D. Briggs and B.J. Briscoe, J. Phys. D: Appl. Phys. **10**, 2453 (1977).
36. K.N.G. Fuller and A.D. Roberts, J. Phys. D: Appl. Phys. **14**, 221 (1981).
37. S. Zilberman and B.N.J. Persson, J. Chem. Phys. **118** 6473 (2003).
38. V.N. Samoilov, I.M. Sivebaek, and B.N.J. Persson, J. Chem. Phys. **121** 9639 (2004).
39. J.N. Israelachvili, *Intermolecular and Surface Forces* (Academic Press, London, 1995).
40. K.L. Johnson, K. Kendall and A.D. Roberts, Proc. Roy. Soc. Lond. A**324**, 301 (1971).
41. G. Sperling, PhD thesis, Karlsruhe Technical University (1964).
42. N.P. Padture, M. Gell, and E.H. Jordan, Science, **296**, 280 (2002).
43. K.T. Turner, S. M. Spearing, J Appl. Phys., **92** (12), 7658 (2002).
44. L.D. Piveteau, B. Gasser, L. Schlapbach, Biomaterials, **21**, 2193 (2000).
45. S. Gorb, B.N.J. Persson, J. Chem. Phys., **119** (21), 11437 (2003).
46. B.N.J. Persson, J. Chem. Phys. 118, 7614 (2003).
47. A.K. Geim, S.V. Dubonos, I.V. Gricorieva, K.S. Novoselov, A.A. Zhukov and S. Yu. Shapoval, Nature Materials, **2**, 461 (2003).
48. A. Peressadko and S.N. Gorb, Journal of Adhesion **80**, 247 (2004).
49. K.L. Johnson, Tribology International **31**, 413 (1998); L. Zhang and Y.-P. Zhao, J. Adhesion Sci. Technol. **18**, 715 (2004).
50. B.N.J. Persson, Wear **254**, 832 (2003).
51. H.J. Gao and H.M. Yao, Proceedings of the National Academy of Sciences of the United States of America **101**, 7851 (2004).
52. B.N.J. Persson and F. Mugele, J. Phys. Condens. Matter **16**, R295 (2004).
53. See *http://finix.mpi-stuttgart.mpg.de/arzt/mf_mpg_abteilungen_e/mf_mpg_abteilungen_arzt/bio/topics.html* for information involving surface roughness in biological "applications".

15 A Multiscale Molecular Dynamics Approach to Contact Mechanics and Friction: From Continuum Mechanics to Molecular Dynamics

U. Tartaglino[1,2], C. Yang[1], and B.N.J. Persson[1]

[1] IFF, FZ-Jülich, 52425 Jülich, Germany
[2] Democritos National Simulation Center, Via Beirut 2, 34014 Trieste, Italy

Adhesion and friction between solid surfaces are common phenomena in nature and of extreme importance in biology and technology [1]. Most surfaces of solids have roughness on many different length scales [2, 3], and it is usually necessary to consider many decades in length scale when describing the contact between solids [4]. This makes it very hard to describe accurately the contact mechanics between macroscopic solids using computer simulation methods, e. g., atomistic molecular dynamics, finite element calculations based on continuum mechanics. Here we will review some recent studies of contact mechanics and sliding friction using a multiscale molecular dynamics approach. In Sect. 15.1 we describe a multiscale molecular dynamics approach to contact mechanics and friction. In Sect. 15.2 we study stationary contact and in Sect. 15.3 sliding contact, for clean smooth and rough surfaces.

15.1 Multiscale Molecular Dynamics

15.1.1 Introduction

Consider a solid with a nominally flat surface. Let x, y, z be a coordinate system with the x, y plane parallel to the surface plane. Assume that $z = h(\boldsymbol{x})$ describes the surface height profile, where $\boldsymbol{x} = (x, y)$ is the position vector within the surface plane. The most important property characterizing a randomly rough surface is the surface roughness power spectrum $C(\boldsymbol{q})$ defined by [4, 5]

$$C(\boldsymbol{q}) = \frac{1}{(2\pi)^2} \int \mathrm{d}^2x \langle h(\boldsymbol{x}) h(\boldsymbol{0}) \rangle \, \mathrm{e}^{i\boldsymbol{q}\cdot\boldsymbol{x}} \, . \tag{15.1}$$

Here $\langle \ldots \rangle$ stands for ensemble average and we have assumed that $h(\boldsymbol{x})$ is measured from the average surface plane so that $\langle h \rangle = 0$. In what follows we will assume that the statistical properties of the surface are isotropic, in which case $C(q)$ will only depend on the magnitude $q = |\boldsymbol{q}|$ of the wave vector $\boldsymbol{q}$.

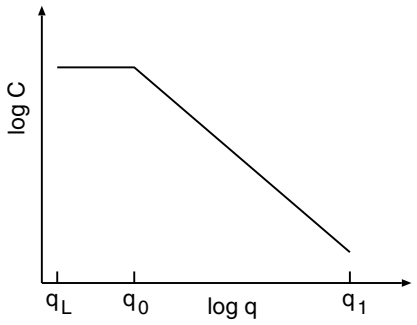


Fig. 15.1. Surface roughness power spectrum of a surface which is self-affine fractal for $q_1 > q > q_0$. The long-distance roll-off wave vector q_0 and the short distance cut-off wave vector q_1 depend on the system under consideration. The slope of the $\log C - \log q$ relation for $q > q_0$ determines the fractal dimension of the surface. The lateral size L of the surface (or of the studied surface region) determines the smallest possible wave vector $q_L = 2\pi/L$

Many surfaces tend to be nearly self-affine fractal. A self-affine fractal surface has the property that if part of the surface is magnified, with a magnification which in general is appropriately different in the direction perpendicular to the surface as compared to the lateral directions, then the surface "looks the same", i. e., the statistical properties of the surface are invariant under this scale transformation [4]. For a self-affine surface the power spectrum has the power-law behavior

$$C(q) \sim q^{-2(H+1)} ,$$

where the Hurst exponent H is related to the fractal dimension D_f of the surface via $H = 3 - D_\mathrm{f}$. Of course, for real surfaces this relation only holds in some finite wave vector region $q_0 < q < q_1$, and in a typical case $C(q)$ has the form shown in Fig. 15.1. Note that in many cases there is a roll-off wavevector q_0 below which $C(q)$ is approximately constant.

Let us consider the contact between an elastic solid with a flat surface and a hard randomly rough substrate. Figure 15.2 shows the contact between the solids at increasing magnification ζ. At low magnification ($\zeta = 1$) it looks as if complete contact occurs between the solids at many *macro asperity* contact regions, but when the magnification is increased smaller length scale roughness is detected, and it is observed that only partial contact occurs at

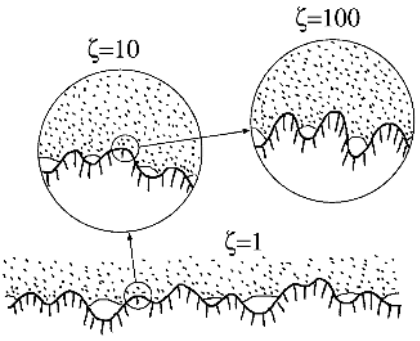


Fig. 15.2. A rubber block (*dotted area*) in adhesive contact with a hard rough substrate (*dashed area*). The substrate has roughness on many different length scales and the rubber makes partial contact with the substrate on all length scales. When a contact area is studied at low magnification ($\zeta = 1$) it appears as if complete contact occurs in the macro asperity contact regions, but when the magnification is increased it is observed that in reality only partial contact occurs

the asperities. In fact, if there would be no short distance cut-off the true contact area would vanish. In reality, however, a short distance cut-off will always exist since the shortest possible length is an atomic distance. In many cases the local pressure at asperity contact regions at high magnification will become so high that the material yields plastically before reaching the atomic dimension. In these cases the size of the real contact area will be determined mainly by the yield stress of the solid.

The magnification ζ refers to some (arbitrary) chosen reference length scale. This could be, e. g., the lateral size L of the nominal contact area, in which case $\zeta = L/\lambda$, where λ is the shortest wavelength roughness which can be resolved at magnification ζ. In this paper we will instead use the roll-off wavelength $\lambda_0 = 2\pi/q_0$ as the reference length so that $\zeta = \lambda_0/\lambda$.

Recently, a very general contact mechanics theory [5] has been developed. This theory can be applied to both stationary and sliding contact for viscoelastic solids (and for elastic solids as a special case). The theory was originally developed in order to describe rubber friction on rough substrates. For elastic solids the theory can also be applied when the adhesive interaction is taken into account [6]. In contrast to earlier contact mechanics approaches, the one presented in Ref. [5,6] is particularly accurate close to complete contact, as would be the case for, e. g., rubber on smooth surfaces. The basic idea behind it is to study the contact at different magnifications. In particular, the theory describes the change in the stress distribution $P(\sigma, \zeta)$ as the magnification ζ increases. Here

$$P(\sigma, \zeta) = \langle \delta(\sigma - \sigma(\boldsymbol{x}, \zeta)) \rangle \tag{15.2}$$

is the stress distribution at the interface when the surface roughness with wavelength smaller than $\lambda = \lambda_0/\zeta$ has been removed. In (15.2), $\sigma(\boldsymbol{x}, \zeta)$ is the perpendicular stress at the interface when surface roughness with wavelength shorter than $\lambda = \lambda_0/\zeta$ has been removed. It is clear that as the magnification ζ increases, the distribution $P(\sigma, \zeta)$ will be broader and broader and the theory describes this in detail. The (normalized) area of real contact (projected on the xy-plane) at the magnification ζ can be written as

$$\frac{A(\zeta)}{A_0} = \int_{0+}^{\infty} \mathrm{d}\sigma \; P(\sigma, \zeta) \; . \tag{15.3}$$

where the lower integration limit 0^+ indicates that the delta function at the origin $\sigma = 0$ (arising from the non-contact area) should be excluded from the integral. The rubber friction theory described in Ref. [5] depends on the function $A(\zeta)/A_0$ for *all magnifications*. This just reflects the fact that the friction force results from the viscoelastic deformations of the rubber on all length scales, and when evaluating the contribution to the friction from the viscoelastic deformations on the length scale λ, it is necessary to know the contact between the rubber and the substrate at the magnification $\zeta = \lambda_0/\lambda$.

Thus, not just the area of real (atomic) contact is of great interest, but many important applications require the whole function $A(\zeta)$, and the pressure distribution $P(\sigma, \zeta)$.

In order to accurately reproduce the contact mechanics between elastic solids, it is in general necessary to consider solid blocks which extend a similar distance in the direction normal to the nominal contact area as the linear size of the contact area. This leads to an enormous number of atoms or dynamical variables already for relatively small systems. In this paper we use a multiscale approach to contact mechanics where the number of dynamical variables scales like $\sim N^2$ rather than as $\sim N^3$, where $N \times N$ is the number of atoms in the nominal contact area.

15.1.2 Multiscale Molecular Dynamics

Let us discuss the minimum block-size necessary in a computer simulation for an accurate description of the contact mechanics between two semi-infinite elastic solids with nominally flat surfaces. Assume that the surface roughness power spectrum has a roll-off wavevector $q = q_0$ corresponding to the roll-off wavelength $\lambda_0 = 2\pi/q_0$. In this case the minimum block must extend $L_x \approx \lambda_0$ and $L_y \approx \lambda_0$ along the x and y-directions. Furthermore, the block must extend at least a distance $L_z \approx \lambda_0$ in the direction perpendicular to the nominal contact area. The latter follows from the fact that a periodic stress distribution with wavelength λ acting on the surface of a semi-infinite elastic solid gives rise to a deformation field which extends a distance $\sim \lambda$ into the solid. Thus, the minimum block is a cube with the side $L = \lambda_0$.

As an example, if λ_0 corresponds to 1000 atomic spacings, one must at least consider a block with 1000×1000 atoms within the xy-contact plane, i. e., one would need to study the elastic deformation in a cubic block with at least 10^9 atoms. However, it is possible to drastically reduce the number of dynamical variables without loss of accuracy if one notes that an interfacial roughness with wavelength λ will give rise to a deformation field in the block which extends a distance λ into the solid, and which varies spatially over a typical length scale λ. Thus when we study the deformation a distance z into the block we do not need to describe the solid on the atomistic level, but we can coarse-grain the solid by replacing groups of atoms with bigger "atoms" as indicated schematically in Fig. 15.3. If there are $N \times N$ atoms in the nominal contact area one need $n \approx \ln N$ "atomic" layers in the z-direction. Moreover the number of atoms in each layer decreases in a geometric progression every time the coarse graining procedure is applied, so that the total number of particles is of the order of N^2 instead of N^3. This results in a huge reduction in the computation time for large systems. This multiscale approach may be implemented in various ways, and in Ref. [7] we outline the procedure we have used in this paper which we refer to as the *smartblock*. Another implementation similar to our approach can be found in Ref. [8].

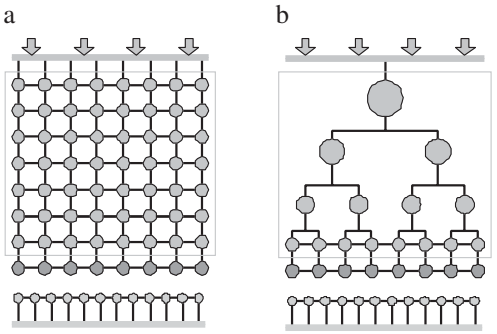


Fig. 15.3. Schematic structure of the model. **a** The fully atomistic model. **b** The multiscale *smartblock* model, where the solid in **a** is coarse grained by replacing groups of atoms with bigger "atoms"

The smartblock model should accurately describe the deformations in the solids as long as the deformations vary slowly enough with time. However, the model cannot accurately describe the propagation of short wavelength phonons. This is, in fact, true with all forms of Hamiltonian multiscale descriptions of solids, because of the energy conservation together and the unavoidable loss of information in the coarse grained region. In principle it should be possible to prevent the back reflection of short wavelength phonons by describing the coarse grained region as a continuum, where the numerical calculation can be carried on through a Finite Element scheme [9–12]. This indeed would require no coarse graining at all in the region treated with molecular dynamics, and a proper choice of the matching conditions between the atomistic and the continuum region. However, with respect to contact mechanics and adhesion the back reflection of short wavelength phonons is not an important limitation. With respect to sliding friction it may be a more severe limitation in some cases.

Figure 15.3 illustrates a case where the block is in the form of a cube with atomically flat surface. It is possible to obtain curved surfaces of nearly arbitrary shape by "gluing" the upper surface of the block to a hard curved surface profile. This was described in detail in Ref. [5]. The elastic modulus and the shear modulus of the solid can be fixed at any value by proper choices of the elongation and bending spring constants for the springs between the atoms (see Ref. [5]). The upper surface of the smartblock can be moved with arbitrary velocity in any direction, or an external force of arbitrary magnitude can be applied to the upper surface of the smartblock. We have also studied sliding friction problems where the upper surface of the smartblock is connected to a spring which is pulled in some prescribed way. The computer code also allows for various lubricant fluids between the solid surfaces of the block and the substrate. Thus the present model is extremely flexible and can be used to study many interesting adhesion and friction phenomena, which we will report on elsewhere.

We note that with respect to contact mechanics, when the slopes of the surfaces are small, i. e. when the surfaces are almost horizontal, one of the

two surfaces can be considered flat, while the profile of the other surface has to be replaced by the difference of the two original profiles [13]. Thus, if the substrate has the profile $z = h_1(\boldsymbol{x})$ and the block has the profile $z = h_2(\boldsymbol{x})$, then we can replace the actual system with a fictive one where the block has an atomically smooth surface while the substrate profile $h(\boldsymbol{x}) = h_2(\boldsymbol{x}) + h_1(\boldsymbol{x})$. Furthermore, if the original solids have the elastic modulus E_1 and E_2, and the Poisson ratio ν_1 and ν_2, then the substrate in the fictive system can be treated as rigid and the block as elastic with the elastic modulus E and Poisson ratio ν chosen so that $(1-\nu^2)/E = (1-\nu_1^2)/E_1 + (1-\nu_2^2)/E_2$.

The results presented below have been obtained for a rigid and rough substrate. The atoms in the bottom layer of the block form a simple square lattice with lattice constant a. The lateral dimensions $L_x = N_x a$ and $L_y = N_y a$.

The atoms at the interface between the block and the substrate interact with the potential

$$U(r) = 4\epsilon \left[\left(\frac{r_0}{r}\right)^{12} - \alpha \left(\frac{r_0}{r}\right)^{6} \right] , \tag{15.4}$$

where r is the distance between a pair of atoms. When $\alpha = 1$, Eq. (15.4) is the standard Lennard-Jones potential. The parameter ϵ is the binding energy between two atoms at the separation $r = 2^{1/6} r_0$. When we study contact mechanics without adhesion we put $\alpha = 0$.

15.1.3 Self Affine Fractal Surfaces

In our calculations we have used self-affine fractal surfaces generated as outlined in Ref. [4]. Thus, the surface height is written as

$$h(\boldsymbol{x}) = \sum_{\boldsymbol{q}} B(\boldsymbol{q})\, \mathrm{e}^{i[\boldsymbol{q}\cdot\boldsymbol{x} + \phi(\boldsymbol{q})]} , \tag{15.5}$$

where, since $h(\boldsymbol{x})$ is real, $B(-\boldsymbol{q}) = B(\boldsymbol{q})$ and $\phi(-\boldsymbol{q}) = -\phi(\boldsymbol{q})$. If $\phi(\boldsymbol{q})$ are independent random variables, uniformly distributed in the interval $[0, 2\pi[$, then one can easily show that higher order correlation functions involving $h(\boldsymbol{x})$ can be decomposed into a product of pair correlations, which implies that the height probability distribution $P_h = \langle \delta(h - h(\boldsymbol{x})) \rangle$ is Gaussian [4]. However, such surfaces can have *arbitrary surface roughness power spectrum*. To prove this, substitute (15.5) into (15.1) and use that

$$\langle \mathrm{e}^{i\phi(\boldsymbol{q}')}\, \mathrm{e}^{i\phi(\boldsymbol{q}'')} \rangle = \delta_{\boldsymbol{q}', -\boldsymbol{q}''}$$

gives

$$\begin{aligned} C(\boldsymbol{q}) &= \frac{1}{(2\pi)^2} \int \mathrm{d}^2 x \, \sum_{\boldsymbol{q}'} |B(\boldsymbol{q}')|^2 \, \mathrm{e}^{i(\boldsymbol{q}-\boldsymbol{q}')\cdot\boldsymbol{x}} \\ &= \sum_{\boldsymbol{q}'} |B(\boldsymbol{q}')|^2 \delta(\boldsymbol{q} - \boldsymbol{q}') \, . \end{aligned}$$

Replacing

$$\sum_{\boldsymbol{q}} \rightarrow \frac{A_0}{(2\pi)^2} \int \mathrm{d}^2 q \,,$$

where A_0 is the nominal surface area, gives

$$C(\boldsymbol{q}) = \frac{A_0}{(2\pi)^2} |B(\boldsymbol{q})|^2 \,.$$

Thus, if we choose

$$B(\boldsymbol{q}) = (2\pi/L)[C(\boldsymbol{q})]^{1/2} \,, \tag{15.6}$$

where $L = A_0^{1/2}$, then the surface roughness profile (15.5) has the surface roughness power density $C(\boldsymbol{q})$. If we assume that the statistical properties of the rough surface are isotropic, then $C(\boldsymbol{q}) = C(q)$ is a function of the magnitude $q = |\boldsymbol{q}|$, but not of the direction of $\boldsymbol{q}$.

15.2 Stationary Contact

In this section we illustrate our multiscale molecular dynamics (MD) approach by some applications. We first compare the MD results to two known contact mechanics results from continuum mechanics. Next we discuss contact mechanics for randomly rough surfaces both with and without adhesion.

The randomly rough substrate surfaces used in our numerical calculations where generated using (15.5) and (15.6) and assuming that the surface roughness power spectra have the form shown in Fig. 15.1, with the fractal dimension $D_\mathrm{f} = 2.2$ and the roll-off wavevector $q_0 = 3q_L$, where $q_L = 2\pi/L_x$. We have chosen $q_0 = 3q_L$ rather than $q_0 = q_L$ since the former value gives some self-averaging and less noisy numerical results. We also used $q_1 = 2\pi/b \approx 216 q_0$ (topography (a) in Fig. 15.4), b being the lattice constant of the substrate, and some surfaces with several smaller values for q_1 (Fig. 15.4b shows the topography when $q_1 = 4q_0$), corresponding to lower magnification.

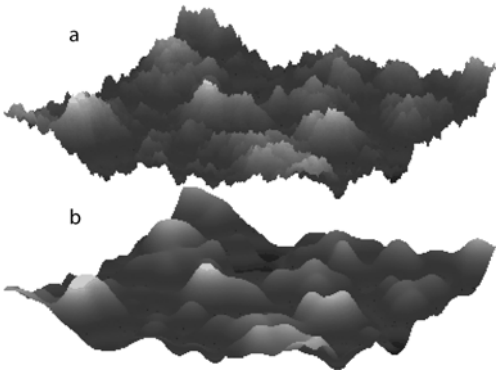


Fig. 15.4. a Fractal surface with the large wavevector cut-off $q_1 = 2\pi/b \approx 216\, q_0$. **b** The same surface as in **a** but at lower resolution with $q_1 = 4q_0$. For a square 1040 Å × 1040 Å surface area. The fractal dimension $D_\mathrm{f} = 2.2$ and the root-mean-square roughness amplitude is 10 Å

For the block, $N_x = 400$ and $N_y = 400$. Periodic boundary conditions are applied in the xy plane. The lateral size of the block is equal to that of substrate, but we use different lattice constant $a \approx b\varphi$, where $\phi = (1 + \sqrt{5})/2$ is the golden mean, in order to avoid the formation of commensurate structures at the interface. The mass of a block atom is 197 a.m.u. and the lattice spacing of the block is $a = 2.6$ Å, reproducing the atomic mass and the density of gold. We consider solid blocks with two different Young's moduli: a hard solid with $E = 77$ GPa, like in gold, and a soft one with 0.5 GPa. The corresponding shear moduli were $G = 27$ GPa and 0.18 GPa, respectively. In the calculations presented below we have used $r_0 = 3.28$ Å and $\epsilon = 18.6$ meV, which (when $\alpha = 1$) gives an interfacial binding energy (per unit area) [14] $\Delta\gamma \approx 4\epsilon/a^2 \approx 11\,\text{meV/Å}^2$.

15.2.1 Test Cases: Hertz Contact and Complete Contact

In 1881 Hertz presented an exact solution for the contact between two perfectly elastic solids with local quadratic profiles. The results were derived using the elastic continuum model and neglecting the adhesion between the solids. In addition, Hertz assumed that the interfacial friction vanishes so that no shear stress can develop at the interface between the solids. When a spherical asperity is squeezed against a flat surface a circular contact area (radius r_H) is formed, where the pressure decreases continuously from the center $r = 0$ to the periphery $r = r_\text{H}$ of the contact according to

$$\sigma = \sigma_\text{H} \left[1 - \left(\frac{r}{r_\text{H}} \right)^2 \right]^{1/2} . \tag{15.7}$$

Let us compare the prediction of our atomistic model with the Hertz theory. We use the Lennard-Jones potential with $\alpha = 0$, i. e. without the attractive term. In Fig. 15.5 we compare the Hertz contact pressure (line) with our numerical data (data points). The numerical data were obtained for a rigid spherical tip squeezed against a flat elastic surface. Note that the pressure obtained from the MD calculation has a tail beyond the Hertz contact radius r_H. Similar "pressure tails" were recently observed in molecular dynamics simulations by Luan and Robbins [15]. The tail reflects the non-zero extent of the atom-atom interaction potential. The deviation between the molecular dynamics results and the continuum mechanics results should decrease continuously as the size of the system increases.

At the atomic level there is no unique way to define when two solids are in contact, and one may use several different criteria. One method is based on the force acting between the atoms at the interface and works best when the adhesive interaction is neglected. Thus, when two surfaces approach each other, the repulsive force between the atoms increases. We may define contact when the repulsive force is above some critical value. When adhesion

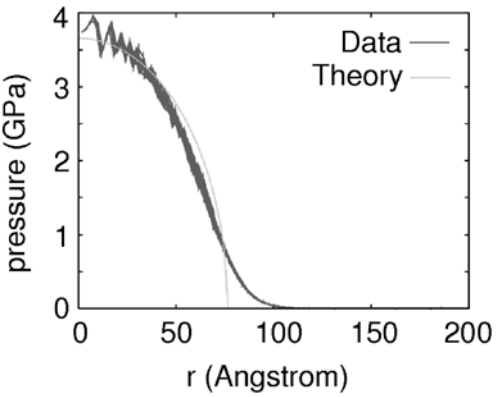


Fig. 15.5. The pressure in the contact region between a spherical tip and a flat elastic surface. We show the simulation data and the theoretical Hertz result. The spherical tip has the radius of curvature $R = 1040\,\text{Å}$ and the loading force $4.6 \cdot 10^{-7}\,\text{N}$

is included the interaction between the wall atoms becomes more long-ranged and it is less obvious how to define contact based on a force criterion, and we find it more convenient to use a criteria based on the nearest neighbor distance between atoms on the two surfaces. Thus, when the separation between two opposing surface atoms is less than some critical value, contact is defined to occur. However, we have found that neither of these two criteria gives fully satisfactory results. The reason is that if the critical force or the critical distance used to define when contact occur is determined by fitting the Hertz pressure profile (15.7) to the numerical data as in Fig. 15.5, then the resulting values depend on the radius of curvature of the asperity. For example, for the Hertz contact in Fig. 15.5 the contact area deduced from the atomistic MD calculation agrees with the Hertz theory if we choose the cut-off pressure $p_c \approx 0.7\,\text{GPa}$. However, if the radius of curvature of the asperity is 10 times smaller ($R = 104\,\text{Å}$) then, for the same penetration, the cut-off would be $p_c \approx 2.5\,\text{GPa}$, i. e., more than three times larger. On the other hand randomly rough surfaces have a wide distribution of curvatures and it is not clear how to choose the optimum cut-off distance or force. In this paper we have therefore used another way of determining the contact area which turned out to be more unique. We will now describe this method.

Let us consider the pressure distribution $P(\sigma, \zeta)$ at the interface. For Hertz contact we get the pressure distribution

$$P(\sigma) = \frac{1}{A_0} \int_{A_0} \mathrm{d}^2x\, \delta(\sigma - \sigma(\boldsymbol{x})) \tag{15.8}$$

Using $\sigma(\boldsymbol{x})$ from (15.7) for $r < r_\text{H}$ and $\sigma(\boldsymbol{x}) = 0$ for $r > r_\text{H}$ gives

$$P(\sigma) = \left(1 - \frac{A}{A_0}\right) \delta(\sigma) + \frac{2\sigma}{\sigma_\text{H}^2} \frac{A}{A_0} \tag{15.9}$$

where $A = \pi r_\text{H}^2$ is the Hertz contact area. In Fig. 15.6 we show the pressure distribution in the contact region between a hard spherical tip and an elastic

solid with a flat surface. The two curves show the simulation data and the theoretical Hertz result obtained by a suitable choice of A in Eq. (15.9). Note that while the Hertz solution and the atomic MD simulation results agree very well for large pressure, there is a fundamental difference for small pressure. Thus, for the Hertz solution, for small pressure $\sigma \to 0$, $P(\sigma) \sim \sigma$, while in the atomistic model $P(\sigma)$ increases monotonically as $\sigma \to 0$. This difference is due to the long-range interaction between the solid walls in the atomistic model, which is absent in the Hertz model. When the long range wall-wall interaction is taken into account the delta function at $\sigma = 0$ in the Hertz solution (15.9) will broaden, resulting in a $P(\sigma)$ which (for the small systems considered here) will decay monotonically with increasing σ as observed for the atomistic model. Note that this effect is of exactly the same origin as the "pressure tail" for $r > r_{\mathrm{H}}$ in Fig. 15.5.

The fact that $P(\sigma, \zeta)$ vanishes linearly with σ as $\sigma \to 0$ is an exact result in continuum mechanics with contact interaction (no long range wall-wall interaction), and is valid not just for the Hertz contact case, but holds in general [16]. However, as explained above, this effect will never be observed in the atomistic model if the wall-wall interaction is long-ranged.

Note that the contact area A can be determined directly by fitting the analytical expression for $P(\sigma)$ for the Hertz contact (Eq. (15.9)) to the numerical MD results for large enough pressure (see Fig. 15.6). In the present case, for $F_{\mathrm{N}} = 4.6 \cdot 10^{-7}$ N (Fig. 15.6a) this gives a contact area $A = \pi r_{\mathrm{H}}^2$ which is

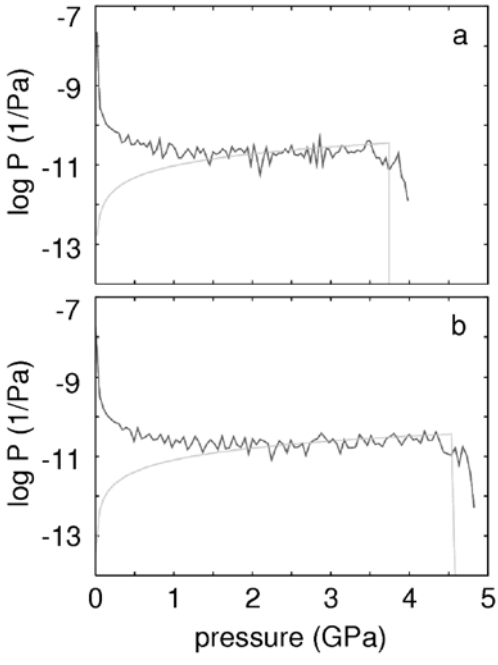


Fig. 15.6. The pressure distribution in the contact region between a spherical tip and a flat surface. We show the simulation data and the theoretical Hertz result. Loading force in **a** is $4.6 \cdot 10^{-7}$ N and in **b** $7.3 \cdot 10^{-7}$ N

nearly identical to the one deduced from the fit in Fig. 15.5. A similar procedure can be used to determine the contact area between randomly rough surfaces using the following analytical expression derived from the contact mechanics theory of Persson (see Eq. (15.10) below):

$$P(\sigma,\zeta) = \frac{1}{2(\pi G)^{1/2}} \left(\mathrm{e}^{-(\sigma-\sigma_0)^2/4G} - \mathrm{e}^{-(\sigma+\sigma_0)^2/4G} \right) ,$$

where σ_0 is the nominal contact stress, and where the fitting parameter $G = G(\zeta)$ can be related to the contact area using Eq. (15.3). Thus, if $A/A_0 \ll 1$ we have $G = (\sigma_0^2/\pi)(A/A_0)^{-2}$. We have found (see below) that this expression for $P(\sigma,\zeta)$ can fit the numerical MD data very well (lending support for the accuracy of the Persson theory), and we have used this method to determine the contact area as a function of the squeezing force for randomly rough substrates.

Let us consider the pressure distribution at the interface between a rigid randomly rough substrate and a flat elastic surface when the solids are in complete contact. Complete contact can result either by squeezing the solids together by high enough force, or if the adhesive interaction between the solids is high enough (or the elastic modulus small enough). However, when complete contact occurs the pressure distribution is the same.

For an elastic solid with a flat surface in perfect contact with a hard randomly rough surface, continuum mechanics predicts a Gaussian pressure distribution of the form [7]:

$$P(\sigma) = \frac{1}{(2\pi)^{1/2}\sigma_{\mathrm{rms}}} \mathrm{e}^{-(\sigma-\sigma_0)^2/2\sigma_{\mathrm{rms}}^2}$$

where the root-mean-square width σ_{rms} is determined by the power spectrum:

$$\sigma_{\mathrm{rms}}^2 = \langle\sigma^2\rangle = \frac{\pi}{2}\frac{E^2}{(1-\nu^2)^2}\int_{q_0}^{q_1} \mathrm{d}q\, q^3 C(q)$$

In Fig. 15.7 we compare the theoretical pressure distribution (dashed-dotted curve) with the pressure distribution obtained from the atomistic model for the case where the complete contact results from the adhesive interaction between the solids. The MD data are well fitted by a Gaussian curve, but the width of the curve is slightly larger than expected from the continuum mechanics theory $\sigma_{\mathrm{rms}}(\mathrm{MD}) = 0.229\,\mathrm{GPa}$ while $\sigma_{\mathrm{rms}}(\mathrm{theory}) = 0.164\,\mathrm{GPa}$. The randomly rough surface used in the MD calculation is self-affine fractal the whole way down to the atomic distance, and one can therefore not expect the continuum mechanics result for $P(\sigma)$, which assumes "smooth" surface roughness, to agree perfectly with the MD result.

15.2.2 Contact Mechanics without Adhesion

Here we study contact mechanics without adhesion as obtained with $\alpha = 0$ in Eq. (15.4), corresponding to purely repulsive interaction between the

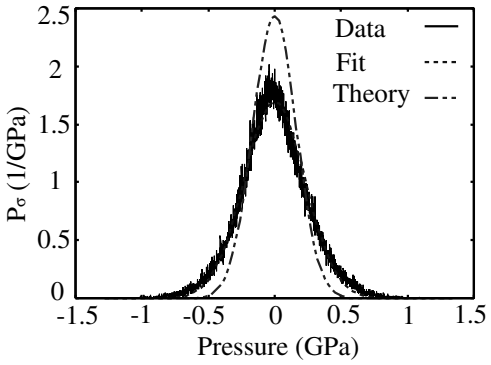


Fig. 15.7. The normalized pressure distribution $P(\sigma)$ at the interface between an elastic block (elastic modulus $E = 0.5$ GPa) with a flat surface and a rigid randomly rough substrate. Because of adhesion complete contact occurs at the interface. The *continuous curve* is the simulation result and the *dotted line* is the Gaussian fit to the simulation data with the root-mean-square width $\sigma_{\rm rms} = 0.229$ GPa. The *dashed-dotted line* is the theoretical Gaussian distribution obtained using continuum mechanics (see Ref. [7]). The theoretical rms width $\sigma_{\rm rms} = 0.164$ GPa

walls. Figure 15.8 shows the contact morphologies at different magnifications ζ for the same load. The red and blue color indicate the contact area at low ($\zeta = 4$) and high ($\zeta = 216$) magnification, respectively. Note that with increasing magnification the contact area decreases, and the boundary line of the contact islands becomes rougher. In Ref. [17] and [18] it has been shown that the statistical properties of the contact regions exhibit power-law scaling behavior. At low magnification ($\zeta = 4$) it looks as if complete contact

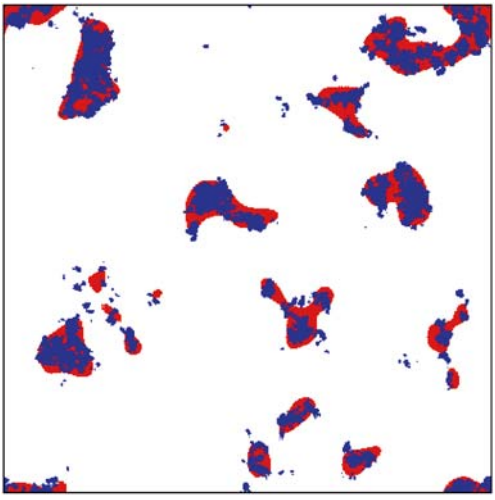


Fig. 15.8. Contact morphology for two different magnifications. The *red color* denotes contact regions for the low magnification $\zeta = 4$, while the *blue color* corresponds to the contact regions for the high magnification $\zeta = 216$

occurs between the solids at asperity contact regions. However, when the magnification is increased, smaller length scale roughness is detected and it is observed that only partial contact occurs at the asperities. In fact, if there were no short distance cut-off in the surface roughness, the true contact area would eventually vanish. But in reality a short distance cut-off always exists, e. g. the interatomic distance.

Figure 15.9 shows the pressure distribution in the contact area for two different magnifications. When we study contact on shorter and shorter length scale, which corresponds to increasing magnification ζ, the pressure distribution becomes broader and broader.

Figure 15.10 shows that the contact area varies (approximately) linearly with the load for the small load at two different magnifications $\zeta = 4$ and 32. The contact area was determined as described in Sec. 15.2.1 by fitting the pressure distribution to a function of the form (15.10). The pressure distributions and the fitting functions are shown in Figs. 15.11 and 15.12 for $\zeta = 4$ and 32, respectively. The slope of the lines in Fig. 15.10 is only a factor 1.14 larger than predicted by the contact theory of Persson (see Sec. 15.2.4).

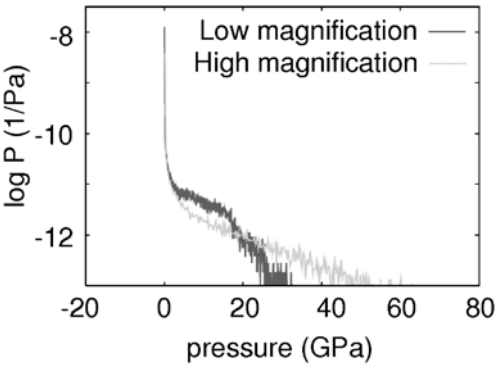


Fig. 15.9. The pressure distribution in the contact area for two different magnifications. The *dark line* corresponds to the pressure distribution for low magnification $\zeta = 4$, while the *light line* is for high magnification $\zeta = 216$

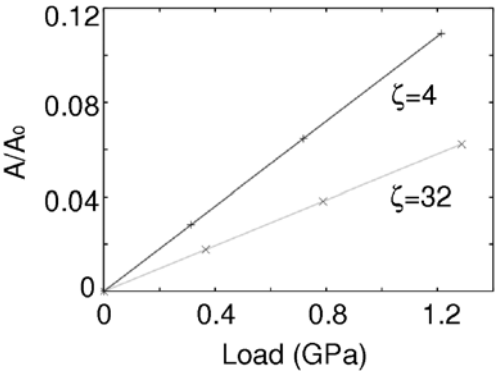


Fig. 15.10. The relative contact area A/A_0, as a function of applied stress F_N/A_0. Results are presented for two different magnifications $\zeta = \lambda_0/\lambda = 4$ and 32. The fractal dimension is $D_f = 2.2$

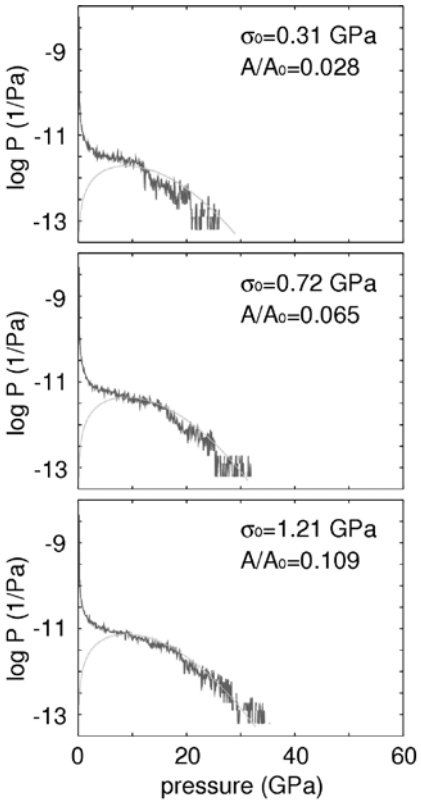


Fig. 15.11. The stress distribution for $\zeta = 4$ for three different nominal pressure

In Fig. 15.13 we show the variation of the contact area with the nominal squeezing pressure for the highest magnification case $\zeta = 216$. In this case we have defined contact to occur when the separation between the surfaces is below some critical value $r_c = 4.3615\,\text{Å}$. In contrast to the definition used above, this definition does not give a strict linear dependence of the contact area on the load for small load as found above when the contact area is defined using the stress distribution.

15.2.3 Contact Mechanics with Adhesion

In this section we include the adhesive interaction i. e. we put $\alpha = 1$ in Eq. (15.4). Figure 15.14 presents the contact morphology both with and without the adhesion at the highest magnification ($\zeta = 216$). The regions with blue color denote the contact area without adhesion. The red color regions denote the *additional* contact area when adhesion is included. The contact area with adhesion is, of course, larger than that without adhesion since the attractive adhesional interaction will effectively increase the external load [19–21].

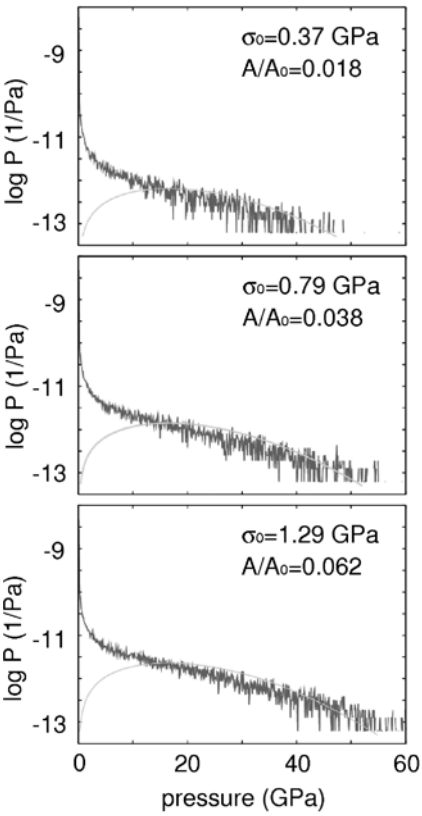


Fig. 15.12. The stress distribution for $\zeta = 32$ for three different nominal pressure

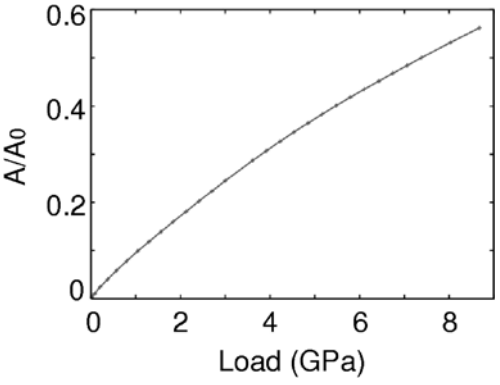


Fig. 15.13. The relative contact area A/A_0, as a function of applied stress F_N/A_0. Results are presented for the highest magnification $\zeta = 216$. Contact is defined when the separation between the surfaces is below a critical value. The fractal dimension is $D_f = 2.2$

Figure 15.15 shows the pressure distribution $P(\sigma, \zeta)$ at high magnification with and without adhesion. When adhesion is neglected (corresponding to the $\alpha = 0$ in (15.4)), the pressure is positive in the contact area and $P(\sigma, \zeta) = 0$ for $\sigma < 0$. When the adhesive interaction is included, the stress becomes tensile close to the edges of every contact region and $P(\sigma, \zeta)$ is in general finite also for $\sigma < 0$.

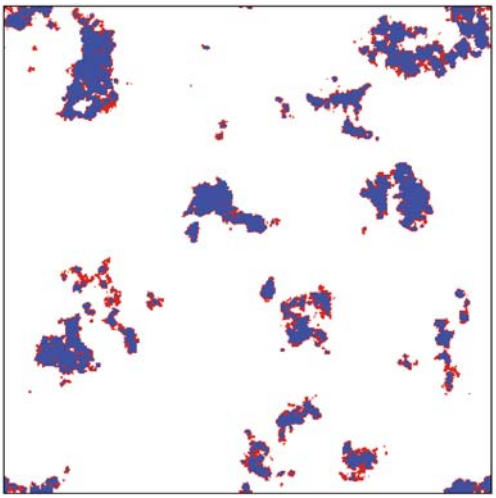


Fig. 15.14. Contact morphology with adhesion and without adhesion. The *blue color region* denotes the contact without adhesion. The *red color region* denote the *additional* contact area when the adhesional interaction is included

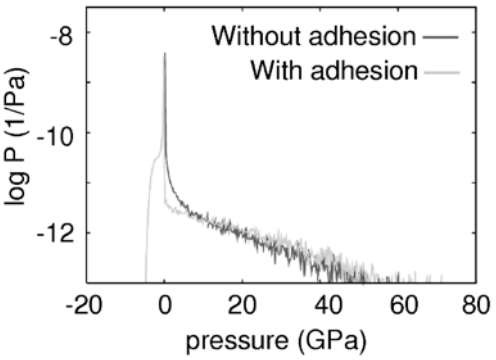


Fig. 15.15. The pressure distribution with and without adhesion. The *light curve* denotes the pressure distribution with adhesion while the *dark curve* is without adhesion

15.2.4 Discussion

Several analytical theories, based on continuum mechanics, have been developed to describe the contact between elastic bodies both with and without the adhesive interaction. Here we will compare the results presented above with the predictions of some of these theories.

Persson [5, 6] has developed a contact mechanics theory where the surfaces are studied at different magnification $\zeta = \lambda_0/\lambda$, where λ_0 is the roll-off wavelength and λ the shortest wavelength roughness which can be observed at the magnification ζ. In this theory [5] the stress distribution $P(\sigma, \zeta)$ at the interface between the block and the substrate has been shown to obey (approximately) a diffusion-like equation where time is replaced by magnification and spatial coordinate by the stress σ. When the magnification is so small that no atomic structure can be detected, the surface roughness will be smooth (no abrupt or step-like changes in the height profile) and one can then show [16] that in the absence of adhesion $P(0, \zeta) = 0$. Using this bound-

ary condition the solution to the diffusion-like equation gives the pressure distribution at the interface ($\sigma > 0$):

$$P(\sigma,\zeta) = \frac{1}{2(\pi G)^{1/2}} \left(\mathrm{e}^{-(\sigma-\sigma_0)^2/4G} - \mathrm{e}^{-(\sigma+\sigma_0)^2/4G} \right) \tag{15.10}$$

where

$$G = \frac{\pi}{4} \left(\frac{E}{1-\nu^2} \right)^2 \int_{q_L}^{\zeta q_0} \mathrm{d}q \; q^3 C(q) \; . \tag{15.11}$$

The relative contact area

$$\frac{A}{A_0} = \int_0^\infty \mathrm{d}\sigma \; P(\sigma,\zeta) \; . \tag{15.12}$$

Substituting (15.10) into (15.12) gives after some simplifications

$$\frac{A}{A_0} = \frac{1}{(\pi G)^{1/2}} \int_0^{\sigma_0} \mathrm{d}\sigma \, \mathrm{e}^{-\sigma^2/4G} \; . \tag{15.13}$$

Thus, for small nominal squeezing pressure $\sigma_0 \ll G^{1/2}$ we get

$$\frac{A}{A_0} = \frac{\sigma_0}{(\pi G)^{1/2}} \; . \tag{15.14}$$

Since the squeezing force $F_\mathrm{N} = \sigma_0 A_0$ we can also write

$$A = \kappa \frac{F_\mathrm{N}}{E^*} \left(\int \mathrm{d}^2 q \; q^2 C(q) \right)^{-1/2} \tag{15.15}$$

where $E^* = E/(1-\nu^2)$ and $\kappa = (8/\pi)^{1/2}$. Thus, for small squeezing force F_N the theory predicts a linear dependence of the area of real contact on the load.

For very high squeezing force $\sigma_0 \gg G^{1/2}$ complete contact will occur at the interface. In this case the second term on the rhs in (15.10) can be neglected, so the pressure distribution is a Gaussian centered at σ_0 and with the root-mean-square width $\sigma_\mathrm{rms} = (2G)^{1/2}$. This result is exact (see Ref. [7]). Thus, the theory of Persson is expected to give a good description of the contact mechanics for all squeezing forces. All other analytical contact mechanics theories are only valid when the squeezing force is so small that the area of real contact is (nearly) proportional to F_N. But in many important applications, e. g., in the context of rubber friction and rubber adhesion, the area of real contact for smooth surfaces is often close to the nominal contact area.

The standard theory of Greenwood and Williamson [22] describes the contact between rough surfaces (in the absence of adhesion), where the asperities are approximated by spherical cups with equal radius of curvature but with Gaussian distributed heights. In this theory the area of real contact dependent (slightly) non-linearly on the load for small load, and can therefore not be

directly compared with the Persson result (15.15). Bush et al. [23] developed a more general and accurate contact theory. They assumed that the rough surface consists of a mean plane with hills and valleys randomly distributed on it. The summits of these hills are approximated by paraboloids, whose distributions and principal curvatures are obtained from the random process theory. As a result of more random nature of the surface, Bush et al. found that at small load the area of contact depends linearly on the load according to (15.15) but with $\kappa = (2\pi)^{1/2}$. Thus the contact area of Persson's theory is a factor of $2/\pi$ smaller than that predicted by Bush. Both the theory of Greenwood and Williamson and the theory of Bush et al. assume that the asperity contact regions are independent. However, as discussed in [16], for real surfaces (which always have surface roughness on many different length scales) this will never be the case even at a very low nominal contact pressure, which may be the origin of difference of $2/\pi$ between Persson's theory and Bush's theory.

Hyun et al. performed a finite-element analysis of contact between elastic self-affine fractal surfaces [18]. The simulations were done for rough elastic surface contacting a perfectly rigid flat surface. They found that the contact area varies linearly with the load for small load. The factor κ was found to be between the results of the Bush and Persson theories for all fractal dimensions $D_{\rm f}$. For $D_{\rm f} = 2.2$ (corresponding to $H = 0.8$) they found that κ was only $\sim 14\%$ larger than predicted by the Persson theory.

The dark curves in Fig. 15.16 shows the pressure distribution from the simulations for several different values of the magnification $\zeta = q_1/q_0 = 4$, 8, 32 and 216, neglecting the adhesion. In the simulations the nominal squeezing pressure $\sigma_0 = 800\,\text{MPa}$. The best fit (light curves in Fig. 15.16) of the analytical pressure distribution given by Eq. (15.10) to the numerical results is obtained if $G^{-1/2}$ is taken to be a factor 1.14 larger than predicted by the Persson theory via Eq. (15.11), corresponding to a contact area which is 14% larger than predicted by the analytical theory, in good agreement with the results obtained by Hyun et al.

Our simulations show that the contact area varies linearly with the load for the small load, see Fig. 15.10. Figures 15.10 and 15.16 show that the slope $\alpha(\zeta)$ of the line $A = \alpha(\zeta)F$ decreases with increasing magnification ζ, as predicted by the analytical theory [16, 17]. Thus, while $A/A_0 = 0.072$ for $\zeta = 4$ we get $A/A_0 = 0.038$ for $\zeta = 32$, which both are 14% larger than predicted by Eq. (15.13).

15.2.5 On the Philosophy of Contact Mechanics

Recently it has been argued that contact mechanics models based on continuum mechanics cannot be used to determine the pressure distribution at the interface with atomistic resolution, or the adhesion or friction between the contacting solids, since these properties depend on atomic scale surface

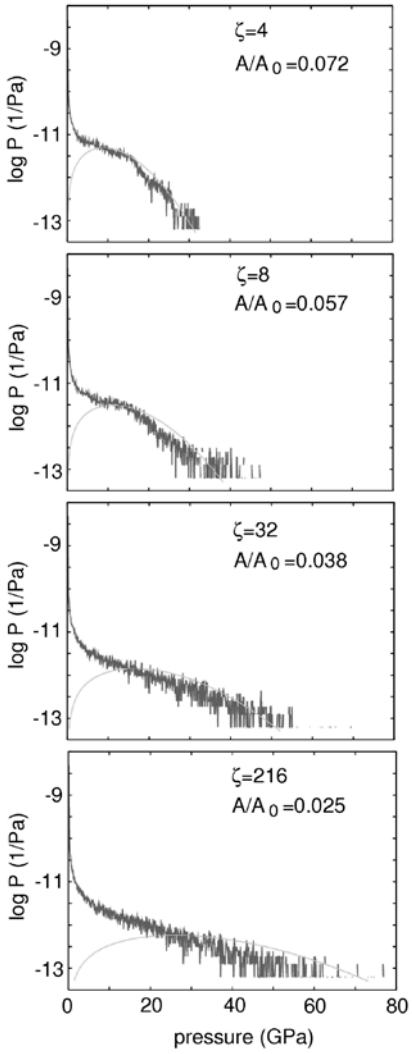


Fig. 15.16. The pressure distribution at four different magnifications $\zeta = q_1/q_0 =$ 4, 8, 32 and 216 for the squeezing pressure $\sigma_0 = 800$ MPa. The *dark curves* are the pressure distributions obtained from the computer simulation, while the *light curves* come from the analytical theory assuming that $G^{-1/2}$ (and hence the relative contact area) is a factor of 1.14 larger than the value predicted by the analytical theory

roughness (surface adatoms, steps, and so on) [15]. But the aim of continuum mechanics was never to describe the physics at the atomistic level but only at larger length scales. Thus, a fundamental approach to the problems mentioned above consists of applying continuum mechanics down to a length scale of the order of a few nanometer. At this length scale or resolution continuum mechanics will correctly predict the contact area and the pressure distribution. At shorter length scale other methods such as molecular dynamics or quantum mechanical methods must be employed, taking into account the atomistic and chemical nature of real surfaces. However, in most practical applications such a detailed approach will not be possible because

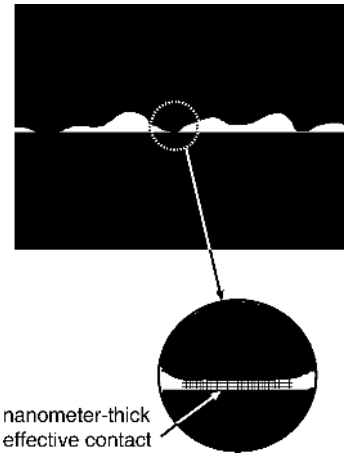


Fig. 15.17. A fundamental approach to contact mechanics, adhesion and friction consist of applying continuum mechanics down to a length scale of the order of a few nanometer. At this length scale or resolution continuum mechanics will correctly predict the contact area and the pressure distribution. At shorter length scale other methods such as molecular dynamics or quantum mechanical methods must be employed taking into account the atomistic and chemical nature of real surfaces

too little is usually known about the actual interface, e. g., thin (nanometer or less) contamination layers may dominate the physics at the atomistic level. Nevertheless, even in these cases the continuum mechanics approach may be extremely useful as it determines the size and the pressure distribution in the *effective* contact areas where most of the interesting physics takes place, see Fig. 15.17.

15.2.6 Comment on Numerical Studies of Contact Mechanics

Many numerical studies based on the finite element method, or other numerical methods, have been presented for the elastic or elastoplastic contact between randomly rough (e. g., self-affine fractal) surfaces. However, numerical methods cannot be used to study the influence of roughness on the contact mechanics for macroscopic systems involving about 7 decades in length scales (from nanometers to centimeters). The required computational time and memory space scale at least linearly with the number of grid points $N \propto (L/a)^2$ in the xy-plane, where a is the grid size and L the linear size of the system studied. Thus the computational cost is at least quadratic with the size of the system (such quadratic behaviour can be achieved with a multiscale approach, see e. g., Ref. [7]). At present reliable numerical studies are limited to systems with 2 or maximum 3 decades in length scale. Nevertheless, numerical studies even on relatively small systems may be useful to test the accuracy of analytical contact mechanics theories.

All numerical studies of contact mechanics for randomly rough surfaces we are aware of are *non-converged* in the following sense. Assume that the wavelength of the shortest surface roughness component is λ_1. In order to properly describe contact mechanics it is necessary to have at least 10×10 grid points and preferably much more (maybe 100×100) within one asperity contact region which may be smaller than, or of the order of $\lambda_1 \times \lambda_1$. Only

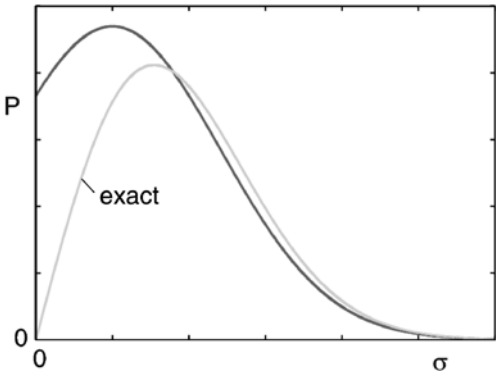


Fig. 15.18. The exact and approximate stress distribution $P(\sigma, \zeta)$. The exact $P(\sigma, \zeta)$ vanish for $\sigma = 0$, but this result is not observed in numerical calculations owing to a too sparse integration grid (see text for details). The two stress distribution functions correspond to the same load, but the approximate stress distribution gives 22% larger area of contact

when this smallest Hertzian-like contact problem is accurately described one can obtain accurate results for the area of real contact. It is easy to show that when too few grid points are used, the numerical study will give an overestimation of the area of real contact. The reason for this is that the true stress distribution vanishes at zero stress (see Sec. 15.2.4), and in order to obtain this result (approximately) in numerical studies it is necessary to have such a high density of grid points that the contact region close to the non-contact area is accurately treated (it is this region which determine the boundary condition $P(0, \zeta) = 0$). In all reported studies we are aware of this is not the case and the stress distribution remains finite (and large) at $\sigma = 0$ as indicated in Fig. 15.18. However, since the applied load $F_{\rm N}$ is determined by the first moment of the pressure distribution,

$$A_0 \int_0^\infty {\rm d}\sigma \; \sigma P(\sigma, \zeta) = F_{\rm N}$$

it follows that the numerically determined $P(\sigma, \zeta)$ curve must be below the exact stress distribution curve for larger σ in order for $F_{\rm N}$ to be the same. But since the area of real contact A is given by

$$A = A_0 \int_0^\infty {\rm d}\sigma \; P(\sigma, \zeta)$$

it follows that the numerically determined contact area will be larger than the actual one.

15.3 Sliding Contact

15.3.1 Introduction

At the most fundamental level friction is (almost) always due to elastic instabilities at the sliding interface. At low sliding velocity an elastic instability

first involves (slow) elastic loading, followed by a rapid rearrangement, where the speed of the rearrangement is much faster than, and unrelated to, the loading (or sliding) velocity. During the fast rearrangement the elastic energy gained during the loading phase, is converted into irregular heat motion. The exact way of how the energy is "dissipated" has usually a negligible influence on the sliding friction force, assuming that the dissipation occurs so fast that no memory of it remains during the next elastic loading event. There are many possible origins of elastic instabilities, e. g., it may involve individual molecules or, more likely, group of molecules or "patches" at the interface which we have denoted by stress domains [24–27]. The most fundamental problem in sliding friction is to understand the physical origin and nature of the elastic instabilities.

Elastic instabilities occur only if the lateral corrugation of the interaction potential between the solid walls is high enough, or the elastic modulus of the solids small enough. Roughly speaking, elastic instabilities can only occur if a characteristic elastic energy is smaller than a characteristic binding (or rather barrier height) energy. To understand this, consider the simple model illustrated in Fig. 15.19. In (a) a particle or atom is moving in a corrugated (substrate) potential. Connected to the particle there is a spring (spring constant k) which is pulled with the velocity v. If the spring is soft enough, or the potential barrier height U is high enough, i. e., $U > ka^2$, the particle will perform stick-slip motion [Fig. 15.19a,b], involving slow elastic loading followed by rapid slip and dissipation of the (elastic) spring energy. In this case the (time averaged) force on the particle is independent of v. However, in the opposite case $U < ka^2$ [Fig. 15.19c], the particle will follow the drive with a velocity which is always comparable to v. In particular, when the drive is on-top of the barrier so will the particle be [Fig. 15.19c]. In this case no rapid motion will occur and the (time averaged) friction force acting on the particle is proportional to v.

In a more realistic situation one must consider the whole interface. In this case, depending on the elasticity and lateral barriers and the size of the contact area, elastic instabilities may or may not occur [28]. Assume first that an elastically very stiff solid slides on a rigid corrugated substrate, [Fig. 15.19d]. In this case the atoms at the bottom surface cannot adjust to the corrugated substrate potential, and (for an incommensurate system) as some atoms move downhill other atoms move uphill in such a way that the total energy is constant. Thus, no elastic instabilities will occur during sliding, resulting in a very low sliding friction; this state has been termed *superlubric* [29]. However, when the block is elastically soft [Fig. 15.19e], the atoms can rearrange themselves so that at any moment in time almost all the atoms occupy positions close to the minima of the substrate potential. During sliding rapid jumps will occur from time to time where a particle changes potential well. In this case the friction is high and (at zero temperature) remains finite as the sliding velocity $v \to 0$.

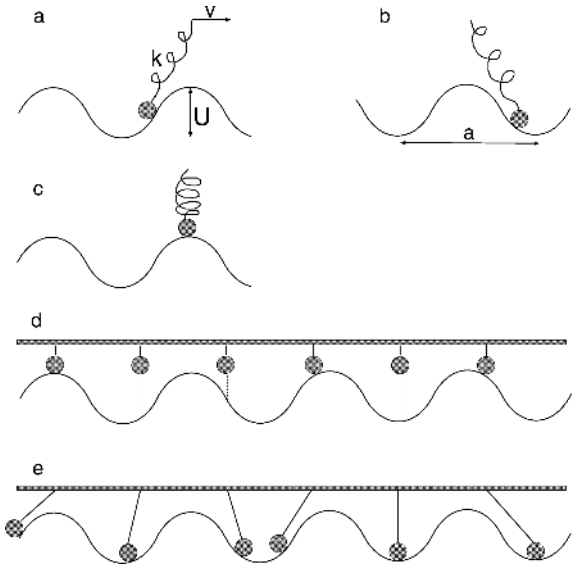


Fig. 15.19. a A particle or atom moving in a corrugated potential. The particle is connected to a spring (spring constant k) which is pulled with the velocity v. **b** If the spring is soft enough, or the potential barrier U is high enough, i. e., $U > ka^2$, the particle will perform stick-slip motion, involving slow elastic loading followed by rapid slip and dissipation of (elastic) spring energy. In this case the (time averaged) force on the particle is independent of v. **c** In the opposite case $U < ka^2$ the particle will follow the drive with a velocity which is always comparable to v. In particular, when the drive is on-top of the barrier so will the particle be. In this case no rapid motion will occur and the friction on the particle is proportional to v. **d** When an elastically stiff solid slides on a rigid corrugated substrate, the atoms at the bottom surface cannot adjust to the corrugated substrate potential. For an incommensurate system, as some atoms move downhill other atoms move uphill in such a way that the total energy is constant. **e** When the block is elastically soft, the atoms can rearrange so that at any moment in time almost all the atoms occupy positions close to the minima of the substrate potential. During sliding rapid jumps will occur from time to time where a particle changes well. In this case the friction is high and (at zero temperature) remains finite as the sliding velocity $v \to 0$

It is well known that elastically hard solids tend to exhibit smaller sliding friction than (elastically) soft materials [30]. One extreme example is diamond which under normal circumstances exhibits very low kinetic friction coefficient, of the order of 0.01, when diamond is sliding on diamond. This can be explained by the nearly absence of elastic instabilities because of the elastic hardness of the material. However, if clean diamond is sliding on clean diamond in ultrahigh vacuum, a huge friction (of the order of $\mu \approx 10$) is observed [31]. The reason is that the clean surfaces have dangling bonds (which are passivated by hydrogen and oxygen in the normal atmosphere) so that

the interaction between the two diamond surfaces is very strong and elastic instabilities (and wear processes) can occur resulting in a very large friction.

It is important to note that even if solids are too stiff for elastic instabilities to occur on short length scale, the ratio between the effective elasticity and the amplitude of the lateral corrugation of the binding potential may decrease when the system is studied at a longer length scale, which may make elastic instabilities possible on a longer length scale [32,33]. To illustrate this, in Fig. 15.20 we show a one-dimensional (1D) case, where an elastic bar (cross section area A) with two binding sites (bumps) is in contact with a substrate with two binding sites (cavities). When the binding sites on the elastic bar overlap with the binding sites (cavities) on the (rigid) substrate, the binding energy U is gained. In order to gain this binding energy the segment of the elastic bar between the bumps (length L) must elongate by the distance b. Thus the strain in the segment is b/L and the elastic energy stored in the elongated segment is $U_{\rm el} = VE(b/L)^2/2$ where the volume $V = LA$. Thus, $U_{\rm el} = AEb^2/(2L)$ which decreases as the length of the segment L increases. It follows that only when $L > AEb^2/2U$ will the bound state have a lower energy than the non-bound state. Thus, only on large enough length scale will the solid be elastically soft enough for elastic instabilities to occur. In most practical cases one is not interested in a 1D situation but rather in semi-infinite solids, which are intermediate between the 2D and 3D case. For surfaces with randomly distributed binding centers this situation is much more complex than for the 1D case because the effective elasticity changes as quickly with the lateral length scale as does the effective amplitude of the lateral corrugation of the binding potential [32,33] (which from random walk arguments [1] scales as L). A detailed analysis of this situation indicates, however, that if no elastic instability can occur at short length scale it is very unlikely that elastic instabilities will occur on any length scale of practical importance, except perhaps in the context of earthquakes [32,33]. If instead

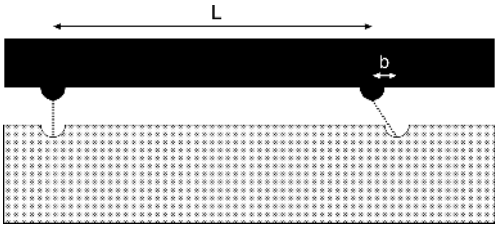


Fig. 15.20. An elastic bar (cross section area A) with two binding sites (bumps). When the binding sites overlap the binding sites (cavities) on the (rigid) substrate, the binding energy U is gained. In order to gain this binding energy the segment of the elastic bar between the bumps (length L) must elongate by the distance b. Thus the strain in the segment is b/L and the elastic energy stored in the elongated segment is $U_{\rm el} = VE(b/L)^2/2$ where the volume $V = LA$. Thus, $U_{\rm el} = A\ Eb^2/(2L)$ which decreases as the length of the segment L increases

of randomly distributed binding sites one assumes incommensurate surfaces, one would expect even weaker pinning effects, and it can be argued that in this case the ratio between the effective elasticity and the amplitude of the lateral corrugation of the binding potential *increases* as $\sim L$ so that if no elastic instabilities occur at short length scale they cannot occur at any length scale [34]. Below we will present numerical results where elastic instabilities do occur also for (nearly) incommensurate structures, but in these cases one of the solids is elastically very soft so that instabilities can occur on a short length scale.

The discussion above has focused on clean surfaces and zero temperature. Temperature is unlikely to have any drastic influence on superlubricity. However, it may have a strong influence on sliding dynamics when elastic instabilities occur. As soon as $T > 0\,\mathrm{K}$, thermal noise is able to activate jumps over the barrier, i. e., to provoke *premature jumps* before the (zero temperature) instability point is reached. It has been shown experimentally [35, 36] and theoretically [37, 38] that this has a crucial influence on friction dynamics at low sliding velocity. Similarly, weakly bound adsorbed atoms and molecules have a large influence on the sliding dynamics, and may strongly increase the friction force [39] as the mobile adsorbates can adjust themselves in the corrugated potential between the block *and* the substrate, giving rise to strong pinning effects. Here we will not address the role of adsorbates or non-zero temperature, but we will focus on the simplest case of clean surfaces at zero temperature.

Recently, superlubricity has been observed during sliding of graphite on graphite [40]. In the experiment a tungsten tip with a graphite flake attached to it was slid on an atomically flat graphite surface. When the flake is in registry with the substrate stick-slip motion and large friction are observed. When the flake is rotated out of registry, the forces felt by the different atoms start to cancel each other out, causing the friction force to nearly vanish, and the contact to become superlubric.

Graphite and many other layered materials are excellent dry lubricants. The most likely reason for this is that the solid walls of the sliding objects get coated by graphite flakes or layers with different orientation so a large fraction of the graphite-graphite contacts will be in the superlubric state. This will lead to a strong reduction in the average friction. However, the coated solid walls are unlikely to be perfectly flat and it is important to address how surface roughness may influence the superlubric state. In this paper we will show that even a relatively small surface roughness may kill the superlubric state.

Lubrication by graphite flakes may even occur for diamond-like carbon (DLC) coatings, which may exhibit very low friction. Thus, Liu et al. [41] have observed that a graphitized tribolayer is formed on top of diamond-like carbon coatings. Thus, also the excellent lubrication properties of DLC films might be caused by superlubric graphite contacts. We also note that DLC films are very hard which will reduce the chance for elastic instabilities to occur [42].

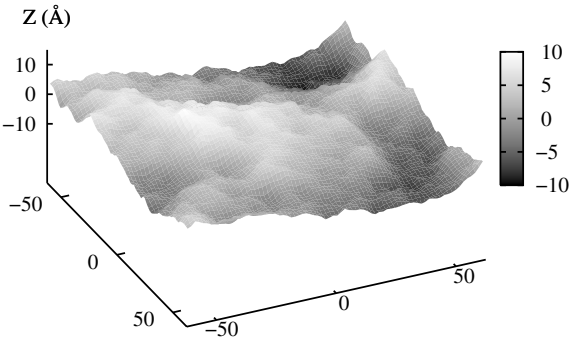


Fig. 15.21. Fractal surface of the substrate with the large wavevector cutoff $q_1 = 2\pi/b = 78\,q_0$. For a square 124.8 Å × 124.8 Å surface area. The fractal dimension $D_\mathrm{f} = 2.2$ and the root-mean-square roughness amplitude is 3 Å

We will now present atomistic Molecular Dynamics calculations of the sliding dynamics for contacting elastic solids with (nearly) incommensurate surface lattice structures. We consider both flat surfaces and surfaces with roughness. We consider the dependence of the sliding friction on the elastic modulus of the solids. For elastically hard solids with planar surfaces with incommensurate surface structures we observe extremely low friction (superlubricity), which very abruptly increases as the elastic modulus decreases. We show that even a small surface roughness may completely kill the superlubricity state. In order to study large systems we use the multiscale approach [7] to contact mechanics described in Sect. 15.1.

The randomly rough substrate surfaces used in our numerical calculations were generated using (15.5) and (15.6) and assuming that the surface roughness power spectra have the form shown in Fig. 15.1, with the fractal dimension $D_\mathrm{f} = 2.2$ and the roll-off wavevector $q_0 = 3q_L$, where $q_L = 2\pi/L_x$. We have chosen $q_0 = 3q_L$ rather than $q_0 = q_L$ since the former value gives some self-averaging and less noisy numerical results. We also used $q_1 = 2\pi/b = 78q_0$. The topography of the substrate with the root-mean-square roughness amplitude 3 Å used in our numerical calculations is shown in Fig. 15.21.

As an illustration, in Fig. 15.22 we show the contact between a flat elastic block (top) and a randomly rough *rigid* substrate (bottom). Only the interfacial block and substrate atoms are shown. The substrate is self-affine fractal with the root-mean-square roughness 3 Å. Note the elastic deformation of the block, and that non-contact regions occur in the "deep" valleys of the substrate. Thus, the real contact area is smaller than the nominal contact area.

15.3.2 Numerical Results

In this section we present the results of molecular dynamics calculations of sliding of elastic blocks on rigid substrates. In all cases, unless otherwise stated, the upper surface of the block moves with the velocity $v = 0.1\,\mathrm{m/s}$,

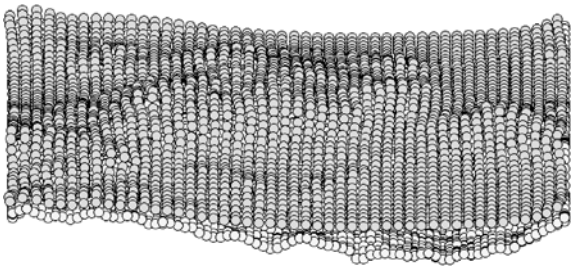


Fig. 15.22. The contact between an elastic block with a flat surface and a rough *rigid* substrate. Only the interfacial layers of atoms are shown. The elastic modulus of the block was $E = 100\,\mathrm{GPa}$. The substrate is self-affine fractal with the root-mean-square roughness 3 Å, fractal dimension $D_\mathrm{f} = 2.2$ and roll-off wave vector $q_0 = 3q_L$, where $q_L = 2\pi/L_x$. The substrate and block interfacial atomic layers consisted of 78×78 and 48×48 atoms, respectively. The applied pressure $p = 10\,\mathrm{GPa}$. Note the elastic deformation of the block, and that the real contact area is smaller than the nominal contact area

and the (nominal) squeezing pressure p is one tenth of the elastic modulus E of the block, i. e., $p = 0.1E$. The reason for choosing p proportional to E is twofold: First, we consider solids with elastic modulus which varies over several order of magnitudes, and it is not possible to use a constant p as this would result in unphysical large variations in the elastic deformation of the block. Second, if two elastic solids are squeezed together with a given load, then as long as the area of real contact is small compared to the nominal contact area, the pressure in the contact areas will be proportional to the elastic modulus of the solids [5].

Let us first assume that both the block and the substrate have atomically smooth surfaces. Figure 15.23 shows the static and the kinetic friction coefficients as a function of the elastic modulus E of the block. Note the

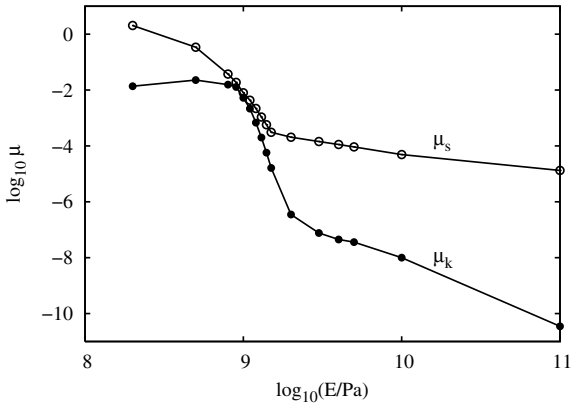


Fig. 15.23. The static and kinetic friction coefficients as a function of the elastic modulus E of the block, for the flat substrate. In the calculation we have assumed the squeezing pressure $\sigma_0 = 0.1E$ and the sliding velocity $v = 0.1\,\mathrm{m/s}$

relatively abrupt decrease in the friction when the elastic modulus changes from $E_1 \approx 0.7\,\mathrm{GPa}$ to $E_2 \approx 2\,\mathrm{GPa}$. For $E > E_2$ practically no instabilities occur and the friction is extremely small, while for $E < E_1$ relatively strong elastic instabilities occur at the sliding interface, and the friction is high. For $E = 0.2\,\mathrm{GPa}$ the static friction $\mu_\mathrm{s} > 2$. This calculation illustrates that the transition from high friction to *superlubricity* can be very abrupt; in the present case an increase in the elastic modulus by only a factor 3 (from 0.7 to 2.1 GPa) decreases the kinetic friction by a factor $\sim 10^{-5}$.

In Fig. 15.24 we show the variation of the shear stress as a function of time when the elastic modulus of the block equals (a) $E = 0.8\,\mathrm{GPa}$ and (b) $E = 2\,\mathrm{GPa}$. The elasticity of the stiffer solid is above the superlubricity threshold, and no (or negligible) elastic instabilities occur, and the stress is a periodic function of time, with the period corresponding to the displacement 0.2 Å. For the softer solid strong elastic instabilities occur during sliding, and the shear stress is less regular (and the arrangement of the interfacial block atoms more disordered) than for the stiffer solid, and the (average) period is *longer* than 0.2 Å.

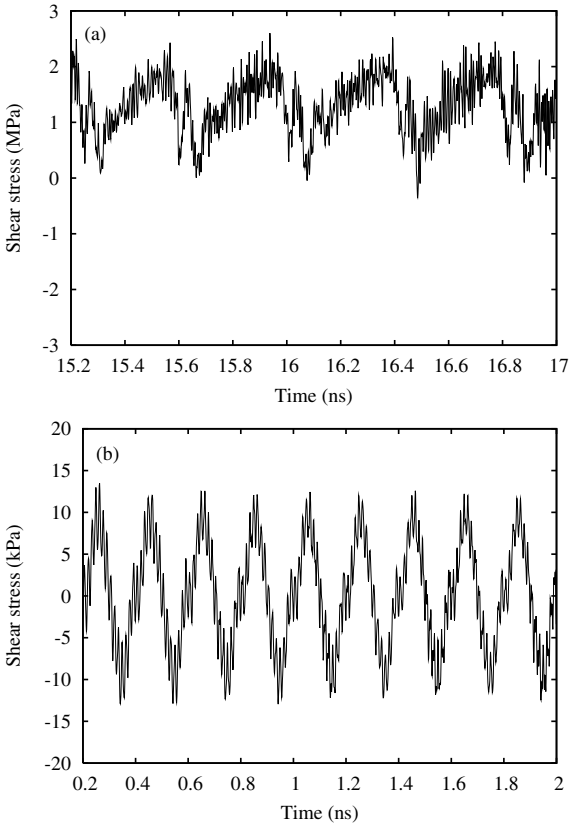


Fig. 15.24. a The shear stress as a function of time for the flat substrate. The elastic modulus of the block was $E = 0.8\,\mathrm{GPa}$. **b** The same as **a** but for the elastic modulus of the block $E = 2\,\mathrm{GPa}$

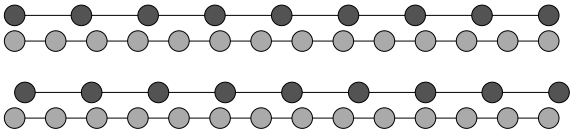


Fig. 15.25. Commensurability ratio 8/13 between 1-dimensional chains. The upper image shows the period of 8 block atoms, i. e., 13 substrate atoms. The lower image is obtained by shifting the block for 1/13 of its lattice spacing. Block atoms occupy the same positions relatively to the substrate's hollows

This result in Fig. 15.24b can be understood as follows. For our system, in the sliding direction there are 8 block atoms for every 13 substrate atoms. Assume first that the block (and the substrate) are perfectly stiff. In this case, the position of the 8 block atoms will take 8 uniformly spaced positions within the substrate unit cell (lattice constant b), see Fig. 15.25. Thus, a shift of the block with the distance $b/8$ will take the system to a geometrically equivalent configuration. Hence, since $b = 1.6\,\text{Å}$ we expect the periodicity of the shear stress to be $b/8 = 0.2\,\text{Å}$. When the block has a finite elasticity but above the superlubricity threshold, the atoms will relax somewhat in the substrate potential, but the configuration of the system will still repeat itself with the same period $b/8$.

Let us now consider the influence of surface roughness on the sliding dynamics. In Fig. 15.26 we show the kinetic friction coefficients for an elastic block sliding on a rough substrate, as a function of the logarithm of elastic modulus E of the block. The curves from top to bottom correspond to the substrate root-mean-square roughness amplitudes $3, 1, 0.3, 0.1\,\text{Å}$ and 0 (flat substrate). For the substrate with the largest roughness, no superlubricity state can be observed for any elastic modulus up to $E = 10^{12}\,\text{Pa}$.

In Fig. 15.27 we show the kinetic friction coefficient as a function of the root-mean-square roughness amplitude of the substrate. The elastic modulus

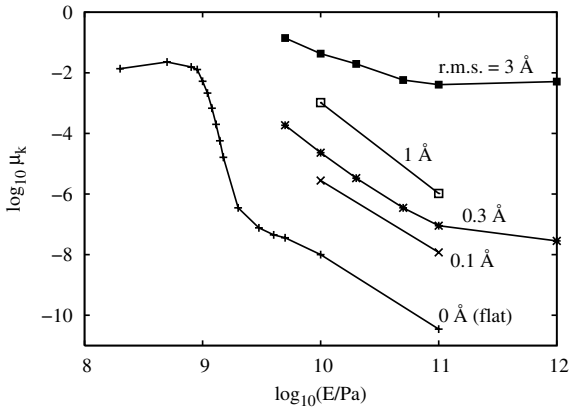


Fig. 15.26. The kinetic friction coefficients for an elastic block sliding on rough substrates, as a function of the logarithm of elastic modulus E of the block. The curves correspond to the root-mean-square roughness amplitudes of the fractal substrate $3, 1, 0.3, 0.1\,\text{Å}$ and 0 (flat substrate)

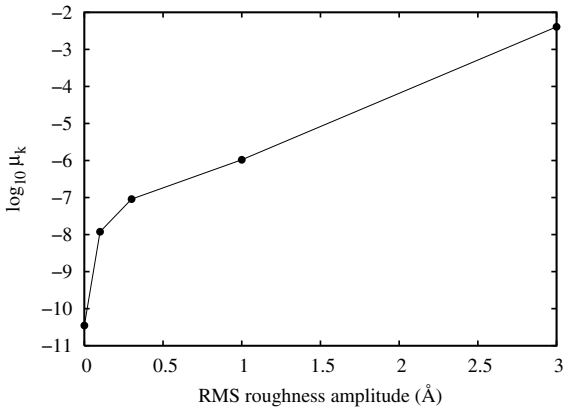


Fig. 15.27. The kinetic friction coefficient as a function of the root-mean-square roughness amplitude of the substrate. The elastic modulus of the block $E = 100\,\mathrm{GPa}$

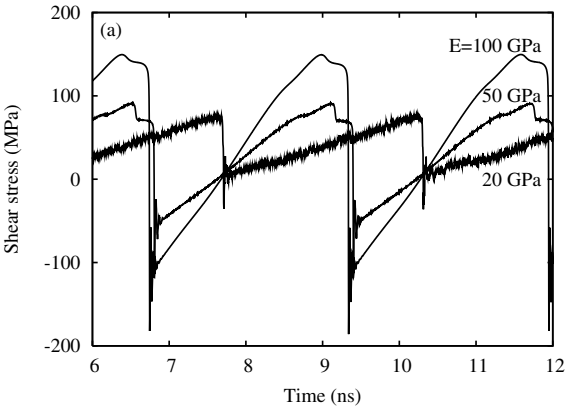
of the block $E = 100\,\mathrm{GPa}$. Note the strong decrease in the friction when the root-mean-square roughness amplitude decreases below $0.3\,\mathrm{Å}$, which corresponds to a peak-to-peak roughness of roughly one atomic lattice spacing.

Figure 15.28a shows the average (or nominal) shear stress as a function of time for the rough substrate with the root-mean-square roughness $3\,\mathrm{Å}$, and for the elastic modulus of the block $E = 100, 50$ and $20\,\mathrm{GPa}$. Note that in addition to major slip events, several small slip events occur in all cases. These events correspond to local slip at some asperity contact regions before the major slip involving the whole contact area. In all cases, the time dependence of the shear stress remains periodic with the period $2.6\,\mathrm{Å}$, which corresponds to the lattice spacing of the block. Note also that for the elastically softer block ($E = 20\,\mathrm{GPa}$), the stress-noise increases after each major slip event; this is caused by the elastic waves (heat motion) excited during the (major) rapid slip events.

Figure 15.28b shows the same as in (a) but now for the elastic modulus of the block $E = 10$ and $5\,\mathrm{GPa}$. In this case the decrease of the elastic modulus of the block results in the increase of both the static and kinetic friction.

Figure 15.29 shows the average displacement of the interfacial atoms of the block (in the sliding direction) as a function of time. The root-mean-square roughness amplitude for the substrate was $3\,\mathrm{Å}$. The elastic modulus of the block was $E = 100, 10$ and $5\,\mathrm{GPa}$. Note that the slip distance for the major slip events increases as the elastic modulus of the block decreases, and that for the elastically hardest solid ($E = 10^{11}\,\mathrm{Pa}$) about half of the forward displacement occurs between the major slip events.

Figure 15.30 shows the average position of the interface block atoms in the z-direction (perpendicular to the sliding direction), as a function of time. Results are shown for the rough substrate with the root-mean-square roughness amplitude $3\,\mathrm{Å}$. The elastic modulus of the block was $E = 100, 50, 20, 10$ and $5\,\mathrm{GPa}$. When the elastic modulus decreases, because of the adhesive interaction the block interfacial atoms come (on the average) more close to the



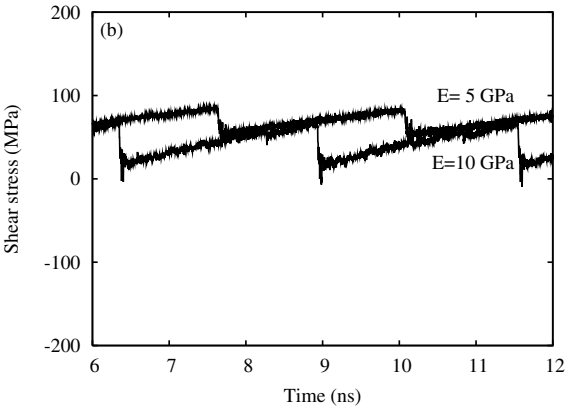


Fig. 15.28. a The shear stress as a function of time for the rough substrate with the root-mean-square roughness amplitude 3 Å. The elastic modulus of the block is $E = 100, 50$ and 20 GPa. **b** The same as **a** but for the elastic modulus of the block $E = 10$ and 5 GPa

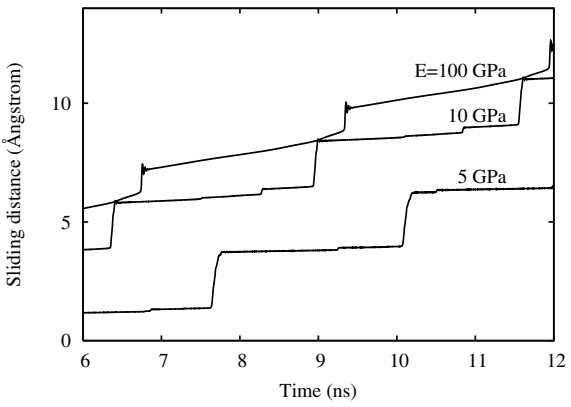


Fig. 15.29. The average displacement of the interfacial atoms of the block as a function of time. The root-mean-square roughness amplitude for the substrate is 3 Å. The elastic modulus of the block is $E = 100, 10$ and 5 GPa

rigid substrate, embracing the substrate asperities. This increases the real area of contact between the surfaces and result in a higher friction.

Figure 15.31a shows the shear stress as a function of time for the rough substrate (root-mean-square amplitude 3 Å) and for the stiff block

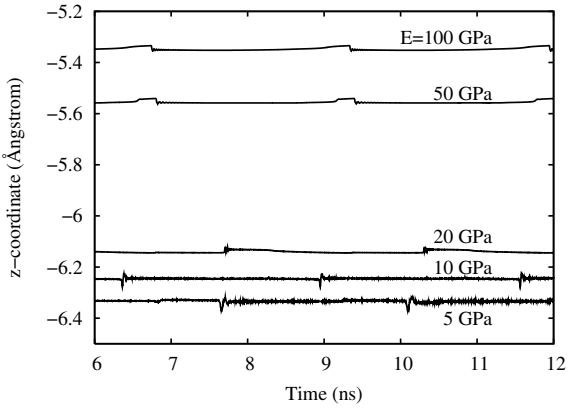


Fig. 15.30. The average position of the interface block atoms in the z-direction (perpendicular to the sliding direction), as a function of time during sliding. The root-mean-square roughness amplitude of the substrate is 3 Å. The curves correspond to the elastic modulus of the block $E = 100, 50, 20, 10$ and 5 GPa

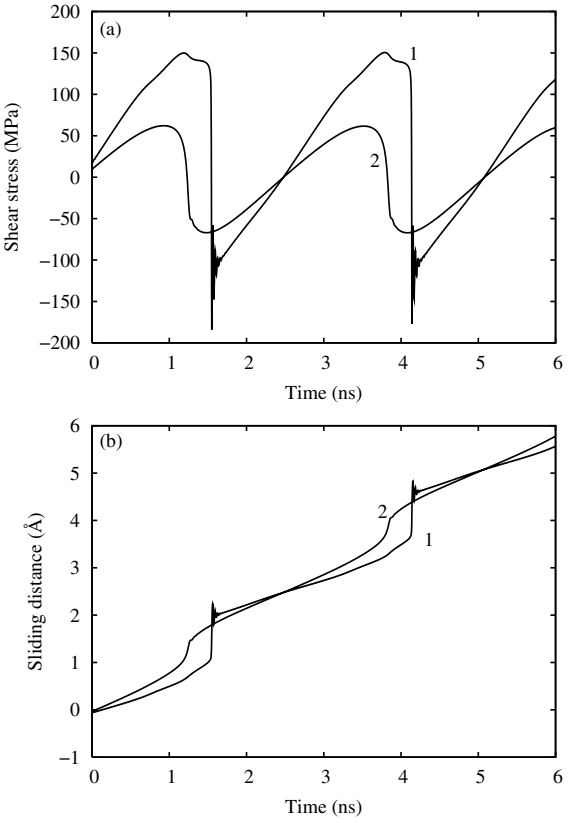


Fig. 15.31. a The shear stress as a function of time calculated for the rough substrate (root-mean-square amplitude 3 Å) and for the stiff block ($E = 100$ GPa). The *curve 1* is including adhesion, while the *curve 2* is obtained without the attractive part in the Lennard-Jones potential, i. e., with $\alpha = 0$ in Eq. (15.4). **b** The average displacement of the interface block atoms (in the sliding direction) as a function of time for the same systems as in **a**

($E = 100$ GPa). The curve 1 is with the adhesion included, while the curve 2 is without the attractive part in the Lennard-Jones potential, i. e., with $\alpha = 0$ in Eq. (15.4). Note that without adhesion the major slip is not so pronounced as for the case with adhesion. Still the time dependence of the shear stress

remains periodic with the same period 2.6 Å, corresponding to the lattice spacing of the block. Without adhesion, the shear stress curve is nearly symmetric around the zero-stress axis, and the kinetic friction coefficient (determined by the average shear stress divided by the squeezing pressure) is about 150 times smaller than when the adhesive interaction is included.

In Fig. 15.31b we show the average displacement of the interface block atoms (in the sliding direction) as a function of time for the same systems as in (a). For the case without adhesion the major slip is not as abrupt as when adhesion is included. At every moment there is some lateral motion of the block interfacial atoms.

Figure 15.32a shows the shear stress as a function of time for the rough substrate (root-mean-square amplitude 3 Å) and for the elastic block with elastic modulus $E = 10\,\text{GPa}$. The curve 1 is with adhesion included, while the curve 2 is without the attractive part in the Lennard-Jones potential, i. e., with $\alpha = 0$ in Eq. (15.4). Figure 15.32(b) shows the average displacement of the interface block atoms (in the sliding direction) as a function of time for the same systems as in (a). For the case without adhesion the major slip is

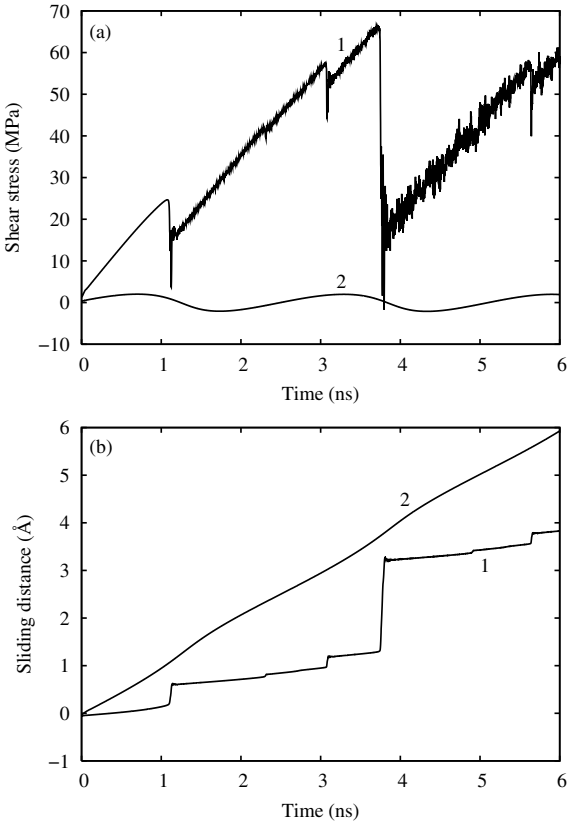


Fig. 15.32. a The shear stress as a function of time for the rough substrate (root-mean-square amplitude 3 Å) and for the block elastic modulus $E = 10\,\text{GPa}$. The *curve 1* is including adhesion, while *curve 2* is without the attractive part in the Lennard-Jones potential, i. e., with $\alpha = 0$ in Eq. (15.4). **b** The average displacement of the interface block atoms (in the sliding direction) as a function of time for the same systems as in **a**

not as abrupt as for the case with adhesion, and the sliding motion is nearly steady. In both cases, the time dependence of shear stress remains periodic with the period 2.6 Å determined by the lattice spacing of the block. For the case with adhesion two small slips and a major slip can be observed in each period, and the kinetic friction is high. For the case without adhesion no elastic instability occurs, and the kinetic friction is very small.

15.3.3 Pressure Dependence of the Frictional Stress

During sliding, the atoms at the sliding interface will experience energetic barriers derived from both the adhesive interaction between the atoms on the two opposing surfaces, and from the applied load. Thus, we may define an *adhesion pressure* $p_{\rm ad}$, and as long as $p_{\rm ad} \gg p$, where p is the pressure in the contact area derived from the external load, the frictional shear stress will be nearly independent of the applied load. Let us illustrate this with the system studied in Sec. 15.3.2. Let us first consider the limiting case where the elastic modulus of the block is extremely small. In this case, in the initial pinned state (before sliding) all the block atoms will occupy hollow sites on the substrate, as indicated by atom **A** in Fig. 15.33. During sliding along the x-direction, the atom **A** will move over the bridge position **B** and then "fall down" in the hollow position **C** (we assume overdamped motion). The minimum energy for this process is given by the barrier height $\delta\epsilon$ (the energy difference between the sites **B** and **A**) plus the work $pa^2\delta h$ against the external load, where a is the block lattice constant and δh the change in the height between sites **B** and **A** (which depends on p). Thus the frictional shear stress $\sigma_{\rm f}$ is determined by $\sigma_{\rm f} a^2 b = \delta\epsilon + pa^2\delta h$, or

$$\sigma_{\rm f} = \delta\epsilon/(ba^2) + p\delta h/b = (p_{\rm ad} + p)\delta h/b\,,$$

where we have defined the adhesion pressure $p_{\rm ad} = \delta\epsilon/(a^2\delta h)$.

In our case $\delta\epsilon \approx 3\,{\rm meV}$ and $\delta h \approx 0.008\,$Å giving $p_{\rm ad} \approx 10^{10}\,{\rm Pa}$. Thus, in the present case, only when the local pressure in the contact regions becomes of the order of $\sim 10\,{\rm GPa}$, or more, it will start to influence the shear

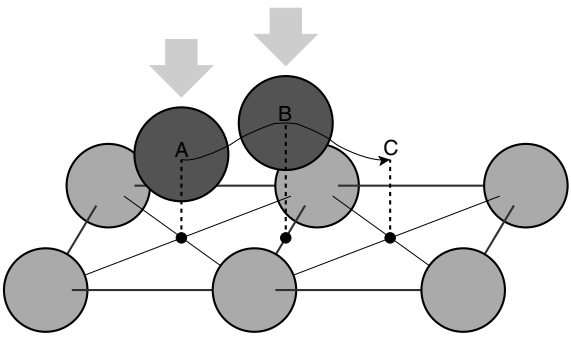


Fig. 15.33. A block atom moving (or jumping) from the hollow site (**A**) over the bridge site (**B**) to the hollow site (**C**). The maximum energy position along the trajectory is at site **B**

stress. This result is in accordance with our simulation results. Thus, for smooth surfaces, the shear stress acting on the block with the elastic modulus $E = 0.5\,\mathrm{GPa}$, squeezed against the substrate with the pressure $p = 50$ and $150\,\mathrm{MPa}$, is identical ($\approx 1\,\mathrm{MPa}$) within the accuracy of the simulations.

For inert materials such as rubber the adhesive pressure may be of similar magnitude as obtained above. Since the contact pressure for rubber in most cases is below $10\,\mathrm{MPa}$, one may expect that the shear stress in the areas of real contact will be independent on the load. Recently, a strong dependence of the (apparent) shear stress on the squeezing pressure was observed for smooth Plexiglas balls sliding on very smooth silicon wafers covered by silane layers [43]. However, as one of us has argued elsewhere [44], this does not reflect a fundamental dependence of the shear stress on the squeezing pressure, but has another origin.

15.4 Summary and Conclusion

In this paper we have shown how multiscale Molecular Dynamics (MD) can be used to study the contact between surfaces which are rough on many different length scales. We have studied the contact morphologies both at high and low magnification, with and without adhesion. We have shown that in atomistic models it is a non-trivial problem how to define the area of real contact between two solids. Our study shows that the area of real contact is best defined by studying the interfacial pressure distribution, and fitting it to an analytical expression. The numerical results are consistent with the theoretical results that the contact area varies linearly with the load for small load, where the proportionality constant depends on the magnification L/λ. For a randomly rough surfaces with the fractal dimension $D_\mathrm{f} = 2.2$ (which is typical for many real surfaces, e. g., produced by fracture or by blasting with small particles) we have found that for small load (where the contact area is proportional to the load) the numerical study gives an area of atomic contact which is only $\sim 14\%$ larger than predicted by the analytical theory of Persson. Since the Persson's theory is exact in the limit of complete contact, it is likely that the Persson theory is even better for higher squeezing loads.

We have also used multiscale MD to study the sliding of elastic solids in adhesive contact with flat and rough interfaces. We considered the dependence of the sliding friction on the elastic modulus of the solids. For elastically hard solids with planar surfaces with incommensurate surface structures we observe extremely low friction (superlubricity), which very abruptly increases as the elastic modulus decreases. Thus, at the superlubricity threshold, an increase in the elastic modulus by a factor ~ 3 resulted in the decrease in the frictional shear stress by a factor $\sim 10^{-5}$. We have shown that even a relatively small surface roughness may completely kill the superlubricity state. For flat surfaces the shear stress is independent of the perpendicular (squeez-

ing) pressure as long as the pressure p is below the adhesion pressure $p_{\rm ad}$, which typically is of the order of several GPa.

Acknowledgement. A part of the present work was carried out in frames of the ESF program "Nanotribology (NATRIBO)". U.T. acknowledge support from IFF, FZ-Jülich, hospitality and help of the staff during his research visits. The authors also thank V.N. Samoilov for his contribution to this paper. This work was partly sponsored by MIUR COFIN No. 2003028141-007 MIUR COFIN No. 2004028238-002, MIUR FIRB RBAU017S8 R004, and MIUR FIRB RBAU01LX5H.

References

1. B.N.J. Persson, *Sliding Friction: Physical Principles and Applications*, 2nd ed., Springer, Heidelberg, 2000
2. J. Krim, I. Heyvaert, C. Van Haesendonck and Y. Bruynseraede, Phys. Rev. Lett. **70**, 57 (1993)
3. R. Buzio, C. Boragno, F. Biscarini, F.B. de Mongeot and U. Valbusa, Nature Materials **2**, 233 (2003)
4. B.N.J. Persson, O. Albohr, U. Tartaglino, A.I. Volokitin and E. Tosatti, J. Phys. Condens. Matter **17**, R1 (2005)
5. B.N.J. Persson, J. Chem. Phys. **115**, 3840 (2001)
6. B.N.J. Persson, Eur. Phys. J. E**8**, 385 (2002)
7. C. Yang, U. Tartaglino and B.N.J. Persson, Eur. Phys. J. E., **19**, 47 (2006)
8. S. Curtarolo and G. Ceder, *Phys. Rev. Lett.* **88**, 255504 (2002)
9. X.B. Nie, S.Y. Chen, W.N. E and M.O. Robbins, *J. Fluid Mech.* **500**, 55 (2004)
10. W. E and Z. Huang, *Phys. Rev. Lett.* **87**, 135501 (2001)
11. W.A. Curtin and R.E. Miller *Modelling Simul. Mater. Sci. Eng.* **11**, R33 (2003)
12. W. Cai, M. de Koning, V.V. Bulatov and S. Yip, *Phys. Rev. Lett.* **85**, 3213 (2000)
13. K.L. Johnson, *Contact Mechanics* (Cambridge University Press, Cambridge, 1985)
14. J. Israelachvili, *Intermolecular and Surface Forces* (Academic Press, London, 1992)
15. B. Luan and M.O. Robbins, Nature **435**, 929 (2005)
16. B.N.J. Persson, F. Bucher, B. Chiaia Phys. Rev. B**65**, 184106 (2002)
17. M. Borri-Brunetto, B. Chiaia, M. Ciavarella, *Comput. Methods Appl. Mech. Eng.* **190**, 6053 (2001)
18. S. Hyun, L. Pei, J.F. Molinari, M.O. Robbins, Phys. Rev. E**70**, 026117 (2004)
19. K.L. Johnson, K. Kendall and A.D. Roberts, Proc. R. Soc. A**324**, 301 (1971)
20. K. Kendall, *Molecular Adhesion and its Applications* (New York, Kluwer, 2001)
21. K.N.G. Fuller and D. Tabor, Proc. R. Soc. A**345**, 327 (1975).
22. J.A. Greenwood and J.B.P. Williamson, Proc. Roy. Soc. London, Ser. A**295**, 300 (1966). See also, J.F. Archard, Proc. Roy. Soc. London, Ser. A**243**, 190 (1957)
23. A.W. Bush, R.D. Gibson and T.R. Thomas, Wear **35**, 87 (1975); A.W. Bush, R.D. Gibson and G.P. Keogh, Mech. Res. Commun. **3**, 169 (1976)

24. B.N.J. Persson, O. Albohr, F. Mancosu, V. Peveri, V.N. Samoilov and I.M. Sivebaek, Wear **254**, 835 (2003)
25. B.N.J. Persson, Phys. Rev. B**51**, 13568 (1995)
26. C. Caroli and P. Nozieres, Eur. Phys. J. B**4**, 233 (1998). T. Baumberger and C. Caroli, cond-mat/0506657 v1 (24 Jun. 2005)
27. C.Y. Yang and Y.P. Zhao, J. Chem. Phy. **120**, 5366 (2004)
28. J. Aubry, J. Phys. (Paris) **44**, 147 (1983)
29. K. Shinjo and M. Hirano, Surf. Sci. **283**, 473 (1993)
30. E. Riedo and H. Brune, Applied Physics Letters **83**, 1986 (2003)
31. R.J.A. van der Oetelaar and C.F.J. Flipse, Surf. Sci. **384**, L828 (1997)
32. B.N.J. Persson and E. Tosatti, Solid State Communications **109**, 739 (1999)
33. C. Caroli and P. Nozieres, in *Physics of Sliding Friction*, ed. by B.N.J. Persson and E. Tosatti, Kluwer, Dordrecht (1996)
34. M.H. Müser, Europhys. Lett **66**, 97 (2004)
35. T. Baumberger and C. Caroli, arXiv:cond-mat/0506657 v1.
36. E. Gnecco, R. Bennewitz, T. Gyalog, Ch. Loppacher, M. Bammerlin, E. Meyer and H.-J. Güntherodt, Phys. Rev. Lett. **84**, 1172 (2000); E. Riedo, E. Gnecco, R. Bennewitz, E. Meyer and H. Brune, Phys. Rev. Lett. **91**, 084502 (2003).
37. B.N.J. Persson, Phys. Rev. B**51**, 13568 (1995)
38. Y. Sang, M. Dube and M. Grant, Phys. Rev. Lett. **87**, 174301 (2001)
39. G. He, M.H. Müser and M.O. Robbins, Science **284**, 1650 (1999)
40. M. Dienwiebel, G.S. Verhoeven, N. Pradeep, J.W.M. Frenken, J.A. Heimberg and H.W. Zandbergen, Phys. Rev. Lett. **92**, 126101 (2004)
41. Y. Liu, A. Erdemir and E.I. Meletis, Surf. Coat. Technol. **86-87**, 564 (1996)
42. The properties of diamond like carbon (DLC) films depends strongly on the preparation method and operation conditions. Thus, only DLC films produced from discharge plasmas containing much hydrogen will exhibit a low friction ($\mu \sim 0.001 - 0.003$). This is believed to result from the passivation of carbon dangling bonds by the hydrogen atoms. Without hydrogen, in an inert atmosphere the friction is huge (of order 1) because of a high concentration of very reactive carbon dangling bonds. In the normal atmosphere, most dangling bonds are passivated and the friction lower but still much higher than for diamond or for DLC films produced from plasmas containing much hydrogen. See, A. Erdemir, Surface and Coatings Technology **146-147**, 292 (2001)
43. L. Bureau, T. Baumberger and C. Caroli, arXiv:cond-mat/0510232 v1
44. B.N.J. Persson, to be published

16 The Role of Nanoroughness in Contact Mechanics

Renato Buzio[1] and Ugo Valbusa[2]

[1] CNR-INFM Unità di Genova, Dipartimento di Fisica, Via Dodecaneso 33, 16146 Genova, Italy buzio@fisica.unige.it
Fax: +39 010 311 066, Tel.: +39 010 353 6289
[2] Dipartimento di Fisica, Università degli Studi di Genova, Via Dodecaneso 33, 16146 Genova, Italy valbusa@fisica.unige.it
Fax: +39 010 311 066, Tel.: +39 010 353 6261

16.1 The Role of Nanoroughness in Depth-Sensing Indentation

Understanding the mechanical response of solid bodies under contact represents one of the most fascinating subjects of materials science and engineering and a necessary step for systematic investigations in novel multidisciplinary areas of biology, like mechanochemical transduction in biomolecules and cells [1] and adhesion mechanisms of small living objects [2, 3].

A considerable advancement of knowledge has been achieved in the past by assessing materials resistance to local deformations by indentation hardness tests, performed either with macro or nano-indentors. In conventional depth-sensing indentation experiments a hard smooth indenter is pressed into the sample and the indentation depth is recorded for different values of load: by analyzing load-displacement curves, both elastic modulus and indentation hardness can be readily estimated (Fig. 16.1a) [4].

Nanoindentors test the mechanical response of bodies by producing deformations in the range of tens or even few nanometers: scaling down displacement sensitivity of indentation techniques has been achieved by introducing new devices, particularly Atomic Force Microscopy (AFM) [5]. Apart from a direct mechanical characterization of ultrathin films and coatings, nanoindentors represent promising tools for exploring novel materials (e. g. nanomaterials) [6, 7] and solving scientific puzzles in solid-state physics (characterization of atomic-scale defects and metallic adhesion at contact [8, 9]) and molecular biology [1, 10]. The degree of complexity of the latter poses serious limitations to indentation experiments dealing with smooth surfaces, bulk deformations, perfectly-plastic behaviour and elasticity in the GPa range; at the same time we are led, on the theoretical side, to partially relax the oversimplifying Hertzian assumptions [5] and include surface roughness, size-dependent viscoelasticity and plasticity, physical and chemical adhesion as key elements for powerful predictive calculations (Fig. 16.1b).

An intuitive picture on the role of surface morphology in contact mechanics emerges from the qualitative arguments suggested by Bowden and Tabor

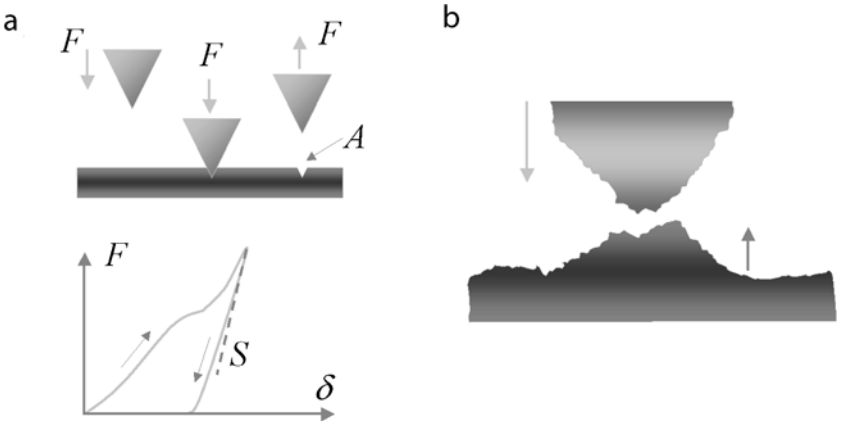


Fig. 16.1. Schematic diagrams of depth-sensing indentation experiments. **a** A conical tip is pressed against a smooth surface with force F, producing an indentation with projected area A. Recording load-displacement curves allows to detect the elasto-plastic response of the specimen and to estimate the contact stiffness $S$and indentation hardness H. **b** At the nanoscale idealized assumptions concerning materials' morphology and mechanical response have to be relaxed: in particular surface roughness and adhesion dominate incipient contact. Arrows indicate surfaces' deformation, eventually induced out-of-contact, due to long-range adhesive forces

at the beginning of the 1900s [11] (Fig. 16.2): morphology reduces contact between surfaces to a small fraction of the nominal contact area (formed by randomly-distributed discrete spots) where high contact pressures localize and plastic yield may eventually occur. Despite its trivial character, this picture still fails to be fully proved on the experimental level, for the difficulty to evaluate separately the relevant deformation phenomena and avoid the appearance of complex third-bodies (contamination, wear particles) effects.

Nanoindentation process is more prone to detect the role of nanoscale roughness with respect to its macro/microscopic counterpart, therefore it can provide deeper insight into contact phenomena: in fact indentors used for macroscopic (microscopic) experiments are not able to appreciate the deformation of the surface layer because of their limited resolution in load and displacement, thus they basically test bulk properties in a single-asperity regime. On the contrary nanoindentors can probe nanoscale roughness in a truly multi-asperity regime, being sensitive both to compressive and Van der Waals adhesive forces. The correctness of this observation has been recently proved by Luan and Robbins through refined Molecular Dynamics (MD) simulations [12]: the authors treated crystalline and amorphous tips indenting an atomically smooth elastic substrate, explicitly comparing atomistic and continuum mechanics calculations. Results demonstrate that even atomic-scale roughness can produce large deviations in local pressure dis-

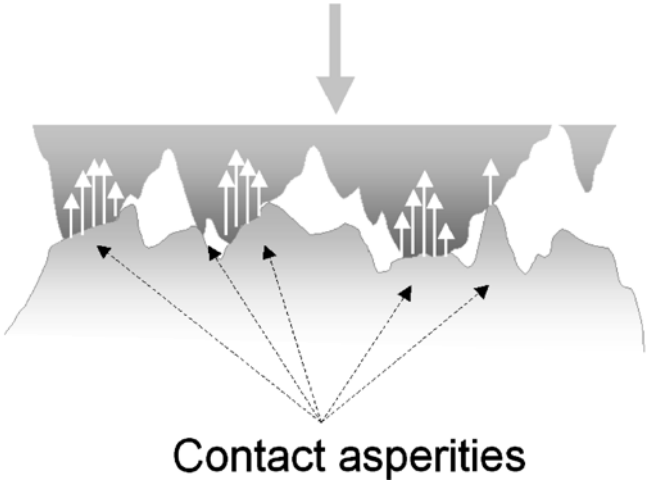


Fig. 16.2. Contact interface of rough surfaces consists of multiple asperities supporting normal forces (*white arrows*)

tributions from continuum predictions assuming smooth contact junctions; therefore any roughness information has to be properly included into continuum calculations to avoid underestimating contact areas and yield stresses, that is, overestimating friction and contact stiffness.

Nanoindentation sensitivity to surface roughness offers a promising approach to validate contact mechanics models dealing with self-affine fractal surfaces [13–16]. Such theories are supported by the experimental observation that the surface of real solids is usually randomly rough and multiscale in nature, i. e., it is a fractal. In detail roughness can be modelled as a statistically non-stationary random process, the scaling $\sigma \propto L^{\alpha}$ (with σ saturated surface roughness, L the unit of measurement and α a real positive exponent) being usually encountered and related to the specific nature of the chosen interface [17]. Computer simulations treating the indentation of Weierstrass-Mandelbrot (WM) fractal surfaces have shown that both the mean hardness and scatter are influenced by roughness whenever penetration depth is smaller or comparable to the interface width, since in such case a multi-asperity contact is actually created between the probe and the sample [18].

16.2 Probing Contact Mechanics at the Nano- and Meso-Scale

Contact mechanics at the nanoscale is studied mostly by AFM. Conventional AFM tips are routinely used in force-spectroscopy experiments to test carbon-based materials and nanoparticles, layered compounds (graphite, mica), metals, polymeric thin films, self-assembling monolayers, macromolecules and

cells [5, 19]. In several cases a non-linear relationship relates applied force, contact area and deformation. Experimental data are routinely analyzed within continuum models for a perfectly elastic and *smooth* contact subject to compressive and adhesive forces, F and F_{adh} [20]. Such models predict the Hertzian scaling $A \approx F_{\mathrm{eff}}^{2/3}$, where A is the contact area and $F_{\mathrm{eff}} = F_{\mathrm{eff}}(F, F_{\mathrm{adh}})$: for example, in the case of the Derjaguin–Muller–Toporov theory, we simply have $F_{\mathrm{eff}} = F + F_{\mathrm{adh}}$ [21]. The power-law scaling of these nanoscale contacts critically depends on the control achieved on the tip shape [22, 23] and roughness [24] as well as on the degree of interface contamination [25]. However, due to the difficulty of monitoring nanoscale roughness during depth-sensing indentation experiments, arguments relating junctions response to surface morphology are still largely speculative at this length scale.

Contact mechanics investigations on the micrometer scale are mainly performed by means of AFM colloidal probes, intentionally blunt/worn AFM tips or custom designed probes [26–39]. Sharp AFM tips satisfyingly access probe geometry and substrate morphology, therefore mechanical response and roughness can in principle be related with greater accuracy.

The role of deterministic roughness on microscale adhesion has been investigated by Ando and Ino using flat AFM tips [34]. The authors have used a focused ion beam to produce two dimensional arrays of peak asperities on atomically flat Si surfaces and to prepare AFM probes having a flat square surface. On all asperities friction and adhesion forces have been measured, demonstrating that they both are proportional to the radius of curvature of the peaks and that capillary forces represent the predominant contribution to adhesion force in ambient air.

Flat AFM tips have been used by Buzio and coworkers to access the mechanical response of self-affine fractal thin films under plastic deformation [37–39]. In their experimental set-up an AFM was operated under standard laboratory conditions, being equipped with rectangular-shaped cantilevers with integrated flat tips. The flat tips had an octagonal shape with edges of a few micrometer in size (Fig. 16.3).

The contact mechanics of a micrometric flat tip squeezed on a rough surface basically differs from the single-asperity contact of sharp tips. The first consequence of tip lateral extension is that the number of contact spots formed at the tip-sample interface is not constant but varies according to the applied load and to the statistical properties of surface roughness. Therefore a micrometric flat tip represents a probe with which structural properties and morphological effects become simultaneously accessible. In order to perform local force-spectroscopy, the tip was first aligned relative to the chosen sample by scanning it at low loads and correcting the tilt of the AFM head in order to visualize without distortions the octagonal shape, which appears on topographies because of the strong convolution of the flat probe with the sharp surface features. Tip lateral motion was then stopped on a given posi-

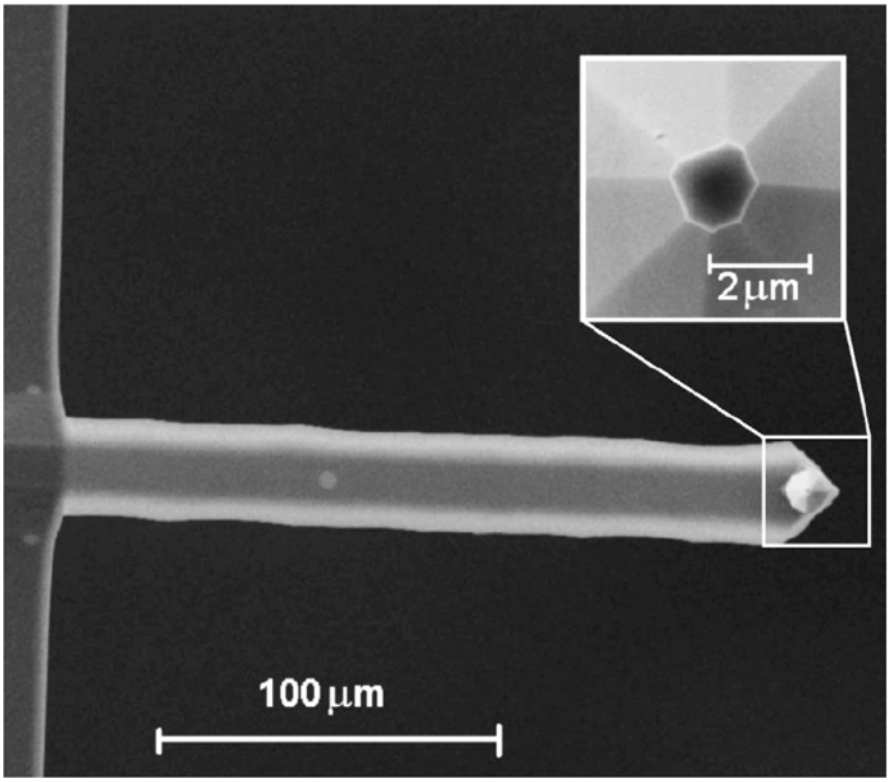


Fig. 16.3. Scanning electron microscopy images of a rectangular AFM cantilever with an integrated flat tip. Top view magnification in the inset: the flat tip has an octagonal shape with an apex area of about $4.8\,\mu m^2$ (from Buzio R et al. (2003), Wear 254:917 [39], Copyright (2003) by Elsevier, *reproduced with permission*)

tion and the vertical deflection of the cantilever was recorded as a function of the scanner displacement z. On each position only the first loading cycle was acquired in view of the plastic deformation expected under repeated indentations. The procedure was iterated for each film on many different regions. The mechanical response of the samples was finally accessed through the analysis of the contact region on the approaching curve. Experiments consisted in acquiring AFM force-distance curves on organic thin films, namely nanostructured carbon films and sexithienyl films, having a large range of fractal dimensions D_f.

Experimental results revealed that: *i*) films with different morphology have a different compliance; *ii*) surface roughness and fractal dimension represent two independent parameters which are capable of considerably decreasing surface stiffness; *iii*) surfaces with a higher fractal dimension are more compliant than less fractal surfaces.

A perfectly–plastic contact theory was used to confirm the reported experimental results on a quantitative basis. The externally applied load F and the indentation depth δ are related through the expression

$$F = HA(D_{\mathrm{f}}, \sigma; \delta + \delta_0) - HA_0 \tag{16.1}$$

where H is the indentation hardness, A the real contact area and A_0, δ_0 respectively the contact area and the penetration depth induced by adhesion [17]. To calculate A, thin films were modelled by 3D WM wave functions, with topographical parameters matching those of the analyzed samples. For each fractal the authors calculated the Abbott–Firestone bearing curve, representing the surface area A lying at the elevation δ below the maximum surface height. Equation (16.1) was finally employed to fit the experimental data and obtain estimates for the unknown parameter H.

Buzio and coworkers explored the possibility to extend similar results to different systems, like ceramic substrates [40]. Indentations were performed with a diamond Berkovich pyramid probe having curvature radius in the range 100 nm–200 nm at the very end of the tip. The analyzed specimens consisted in single-crystals of $SrTiO_3$, commonly used as substrate for epitaxial growth of oxides. Force-displacement curves were acquired on two samples, exposing to the indenter respectively an atomically smooth and a micrometric rough surface. Surface morphology was first characterized by AFM, then load-displacement curves were acquired for each sample on many different locations and results were averaged to obtain the mean mechanical response of the surfaces.

We show in Fig. 16.4 indentation data respectively for the smooth and the rough substrate. For the atomically-smooth $SrTiO_3$ specimen (Fig. 16.4a), the average load-displacement curve reveals the elasto-plastic deformation of the surface for loads up to 500 μN, with a maximum contact depth of about 40 nm and a residual impression depth slightly less than 30 nm. By analyzing the unloading part of the curve [4], the reduced Young modulus $E$is estimated to be $E_1 = 360$ GPa and the hardness $H_1 = 14$ GPa. The average load-displacement curve for the rough sample (Fig. 16.4b) reveals again the elasto-plastic response of the probed surface, the maximum contact depth being now $d = 85$ nm for the maximum compressive load of 500 μN. Such observation suggests the smooth sample to have a *stiffer* surface than the rough sample. In order to understand the physical reasons underlying such effect the mean loading curve for the rough interface was fitted with an elasto-plastic deformation model [40]. The authors assumed the deformed surface to be the sum surface of the tip and the rough sample. Sample morphology was characterized by calculating the height-height correlation function on AFM topographies, thus obtaining a fractal dimension $D_{\mathrm{f}} = (2.41 \pm 0.05)$ and a saturated surface roughness $\sigma = (400 \pm 10)$ nm. The estimated values for fractal parameters were used to numerically generate WM surfaces having the same topographical parameters of the analyzed samples. The Berkovich indenter

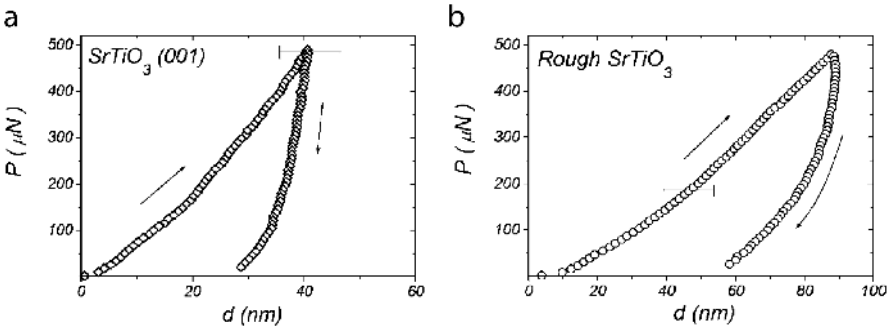


Fig. 16.4. Averaged load-displacement curves acquired respectively for atomically smooth (**a**) and rough (**b**) ceramic substrates by a commercial nanoindenter. Larger penetration depth are reported for the rough interface with respect to the smooth one, whereas the applied loads vary in the same range for both specimen (from Buzio R et al. (2003), Proceedings of the 2003 STLE/ASME Joint International Tribology Conference, Ponte Vedra Beach, Florida USA, 2003 Trib:279 [40], Copyright (2003) by American Society of Mechanical Engineers, *reproduced with permission*)

was modelled as an equivalent conical indenter of included angle 140°. The radius of curvature R of the tip was fixed before each simulation and the indentation hardness H and equivalent elastic modulus E were estimated as fitting parameters. A fully-plastic approximation allowed to estimate an indentation hardness $H_2 = 50\,\mathrm{GPa}$ for the rough surface, with good agreement of data with model.

Data reported in Fig. 16.4 are thus compatible with the following picture: on the smooth $SrTiO_3$ surface the indentation process occurs at the tip apex of the indenter (single-asperity contact) and the values of the Young modulus E_1 and indentation hardness H_1 are due to the crystalline structure of the specimen. On the rough $SrTiO_3$ sample, *because of surface roughness*, the contact often occurs over indenter lateral surface more than at the tip apex (multi-asperity contact) while contact area is definitely smaller than expected at similar values of penetration depth d for the smooth (001) surface. This causes the rough sample to appear more compliant than the smooth one. No differences are of course expected to occur in the mechanical response of the two specimens for micro-indentations operating above the interface width ($d \gg \sigma$).

16.3 Depth-Sensing Indentation with Elastomer AFM Probes

Experimental results described above prove that nanoroughness can significantly affect materials' response under severe *plastic* conditions. This is certainly a crucial feature for any incipient contact between rough surfaces and

might have a role for designing technologically-relevant coatings [41,42]. Nevertheless, it covers just one aspect of the contact process, i. e., irreversible deformations of contact spots. Studying the role of nanoroughness for *elastically*-deformed junctions appears definitely rich and fruitful, with profound implications on adhesive science and applied fields like automotive tire friction [43], sealing effects and adhesion in biology [2,3].

Depth-sensing indentation of perfectly-elastic, adhesive and rough micro-/nano-junctions represents a challenging problem in experimental contact mechanics: it could be solved by implementing on the microscale the experimental set-up originally used by Jonhson, Kendall and Roberts (JKR) [44], i. e., investigating junctions dynamics when an *elastomer* and *smooth* AFM probe is pressing against a hard and rough countersurface. Such system decouples the elastic and morphological properties (localized on opposed surfaces) and allows exploiting the most common nanostructuring techniques (like optical and electron beam lithography, focused ion beam milling or defocused ions sputtering) to tailor surface morphology; moreover, it can be directly compared with refined theoretical models and calculations dealing with deterministic or randomly rough surfaces [15,45,46]. Up to now only few experimental investigations involved the use of elastomer AFM tips on *atomically smooth* surfaces: some efforts are summarized in the following, emphasizing their limitations as well as their peculiar and unique features.

Wang X et al. [47] and Zou J et al. [48] and coworkers fabricated polyimide microcantilevers with poly(dimethylsiloxane) (PDMS) tips. The overall fabrication process involves bulk etching (for forming molds of tips), surface micromachining (for sacrificial release) and bonding (Fig. 16.5a).

A Si wafer is first oxidized and patterned, followed by anisotropic wet silicon etching to form inverted cavities (these pits serve as molds to create elastomeric probe tips); an aluminium sacrificial layer is than deposited on top of the Si wafer. A liquid PDMS prepolymer is applied to fill the pits, then cured to obtain a solid probe. A thin polyimide layer is spin coated, patterned, and cured to form probe beams. Finally, sacrificial etching of the aluminium layer causes probes release. The PDMS probes were successfully used to implement an AFM-based microcontact printing, i. e., depositing chemical materials on substrates in two-dimensional sub-micrometric patterns.

Cho and coworkers described their efforts to monitor the frictional properties of organic monolayers by using an AFM cantilever with a soft PDMS probe at its apex [49] (Fig. 16.5b). A standard AFM tipless cantilever is first treated by oxygen plasma and then chemically modified with trichloro-(1H,1H,2H,2H-perfluorooctyl) silane. The resulting cantilever is then placed on a homemade holder and coated by a prepolimer PDMS drop using a sharp needle through optical microscopy, in order to obtain a micrometric PDMS lens. The PDMS-coated AFM cantilever is finally cured, to speed-up and terminate the cross-linking reaction of the PDMS network.

Both nanotribological and micro-contact-printing studies reported above require, indeed, an accurate characterization of probe viscoelasticity, causing

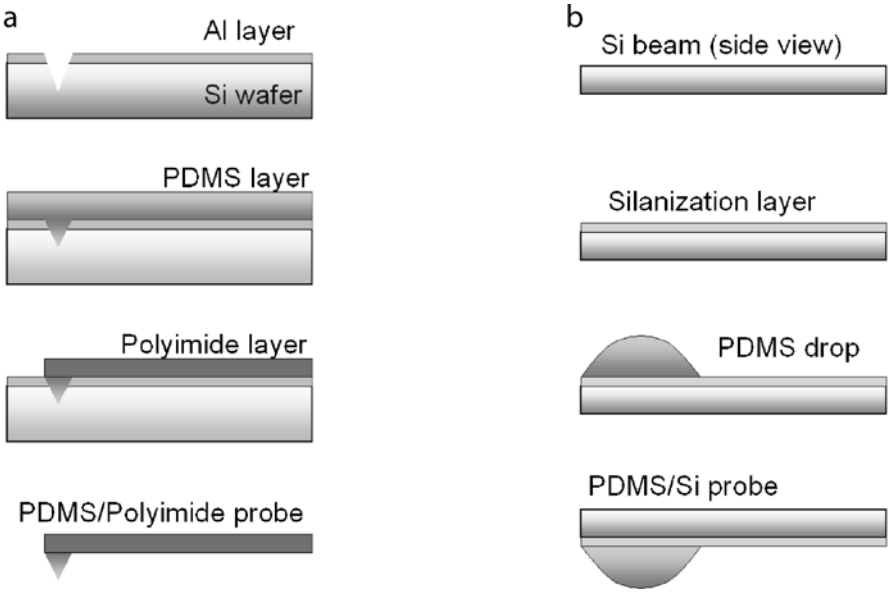


Fig. 16.5. Schematic diagrams showing processes to obtain PDMS AFM tips. **a** The mould-and-transfer method developed by Wang et al. [47] and Zou et al. [48] producing nanometric PDMS tips; **b** the deposition method proposed by Cho et al. to obtain micrometric PDMS tips [49]

a remarkable dependence of interface deformation, contact area and pull-off force on dwell time, loading and detachment rates; it is clear that these aspects have to be carefully analyzed before entering the crucial problem of nanoscale roughness in contact mechanics.

Vakarelsky et al. [50] reported an accurate contact mechanics investigation of microscale viscoelasticity for a colloidal AFM probe with microbeads of 98 wt % poly(diethylhexyacrylate) and 2 wt % poly(acrylmethacrylate). The former belongs to a cross-linked amorphous polymer and the glass transition temperature is $-55\,^{\circ}\mathrm{C}$, therefore it behaves as an elastomer of low elasticity modulus above the glass transition temperature. The authors acquired force-distance curves while indenting an atomically-flat mica surface in water and separately evaluated the role of contact time t_{cm} and retracting velocity V on unloading curves. In Fig. 16.6a we show two typical loading/unloading cycles acquired with scan rates $V = 45\,\mathrm{nm/s}$ and $V = 800\,\mathrm{nm/s}$ respectively and a contact time $t_{\mathrm{cm}} = 80\,\mathrm{s}$; in the same figure we report the dependence of the adhesion force F_{adh}on V.

These figures indicate first a marked *adhesion hysteresis* between loading and unloading curves, second that F_{adh} increases with V at a fixed contact time, as expected by numerous previous macroscopic reports on elastomer materials [51–54]. It can be assumed that the effective interface energy w depends on the displacement velocity of the contact line v_{p}, which is the

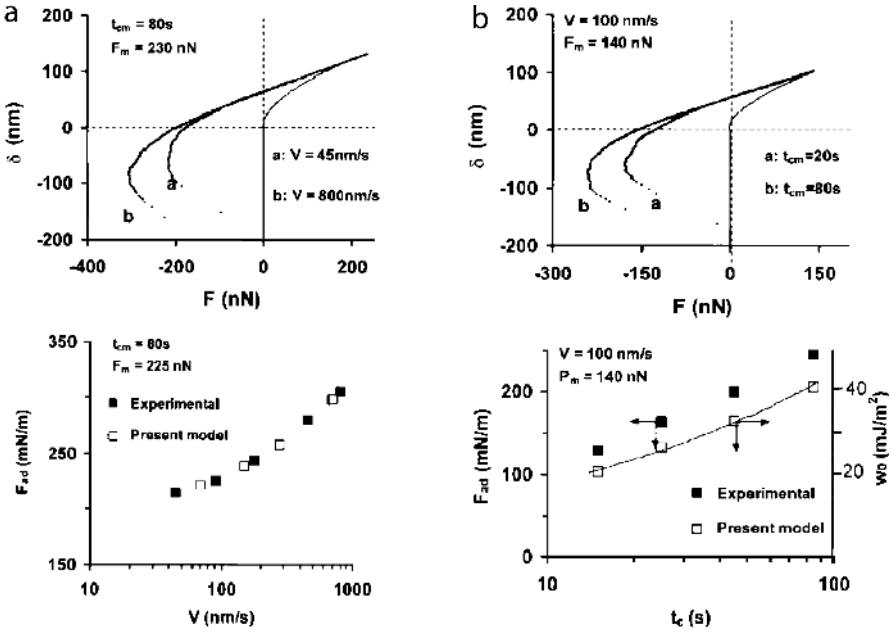


Fig. 16.6. Experimental AFM data describing the adhesive interaction of a micrometric elastomer bead with atomically smooth mica under water. **a** Averaged load-displacement curves (*top*) and adhesion force vs unloading velocity (*bottom*); **b** Averaged load-displacement curves (*top*) and adhesion force vs contact time (*bottom*). White squares indicate theoretical predictions based on the Maugis and Barquins thermodynamic model (from Vakarelski IU et al. (2001), Langmuir 17:4739, Copyright (2003) by American Chemical Society [50], *reproduced with permission*)

velocity of peeling off two surfaces:

$$w = w_0 \left[1 + \alpha\left(T\right) v_{\mathrm{p}}^{n}\right] \tag{16.2}$$

where w_0 is the thermodynamic work of adhesion at $v_{\mathrm{p}} = 0$ and $\alpha(T)$ and n are constants describing the elastomer material. The second term in the bracket characterizes the dissipation energy (viscoelastic loss) at the tip of the crack during the detachment. Since v_{p} increases with V, Eq. (16.2) is consistent qualitatively with the increases of F_{adh} with V. The experimental behaviour of Fig. 16.6a can be described in greater detail using the Maugis and Barquins (MB) thermodynamic approach [55]. When the stress energy release rate G is introduced, Eq. (16.2) gives us an explicit expression for the contact line velocity:

$$v_{\mathrm{p}} = \left|\frac{\mathrm{d}a}{\mathrm{d}t}\right| = \left(\frac{G - w_0}{\alpha\left(T\right) w_0}\right)^{1/n} = \left(\frac{3a^3 E}{8\pi R^2 \alpha\left(T\right) w_0}\left[1 - \left(\frac{R\delta}{a^2}\right)\right]^2 - \frac{1}{\alpha\left(T\right)}\right)^{1/n} \tag{16.3}$$

where a is the contact radius and G is deduced from the JKR theory to be:

$$G = \left(\frac{3a^3E}{8\pi R^2}\right)\left[1 - \left(\frac{R\delta}{a^2}\right)\right]^2 \tag{16.4}$$

with E the effective Young modulus and δ the elastic displacement. Numerical integration of Eq. (16.3) provides the time evolution of the contact radius and contact area between probe and surface. From the JKR theory we also have:

$$\delta = \frac{a^2}{3R} + \frac{2F}{3aK} \tag{16.5}$$

thus deformation δ and normal force F can be evaluated at every time step to predict force-distance curves [56]. We show in Fig. 16.7 unloading curves numerically simulated according to the described procedure: they refer to an elastomer bead of Young modulus $E = 1\,\text{MPa}, R = 5\,\mu\text{m}, n = 0.6, \alpha = 45{,}000\,\text{SI units}$, $w_0 = 40\,\text{mJ/m}^2$, attached to a cantilever with stiffness $k_c \gg k_{\text{contact}}$, where $k_{\text{contact}} = 3aE/2 \approx 14\,\text{N/m}$.

Vakarelsy et al. investigated also the effects of t_{cm}on the unloading curve F_{adh} at a fixed detachment rate $V = 100\,\text{nm/s}$, shown in Fig. 16.6b. From the

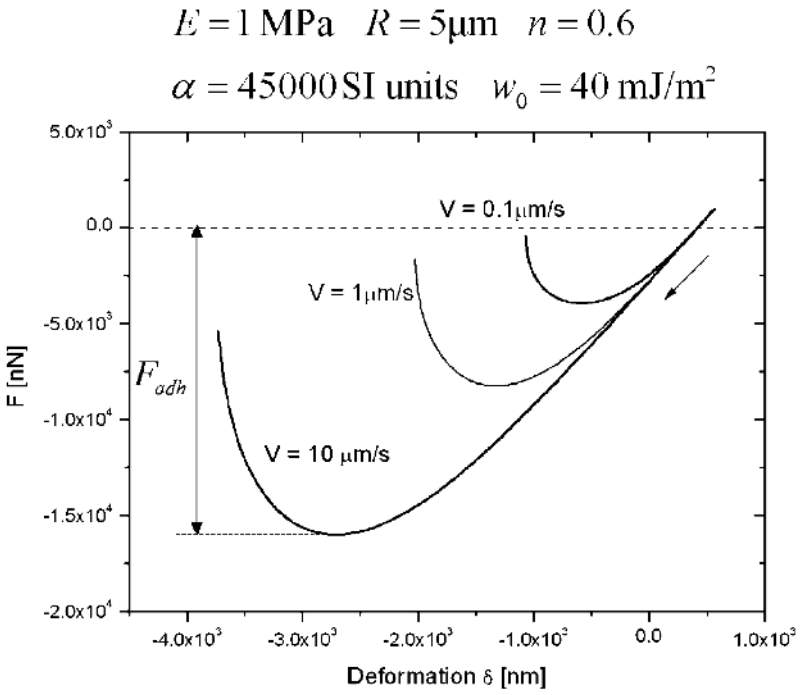


Fig. 16.7. Load-displacement curves theoretically predicted by Maugis and Barquins thermodynamic model. The marked dependence of adhesion force on unloading velocity is due to viscoelastic losses at the crack tip

two unloading curves for $t_{\rm cm} = 20\,{\rm s}$ and $t_{\rm cm} = 80\,{\rm s}$, and the corresponding dependence of $F_{\rm adh}$ on $t_{\rm cm}$, it is clear that the value of $F_{\rm adh}$ increases with $t_{\rm cm}$. Hence, we consider that w_0 in Eq. (16.2) must be an increasing function of $t_{\rm cm}, w_0 = w_0(t_{\rm cm})$. This increase is attributable to the fact that the bond formation between surfaces in contact increases with the contact time, and can be related to surface rearrangement or even chemical reactions occurring at the polymer-substrate interface [57].

Recently Buzio and coworkers joined the widespread technological interest on silicone elastomers with the cheap and quick preparation methods of colloidal AFM probes, developing PDMS beads for contact mechanics investigations at the microscale (Fig. 16.8). The PDMS particles are prepared starting from commercial liquid prepolymers and offer several properties instrumental to contact mechanics investigations: PDMS is a soft and durable elastomer, with a surface that is low in interfacial free energy (ca. $22\,{\rm mJ/m^2}$) and chemically inert; PDMS is homogeneous, isotropic and optically transparent. Moreover its surface properties can be readily modified by treatment with plasma followed by the formation of self-assembling-monolayers to give appropriate interfacial interactions with materials that themselves have a wide range of interfacial free energies [58, 59].

Previous results demonstrate that a considerable progress has been done at the micro- and nano-scale by indentation techniques for elastically deformable interfaces. At present these studies require a consistent and robust characterization of polymers bulk viscoelasticity and interfacial stability, but it is plausible that the main efforts will be oriented, in a near future, to systematically access and characterize the role of nanoroughness in contact mechanics.

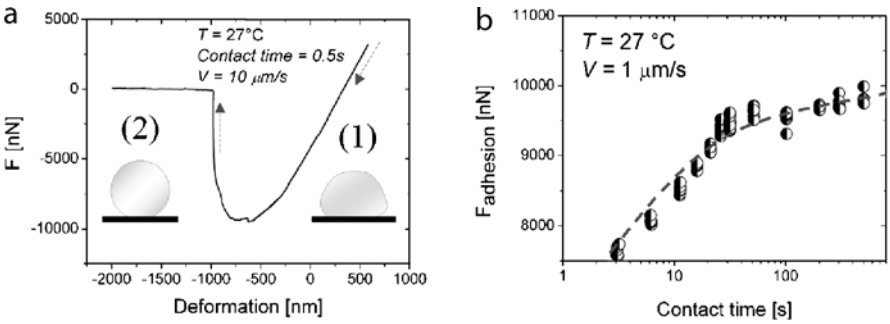


Fig. 16.8. Experimental AFM data describing the adhesive interaction of a micrometric PDMS bead with atomically smooth mica in dry air. **a** Load-displacement curves show good qualitative agreement with theoretical predictions from the thermodynamic MB model: cantilever force decreases linearly due to bead deformation at constant contact area (1) until crack opens and contact area decreases (2); **b** adhesion force depends on contact time, indicating a time variation of interface energy w [from Buzio R, Bosca A, Krol S, Valbusa U, *private communication*]

16.4 Conclusions

Depth-sensing indentation offers the unique opportunity to understand and evaluate the role of nanoroughness in contact mechanics. While experimental difficulties still persist for the investigation of rough nanosized junctions, multi-asperity contacts are routinely formed by pressing colloidal or custom-designed AFM probes against rough substrates: in such case a significant correlation can be appreciated between contact stiffness and morphological properties at incipient contact. We present AFM investigations using elastomer probes: in the near future they will offer the opportunity to evaluate the impact of nanoscale roughness on interface dynamics, adhesion and friction forces under purely elastic deformation conditions. Such studies will certainly represent an invaluable base for testing advanced theoretical models and provide inspiration for development of innovative micro-electro-mechanical devices.

Acknowledgement. Authors acknowledge support by CNR-INFM project FIRB "Carbon based micro- and nano-structures" and MIUR projects PRIN "Nanotribology" and FIRB "NANOMED".

References

1. Bao G and Suresh S (2003), Nat. Mat. 2:715
2. Autumn K, Sitti M, Liang YA, Peattie AM, Hansen WR, Sponberg S, Kenny TW, Fearing R, Israelachvili JN, Full RJ (2002), Proc. Natl. Acad. Sci. USA 99:12252
3. Gao H and Yao H (2004), Proc. Natl. Acad. Sci. USA 101:7851
4. Oliver WC, Pharr GM (1992), J. Mat. Res., 7:1564
5. Cappella B, Dietler G (1999), Surf. Sci. Rep. 34:1
6. Suresh S, Nieh TG, Choi BW (1999), Scripta Materialia 41:951
7. Buzio R, Gnecco E, Boragno C, Valbusa U (2002), Carbon 40:883
8. Rodriguez de la Fuente O, Zimmerman JA, González MA, de la Figura J, Hamilton JC, Woei Wu Pai, Rojo JM (2002), Phys. Rev. Lett. 88:36101
9. Rubio-Bollinger G, Joyez P, Agrait N (2004), Phys. Rev. Lett. 93:116803
10. Bowen WR, Lovitt RW, Wright CJ (2000), Colloids Surf. A 173:205
11. Bowden FP, Tabor D (1954), Friction and Lubrication of Solids, Oxford University Press, London
12. Luan B and Robbin MO (2005), Nature 435:929
13. Majumdar A, Bhushan B (1991), ASME J. Tribol. 113:1
14. Persson BNJ (2001), Phys. Rev. Lett. 87:116101
15. Persson BNJ (2002), Phys. Rev. Lett. 89:245502
16. Persson BNJ, Albohr O, Creton C, Peveri V (2004), J. Chem. Phys. 120:8779
17. Barabasi AL, Stanley HE (1995), Fractal concepts in surface growth, Cambridge University Press, Cambridge
18. Bobji MS, Biswas SK, Pethica JB (1997), Appl. Phys. Lett. 71:1059
19. Butt HJ, Cappella B, Kappl M(2005), Surf. Sci. Rep. 59:1

20. Schwarz UD (2003), J. Coll. Interf. Sci. 261:99
21. Unertl WN (1999), J. Vac. Sci. Technol. A 17:1779
22. Carpick RW, Agrait N, Ogletree DF, Salmeron M (1996), J. Vac. Sci. Technol. B 14:1289
23. Schwarz UD, Zwoerner O, Koester P Wiesendanger R (1997), Phys. Rev. B 57:6987
24. Cohen SR, Neubauer G, McClelland GM (1990), J. Vac. Sci. Technol. A 8:3449
25. Putman CAJ, Igarashi M, Kaneko R (1995), Appl. Phys. Lett. 66:3221
26. Ducker WA, Senden TJ, Pashley RM (1991), Nature 353:239
27. Toikka G, Hayes RA, Ralston J (1996), J. Coll. Interf. Sci. 180:329
28. Considine RF, Hayes RA, Horn RG (1999) Langmuir 15:1657
29. Kohonen MM, Karaman ME, Pashley RM (2000) Langmuir 16:5749
30. Heim L, Blum J, Preuss M, Butt HJ (1999) Phys. Rev. Lett. 83:3328
31. Mohideen U, Anushree R (1998), Phys. Rev. Lett. 81:4549
32. Bhushan B, Sundararajan S (1998), Acta Mater. 46:3793
33. Biggs S, Cain RG, Dagastine RR, Page NW (2002), J. Adhes. Sci. Technol. 16:869
34. Ando Y, Ino J(1996) Sensors and Actuators A 57:83
35. Butt HJ (1991) Biophys. J. 60:1438
36. Hsu SM, McGuiggan OM, Zhang J, Wang Y, Yin F, Yeh YP, Gates RS (2001) In: Bhushan B (ed) Fundamental of tribology and bridging the gap between the Macro and Micro/Nanoscales, Kluwer, Dordrecht, p 691
37. Buzio R, Buatier F, Boragno C, Valbusa U (2003) Thin Solid Films 133:111
38. Buzio R, Boragno C, Biscarini F, Buatier F, Valbusa (2003), Nat. Mat. 2:233
39. Buzio R, Boragno C, Valbusa U (2003), Wear 254:917
40. Buzio R, Malyska K, Rymuza Z, Boragno B, Bautier B, Valbusa U (2003), Proceedings of the 2003 STLE/ASME Joint International Tribology Conference, Ponte Vedra Beach, (Florida) USA
41. Buzio R, Gnecco E, Boragno C, Valbusa U, Piseri P, Barborini E, Milani P (2000), Surf. Sci. 444: L1
42. Buzio R, Chierichetti A, Bianchi G, Valbusa (2006), Surf. Coat. Technol. 200:6430
43. Persson BNJ, Tartaglino U, Albohr O., Tosatti E (2004), Nat. Mat. 3:882
44. Johnson KL, Kendall K, Roberts AD (1971), Proc. Roy. Soc. Lond. A 324:301
45. Carbone G, Mangialardi L (2004), J. Mech. Phys. Solids 52:1267
46. Ciavarella M, Murolo G, Demelio G, Barber JR (2004), J. Mech. Phys. Solids 52:1247
47. Wang X, Ryu K, Bullen D, Zou J, Mirkin CA, Liu C (2003), Proc. 12th Int. Conf. on Solid-State Sensors, Actuators and Microsystems, (Boston, MA, 2003).
48. Zou J, Wang X, Bullen D, Ryu K, Liu C, Mirkin CA (2004), J. Micromech. Microeng. 14:204
49. Cho JH, Lee DH, Shin HS, Pattanayek SK, Ryu CY and Cho K (2004), Langmuir 20:11499
50. Vakarelski IU, Toritani A, Nakayama M, Higashitani K (2001), Langmuir 17:4739
51. Gent AN, Schultz, J (1972), J. Adhes. 3:281
52. Andrews EH, Kinloch AJ (1973), Proc. R. Soc. London A 332:385
53. Kendall KJ (1973), Phys. D 6:1782
54. Fuller KG, Roberts AD (1980), J Phys. D 14:221

55. Maugis D, Barquins M (1978), J. Phys. D 11:1989
56. Barquins M, Maugis D (1981), J. Adhes. 13:53
57. Pickering JP, Van Der Meer DW, Vancso GJ (2001), J. Adhesion Sci. Technol. 15:1429
58. Xia Y, Whitesides GM (1998), Angew. Chem. Int. Ed. 37:550
59. Xia Y, Whitesides GM (1998), Annu. Rev. Mater. Sci. 28:153

17 Energy Dissipation and Nanoscale Imaging in Tapping Mode AFM

Ricardo García, Nicolás F. Martínez, and Carlos J. Gómez,
and Antonio García-Martín

Instituto de Microelectrónica de Madrid, CSIC, c/ Isaac Newton 8,
28760 Tres Cantos, Madrid rgarcia@imm.cnm.csic.es

17.1 Introduction

Amplitude modulation atomic force microscopy (AM-AFM) also known as *tapping mode* AFM is arguably the dominant technique for nanometer-scale characterization of surfaces in air and liquids. High resolution images of semiconductors, polymers, biomolecules or nanocrystals have been demonstrated. Its unique flexibility allows the simultaneous mapping of topography and other material properties such as electrostatic, magnetic or mechanical properties. The instrument has also given rise to prominent nanolithographies such as local oxidation nanolithography. Because their robustness, versatility and nanoscale resolution, AM-AFM instruments can be found in both academic oriented laboratories and technological facilities.

In AM-AFM the cantilever-tip ensemble is excited at a fixed frequency, usually near or at the free resonance frequency while the oscillation amplitude is used as a feedback parameter to control the operation of the microscope. The non linear character of the interaction forces between tip and surface gives rise, in many cases of interest, to the coexistence of two oscillation states for the same external conditions. This genuinely nonlinear dynamic behavior of the AFM is formally described in terms of the coexistence of two oscillation branches, low and high amplitude respectively. However, it is more common and intuitive to use a description based on terms of two dominant interactions regimes, *attractive* and *repulsive*. In the attractive interaction regime, a net attractive force dominates the amplitude reduction ($\langle F_{\rm ts} \rangle \leq 0$) while in the repulsive regime the amplitude reduction is dominated by a net repulsive force ($\langle F_{\rm ts} \rangle \geq 0$). Although the description of oscillation states in terms of attractive and repulsive regimes does not capture the whole complexity of AM-AFM operation, it can be applied to many experimental situations of interest, so both descriptions will be used indistinctly here.

In recent years most of the effort in tapping mode AFM has been placed on either demonstrating new applications or in developing a precise theoretical framework to understand its performance [1–26]. So it is not a surprise that the issue of energy dissipation in AM-AFM has received little interest.

Yet energy dissipation is very akin to AM-AFM operation in two different aspects. First, because it is directly related to either sample or tip damage and deformation, and secondly, because it is closely related to *phase imaging*

which is one of the more powerful methods for qualitative compositional contrast at the nanoscale usually known as phase imaging.

There are two different sources of dissipation in AM-AFM, the hydrodynamic damping of the cantilever motion with the molecules of the medium and the loss of mechanical energy of the cantilever due to the tip-surface interactions. This is the energy dissipation of relevance here. Let's also remark that dissipation process of interest in this chapter do not necessarily involve either tip or sample damage.

17.2 Energy Balance in Amplitude Modulation AFM

Amplitude modulation AFM operation implies a steady-state situation where the energy per cycle supplied by an external power source ($\overline{E}_{\rm ext}$) must be converted either into hydrodynamic damping in the medium ($\overline{E}_{\rm med}$) and energy dissipated in the sample ($\overline{E}_{\rm dis}$) [27–29],

$$\overline{E}_{\rm ext} = \bar{E}_{\rm med} + \bar{E}_{\rm dis} \tag{17.1}$$

where,

$$\overline{E}_{\rm ext} = \oint F_{\rm ext} \frac{{\rm d}z}{{\rm d}t}\,{\rm d}t = \oint F_0 \cos\omega t \frac{{\rm d}z}{{\rm d}t}\,{\rm d}t \tag{17.2}$$

$$\overline{E}_{\rm med} = \oint -F_{\rm med} \frac{{\rm d}z}{{\rm d}t}\,{\rm d}z \tag{17.3}$$

$$\overline{E}_{\rm dis} = \oint F_{\rm ts} \frac{{\rm d}z}{{\rm d}t}\,{\rm d}t \tag{17.4}$$

whenever the tip motion is described by a sinusoidal function $z = z_0 + A\cos(\omega t - \phi)$, the external energy can be expressed in analytical form as

$$\overline{E}_{\rm ext} = \frac{\pi k A A_0 \sin\phi}{Q} \tag{17.5}$$

If the hydrodynamic damping is modelled by a linear viscous damping law $F_{\rm med} = -(m\omega_0/Q)\dot{z}$, then

$$\overline{E}_{\rm med} = \frac{\pi k A^2 \omega}{Q\omega_0} \tag{17.6}$$

By combining Eqs. (17.5) and (17.6) the energy dissipated by tip-surface interactions forces is given by

$$\overline{E}_{\rm dis} = \overline{E}_{\rm ext} - \overline{E}_{\rm med} = \frac{\pi k A A_0 \sin\phi}{Q} - \frac{\pi k A^2 \omega}{Q\omega_0} \tag{17.7}$$

where Q is the quality factor of the free cantilever and A_0 is the free amplitude.

The assumption that the tip motion in tapping mode AFM is a sinusoidal function $z = z_0 + A\cos(\omega t - \phi)$ is a rather good approximation. Numerical simulations have shown that under the most relevant experimental situations the fast fourier transform (FFT) of the oscillation reveals a large peak at the

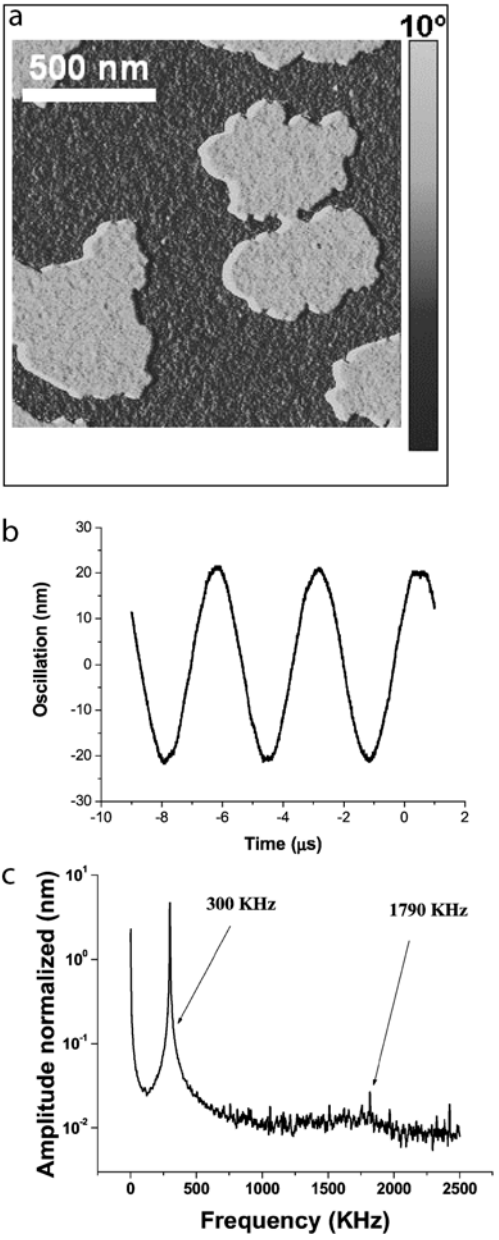


Fig. 17.1. a Phase image of several monolayer islands (T6) on a silicon surface. **b** Oscillation amplitude evolution while imaging. **c** FFT of the oscillation shown in **b**

excitation frequency. This peak is about 10^3–10^4 larger than the second peak of the FFT [30].

Figure 17.1a shows a phase shift image of several sexithienyl molecules (T6) deposited onto silicon substrates. Figure 17.1b shows the experimental oscillation amplitude as a function of time while taking the image. The fast Fourier transform (FFT) (Fig. 17.1c) shows a single and strong peak centred at the excitation frequency $\omega = \omega_0$. A small bump is observed at 1.79 MHz, i. e., a frequency that matches the second eigenmode of cantilever. Notice that the first resonance is about 10^3 larger than the second one.

17.3 Determination of the Dissipated Energy on the Sample

Figure 17.2 shows amplitude and phase shift vs. tip displacement (z piezo displacement) curves taken on a silicon surface. The silicon face Si(100) is covered by a native oxide of about 1 – 2 nm in thickness. The sudden jump in the curves reveals a transition between the attractive (low amplitude) and repulsive regimes (high amplitude). At the transition point, the amplitude shows an increase of about 0.5 nm (Fig. 17.2a and b while the phase shift shows a remarkable reduction of 40° (Fig. 17.2c). The phase changes from values above 90° (attractive regime or low amplitude branch) to values below 90° (repulsive regime or high amplitude branch).

Equation (17.7) allows to transform phase shift data into values of the dissipated energy. Figure 17.3 shows the energy dissipation curve as a function of the amplitude ratio A/A_0 for the data of Fig. 17.2c. The dissipation shows a slight increase with A/A_0 reduction until it drops to zero for amplitude ratios below 0.1. The maximum of the dissipated energy is about 10 eV per cycle. It happens in the section of the curve that belongs to the repulsive regime. The transition between attractive an repulsive regimes is accompanied by a small increase of the dissipated energy, in this case of about 2 eV.

In the same experiment ($A/A_0 = 0.6$) the external energy and the hydrodynamic damping amounts 109 eV and 99.5 eV respectively. The comparison reveals that the dissipated energy in the sample is a small fraction of the energy supply to the cantilever. Furthermore, dissipation in the sample should not be associated with sample irreversible deformation. This is easily explained by noticing that for contact radii of 2 nm, a total dissipation of 10 eV implies that the energy dissipated per atom is roughly 0.04 eV respectively.

17.4 Sources of Dissipation

Energy dissipation at atomic and nano scales is an active area of research because its practical implications in areas such as lubrication, tribology, high

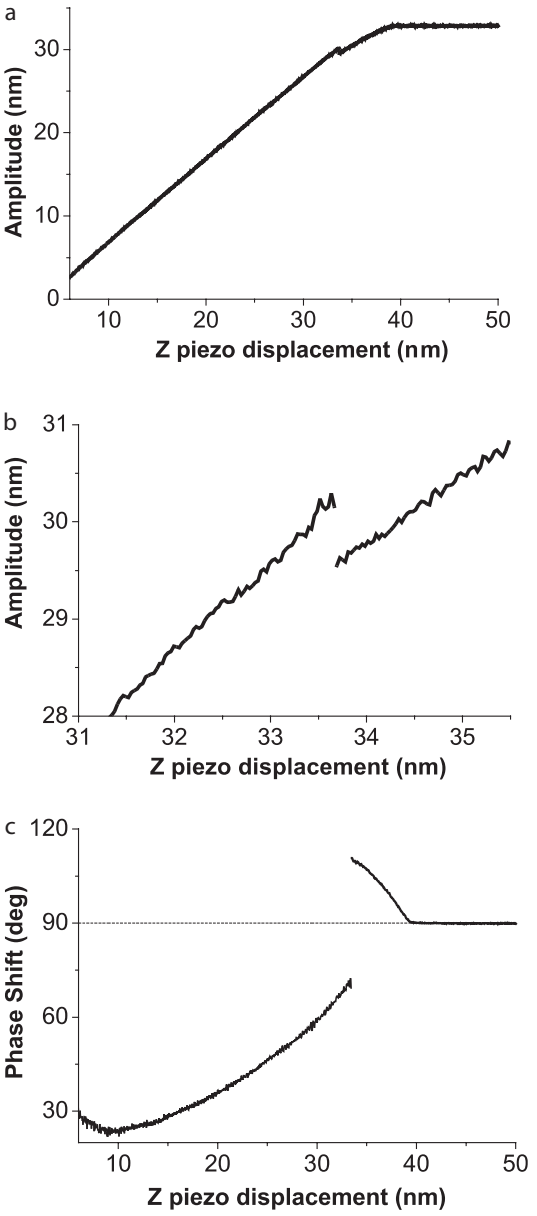


Fig. 17.2. Amplitude (**a**, **b**) and phase shift (**c**) curves. The steplike jump in both curves marks the transition between attractive and repulsive interaction regimes. Because of the direct dependence of the phase shift with the sign of the interaction force, the transition between regimes is better observed in the phase shift curves than in the amplitude curves. The zero in the horizontal axis has been chosen arbitrarily although is close to the zero tip-surface distance. **b** Amplitude curve in the vicinity of a transition point

resolution imaging and because its fundamental relevance. Equation (17.7) allows to transform experimental data into energy dissipation in AM-AFM, however, it does not inform us about the physical nature of the processes involved in the energy loss. Dissipation in atomic force microscopy can be described at the level of atoms and molecules or at the nanoscale [31–38].

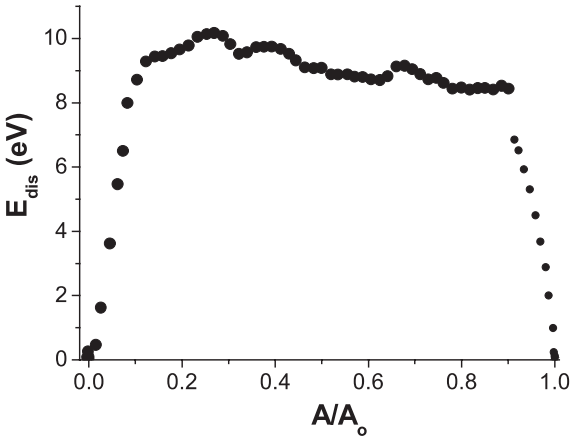


Fig. 17.3. Energy dissipation curves deduced by applying Eq. (17.7) to the data presented in Fig. 17.3. The curves show a maximum with the amplitude ratio. The transition between attractive and repulsive regimes is usually marked by a small jump in the dissipated energy. Data obtained on a silicon region

Both descriptions are complementary. In the first case, the emphasis is placed on the specific atomic processes that allow the transformation of mechanical energy from the tip in atomic and molecular motions in the sample. Several mechanism have been proposed. They range from tip-induced motion of image charges and its associated ohmic losses to reversible structural changes in the tip-surface junction or the excitation of phonons in the sample due to the strength of the short-range interactions forces. At the nanoscale, which is the focus of this chapter, the emphasis is placed on the overall shape of the dissipation process and its relationship with macroscopic quantities such as viscosity, surface energies, hysteresis in the energy of adhesion or mechanical properties in general.

Hysteresis in the energy of adhesion or in the long-range attractive forces is believed to dominate the mechanism of dissipation at the nanoscale with stiff surfaces. Adhesion hysteresis implies that the nanoscale landscape ob-

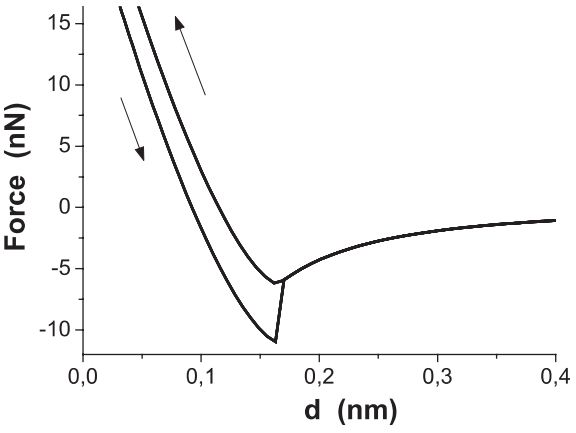


Fig. 17.4. Simulated force distance curves with hysteresis in the surface energy. The area enclosed by approaching and retracting curves is the energy loss per cycle

served by the tip depends whether the tip is moved towards or from the surface. Several mechanisms could account for energy adhesion hysteresis such as (i) exchange of different adsorbates between tip and sample, (ii) reorientation of polar groups at surfaces without involving any translational displacement or (iii) rearrangement and/or interdigitation of surface molecules [39]. A combination of those processes or any of them acting independently will result in a lower surface free energy during the approach than on the retracting cycle and thus will lead to energy dissipation. Those processes will give rise to force curves as the one shown in Fig. 17.4. The area enclosed by the force distance curve represents the dissipated energy. In this case the interaction forces includes long-range van der Waals forces and contact forces described by the DMT model (see Ref. [17]). The simulation was performed for a tip of radius $R = 10\,\text{nm}$, force constant of k=2 N/m, resonant frequency $\omega = \omega_0 = 59\,\text{kHz}$, quality factor $Q = 150$, free amplitude $A_0 = 32.5\,\text{nm}$ and average tip-surface separation of $z_c = 20\,\text{nm}$. Dissipation arises by us-

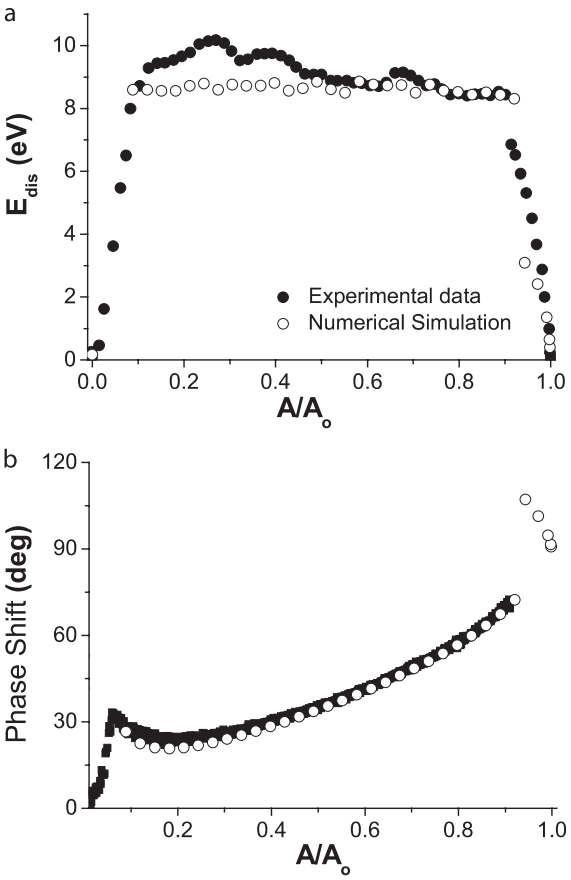


Fig. 17.5. Comparison theory (*open circles*) and experiment (*filled circles*). **a** The agreement between theory and experiment in the determination of the dissipated energy shows that hysteresis in the surface energy is the major source of dissipation in AM-AFM. **b** Experiment and simulated phase shift as a function of the amplitude ratio

ing different values of the surface energy in the approaching and retracting half cycles, 100 and 175 $\mathrm{mJ/m^2}$ respectively. Indeed the excellent quantitative agreement between simulations and experiments presented in Fig. 17.5 supports the role of adhesion hysteresis as a dominant mechanism of energy dissipation in AM-AFM.

Needless to say that the dominant channel for dissipation at the nanoscale will also depend on the properties of the sample. For example, viscoelastic properties could also contribute significantly to dissipation while imaging polymers.

17.5 Phase Imaging

Recording the phase shift between the excitation force and the tip response has been proposed and used to image compositional variations in heterogeneous samples [41–55]. In fact *phase imaging*, as it is known, has become a powerful method for mapping at high spatial resolution variations in composition, friction, viscoelasticity and adhesion of the sample surface with independence of the topographic features. However, compositional contrast has not yet been transformed into quantitative information about material properties. This is probably due to the still emerging understanding of the various factors and interactions that affect and modify the phase shift during AM-AFM operation. Among those factors are the ratio between excitation and resonance frequencies, sample elastic and viscoelastic properties, tip's radius, friction, lever tilt angle or cross-talking between morphological and compositional contributions. Nonetheless, numerical simulations as well as analytical expressions have clearly established that phase shift variations at a fixed feedback amplitude are directly linked to energy dissipation process [27, 28, 41].

Figure 17.6 gives a demonstration of the ability of phase imaging to distinguish between materials of different properties at the nanoscale. The topography (Fig. 17.6a shows a dot while the phase image (Fig. 17.6b) reveals the presence of 16 single-molecule magnets forming a ring around a local oxide dot [55].

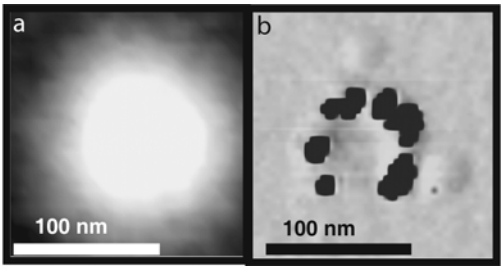


Fig. 17.6. Compositional contrast between single-molecule magnets and a silicon oxide dot. **a** Topography and **b** phase image. The material property contrast provided by phase images reveal several the presence of several single-molecule magnets that were concealed by topographic features in **a**

To develop *phase imaging* as a tool for quantitative analysis of material properties at the nanoscale, requires, among other things, a precise theoretical and experimental understanding of the dependence of phase shifts with operational parameters such as the driving force and feedback amplitude, cantilever parameters as well as tip-surface energy dissipation processes.

17.6 Summary

In this contribution we have examined the relationship existing between energy dissipation and nanoscale imaging in amplitude modulation AFM. Higher resolution and contrast image are obtained by plotting the phase shift dependence across the sample surface instead of amplitude variations. Phase shifts are directly related to energy dissipation processes on the sample surface. We also show that one of the major sources of energy dissipation is the hysteresis in the adhesion energy.

Acknowledgement. This work was financially supported by the European Commission (FORCETOOL, NMP4-CT-2004-013684).

References

1. A. Terawaki, Y. Otsuka, H.Y. Lee, T. Matsumoto, H. Tanaka, T. Kawai, Appl. Phys. Lett. 86, 113901 (2005)
2. N. Rehse, S. Marr, S. Scherdel, and R. Magerle, Adv. Mater. 17, 2203 (2005)
3. N.H. Thomson, J. Microscopy 217, 193 (2005)
4. P. Samori, Chem. Soc. Rev. 34, 551 (2005)
5. G. Leveque, P. Cadet, and R. Arinero, Phys. Rev. B 71, 205419 (2005)
6. M. Tello, R. Garcia, J.A. Martín-Gago, N.F. Martinez, M.S. Martin-Gonzalez, L. Aballe, A. Barinov, and L. Gregoratti, Adv. Mater. 17, 1480 (2005).
7. C.M. Stroch, A. Ebner, M. Geretschlager, G. Freudenthaler, F. Kienberger, A.S.M. Kamruzzahan, S.J. Smith-Gil, H.J. Gruber and P. Hinterdorfer, Biophys. J 87, 1981 (2004)
8. O. Sahin, C.F. Quate, O. Sogaard, and A. Atalar, Phys. Rev. B 69, 165416 (2004)
9. T. Fukuma, K. Kimura, K. Kobayashi, K. Matsushige and H. Yamada, Appl. Phys. Lett. 85, 6287 (2004).
10. L. Chen, C.L. Cheung, P.D. Ashby, C.M. Lieber, Nano Lett. 4, 1725 (2005)
11. R.J. Colton, J. Vac. Sci. technol. B 22, 1609 (2004)
12. T.R. Rodriguez and R. Garcia, Appl. Phy. Lett. 449, 84 (2004)
13. S.I. Lee, S.W. Howell, A. Raman, R. Reifenberger, C.V Nguyen, M. Meyyappan, Nanotechnology 15, 416-421 (2004)
14. S.D. Solares, Y. Matsuda, W.A. Goddard, J. Phys. Chem. B 109 (35), 16658 (2005)
15. C. Su, L. Huang, K. Kjoller and K. Babcock, Ultramicroscopy 97, 135 (2003)

16. M.F. Yu, T. Kowalewski, R.S. Ruoff, Phys. Rev. Lett. 86, 87 (2001)
17. R. Garcia and R. Perez, Surf. Sci. Rep. 47, 197 (2002)
18. B. Pignataro, L. Chi, S. Gao, B. Anczykowski, C. Niemeyer, M. Adler and H. Fuchs, Appl. Phys. A 74, 447 (2002)
19. M. Stark, R. W. Stark, W. M. Heckl, and R. Guckenberger, Proc. Natl. Acad. Sci. USA 99, 8473 (2002).
20. G. Reiter, G. Castelein, J.U. Sommer, A. Röttele and T. Thurn-Albrecht, Phys. Rev. Lett. 87, 2261 (2001)
21. L. Nony, R. Boisgard, and J.P. Aimé, Biomacromolecules 2, 827 (2001)
22. J. Tamayo, ADL Humpris, and M. Miles, Appl. Phys. Lett. 77, 582 (2000)
23. R. Garcia and A. San Paulo, Phys Rev B 61, R13381 (2000)
24. R. Garcia and A. San Paulo, Phys. Rev. B 60, 4961 (1999)
25. A. Kühle, A.H. Soerensen and J. Bohr, J. Appl. Phys. 81, 6562 (1997)
26. A. S. Paulo and R. Garcia, Phys. Rev. B 66, 041406 (2002)
27. B. Anczykowski, J. Cleveland, D. Krüger, V. Elings and H. Fuchs, Appl. Phys. A 66, S885 (1998)
28. R. García, J. Tamayo, A. San Paulo, Surf. Interf. Anal. 27, 312 (1999)
29. N.F. Martinez and R. Garcia, Nanotechnology (2006)
30. T.R. Rodriguez and R. Garcia, Appl. Phys. Lett. (2002)
31. P. Martin, S. Marsaudon, J.P. Aimé, B. Bennetau, Nanotechnology 16, 901 (2005)
32. P.M. Hoffmann, S. Jeffery, J.B. Pethica, H. Ozgür, A. Oaral, Phys. Rev. Lett. 87, 265502 (2001)
33. T. Trevethan and L. Kantorovich, *N*anotechnology 16, S79 (2005)
34. E. Meyer, H.J. Hug and R. Bennewitz, Scanning Probe Microscopy Springer-Verlag Berlin (2004)
35. M. Kageshima, S. Takeda, A. Ptak, C. Nakamura, S.P. Jarvis, H. Tokumoto and J. Miyake, Jap J. Appl. Phys. 43, L1510 (2004)
36. A. Schirmeisen and H. Hölscher, Phys. Rev. B 72, 045431 (2005)
37. N. Sasaki and M. Tsukada, Jap. J. Appl. Phys. 39, L1334 (2000)
38. B. Gotsmann, C. Seidel, B. Anczykowski, and H. Fuchs, Phys. Rev. B 60, 11051 (1999)
39. H. Yoshizawa, Y.-L. Chen and J. Israelachvili, J. Phys. Chem. 97, 4128 (1993)
40. J. Tamayo and R. Garcia, Appl. Phys. Lett. 71, 2394 (1997)
41. J. Tamayo and R. Garcia, Appl. Phys. Lett. 73, 2926 (1998)
42. G. Bar, Y. Thomann, M.H Whangbo, Langmui*r* 14, 1219 (1998)
43. O.P. Behrend, L. Odoni, J.L. Loubet and N.A. Burnham, Appl. Phys. Lett. 75, 2551 (1999)
44. M. Stark, C. Möller, D.J. Müller and R. Guckenberger, Biophys. J. 80, 3009 (2001)
45. A. San Paulo and R. Garcia, Phys. Rev. B 64, 193411 (2001)
46. G. Kraush and R. Magerle, Adv. Mater. 14, 1579 (2002)
47. B. Bhushan and J. Qi, Nanotechnology 14, 886 (2003)
48. X. Chen, C.J. Roberts, J. Zhang, M.C. Davies and S.J.B. Tendler, Surf. Sci. 519, L593 (2002)
49. M. Balantekin and A. Atalar, Phys. Rev. B 67, 193404 (2003)
50. M.J. D'Amato, M.S. Markus, D.Y. Sasaki and R.W. Carpick, Appl. Phys. Lett. 85, 4738 (2004)

51. H. Bodiguel, H. Montes and C. Fretigny, Rev. Sci. Intrum. 75, 2529 (2004)
52. S. Crittenden, A: Raman, and R. Reifenberger, Phys. Rev. B 72, 235422 (2005).
53. P.D. Ashby and C.M. Lieber, J. AM. Chem. Soc. 127, 6814 (2005)
54. W. Wu, K. Matyjaszewski, and T. Kowalewski, Langmuir 21, 1143 (2005).
55. R.V. Martinez, F. Garcia, R. Garcia, E. Coronado, A. Forment-Aliaga, F.M. Romero and S. Tatay, Adv. Mater. (submitted).

18 Mechanisms of atomic scale dissipation at close approach in dynamic atomic force microscopy

Thomas Trevethan[1] and Lev Kantorovich[2]

[1] Dept. of Physics and Astronomy, Univ. College London, Gower Str., London,WC1E 6BT, U.K.
[2] Department of Physics, Kings College London, Strand, London WC2R 2LS, U.K.

18.1 Introduction

The Non-Contact Atomic Force Microscope (NC-AFM) has become an indispensable tool in nano and surface science that is now capable of routinely achieving atomic and even sub-atomic resolution on a wide variety of insulating as well as conducting surfaces [1, 2] in ultra high vacuum (UHV) conditions. In the NC-AFM, an atomically sharp tip attached to the end of a micro-cantilever is set oscillating at its resonant frequency above a surface in a direction perpendicular to the surface plane. Here, the force on the tip apex over the oscillation cycle is directly related to the change in resonant frequency of the cantilever, and this change can be measured as the surface is scanned, and can be used as the interaction signal. It is usual for NC-AFM experiments to be operated in the *constant frequency shift* mode, where the vertical position of the sample is modulated via a feedback mechanism to maintain the constant resonant frequency of the cantilever as the surface is scanned laterally, leading to the generation of a topography of constant frequency shift.

The cantilever in the NC-AFM is driven by an excitation signal via a feedback mechanism to maintain the cantilever oscillations at resonance and at a constant amplitude. This excitation signal can also be recorded during a scan (the so-called *damping* or *dissipation* image), and quite remarkably, shows atomic scale contrast that does not necessarily correlate directly with the topography or frequency shift [3]. The excitation signal can be related to the damping of the cantilever, suggesting that mechanical energy in the range of $0.01 - 1\,\mathrm{eV}$ is being dissipated to the surface via some *local* interaction. Note that this effect must have some local (atomic scale) nature as the dissipation signal changes on the atomic scale. The origin of these damping images has been the subject of some debate in recent years and several mechanisms have been proposed to explain them, with the hope that they may provide important additional information about surfaces and surface processes on the atomic scale. In this chapter we review the most relevant of these mechanisms to the NC-AFM and describe detailed realistic modelling performed to determine their contribution to the experimental observations.

18.1.1 Experimental Results

Atomic-scale variation in the damping of the cantilever was first observed in 1997 on the Si (111) 7×7 surface [4], and since then many other experiments have also found atomic scale contrast in damping images for various surfaces and under a variety of different conditions. Atomic-scale damping has been observed on semiconductor surfaces, such as Si (001) [5] and Si (111) [6]; a variety of ionic surfaces such as NaCl (001) [7,8], KBr (001) [9,10], CaF_2 [11] (111) and many others.

The corrugation in atomic resolution damping images of perfect flat surfaces typically lie in the range of 0.01 to 0.1 eV per tip oscillation cycle [3,7–9,12]; however, sometimes it can be much less than this (no contrast detected) even though the topography is clear [9,10]. Often significantly larger energies are observed, up to around 1 eV per cycle, over more complex and lower coordinated surface structures, such as defects, vacancies and step and corner edges [7,10,13]. The contrast patterns in the damping images show no generally repeatable, uniform correlation with either the frequency shift (Δf) or the topography, even though the surface periodicity and the positions of other surface features are clear. It has also been observed [3] that dissipation signal appears at tip-surface separations below some critical values at which the short-range chemical interaction between the tip and surface becomes dominant (several Å); this behaviour is opposite to the topography signal that increases gradually with the distance of closest approach.

Taking into account the experimental observations of atomic-scale damping in the NC-AFM, any mechanism proposed to explain it should conform to the following properties:

1. Any tip-surface interaction must be *structurally reversible* and *nondestructive*. The atomic structure of the tip and the surface must be the same on retraction of the tip from the surface as it was on approach for each oscillation cycle. This must be true for any repeatable true atomic-resolution imaging.
2. The damping must vary significantly over different lattice sites and be sensitive to atomic details. This must be true to account for the atomic-scale contrast observed in damping images.
3. The energy dissipated over a single oscillation of the tip, for every oscillation, must be comparable to experimentally observed values, i. e. 0.01 – 1 eV.
4. The damping is more sensitive to the separation at closest approach than the frequency shift.
5. The damping is more sensitive to the exact nature of the tip apex than the frequency shift.

18.1.2 Possible Mechanisms of Atomic Scale Damping

The damping signal in NC-AFM can be directly interpreted as *dissipation* of energy into the surface from the oscillating tip if it is assumed that the control electronics completely linearise the cantilever dynamics so that there are no transient effects that may lead to an apparent and non-physical damping. There are several mechanisms that can lead to energy dissipation from a tip moving relative to a surface in dynamic AFM, however very few are capable of explaining the atomic scale features of damping images observed in NC-AFM experiments.

There are several well established mechanisms of non-contact friction which may be relevant to the NC-AFM. These include Joule dissipation, mediated by an electric or magnetic field, and vacuum (or van der Waals) friction [14]. Joule dissipation occurs when both the tip and the sample are conductors, and currents are induced by the time varying non-uniform electric or magnetic fields generated by the oscillating tip and cantilever [15]. Although this mechanism can be the main source of dissipation in certain conducting or magnetic systems, it is a macroscopic effect and it is unclear how it could be sensitive to atomic details. Vacuum friction arises between any two bodies moving relative to each other [16–19] and is due to the fluctuating electromagnetic dipoles of atoms in the surface and tip, however estimates of its magnitude on the atomic scale [18, 20] show that it is far too small to contribute to the dissipation observed in the NC-AFM. At the same time, this may be the main dissipation mechanism at large tip-surface separations (over about 10 Å).

The two dissipation mechanisms that are widely considered most likely to be responsible for the observed dissipation in NC-AFM, and which are discussed in detail in the next two sections, are the *stochastic friction force* mechanism, which is due to the random thermal fluctuations of atoms in the tip-surface junction, and the *adhesion hysteresis* mechanism, which relies on a reversible structural change in the tip and/or surface on approach and retraction [14]. The two mechanisms of damping are considered in more detail in the following Sects. 18.2 and 18.3.

Note that both the adhesion hysteresis and stochastic friction mechanisms come out naturally from a *unified* non-equilibrium statistical-mechanical treatment [21] in which there is a *soft surface vibrational mode* either intrinsic to the surface (e. g. a flip-flop dimer on the Si(001) surface or a surface defect) or induced by the tip on approach. This way it is possible to derive an equation of motion for the tip that contains both the friction force proportional to the tip momentum P, and the time-dependent tip-surface force $F_{\mathrm{ts}}(Q, t)$ which may be not the same on approach and retraction for the *same* tip trajectory $Q(t)$.

The analysis of the relationship between the driving amplitude and the energy dissipated by the oscillating tip assumes that the cantilever is driven at resonance and that the response of the feedback system is perfect and

instantaneous. In the real experiment, however, the freedback system has a limited response time and the driving signal may not be exactly out of phase with the cantilever oscillations. These additional complexities to the operation of a real experiment may result in an apparent damping signal and thus an image, that is not related to dissipative tip-sample interactions, but is instead due to a transient effect from the finite response of the electronics and the non-linear tip-surface interaction [22]. To investigate this type of effects a complete numerical simulation of the NC-AFM experimental procedure is necessary; it will be discussed in more detail in Sect. 18.4.

18.2 Stochastic Friction Force Mechanism

The stochastic friction force mechanism, or 'Brownian particle' mechanism as it is sometimes referred to, is based on the idea that the moving tip in a NC-AFM experiment experiences a frictional force due to the thermal fluctuations of atoms in the tip-surface junction. In this theory, the tip is treated as a large Brownian particle interacting with the fast random force fluctuations generated by the microscopic thermal motion of atoms in the surface and tip, which act to slow the Brownian particle down, so that its average kinetic energy comes into equilibrium with the system temperature [14].

In this idea, which was first suggested in [23], a relatively slow moving tip (oscillation frequency $f_0 = 10^5$ Hz) samples the relatively fast fluctuating force due to thermal motion of atoms in the junction (vibrational frequencies $10^{12} - 10^{13}$ Hz) and that the resulting force on the tip is calculated as a statistical average over these force fluctuations. It was found that in addition to the conservative force in the static limit, the tip also experiences a frictional force, $-\xi P$, proportional, and acting in the opposite direction to, the tip momentum P. The resuling friction coefficient ξ was found proportional to the time integral of the force-force autocorrelation function, $\langle \Delta F_{\rm ts}(0) \Delta F_{\rm ts}(t) \rangle_{\rm eq}$, of the instantaneous force on the tip, $F_{\rm ts}(t)$. The obtained expression corresponds to a particular case of the fluctuation-dissipation theorem [24]. A more rigorous approach based on an entirely non-equilibrium consideration of the entire tip-surface system, that generalises the surface and tip, was presented for the first time in [25], which is based on the original microscopic non-equilibrium treatment of Brownian motion by Kirkwood in 1946 [26]. This derivation leads to a similar result for the tip friction coefficient which is a function only of the tip position above the surface:

$$\xi(Q) = \frac{1}{k_{\rm B}T} \int_0^\infty \langle \Delta F_{\rm ts}(0) \Delta F_{\rm ts}(t) \rangle_{\rm eq} \, {\rm d}t \, . \tag{18.1}$$

Here, the friction coefficient $\xi(Q)$ is dependent on the tip position Q above the surface. When the tip moves, energy is dissipated to surface phonons via the work done by the friction force. The same basic result has been found in

more detailed treatments that have included microscopic degrees of freedom in the tip as well as the surface [27–29] and quantum consideration for the surface phonons [30–32].

The velocity-dependent friction force results in a hysteresis in the force on the tip during approach and retraction [14, 25]. The friction coefficient, $\xi(Q)$, and thus the energy dissipated by the oscillating tip, will depend on the position of the tip above the surface, i. e. on its instantaneous velocity and trajectory. The energy dissipated over a single oscillation cycle will be [14, 33, 34]:

$$\Delta E = \oint \xi(Q)\dot{Q}\,\mathrm{d}Q = \int_0^{2\pi/\omega} \xi(Q))\dot{Q}^2\,\mathrm{d}t\,. \tag{18.2}$$

Several studies have been conducted to estimate the dependence of the damping coefficient on the tip-surface separation, $\xi(Q)$, using several methods and levels of complexity. In the initial study [23], the calculation assumed a single atom in the surface interacting with a single atom tip with a Lennard-Jones potential. In this case the displacement auto-correlation function of the surface atom can be calculated analytically using the harmonic approximation within the Debye model. These calculations have been extended in [35,36] to include several atoms in the surface interacting with several atoms in a rigid tip, also within the Debye model, however in all these calculations only very low frequency phonons near the $\omega = 0$ limit were considered. A more detailed calculation including induced local phonons and anharmonicity was performed in [28, 29]. Although formally the entire phonon frequency spectrum was taken into account, this treatment was limited to a single atom rigid tip interacting with a single atom in the surface.

The general result of all these studies, using the calculated $\xi(Q)$, Eq. (18.1) and typical experimental parameters for the tip oscillation, is that the energy dissipation due to this mechanism is approximately 10^{-6} eV per cycle at a distance of closest approach of 3 Å, which is five orders of magnitude less than typical experimental observations. In spite of these estimates it is important to properly account for any anharmonic effects and realistic atomistic structure of the tip-surface junction, which may substantially enhance the dissipation, and to determine the extent and form of any atomic scale contrast. This can only be done with a more explicit and realistic model for the NC-AFM system.

An important feature of the expression for the friction coefficient above is that it is an *equilibrium* ensemble average. Since this expression is an equilibrium property, it can be determined via an *equilibrium* Molecular Dynamics (MD) simulation for a fixed tip position above the surface. By repeating this calculation for a series of different tip positions above the surface, the friction coefficient can be evaluated as a function of tip position, $\xi(Q)$. Using this 'friction field' function and typical experimental parameters for the tip oscillation, it is possible to calculate the energy dissipated by the tip due to the stochastic friction force using Eq. (18.2).

In [37,38] $\xi(Q)$, and the energy dissipated by the moving tip, are evaluated for the MgO (001) and NaCl (001) surfaces interacting with the MgO tip using classical ionic pairwise potentials [39]. The atomistic system used for the evaluation of the tip force fluctuations consisted of an oxygen terminated MgO tip and a MgO (001) or NaCl (001) surface cluster. In these simulations the tip and surface cluster consisted of free and frozen regions with the atoms at the boundary of the free region coupled to a stochastic thermostat which is necessary to remove finite size effects [37]. The stochastic thermostat was employed with a damping parameter $\gamma = 10\,\mathrm{ps}^{-1}$ and a temperature $T_{\mathrm{ext}} = 300\,\mathrm{K}$. With the tip at a fixed height and lateral $x - y$ position with respect to the surface, an equilibrium simulation is run for $n = 500{,}000$ steps after an equilibration period of 10,000 steps using a time-step of $\Delta t = 1\,\mathrm{fs}$, giving a production run of 500 ps. During the simulation run, the instantaneous force on the tip is recorded continuously for every step in the simulation, using the tip force definition derived in [40]. The friction coefficient is then calculated by taking the time integral of force fluctuation autocorrelation function after it has converged according to Eq. (18.1).

This calculation was then repeated for a series of positions of the tip above each surface to generate the friction coefficient as a function of tip position. The dissipated energy was then calculated through a numerical integration of Eq. (18.2) for separate lateral points as a function of the distance of closest approach, $Q_0 - A_0$, where Q_0 is the average tip position above the surface during oscillations and A_0 is the oscillation amplitude. The parameters for the cantilever oscillation used in this calculation were $A_0 = 36\,\text{Å}$ and $f_0 = 158\,\mathrm{kHz}$ which are considered to be fairly typical of NC-AFM experiments [7]. Figure 18.1 shows the friction coefficient as a function of tip height and the dissipated energy as a function of the distance of closest approach for the O terminated MgO tip directly above a Na atom in the NaCl (001) surface. Below certain tip-surface separations ($< 5.5\,\text{Å}$ for the case of the NaCl system), jumps of individual atoms can occur between the tip and the

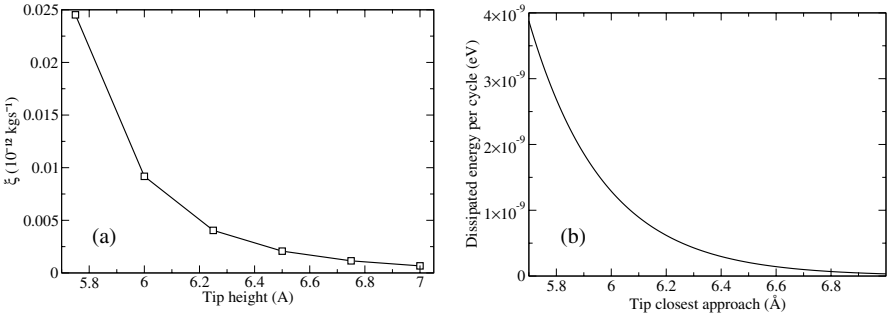


Fig. 18.1. Friction coefficient as a function of tip height (**a**) and dissipated energy per cycle as a function of closest approach (**b**) for the O terminated MgO tip directly above a Na atom in the NaCl surface

surface, which prevents the theory, and thus Eq. (18.1), from being strictly valid since all surface phonons are considered "fast" with respect to the tip oscillation itself. These effects can however contribute to the dissipation which is discussed in detail in the next section.

Although the dissipation due to the stochastic friction force mechanism is sensitive to atomic details on a surface, the dissipated energies for both the MgO and NaCl systems are far too small to account for experimental observations. The largest dissipated energies calculated in these studies were $\approx 4 \cdot 10^{-9}$ eV per cycle, which is almost 8 orders of magnitude smaller than typical experimental atomic scale contrast, but is comparable to some of the previous calculations, mentioned above, that employed simple analytical models. These results allow us to conclude that the stochastic friction force mechanism does not contribute to the formation of dissipation image contrast observed in NC-AFM experiments on stable crystal surfaces.

18.3 Adhesion Hysteresis

The adhesion hysteresis phenomena is well known in studies of contacts on the nano-scale, where below a critical distance, two surfaces snap into contact and require a stronger force upon separation to break the contact, which leads to a hysteresis in the force-separation curves (see, e. g. [41]). Adhesion of an AFM tip to a surface causes the snap-to-contact instability, and is well understood in terms of Johnson-Kendal-Roberts (JKR) theory [42]. This type of instability is however avoided in NC-AFM since the restoring force of the cantilever is much larger than the tip-surface interaction force over the entire tip trajectory. In addition, the tip-surface interaction only weakly perturbs the tip oscillations, and it is clear from frequency shift measurements and modelling that for atomic resolution imaging the tip-surface separation at close approach approximately lies in the range of $3-6$ Å. It is possible however that even if the trajectory of the tip is the same on retraction as it is on approach, the force acting on it may be different due a *reversible* microscopic structural change in the tip-surface junction [12, 25, 43].

It is well known that when two surfaces are separated by less than around two interatomic distances, significant displacements and 'jumps' of atoms can occur due to the onset of strong attractive interatomic interactions [44], and it has been demonstrated in realistic atomistic simulations that this type of process can occur when an AFM tip moves very close to a surface [39, 45]. This type of microscopic 'stick-slip' behaviour, that will most likely consist of a surface atom jumping to the tip from the surface, or vice versa, will lead to a closed hysteresis loop, and hence energy dissipation, if the system returns to its original configuration as the tip retracts (i. e. the atom jumps back).

This type of process, which can result from the superposition of the potential energy surfaces of the surface and tip with respect to a microscopic degree of freedom that characterises an instability in the junction (the *soft*

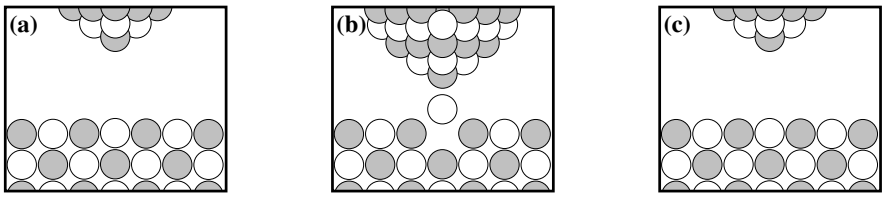


Fig. 18.2. Illustration of the development of a reversible atomic jump from the surface to the tip apex. **a** Approach, **b** the turning point of the trajectory and **c** retraction

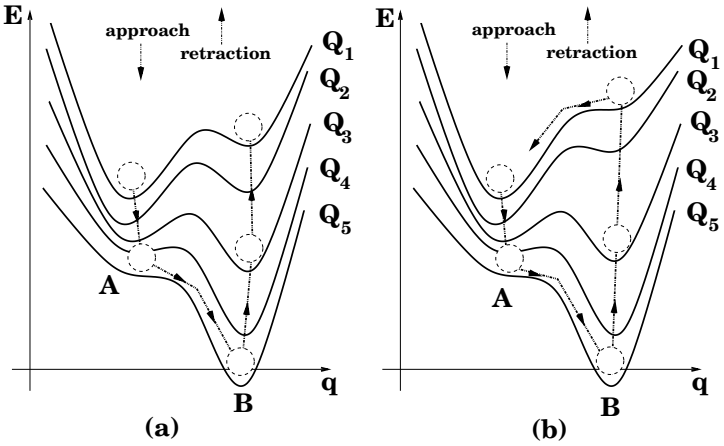


Fig. 18.3. Potential energy surfaces for a local microscopic degree of freedom experiencing instability for a series of tip heights $Q_1 < Q_2 < Q_3 < Q_4 < Q_4 < Q_5$: **a** a structurally reversible and **b** a structurally irreversible processes on tip retraction [14]. Reproduced with permission from Springer, Copyright (2002)

vibrational mode), is illustrated in Fig. 18.2. Figure 18.3 shows an illustration of the potential energy surfaces for the microscopic degree of freedom over a range of tip heights for two different scenarios. When the tip is far from the surface (Q_1), the system sits in the energy minimum A. As the tip approaches, a second energy minimum B becomes more energetically favourable, but the system remains in state A due to the barrier between them (Q_3). Then, as the tip moves closer, the barrier disappears (Q_5) and the system falls into state B (which may consist of a surface atom jump to the tip or vise versa). Then, as tip then retracts, the system remains in minimum B, until the barrier disappears (Q_1) and the system returns to state A (the atom jumps back to the surface), as illustrated in Fig. 18.3b. If the barrier does not dissappear on retraction, as is illustrated in Fig. 18.3a, then the system may remain in state B as the tip retracts and in this case the structural change is *irreversible* (the atom is permanently pulled away from the surface: a tip change).

The process 18.3b described above will lead to a different force on the tip during approach and retraction due to the different states occupied by the system at different stages in the tip trajectory. The work done by the tip will be equal to the area enclosed by the hysteresis curve and this energy will be dissipated into local phonons as the system moves from one state into the other. This is strictly a *statistically* irreversible process, even though it is *structurally* reversible.

The effect of this type of microscopic non-conservative process on the frequency shift and damping signal as a function cantilever height, Q_0, was considered for the first time in [43]. In this study, which used a time-averaging perturbation theory method at zero temperature, it was shown how the damping sets in as a step function and there is a discontinuity in the frequency shift when the distance of closest approach gets small enough that jumps start to occur.

It is clear that this mechanism of dissipation will be sensitive to the tip surface separation and to the atomic details of the surface and tip. In addition, the energies involved in typical atomic-scale stick-slip movements have been measured by Friction Force Microscopy (FFM) and are of the order of $0.1-1\,\mathrm{eV}$ [41]. The effect of a finite temperature will also be critical to this mechanism, since it will effect when the barriers are crossed and even which states are accessible.

18.3.1 Molecular Dynamics Simulations

To investigate the hysteresis for a realistic tip-surface system due to the type of atomic instabilities and structural changes described above, it is possible to perform non-equilibrium molecular dynamics simulations in which the tip is approached and retracted, and the dynamical response is determined in real time during the simulation run [46, 47]. In these simulations the frozen atoms in the tip are moved in concert, sinusoidally in the vertical direction at a given amplitude, A_0, and frequency, f, i. e. the tip height $Q(t) = Q_0 + A\sin(2\pi f)$. The coordinates of the tip atoms are incremented at each step in the simulation, and the hysteresis in the tip force can then be calculated by integrating the area enclosed in the continuous force-distance curves. The energy imparted to the system by the moving tip is removed via the stochastic boundary thermostat [37], which acts as the heat bath.

An example of this type of calculation is shown in Fig. 18.4 which shows the tip-force and the vertical z-coordinates of the tip apex atom and the atom directly below it, for a single approach and retract calculation for a Mg terminated MgO tip directly above an O atom in the CaO (001) surface. In this calculation $A_0 = 5\,\text{Å}$ the distance of closest approach $Q_0 - A_0 = 2.5\,\text{Å}$ and the temperature $T = 300\,\mathrm{K}$. The tip is oscillated at the unrealistically high frequency of $f = 500\,\mathrm{MHz}$ (resulting in an MD run of 1×10^6 steps) due to the computational expense of the calculations. It can be seen that as the tip approaches the surface, the O atom directly below it jumps to the

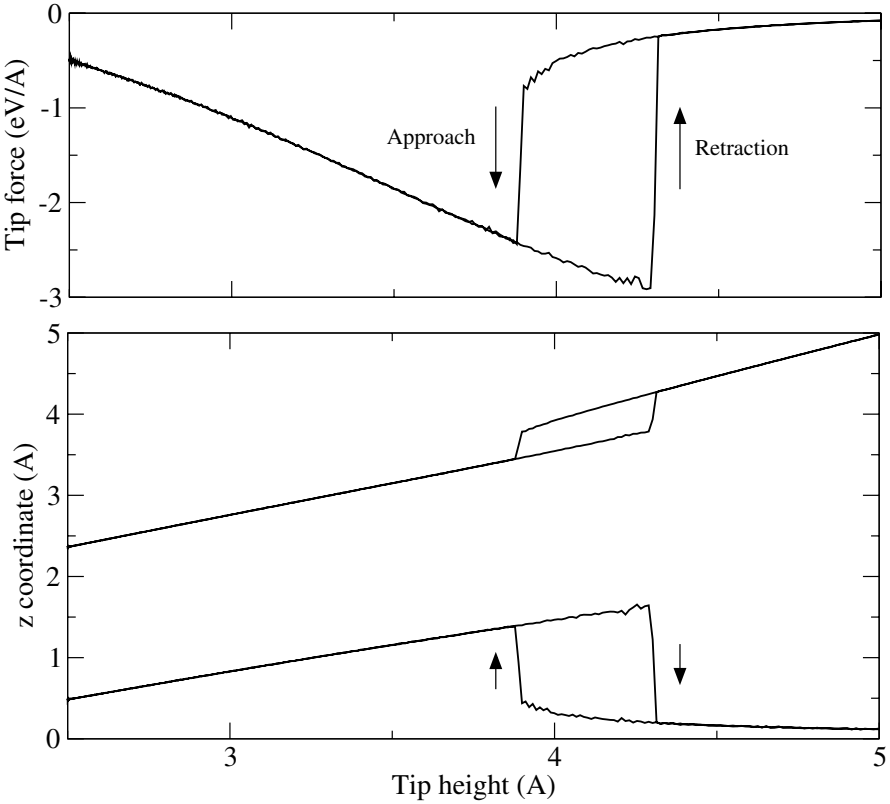


Fig. 18.4. Time smoothed tip force (*top*), tip apex atom z-coordinate (*middle*) and surface O atom z-coordinate for the Mg terminated tip approaching and retracting above an O atom in the CaO surface at 100 K

tip apex at a certain tip surface separation. Then, as the tip retracts from the surface, the atom finally jumps back at a different tip surface separation resulting in a hysteresis in the tip-surface force. The energy dissipated by the moving tip in this case is ≈ 1 eV.

The same type of behaviour is seen in calculations performed on several other systems [46–48]. We find that in general the formation of a single atomic jump is very sensitive to the lateral position and exact chemical identity of the tip apex.

18.3.2 Theory of Dynamical Response

MD simulations have shown how a hysteresis in the tip-surface force can arise due to the system moving between two minima on the microscopic potential energy surface as the tip approaches and retracts. In this type of process the potential energy surface changes as a function of tip height, and at certain

separations displays two minima separated by a barrier which the system can move between.

Note, however, that MD simulations as described in the previous section are quite expensive to be used routinely for images calculations. A different, although closely related method, that implements the adhesion hysteresis mechanism using transition state theory and calculated potential energy surfaces, has been suggested in [49] for practical calculations of the topography and dissipation images of real systems at various temperatures.

To determine the potential energy surface for an atomistic system, a constrained minimisation technique can be used to find the adiabatic minimum energy path between the two minima for a given tip position [45]. For a given lateral position of the tip relative to the surface, this procedure for determining the minima and barrier, is then repeated over a series of fixed tip heights resulting in a sequence of potential energy curves. Figure 18.5 shows potential energy curves for a series of tip heights as a function of the z-coordinate of an O atom (the microscopic degree of freedom associated with the instability) in the CaO (001) surface directly below the Mg apex of the MgO tip. Here it can be seen that when the tip is close to the surface only a single state is accessible, but when the tip is further from the surface there are two states separated by a barrier.

From the curves like those in Fig. 18.5, it is possible to determine the energies of the minima and the barrier as a function of tip height, $E_{\mathrm{A}}(Q), E_{\mathrm{B}}(Q)$ and $E_{\mathrm{T}}(Q)$, which are shown in Fig. 18.6, and the oscillation pre-factors around each minimum $\nu_{\mathrm{A}}(Q)$ and $\nu_{\mathrm{B}}(Q)$. Here it can be seen how the system, initially in state A, will fall into state B as the tip approaches the surface and the barrier vanishes. As the tip retracts, the barrier to return the state A remains, however at a finite temperature the system may be able to cross the barrier and return to its original state.

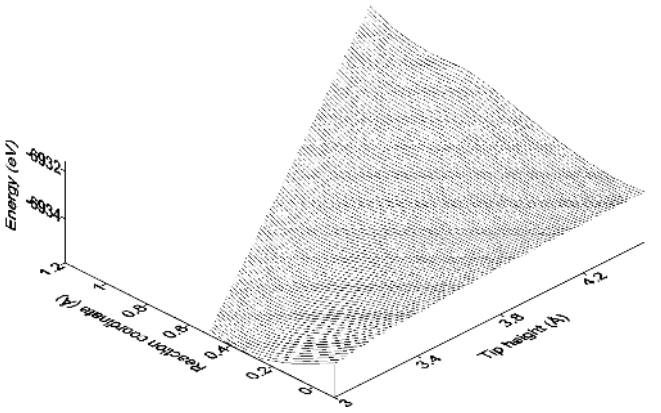


Fig. 18.5. Potential energy curves for a series of tip heights as a function of the z-coordinate of an O atom in the CaO (001) surface, directly below the Mg atom apex of the MgO tip

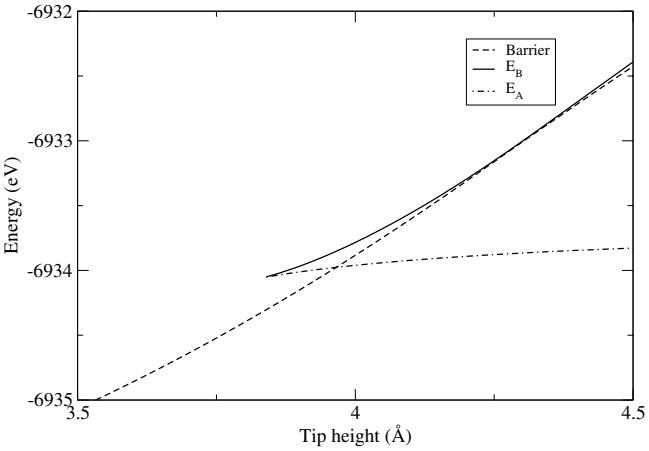


Fig. 18.6. Energies of the two minima $E_{\mathrm{A}}(Q)$ and $E_{\mathrm{B}}(Q)$, and the barrier separating them, $E_{\mathrm{T}}(Q)$, as a function of tip height Q for the Mg terminated MgO tip directly above an O atom in the CaO (001) surface

With details of the energy states and barriers as a function of tip height above the surface, the dynamical response of the system can be determined using the method proposed in [49]. The rate of the system jumping from state A to B, $W_{\mathrm{AB}}(Q)$, and from B to A, $W_{\mathrm{BA}}(Q)$, at any given tip height can be determined from a simplified classical transition state theory [50]. Here, the rate is a function of the tip height above the surface and is determined from the energy barrier, the oscillation pre-factor and the system temperature:

$$W_{\mathrm{AB}}(Q) = \nu_{\mathrm{A}}(Q) \exp\left[-\beta \Delta E_{\mathrm{TA}}(Q)\right] . \tag{18.3}$$

Here, $\beta = 1/k_{\mathrm{B}}T$ is the inverse temperature. The heights of the barriers, $\Delta E_{\mathrm{TA}}(Q) = E_{\mathrm{T}}(Q) - E_{\mathrm{A}}(Q)$ and $\Delta E_{\mathrm{TB}}(Q) = E_{\mathrm{T}}(Q) - E_{\mathrm{B}}(Q)$, can be determined from the microscopic potential energy surface and represented for each lateral tip position by the curves similar to those shown in Fig. 18.6), as functions of tip height above the surface.

The probability of the system being in state A at time t is $P_{\mathrm{A}}(t)$, and the probability that the system is in state B at time t is $P_{\mathrm{B}}(t) = 1 - P_{\mathrm{A}}(t)$. Since the probability of the system staying in state A in the time between t to $t + \mathrm{d}t$ is $P_{\mathrm{A}}(t)(1 - W_{\mathrm{AB}}\,\mathrm{d}t)$ and the probability of it returning to A if it was in B is $P_{\mathrm{B}}(t)W_{\mathrm{BA}}\,\mathrm{d}t$, the evolution of $P_{\mathrm{A}}(t)$ in time is determined by the following "Master" equation:

$$\frac{\mathrm{d}P_{\mathrm{A}}(t)}{\mathrm{d}t} = -P_{\mathrm{A}}(t)W_{\mathrm{AB}}(Q) + P_{\mathrm{B}}(t)W_{\mathrm{BA}}(Q) \tag{18.4}$$

and similarly for the evolution of state B. If it is assumed that in a given cycle when the tip is initially far from the surface, the system is in state A,

i. e. $P_{\mathrm{A}}(0) = 1$, then Eq. (18.4) can be solved analytically to give:

$$P_{\mathrm{A}}(t) = \mathrm{e}^{-h(t)} \left[1 + \int_0^t W_{\mathrm{BA}}(\tau)\, \mathrm{e}^{h(\tau)}\, \mathrm{d}\tau \right] \tag{18.5}$$

where $h(t) = \int_0^t (W_{\mathrm{AB}}(\tau) + W_{\mathrm{BA}}(\tau))\, \mathrm{d}\tau$. With these expressions, the dynamical response of the system can be determined as a function of time, since $W_{\mathrm{AB}}(Q)$ and $W_{\mathrm{BA}}(Q)$ are changed in real time. The microscopic force on the tip can then be calculated at a given time-step as the weighted sum of the forces acting on the tip in each state, $F_{\mathrm{m}}(Q,t) = P_{\mathrm{A}}(t)F_{\mathrm{A}}(Q) + P_{\mathrm{B}}(t)F_{\mathrm{B}}(Q)$. The force on the tip in each state is proportional to the gradient of the energy, i. e. $F_{\mathrm{A}}(Q) = -\frac{\partial E_{\mathrm{A}}(Q)}{\partial Q}$. The total force on the tip is then determined by adding a macroscopic van der Waals force to the microscopic force to give

$$F_{\mathrm{ts}}(Q,t) = [P_{\mathrm{A}}(t)F_{\mathrm{A}}(Q) + P_{\mathrm{B}}(t)F_{\mathrm{B}}(Q)] + F_{vdW}(Q)\,. \tag{18.6}$$

Assuming a sinusoidal trajectory for the tip motion [51], the evolution of P_{A}, and hence P_{B}, is integrated numerically with a finite time-step over a single oscillation cycle. With the same amplitude and frequency as in the previous section, a calculation was performed with the energies and prefactors above for the Mg terminated tip directly above an O atom in the CaO surface. Figure 18.7a shows P_{A} on approach and retraction as a function of the tip height at the temperature $T = 100\,\mathrm{K}$. The corresponding tip-force is shown in Fig. 18.7b, where a macroscopic tip radius of $R = 30\,\mathrm{nm}$ has been used for the macroscopic van der Waals force.

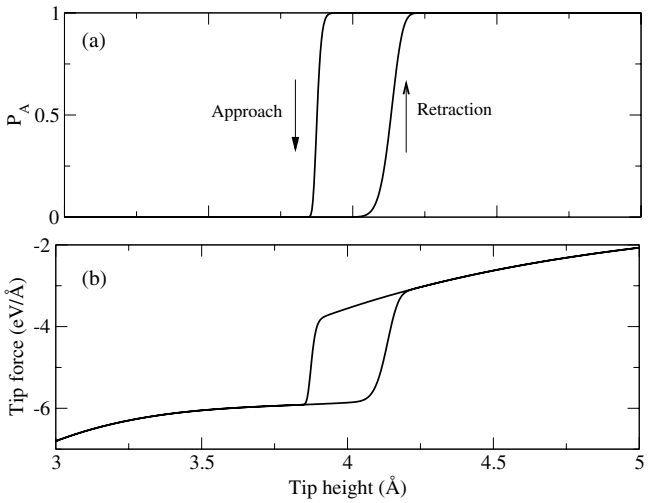


Fig. 18.7. Evolution of P_{A} (**a**) and the tip-force (**b**) as a function of tip height over a single oscillation cycle for the Mg terminated tip directly above an O atom in the CaO surface at 100 K

The frequency shift and the amplitude of the driving signal (which is directly linked to the hysteresis) can be determined if the tip trajectory and driving signal are assumed to be sinusoidal. The equation of motion for the tip can be solved in this case to give (assuming $Q = Q_0 - A_0 \sin \omega t$):

$$\omega^2 - \omega_0^2 = -\frac{\omega}{\pi A_0 m} \int_0^{2\pi/\omega} F_{\mathrm{ts}}(Q_0 - A_0 \sin \omega t, t) \sin(\omega t)\, \mathrm{d}t \tag{18.7}$$

$$A_{\mathrm{exc}} = \frac{\pi m \omega_0 A_0 \omega}{Q_{\mathrm{c}}} - \frac{\omega}{\pi} \int_0^{2\pi/\omega} F_{\mathrm{ts}}(Q_0 - A_0 \sin \omega t, t) \cos(\omega t)\, \mathrm{d}t \tag{18.8}$$

where Q_{c} is the quality factor and m effective mass of the tip.

The above equation for the frequency shift, $\Delta f = (\omega - \omega_0)/2\pi$, cannot be evaluated directly as in the case of a conservative tip-surface force [52, 53], since now the integral over the tip trajectory is dependent on ω itself. This equation is in fact a transcendental algebraic equation that must be solved simultaneously and self-consistently with Eq. (18.6), which is performed numerically. The frequency shift and energy dissipation as a function of the distance of closest approach are shown in Fig. 18.8 for the Mg terminated tip above a O atom in the CaO surface at $T = 100\,\mathrm{K}$. In this calculation $f_0 = 60\,\mathrm{kHz}, Q_{\mathrm{c}} = 10{,}000$ and $k_{\mathrm{c}} = m\omega_0^2 = 6\,\mathrm{Nm}^{-1}$. In the frequency shift curve, a small jump occurs when the tip comes close enough for the jumping to start and the dissipation behaves like a step function with the onset of jumping.

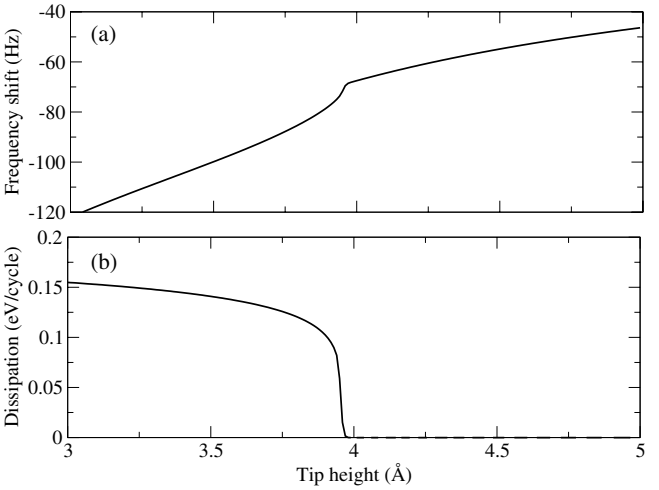


Fig. 18.8. Frequency shift (*top*) and dissipated energy (*bottom*) as a function of the distance of closest approach for the Mg terminated MgO tip directly above an O atom in the CaO surface at 100 K

The theory of dynamical response in NC-AFM can realistically describe the imaging process given parameters from atomistic calculations. The tip surface system moves between two minima, and this leads to a hysteresis in the tip surface force consistent with the more explicit MD simulations. The energy dissipated by the moving tip in the NC-AFM can be calculated, along with the frequency shift, accurately and with relatively small computational expense. The values for the dissipated energies, and behaviour of the model systems are consistent with experimental observations of similar systems.

This approach has now been applied to several systems and by determining the potential energy surfaces for many non-equivalent lateral positions, dissipation images of the model CaO (001) surface have been produced [54]. Due to the way the potential energy states and barriers of the tip-surface system change as the tip lateral position is shifted, and the dependence of the size of the hysteresis on the system temperature, it has been shown how the contrast pattern changes in different temperature and distance regimes [54]. In the MgO-CaO (001) system described above, the dissipation can be either correlated or anti-correlated with the topography depending on the temperature and frequency shift. Experimentally the dissipation images of binary ionic surfaces are often either correlated, anti-correlated or phase-shifted with respect to the topography [3, 9].

Due to the structure and symmetry of the MgO tip used in these calculations, the tip apex atom is very stable in the z direction and does not experience a significant displacement when a bond is formed with the surface, hence it is a surface atom that experiences considerable displacement. With different tip configurations and orientations however, it may be the tip apex atom (or some other degree of freedom in the tip) that is primarily involved in the jumping and therefore the formation of contrast in the dissipation image. Although the general physics should be the same as for the processes investigated here, the contrast patterns in the dissipation may be very different with respect to the topography and this should be investigated further.

18.4 Apparent Damping

The results of the calculations performed in the previous section can be compared to the damping measured experimentally subject to the simplifying assumptions that the tip follows a sinusoidal trajectory and that the feedback loops that control the amplitude and surface position work ideally and instantaneously. However, in the real experiment the finite response of the instrumentation of the system may effect the signals and images obtained. Specifically, it has been suggested [22], that the damping signal observed in NC-AFM is not due to a non-conservative tip surface interaction, but due to an interplay between the tip motion and the finite response of the feedback control loops.

Previous extensive numerical modelling has shown that the sinusoidal approximation to the tip trajectory is appropriate and accurate [55] at a large enough amplitude, with a conservative tip-surface interaction, over a wide range of experimental parameters. However, a consideration of the dynamical response of the entire experimental apparatus is essential to understanding the formation of real images and justifying the simplifying approximations. To this effect, the concept of a Virtual Atomic Force Microscope (VAFM) was introduced [56], that consists of an explicit real-time numerical simulation of the entire experimental set-up of a realistic NC-AFM system, using a known tip-surface interaction.

The basic design of the VAFM is based on the experimental set-up typical to most modern NC-AFM instruments. The numerical description of the VAFM is described in detail in [56] and has been implemented in a self contained simulation code [57] that operates using a finite time-step integration scheme, with the velocity Verlet algorithm [58] propagating the cantilever dynamics and hence the tip trajectory. The VAFM is operated in a very similar way to an actual experiment, where initially the cantilever is allowed to reach a steady state away from the surface before it is approached to the surface to reach the desired frequency shift set-point, after which imaging can begin. The tip moves in a continuous three-dimensional force-field due to the surface, derived (and interpolated) from realistic atomistic modelling, which can be moved laterally in real time during the simulation. This allows for a totally realistic description of the experiment, including the effect of lateral scanning velocity and the associated behaviour of the feedback mechanisms.

The theory of dynamical response as described in Sect. 18.3.2 was incorporated into the VAFM so the response of the the instrument to a non-conservative force-field can be studied. Our recent calculations on the CaO (001) surface imaged with the MgO tip show that atomic scale damping contrast is directly due to a non-conservative tip-surface interaction and not an artifact of the control systems.

The calculations have shown how image contrast is formed using a realistic non-conservative tip-surface interaction. The contrast in the damping image with an entirely *conservative* tip-surface force-field (apparent damping) is typically less than 1 meV, which is much too small to account for experimental observations. When a realistic hysteresis is included in the tip-surface interaction, i. e. a non-conservative interaction via the theory of dynamical response considered in Sect. 18.3.2, the contrast in the damping image shows strong corrugation consistent with the work done by the hysteresis of the tip-surface force. The calculations give dissipated energies that compare very well with the energies observed experimentally for similar surfaces. These results show that the atomic scale damping contrast observed experimentally is almost entirely due to a physical dissipative tip-surface interaction and is not an artifact.

18.5 Conclusion and Discussion

The possible mechanisms for the origin of the damping have been studied for realistic model systems, and only a reversible structural change in the tip-surface junction at close approach is capable of explaining the magnitude of the dissipation and the contrast patterns observed experimentally.

This structural change can arise from the system moving between two minima on the potential energy surface of the microscopic junction as the tip approaches and retracts from the surface. The formation of the soft mode (instability) and dynamical response of this type of process is very sensitive to the exact chemical nature of the tip apex and the surface, as well as the temperature of the system and the tip vertical velocity.

Calculations performed using the theory of dynamical response at close approach have showed how the evolution of the system can be accurately described using the transition rates between states as a function of tip position above the surface. This theory means that the dynamical response and hysteresis can be determined with relatively little computational expense and therefore allow comprehensive spectroscopy and image calculations to be performed given details of the microscopic potential energy surface as a function of tip position. This theory of dynamical response will be essential to calculations performed in order to understand mechanisms of controlled atomic scale manipulation in scanning probe microscopy, and may lead to new insights into how this can be achieved in NC-AFM.

The Virtual AFM, which explicitly simulates the operation of the entire NC-AFM instrument has been used to demonstrate that the damping signal observed in experiments is due to a non-conservative interaction in the tip-surface junction and is not an artifact of the imperfect response of the control loops.

The calculations presented here are general and meant to provide an insight into the dissipation process and image formation. A method has now been developed to model the dissipation imaging which is a basis for any future study of damping images. However a quantitative direct comparison with experimental images is now required in order for this mechanism and method to reach maturity and true acceptance in the field. The achievement of this is the next step for this research and can only be achieved by studying and modelling a wider variety of systems.

References

1. S. Morita, R. Wiesendanger, and E. Meyer, editors, *Non-Contact Atomic Force Microscopy*, Nanoscience and Technology (Springer, 2002).
2. F. J. Giessible, Rev. Mod. Phys. **75**, 957 (2003).
3. H. J. Hug and A. Baratoff, *Non-Contact Atomic Force Microscopy* Nanoscience and Technology (Springer, 2002), chap. 20. Measurement of dissipation induced by tip-sample interactions.

4. R. Lüthi et al., Surf. Rev. Lett. **4**, 1025 (1997).
5. O. Ozer, M. Atabak, and A. Oral, Appl. Surf. Sci. **210**, 12 (2003).
6. M. Guggisberg et al., Surf. Science **461**, 255 (2000).
7. R. Bennewitz et al., Phys. Rev. B **62**, 2074 (2000).
8. M. Bammerlin et al., Appl. Phys. A **66** (1998).
9. R. Hoffman, Magnetic and interatomic forces measured by low temperature scanning force microscopy, Master's thesis, Institute of Physics, University of Basel, 2001.
10. C. Loppacher et al., Phys. Rev. B **62**, 13674 (2000).
11. C. Barth, A. Foster, M. Reichling, and A. Shluger, J. Phys.: Condens. Matter **13**, 2061 (2001).
12. P. Hoffmann, S. Jeffery, J. Pethica, H. Ozgur, and A. Oral, Phys. Rev. Lett. **87**, 265502 (2001).
13. R. Bennewitz and E. Meyer, editors, *Non-Contact Atomic Force Microscopy* Nanoscience and Technology (Springer, 2002), chap. 4. Alkali Halide Surfaces.
14. M. Gauthier, L. Kantorovich, and M. Tsukada, *Non-Contact Atomic Force Microscopy* Nanoscience and Technology (Springer, 2002), chap. 19. Theory of energy dissipation into surface vibrations.
15. W. Denk and D. W. Pohl, Appl. Phys. Lett. **59**, 2171 (1991).
16. I. Dorofeyev, H. Fuchs, G. Wenning, and B. Gotsmann, Phys. Rev. Lett. **83**, 2402 (1999).
17. J. B. Pendry, J. Phys.: Condens. Matter **9**, 10301 (1997).
18. B. N. J. Persson and A. I. Volokitin, Phys. Rev. Lett. **84**, 3504 (2000).
19. A. I. Volokitin and B. N. J. Persson, Phys. Rev. Lett. **94**, No. 086104 (2005).
20. I. Dorofeyev, H. Fuchs, and B. Gotsmann, Phys. Rev. Lett. **84**, 3505 (2000).
21. T. Trevethan and L. Kantorovich, Nanotechnology **15**, S44 (2004).
22. M. Gauthier and M. Tsukada, Phys. Rev. Lett **85**, 5348 (2000).
23. M. Gauthier and M. Tsukada, Phys. Rev. B **60**, 11716 (1999).
24. D. Zubarev, V. Morozov, and G. Ropke, *Statistical mechanics of nonequilibrium processes. Vol. 2: Relaxation and hydrodynamic processes* (Akademie Verlag, Berlin, 1996).
25. L. Kantorovich, J. Phys. Cond. Matter **13**, 945 (2001).
26. J. Kirkwood, J. Chem. Phys. **14**, 80 (1946).
27. L. N. Kantorovich, Phys. Rev. B **64**, 245409 (2001).
28. L. N. Kantorovich, J. Phys.: Cond. Matter **14**, 4329 (2002).
29. L. Kantorovich, Appl. Surf. Science **210**, 27 (2003).
30. L. N. Kantorovich, Phys. Rev. Lett. **89**, 096105 (2002).
31. L. N. Kantorovich, J. Phys.: Cond. Matter **14**, 7123 (2002).
32. L. N. Kantorovich, Surface Science **521**, 117 (2002).
33. B. Anczykowski, B. Gotsmann, H. Fuchs, J. P. Cleveland, and V. B. Elings, Appl. Surf. Science **140**, 376 (1999).
34. B. Gotsmann, C. Seidel, B. Anczykowski, and H. Fuchs, Phys. Rev. B **60**, 11051 (1999).
35. M. Gauthier, Theory of noncontact dissipation force microscopy, Master's thesis, Graduate School of Science, Dept. of Physics, Univ. of Tokyo, 2000.
36. M. Gauthier and M. Tsukada, Surface Science **495**, 204 (2001).
37. T. Trevethan and L. Kantorovich, Phys. Rev. B **70**, 115411 (2004).
38. T. Trevethan, *Theory of atomic scale dissipation in non-contact atomic force microscopy*, PhD thesis, University of London, 2006.

39. A. L. Shluger, A. L. Rohl, D. H. Gay, and R. T. Williams, J. Phys.: Condens. Matter **6**, 1825 (1994).
40. T. Trevethan and L. Kantorovich, Surface Science **540**, 497 (2003).
41. E. Gnecco et al., Phys. Rev. Lett. **84**, 1172 (2000).
42. K. Johnson, K. Kendal, and A. Roberts, Proc. R. Soc. A **324**, 301 (1971).
43. N. Sasaki and M. Tsukada, Jpn. J. Appl. Phys. **39**, L1334 (2000).
44. J. Israelachvili, *Intermolecular and Surface Forces* (Academic Press, London, 1991).
45. A. L. Shluger, L. N. Kantorovich, A. I. Livshits, and M. J. Gillan, Phys. Rev. B **56**, 15332 (1997).
46. T. Trevethan and L. Kantorovich, Nanotechnology **15**, S34 (2004).
47. T. Trevethan and L. Kantorovich, Nanotechnology **16**, S79 (2005).
48. A. Abdurixit, A. Baratoff, and E. Meyer, Appl. Surf. Science **157**, 355 (2000).
49. L. Kantorovich and T. Trevethan, Phys. Rev. Lett. **93**, 236102 (2004).
50. G. H. Vineyard, J. Phys. Chem. Solids **3**, 121 (1957).
51. W. A. Hofer, A. S. Foster, and A. L. Shluger, Rev. Modern Phys. **75**, 1287 (2003).
52. R. Garcia and R. Perez, Surf. Science Rep. **47**, 197 (2002).
53. F. J. Giessibl, Phys. Rev. B **56**, 16010 (1997).
54. T. Trevethan and L. Kantorovich, Nanotechnology, to be published (2006).
55. L. Nony, R. Boisgard, and J. Aime, J. Chem. Phys. **111**, 1615 (1999).
56. J. Polesel-Maris and S. Gauthier, J. Appl. Phys. **97**, 1 (2005).
57. L. N. Kantorovich, J. Polesel-Maris, and T. Trevethan, Virtual atomic force mircoscope simulation code, unpublished.
58. M. P. Allen and D. J. Tildesley, *Computer simulation of liquids* (Clarendon Press, Oxford, 2002).

19 Theory of Noncontact Friction

A.I. Volokitin[1,2] and B.N.J. Persson[1]

[1] Institut für Festkörperforschung, Forschungszentrum Jülich, D-52425, Germany
[2] Samara State Technical University, 443100 Samara, Russia

19.1 Introduction

In the history of physics, studies of the thermal radiation from materials have played a very important role. In this context, we only need mention that quantum mechanics originated from attempts to explain paradoxical experimental results related to black body radiation. In the past, the nonradiative near-field part of electromagnetic radiation usually was ignored, because it plays no role in the far-field properties of emission from planar sources. Nevertheless, recent interest in microscale and nanoscale radiative heat transfer [1–6], together with the development of local-probe thermal microscopy [7,8], has raised new challenges. These topics and recent progress in detecting noncontact friction force at the sub-attonewton level [9–13], as well as the observation of coherent thermal emission from doped silicon and silicon carbide (SiC) gratings [14], are linked in that the nonradiative (evanescent) thermal field plays a substantial role in each.

A great deal of attention has been devoted to the problem of the frictional drag force between two-dimensional quantum wells [15–17] and the friction force between an atomic force microscope tip and a substrate [9–13], because of the importance of noncontact friction for ultrasensitive force detection experiments. This is because the ability to detect small forces is inextricably linked to friction via the fluctuation–dissipation theorem. According to this theorem, the random force that makes a small particle jitter would also cause friction if the particle were to be dragged through the medium. For example, the detection of single spins by magnetic resonance force microscopy [18], which has been proposed for three-dimensional atomic imaging [19] and quantum computation [20], will require force fluctuations (and consequently the friction) to be reduced to unprecedented levels. In addition, the search for quantum gravitation effects at short length scales [21], as well as future measurements of dynamical Casimir forces [22], may eventually be limited by noncontact friction effects.

In noncontact friction, the bodies are separated by a potential barrier thick enough to prevent electrons or other particles with a finite rest mass from tunneling across it, but allowing interaction via the long-range electromagnetic field, which is always present in the gap between bodies and can

have different origins. The presence of an inhomogeneous tip–sample electric field is difficult to avoid, even under the best experimental conditions [11]. For example, even if both the tip and the sample were metallic single crystals, the tip would still have corners, and more than one crystallographic plane would be exposed. The presence of atomic steps, adsorbates and other defects also contribute to the spatial variation of the surface potential. This is referred to as the *patch effect*. The surface potential can also be easily changed by applying a voltage between the tip and the sample. An inhomogeneous electric field can also be created by charged defects embedded in a dielectric sample. The relative motion of the charged bodies will produce friction, which is termed the *electrostatic friction*.

The electromagnetic field can also be created by the fluctuating current density that results from thermal and quantum fluctuations inside the solids. This fluctuating electromagnetic field gives rise to the well-known long-range attractive van der Waals interaction between two bodies [23], and is responsible for radiative heat transfer. If the bodies are in relative motion, the same fluctuating electromagnetic field will give rise to a friction which is frequently called *the van der Waals friction*.

The origin of the van der Waals friction is closely connected with the van der Waals interaction. The van der Waals interaction arises when an atom or molecule spontaneously develops an electric dipole moment due to a quantum fluctuation. The short-lived atomic polarity can induce a dipole moment in a neighboring atom or molecule some distance away. The same is true for extended media, where thermal and quantum fluctuations of the current density in one body induce a current density in the other body: the interaction between these current densities is the origin of the van der Waals interaction. When two bodies are in relative motion, the induced current will lag slightly behind the fluctuating current inducing it, and this is the origin of the van der Waals friction. The van der Waals interaction is mostly determined by an exchange of virtual photons between the bodies (connected with quantum fluctuations), and does not vanish, even at zero temperature. In contrast, the van der Waals friction, at least to the lowest order of perturbation theory and to linear order in the sliding velocity, is determined by the exchange of real photons, and vanishes at zero temperature.

To clarify the origin of the van der Waals friction, let us consider two flat parallel surfaces, separated by a sufficiently wide vacuum gap, which prevents electrons from tunneling across it. If the surfaces are in relative motion (velocity v) a frictional stress will act between them. This frictional stress is related to an asymmetry of the reflection coefficient along the direction of motion; see Fig. 19.1. If one body emits radiation, then these waves are Doppler-shifted in the rest reference frame of the second body, which will result in different reflection coefficients. The same is true for radiation emitted by the second body. The exchange of "Doppler-shifted photons" is the origin of van der Waals friction.

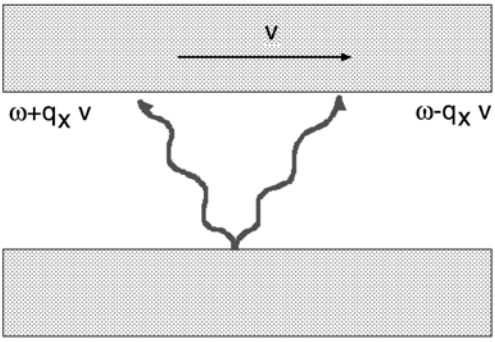


Fig. 19.1. The electromagnetic waves emitted in the opposite direction by one body will experience the opposite Doppler shift in the moving reference frame in which the second body is at rest. Due to the frequency dispersion of the reflection coefficient, these electromagnetic waves reflect differently from the surface of the second body, and this will give rise to the transfer of momentum between the bodies. The origin of van der Waals friction is associated with this momentum transfer

From the point of view of quantum mechanics, the van der Waals friction originates from two types of processes. (a) Photons are created in each body with opposite momenta, and the frequencies of these photons are connected by $vq = \omega_1 + \omega_2$, where q is the momentum. (b) A photon is annihilated in one body and created in another. The first process (a) is possible even at zero temperature, and it gives rise to a friction force which depends cubically on the sliding velocity [24, 25]. The second process (b) is possible only at finite temperatures, and gives rise to a friction which depends linearly on the sliding velocity. Thus, process (b) will provide the main contribution to the friction at sufficiently high temperatures, and at not too large velocities.

In contrast to the van der Waals interaction, for which theory is well-established, the field of van der Waals friction is still controversial. As an example, different authors have studied the van der Waals friction between two flat surfaces in parallel relative motion using different methods, and obtained results which are in sharp contradiction to each other. The first calculation of van der Waals friction was done by Teodorovich [26]. Teodorovich assumed that the force of friction can be calculated as the ordinary van der Waals force between bodies at rest, whose dielectric function depend on the velocity due to the Doppler shift. However, it follows [25] from the dynamical modification of the Lifshitz theory of the van der Waals interaction [27] that this is not true. Later, the same approach was used by Mahanty [28] to calculate the friction between molecules. Both theories predict wrong nonzero friction to linear order in the sliding velocity at absolute zero of temperature. The same nonzero friction at zero temperature was predicted in [29, 30]. In [31], it was shown that the basic equation in [29, 30] is incorrect, and a correct treatment gives a vanishing linear friction at $T = 0$ K. Schaich and Harris

developed a theory [32] which correctly describes the dependence of friction on the temperature T and on the separation d. However, in their calculations they made a series of unphysical approximations, and as a result their final formula for the friction for parallel relative motion has an incorrect dependence on the reflection coefficient. In particular, for the Drude model this formula gives a divergent result. The friction obtained in [33–35] vanishes in the limit of infinite light velocity $c \to \infty$. This result is incorrect because, at least for short distances, one can neglect retardation effects when calculating the van der Waals friction, as well as van der Waals interactions. Pendry [24] obtained the correct formula for friction at zero temperature, in which case the friction depends cubically on the velocity, and retardation effects are neglected. Persson and Zhang [36] obtained the correct formula for friction in the limit of small velocities and finite temperature, again neglecting retardation effects. In [25], Volokitin and Persson developed a theory of the van der Waals friction based on the dynamical modification of the well-known Lifshitz theory [27] of van der Waals interaction. In the nonretarded limit and for zero temperature this theory agrees with the results of Pendry [24]. Similarly, in the nonretarded limit and for small sliding velocity this theory agrees with the study of Persson and Zhang [36]. In [37, 38], the theory was extended to two flat surfaces in normal relative motion. It was shown that the van der Waals friction can be greatly enhanced if the surfaces of the bodies can support localized surface modes like surface plasmon polaritons, surface phonon polariton or adsorbate vibrational modes. Recently it was discovered that resonant photon tunneling between surface plasmon modes gives rise to extraordinary enhancement of the optical transmission through subwavelength hole arrays [39]. The same coupling will enhance the van der Waals friction [37, 38] if the frequency of these modes is sufficiently low as to be excited by thermal radiation. At room temperature, only the modes with frequencies below $\sim 10^{14}$ s^{-1} can be excited. Surface plasmons have frequencies that are much too high for normal metals; at thermal frequencies the dielectric functions of normal metals become nearly purely imaginary, which exclude surface plasmon enhancement of the heat transfer for good conductors. However, surface plasmons for semiconductors are characterized by much smaller frequencies and damping constants, and they can provide an important contribution to the van der Waals friction. On a metal surface, the adsorbate vibrations parallel to the surface can have very low frequencies and they can also provide an important contribution to the van der Waals friction. For the case of resonant photon tunneling between surface localized states, normal motion results in drastic difference from parallel relative motion. When one of the bodies is sufficiently rarefied, the theories [25, 37, 38] give the friction between a flat surface and a small particle, which in the nonretarded limit agrees with the results of Tomassone and Widom [40]. A theory of the van der Waals friction between a small particle and flat surface, which takes into account screening, nonlocal optic effects, and retardation effects, was developed in [31].

Recently Gotsmann and Fuchs [10] reported measurements of a long-range noncontact friction between an aluminum tip and a gold (111) surface. The friction force F acting on the tip was found to be proportional to the velocity v, $F = \Gamma v$. For motion of the tip normal to the surface, the friction coefficient $\Gamma(d) = b \cdot d^{-3}$, where d is the tip–sample spacing and $b = (8.0^{+5.5}_{-4.5}) \times 10^{-35}\,\mathrm{N\,s\,m^2}$ [10]. Later Stipe et al. [11] observed noncontact friction between a gold surface and a gold-coated cantilever as a function of tip–sample spacing d, temperature T, and bias voltage V. For vibration of the tip parallel to the surface, they found $\Gamma(d) = \alpha(T)(V^2 + V_0^2)/d^n$, where $n = 1.3 \pm 0.2$, and $V_0 \sim 0.2\,\mathrm{V}$. At 295 K, for the spacing $d = 100\,\text{Å}$, they found $\Gamma = 1.5 \cdot 10^{-13}\,\mathrm{kg\,s^{-1}}$, which is $\sim$ 500 times smaller that reported in [10] at the same distance using a parallel cantilever configuration. An applied voltage of 1 V resulted in a friction $\Gamma = 3 \cdot 10^{-12}\,\mathrm{kg/s}$ at 300 K with $d = 20\,\mathrm{nm}$. Using the fluctuation–dissipation theorem, the force fluctuations were intepreted in terms of near-surface fluctuating electric fields with static surface charge.

In [11], the noncontact friction was also measured for fused silica samples. Near the silica surface, the friction was found to be an order of magnitude larger than for the gold sample. The silica sample had been irradiated with γ rays, which produce E' centers (Si dangling bonds) at a density of $7 \cdot 10^{17}\,\mathrm{cm^{-3}}$. Although the sample is electrically neutral overall, the E' centers are known to be positively charged, creating enhanced field inhomogeneity and causing the noncontact friction to rise by another order of magnitude.

In a recent letter, Dorofeyev et al. [9] claim that the noncontact friction observed in [9,10] is due to ohmic losses mediated by the fluctuating electromagnetic field. This claim is controversial, however, since the van der Waals friction for good conductors like copper has been shown [3, 25, 31, 41] to be many orders of magnitude smaller than the friction observed by Dorofeyev et al. In [42], it was proposed that the van der Waals friction may be strongly enhanced between the high-resistivity mica substrate and the silica tip. However, the mica substrate and the silica tip were coated with gold films thick enough to completely screen the electrodynamic interaction between the underlying dielectrics.

At a small separation $d \sim 1\,\mathrm{nm}$, resonant photon tunneling between adsorbate vibrational modes on the tip and the sample may increase the friction by seven orders of magnitude in comparison with the surfaces of the good conductors [37, 38]. However, the distance dependence ($\sim 1/d^6$) is stronger than observed experimentally [11].

Recently, a theory of noncontact friction was suggested where the friction arises from ohmic losses associated with the electromagnetic field created by moving charges induced by the bias voltage [43]. In the case of a spherical tip, this theory predicts the same weak distance dependence of the friction as observed in the experiment, but the magnitude of the friction is many orders of magnitude smaller than found experimentally. Recently [44], we have shown that the electrostatic friction can be greatly enhanced if there is

an incommensurate adsorbed layer exhibiting acoustic vibrations. This theory gives a tentative explanation for the unexplained experimental noncontact friction data [11].

As we shall show in Sect. 19.3, the electrostatic friction as well as the van der Waals friction can be greatly enhanced for 2-D systems, e. g. a 2-D electron system or an incommensurate layer of adsorbed ions exhibiting an acoustic vibrations. The origin of this enhancement is related to the fact that screening in the 2-D systems is much less effective than for the 3-D systems. An atomic force microscope tip charged by the bias voltage, or by the spatial variation of the surface potential, and moving close to the metal surface will induce "image" charges in the 2-D system. Because of the finite response time, this "image" charge will lag behind the tip, and this effect results in a force acting on the tip, referred to as the *electrostatic friction*. However, the weaker screening effect in the 2-D system will result in a much weaker restoring force, which occurs when the "image charge" is displaced from the equilibrium position, and this results in a larger lag of the "image" charge in the 2-D systems in comparison with the 3-D systems.

This review is organized as follows. In Sect. 19.2 we present a short overview of the basic idea of Rytov's theory of the fluctuating electromagnetic field [45]. This theory was originally developed for calculating the thermal electromagnetic field, but has found a broad application in theories of the van der Waals interaction [27], radiative heat transfer [3, 46, 47] and the van der Waals friction [25, 38]. In this paper, the theory is applied to the problem of van der Waals friction (Sect. 19.3). Special attention is paid to the possible enhancement mechanism of the van der Waals friction. In Sect. 4 we consider electrostatic friction, and discuss noncontact friction experiments. In Sect. 19.5 we consider friction resulting from phonon emission, and from internal friction. Section 19.6 presents conclusions and an outlook for further development of the field.

19.2 Theory of the Fluctuating Electromagnetic Field

There are two approaches used to study the fluctuating electromagnetic field. One approach, due to Rytov [45], assumes that the fluctuating electromagnetic field is created by the thermal and quantum fluctuations of the current density $\boldsymbol{j}^f$ inside the medium. The average $\langle \boldsymbol{j}^f \rangle = 0$ but $\langle \boldsymbol{j}^f \boldsymbol{j}^f \rangle \neq 0$. The electromagnetic field can be calculated using Maxwell equations with the fluctuating current density as the source. Knowing the fluctuating electromagnetic field, one can calculate the Poynting vector, the stress tensor, etc., and obtain the heat transfer between bodies [2, 3, 46], the van der Waals interaction [23, 27] and the van der Waals friction [25, 38]. Of these problems, the calculations of the van der Waals friction is the most involved because it requires the solution of a complex electromagnetic problem involving moving

boundaries [25,38]. Due to its complexity, the correct results for the van der Waals friction were only obtained recently.

In the second approach the electromagnetic field is described by Green's functions [23], which can be calculated using quantum electrodynamics [48]. Any physical quantities can be calculated using these Green's functions. For equilibrium problems like the van der Waals interaction, both approaches give the same results [23], though the method based on Green's functions is more general. For nonequilibrium problems Rytov's approach is the simplest, and for this reason the correct results for the radiative heat flux [46] and the van der Waals friction between two parallel surfaces [25, 38], were initially obtained using this approach.

Rytov's theory is based on the introduction of a "random" current in the Maxwell equations (just as, for example, one introduces a "random" force in the theory of the Brownian motion of a particle). For a monochromatic field (time factor $\exp(-\mathrm{i}\omega t)$) in a dielectric, nonmagnetic medium, these equations are:

$$\boldsymbol{\nabla} \times \boldsymbol{E} = i\frac{\omega}{c}\boldsymbol{B} \tag{19.1}$$

$$\boldsymbol{\nabla} \times \boldsymbol{H} = -i\frac{\omega}{c}\boldsymbol{D} + \frac{4\pi}{c}\boldsymbol{j}^f \tag{19.2}$$

where the fluctuating current density $\boldsymbol{j}^f$ is associated with thermal and quantum fluctuations. $\boldsymbol{E}$, $\boldsymbol{D}, \boldsymbol{H}$ and $\boldsymbol{B}$ are the electric and the electric displacement field, the magnetic and the magnetic induction fields, respectively. For a nonmagnetic medium $\boldsymbol{B} = \boldsymbol{H}$ and

$$\boldsymbol{D}(\boldsymbol{r}) = \int \mathrm{d}^3\boldsymbol{r}' \overleftrightarrow{\varepsilon}(\boldsymbol{r}, \boldsymbol{r}', \omega)\boldsymbol{E}(\boldsymbol{r}') \,, \tag{19.3}$$

where $\overleftrightarrow{\varepsilon}(\boldsymbol{r}, \boldsymbol{r}')$ is the dielectric dyadic of the surrounded media. According to the fluctuation–dissipation theorem [49], the average value of the product of components of $\boldsymbol{j}^f$ is given by the formulae

$$\left\langle j_i^f(\boldsymbol{r}, \omega) j_k^{f*}(\boldsymbol{r}', \omega') \right\rangle = \left\langle j_i^f(\boldsymbol{r}) j_k^{f*}(\boldsymbol{r}') \right\rangle_\omega \delta(\omega - \omega')$$

$$\left\langle j_i^f(\boldsymbol{r}) j_k^{f*}(\boldsymbol{r}') \right\rangle_\omega = \frac{\hbar}{(2\pi)^2}\left(\frac{1}{2} + n(\omega)\right)\omega^2 \mathrm{Im}\varepsilon_{ik}(\boldsymbol{r}, \boldsymbol{r}', \omega) \tag{19.4}$$

$$n(\omega) = \frac{1}{\mathrm{e}^{\hbar\omega/k_\mathrm{B}T} - 1} \,. \tag{19.5}$$

From the Maxwell equations it follows that the component of the electric field produced by the random current density $\boldsymbol{j}^f$ is given by

$$E_i(\boldsymbol{r}) = \frac{i}{\omega}\int \mathrm{d}^3\boldsymbol{r}' D_{ik}(\boldsymbol{r}, \boldsymbol{r}', \omega) j_k^f(\boldsymbol{r}') \,, \tag{19.6}$$

where Green's functions of the electromagnetic field $D_{ij}(\boldsymbol{r},\boldsymbol{r}',\omega)$ obey the equations

$$\left(\nabla_i\nabla_k-\delta_{ik}\nabla^2\right)D_{kj}(\boldsymbol{r},\boldsymbol{r}',\omega)-(\omega/c)^2\int \mathrm{d}^3x''\epsilon_{ik}(\boldsymbol{r},\boldsymbol{r}'',\omega)D_{kj}(\boldsymbol{r}'',\boldsymbol{r}',\omega)$$
$$=(4\pi\omega^2/c^2)\delta_{ij}\delta(\boldsymbol{r}-\boldsymbol{r}') \tag{19.7}$$
$$\left(\nabla'_j\nabla'_k-\delta_{jk}\nabla'^2\right)D_{ik}(\boldsymbol{r},\boldsymbol{r}',\omega)-(\omega/c)^2\int \mathrm{d}^3x''\epsilon_{kj}(\boldsymbol{r}'',\boldsymbol{r}',\omega)D_{ik}(\boldsymbol{r},\boldsymbol{r}'',\omega)$$
$$=(4\pi\omega^2/c^2)\delta_{ij}\delta(\boldsymbol{r}-\boldsymbol{r}')\,. \tag{19.8}$$

Using Eqs. (19.3)–(19.8), we obtain the power spectral density of the electric field emitted by the body at temperature T [3]:

$$\langle E_i(\boldsymbol{r})E_j(\boldsymbol{r}')\rangle_\omega=\frac{\hbar c^2}{16\pi^2 i\omega^2}\coth(\hbar\omega/2k_{\mathrm{B}}T)\int \mathrm{d}\boldsymbol{S}''$$
$$\cdot\left\{D_{ik}(\boldsymbol{r},\boldsymbol{r}'')\boldsymbol{\nabla}''D^*_{kj}(\boldsymbol{r}'',\boldsymbol{r}')-\boldsymbol{\nabla}''D_{ik}(\boldsymbol{r},\boldsymbol{r}'')D^*_{kj}(\boldsymbol{r}'',\boldsymbol{r}')\right\}, \tag{19.9}$$

where the points $\boldsymbol{r}$ and $\boldsymbol{r}'$ are outside the body. Here we have transformed an integral over the volume of the body into an integral over the surface of the body. For the evanescent waves, the surface of integration can be moved to infinity. Thus, using Eqs. (19.7) and (19.8, and taking into account the fact that the surface integral vanishes in this case, we get

$$\langle E_i(\boldsymbol{r})E_j(\boldsymbol{r}')\rangle^{\mathrm{evan}}_\omega=\frac{\hbar}{2\pi}\coth(\hbar\omega/2k_{\mathrm{B}}T)\,\mathrm{Im}\,D_{ij}(\boldsymbol{r},\boldsymbol{r}')\,. \tag{19.10}$$

A particularly important limiting case is the radiation from a flat surface, when the power spectral density can be written in the form [38]

$$\langle \boldsymbol{E}(\boldsymbol{r})\boldsymbol{E}(\boldsymbol{r}')\rangle_\omega=\coth(\hbar\omega/2k_{\mathrm{B}}T)\int\frac{\mathrm{d}^2q}{(2\pi)^2}\left(\hat{s}w_{ss}\hat{s}+\boldsymbol{K}w_{pp}\boldsymbol{K}^*\right)$$
$$\cdot\,\mathrm{e}^{i(\boldsymbol{q}(\boldsymbol{x}-\boldsymbol{x}')}\,\mathrm{e}^{i(pz-p^*z')}\,, \tag{19.11}$$

where

$$w_{ss}=\frac{\hbar\omega^2}{4c^2|p|^2}[(p+p^*)(1-|R_s|^2)+(p-p^*)(R_s^*-R_s)] \tag{19.12}$$
$$w_{pp}=\frac{\hbar}{4|p|^2}[(p+p^*)(1-|R_p|^2)+(p-p^*)(R_p^*-R_p)]\,, \tag{19.13}$$

where $\boldsymbol{r}=(\boldsymbol{x},z), p=\sqrt{(\omega/c)^2-q^2}, \hat{s}=\hat{z}\times\hat{q}, \boldsymbol{K}=q\hat{z}+p\hat{q}$, and where R_s and R_p are the reflection coefficients for s- and p- polarized electromagnetic waves respectively.

Equation (19.11) describes the temporal and spatial coherence of the thermal radiation of the near field emitted into free space by the plane surface.

Thermal radiation is often presented as a typical example of an incoherent source, in marked contrast to a laser. Whereas a laser is highly monochromatic and very directional, a thermal source has a broad spectrum and is usually quasi-isotropic. However, as is often the case, a different behavior can be expected at a microscopic scale. Thus, it has been shown [14, 50, 51] that the field emitted by a thermal source made of a polar material is enhanced by more than four orders of magnitude and is partially coherent at a distance of the order of 10 – 100 nm. This phenomenon is due to surface electromagnetic waves, and can only be observed on materials that support them. Surface electromagnetic waves are modes which propagate along an interface and decrease exponentially in the perpendicular direction (evanescent waves). The propagation length of these surface waves is typically a hundred wavelengths, resulting in a long-range spatial correlation of the electromagnetic field along the interface [14, 50, 51]. The near-field properties of the thermal electromagnetic field in the presence of surface electromagnetic waves were reviewed recently in [1].

There are different types of optically active surface waves. Surface-phonon polaritons are observed for polar materials such as silicon carbide (SiC), glass, II–IV and III–V semiconductors. They are mechanical vibrations (phonons) propagating in a partially ionic material so that each unit cell becomes an oscillating electric dipole. Surface-plasmon polaritons are longitudinal electron oscillations (acoustic-type waves in an electron gas) that can be observed for metals and doped semiconductors, which generate electromagnetic fields with longitudinal polarization. Surface waves due to excitons and ionic vibrations in an adlayer of alkali atoms have also been observed.

Recent studies have shown that surface-phonon polaritons can couple to a propagating wave through a grating [14]. Through this coupling, the near-field coherence modifies the far-field emission. The spatial coherence leads to a high directionality of the thermal source – the radiation is emitted in angular lobes, as from an antenna. The same effects can be produced with a surface-plasmon polariton excited on doped silicon [52]. Thermal emission mediated by surface plasmons has been demonstrated with gold [54]. In [53], a thermally stimulated midinfrared source was developed that emits radiation within a narrow range of wavelengths ($\delta\lambda/\lambda \leq 0.2$). In this experiment, the silicon wafer was covered by a metal film. A lattice of holes in the metal mediated the coupling of the surface plasmon states to the emitted light. This technology will afford tunable infrared emitters with high power in a narrow spectral band that are critical for sensing, spectroscopy, and thermophotovoltaic applications.

The existence of the evanescent waves gives rise to remarkable difference between the electromagnetic emission spectra in the near- and the far-field zones. In the far-field zone, the Planck energy spectral density $u_{BB}(\omega)$ of this radiation is obtained by multiplying the thermal energy $\Pi(T, \omega) = \hbar\omega/[exp(\hbar\omega/k_{\mathrm{B}}T) - 1]$ of a quantum oscillator by the density

of oscillations (modes) per unit volume $N(\omega) = \omega^2 d\omega/(\pi^2 c^3)$:

$$u_{BB} = \Pi(T,\omega)N(\omega) = \frac{\hbar}{\pi^2 c^3}\frac{\omega^3}{\exp(\hbar\omega/k_{\rm B}T)-1}\,. \tag{19.14}$$

The energy density $u(\boldsymbol{r},\omega)$ of the emitted electric field in the near zone is defined by the formula

$$u(\boldsymbol{r},\omega) = \frac{\langle E_i(\boldsymbol{r})E_i(\boldsymbol{r})\rangle_\omega}{8\pi} = \frac{\Pi(T,\omega)}{8\pi^2}\,{\rm Im}\,D_{ii}(\boldsymbol{r},\boldsymbol{r},\omega)\,, \tag{19.15}$$

where summation over repeated indices i is assumed. For the plane surface, using the Fresnel reflection coefficient for s-polarized electromagnetic waves, we get the following distance-independent contribution to the energy density in the near zone:

$$u_s(\boldsymbol{r},\omega) = \frac{1}{6\sqrt{2}}u_{BB}\sqrt{|\varepsilon|+|\varepsilon'|}\,. \tag{19.16}$$

Similarly, the p-polarized electromagnetic wave contribution

$$u_p(\boldsymbol{r},\omega) = \frac{1}{8}u_{BB}\left(\frac{c}{\omega d}\right)^3\frac{\varepsilon''}{|\varepsilon+1|^2}\,, \tag{19.17}$$

where ε' and ε'' are the real and imaginary parts of the dielectric function $\varepsilon = \varepsilon' + i\varepsilon''$.

For p-polarized electromagnetic waves, the near-field emission spectrum has a maximum near the frequency ω_p defined by the condition $\varepsilon'(\omega_p) = -1$. This effect results from the existence of a large number of surface modes with different wavenumbers but with frequencies $\omega \approx \omega_p$ that are very close to each other. Therefore, if ε'' is not very large at $\omega = \omega_p$, the surface mode density will necessarily display a strong peak at $\omega = \omega_p$. The presence of surface waves is the origin of the peak in the near-field spectrum of SiC at $\omega = 1.78\cdot 10^{14}\,{\rm s}^{-1}$. However, since the surface waves decay exponentially away from the surface, this peak will vanish in the far zone. The presence of a resonance in the density of modes $N(z,\omega)$ is, however, not required to observe spectral changes caused by the loss of evanescent modes. Indeed, in the short distance regime, the spectrum is given by Eqs. (19.16) and (19.17), whereas in the far field, the spectrum is given by Eq. (19.14) multiplied by the emissivity of the surface. Thus, even in the absence of resonant surface waves, the near-field spectrum is different from the far-field spectrum, but the changes are less dramatic.

Equations (19.11)–(19.13) also suggest a new application for near-field spectroscopy. The near-field spectrum at a given distance to the interface gives access to ${\rm Im}\,R_{p(s)}$, and one can hope to get information about reflection coefficients for large wave vectors, similar to the method usually used to obtain $R_{p(s)}$ for propagating electromagnetic waves from reflectivity measurements. With the rapid development of near-field optical microscopy, such near-field spectra can be measured. This could open the way to the new technique of local solid-state spectroscopy.

19.3 Van der Waals Friction

19.3.1 General Formalism

There are two approaches to the theories of the van der Waals interaction and the van der Waals friction. In the first approach, the fluctuating electromagnetic field is considered to be a classical field which can be calculated from Maxwell's equation with the fluctuating current density as the source of the field, and with appropriate boundary conditions. This approach was used by Lifshitz in the theory of van der Waals interaction [27] and by Volokitin and Persson in the theory of van der Waals friction [25, 38]. The calculation of van der Waals friction is more complicated than van der Waals force because it requires a determination of the electromagnetic field with moving boundaries. The solution can be found by writing the boundary conditions on the surface of each body in the rest reference frame of this body. The relation between the electromagnetic fields in the different reference frames is determined by the Lorentz transformation. The advantage of this approach is that, in principle, it can be used to calculate friction at arbitrary relative velocities. However, the calculations become very complicated for bodies with complex geometries. At present, the solutions are known for the van der Waals friction between two parallel plane surfaces [25,38], and between a small particle and a plane surface [31].

In the other approach the electromagnetic field is treated in the framework of quantum field theory. This approach was used in [23] to obtain the van der Waals interaction for an arbitrary inhomogeneous medium where all parts are at rest.

For two bodies in slow uniform relative motion (velocity $\boldsymbol{v}$), the force acting on either body may be written as $\boldsymbol{F} = \boldsymbol{F}_0 - \overleftrightarrow{\boldsymbol{\Gamma}} \cdot \boldsymbol{v}$, where the adiabatic force $\boldsymbol{F}_0$ is independent of $\boldsymbol{v}$, and where $\overleftrightarrow{\boldsymbol{\Gamma}}$, the so-called friction tensor, is defined by

$$\overleftrightarrow{\boldsymbol{\Gamma}} = (k_{\mathrm{B}}T)^{-1}\,\mathrm{Re}\int_0^\infty \mathrm{d}t \left\langle \hat{\boldsymbol{F}}(t)\hat{\boldsymbol{F}}(0)\right\rangle . \tag{19.18}$$

Here $\langle \ldots \rangle$ represents a thermal average of the fluctuating force in the equilibrium state at a fixed separation d between bodies, and $\hat{\boldsymbol{F}}(t)$ is the force operator in the Heisenberg representation. In fact, Eq. (19.18) leads from the fluctuation–dissipation theorem [49]. For the case of interaction between a localized and an extended system, Eq. (19.18) has been derived by several authors (Schaich 1974 [55], d'Agliano et al. 1975 [56], Nourtier 1977 [57]), and is also valid for two extended systems. In the theory of van der Waals friction, Eq. (19.18) was first used by Schaich and Harris [32], but their treatment is not complete.

In the case of the extended system, the fluctuating force operator can be expressed through the operator of the stress tensor $\hat{\sigma}_{ik}$

$$\hat{F}_i = \int \mathrm{d}S_k \hat{\sigma}_{ik} \,, \tag{19.19}$$

where the integration is over the surface of the one of the bodies and

$$\hat{\sigma}_{ik} = \frac{1}{4\pi}\left[E_i E_k + B_i B_k - \frac{1}{2}\delta_{ik}\left(E^2 + B^2\right)\right] \,, \tag{19.20}$$

where E_i and B_i are the electric and magnetic induction field operators, respectively. The calculation of the force–force correlation function can be performed using methods from quantum field theory [48,58]. The advantage of this approach is that it reduces to finding Green's functions of the electromagnetic field for the equilibrium system with fixed boundaries. Thus, this approach can be easily extended to bodies with complex geometries. However, it is restricted to small relative velocities.

19.3.2 Van der Waals Friction Between Two Plane Surfaces in Parallel Relative Motion at Arbitrary Sliding Velocities

The van der Waals frictional stress $\sigma_\parallel$ that acts on the surfaces of the two bodies in parallel relative motion with arbitrary sliding velocity is given by [25]:

$$\begin{aligned}\sigma_\parallel = {} & \frac{\hbar}{8\pi^3}\int_0^\infty \mathrm{d}\omega \int_{q<\omega/c} \mathrm{d}^2 q q_x \\ & \times \left\{\frac{(1-\mid R_{1p}\mid^2)(1-\mid R_{2p}^-\mid^2)}{\mid 1-\mathrm{e}^{2ipd}R_{1p}R_{2p}^-\mid^2}\left(n(\omega-q_x v)-n(\omega)\right)+[R_p\to R_s]\right\} \\ & + \frac{\hbar}{2\pi^3}\int_0^\infty \mathrm{d}\omega \int_{q>\omega/c} \mathrm{d}^2 q q_x \,\mathrm{e}^{-2|p|d} \\ & \times \left\{\frac{\mathrm{Im}\, R_{1p}\,\mathrm{Im}\, R_{2p}^-}{\mid 1-\mathrm{e}^{-2|p|d}R_{1p}R_{2p}^-\mid^2}\left(n(\omega-q_x v)-n(\omega)\right)+[R_p\to R_s]\right\} \,,\end{aligned} \tag{19.21}$$

where $n(\omega) = [\exp(\hbar\omega/k_\mathrm{B}T - 1]^{-1}$ and symbol $[R_p \to R_s]$ denotes the term obtained from the first one by replacing the reflection coefficient $R_p(\omega)$, for p-polarized electromagnetic waves, with the reflection coefficient $R_s(\omega)$ for s-polarized electromagnetic waves, and where $R_{p(s)}^\pm = R_{p(s)}(\omega \pm q_x v)$. Equation (19.21) was obtained by Volokitin and Persson [25] using the dynamical modification of Lifshitz theory [27] for the van der Waals interaction. The first term in Eq. (19.21) represents the contribution to the friction from propagating waves ($q < \omega/c$), and the second term from the evanescent waves ($q > \omega/c$). It can be shown that Eq. (19.21) is valid even when the spatial

dispersion of the dielectric function is taken into account [38]. In the nonretarded case ($c \to \infty$), and for small velocity v, Eq. (19.21) reduces to the formula obtained by Persson and Zhang [36], and in the nonretarded case and at zero temperature it is reduced to the formula obtained by Pendry [24].

There is an important difference related to the types of elementary excitations that play an important role in radiative heat transfer and van der Waals friction. For radiative heat transfer, all kinds of elementary excitations with frequencies that can be thermally excited are equally important. In the case of van der Waals friction, low-frequency excitations which can emit and absorb plenty of momentum are particularly important. Such low-energy excitations can be associated with the vibrations of massive particles. In particular, many adsorbate vibrations parallel to the metal surface are characterized by very low frequencies.

Discussion of General Formula and Limiting Cases

For $d \ll \lambda_T = c\hbar/k_{\mathrm{B}}T$, we can neglect the first term in Eq. (19.21), and the second term can be written as

$$\sigma_{\parallel} = \frac{\hbar}{2\pi^3}\int_{-\infty}^{\infty} \mathrm{d}q_y \int_0^{\infty} \mathrm{d}q_x q_x \, \mathrm{e}^{-2qd} \left\{ \int_0^{\infty} \mathrm{d}\omega [n(\omega) - n(\omega + q_x v)] \right.$$
$$\times \left(\frac{\mathrm{Im}\, R_{1p}^{+} \, \mathrm{Im}\, R_{2p}}{|\, 1 - \mathrm{e}^{-2|p|d} R_{1p}^{+} R_{2p}\, |^2} + (1 \leftrightarrow 2) \right)$$
$$\left. - \int_0^{q_x v} \mathrm{d}\omega [n(\omega) + 1/2] \left(\frac{\mathrm{Im}\, R_{1p}^{-} \, \mathrm{Im}\, R_{2p}}{|\, 1 - \mathrm{e}^{-2qd} R_{1p}^{-} R_{2p}\, |^2} + (1 \leftrightarrow 2) \right) + [p \to s] \right\} . \tag{19.22}$$

For zero temperature, $n(\omega) = 0$ for $\omega > 0$, and the friction is given by [24]:

$$\sigma_{\parallel} = -\frac{\hbar}{4\pi^3}\int_{-\infty}^{\infty} \mathrm{d}q_y \int_0^{\infty} \mathrm{d}q_x q_x \, \mathrm{e}^{-2qd} \Big\{ \int_0^{q_x v} \mathrm{d}\omega$$
$$\times \left(\frac{\mathrm{Im}\, R_{1p}^{-} \, \mathrm{Im}\, R_{2p}}{|\, 1 - \mathrm{e}^{-2qd} R_{1p}^{-} R_{2p}\, |^2} + (1 \leftrightarrow 2) \right) + [p \to s] \Big\} . \tag{19.23}$$

In the local optic approximation, the reflection coefficients are given by Fresnel's formulae . For $d < \lambda_T |\varepsilon(\omega_T)|^{-1/2}$, the reflection coefficient for p-polarized electromagnetic waves can be approximated by

$$R_p \approx \frac{\varepsilon(\omega) - 1}{\varepsilon(\omega) + 1} . \tag{19.24}$$

Substituting (19.24) in (19.23), and using the dielectric function $\varepsilon = 1 + 4\pi i\sigma/\omega$, where σ is independent of ω, in the limiting cases of small and high

velocities:

$$\sigma_{\parallel} = \frac{15}{2^7\pi^2}\left(\frac{\hbar v}{d^4}\right)\left(\frac{v}{4\pi\sigma d}\right)^2 \quad \text{for } v \ll 4\pi\sigma/d \tag{19.25}$$

$$\sigma_{\parallel} = \frac{\hbar(4\pi\sigma)^2}{2\pi^2 d^2 v}\ln\left(\frac{v}{4\pi\sigma d}\right) \quad \text{for } v \gg 4\pi\sigma/d\,. \tag{19.26}$$

For example, for good conductors with conductivity $\sigma = 10^{17}\,\mathrm{s}^{-1}$, $d = 1\,\mathrm{nm}$ at $v = 1\,\mathrm{m/s}$, the friction is extremely small: $\sigma_{\parallel} \approx 10^{-18}\,\mathrm{N}\cdot\mathrm{m}^{-2}$. The friction increases when the conductivity decreases, but there is a limit on the enhancement. Thus, according to the Drude model, the conductivity is given by

$$\sigma = \frac{ne^2\tau}{m}\,, \tag{19.27}$$

where the relaxation time τ cannot be shorter than $\sim 10^{-16}\,\mathrm{s}$. The lowest value of the electron concentration is restricted by the validity of the macroscopic theory, which is valid only if the average separation between electrons is much smaller than the length scale of the variation in the electric field, which is determined by the separation d, and thus $n_{\min} > d^{-3}$. Hence, at $d = 1\,\mathrm{nm}$, the conductivity cannot be smaller than $\sigma_{\min} \approx 10^{12}\,\mathrm{s}^{-1}$ and, consequently, at $v = 1\,\mathrm{m/s}$ the friction cannot be larger than $10^{-13}\,\mathrm{N}\cdot\mathrm{m}^{-2}$.

19.3.3 Van der Waals Friction Between Plane Surfaces at Finite Temperatures and Small Velocities

For $v < dk_{\mathrm{B}}T/\hbar$ (at $d = 1\,\mathrm{nm}$ and $T = 300\,\mathrm{K}$ for $v < 10^3\,\mathrm{m/s}$), the main contribution to the friction comes from the term in (19.21) which depends linearly on the sliding velocity v. In this case, the frictional stress σ which acts on the surfaces of the two bodies in parallel or normal relative motion, to linear order in the relative velocity v, can be written in the form $\sigma = \gamma v$. The friction coefficient $\gamma_{\parallel}$ can be written as $\gamma_{\parallel} = \gamma_{\parallel}^{\mathrm{rad}} + \gamma_{\parallel}^{\mathrm{evan}}$. The contribution to the friction coefficient from the propagating electromagnetic waves for parallel relative motion is given by

$$\gamma_{\parallel}^{\mathrm{rad}} = \frac{\hbar}{8\pi^2}\int_0^{\infty}\mathrm{d}\omega\left(-\frac{\partial n}{\partial\omega}\right)\int_0^{\omega/c}\mathrm{d}q\,q^3 \times \frac{(1-|R_{1p}|^2)(1-|R_{2p}|^2)}{\left|1-\mathrm{e}^{2ipd}R_{1p}R_{2p}\right|^2} + [p\to s]\,, \tag{19.28}$$

and the contribution from the evanescent electromagnetic waves is given by

$$\gamma_{\parallel}^{\mathrm{evan}} = \frac{\hbar}{2\pi^2}\int_0^{\infty}\mathrm{d}\omega\left(-\frac{\partial n}{\partial\omega}\right)\int_{\omega/c}^{\infty}\mathrm{d}q\,q^3\,\mathrm{e}^{-2|p|d} \times \mathrm{Im}\,R_{1p}\,\mathrm{Im}\,R_{2p}\frac{1}{\left|1-\mathrm{e}^{-2|p|d}R_{1p}R_{2p}\right|^2} + [p\to s]\,. \tag{19.29}$$

For normal relative motion the contribution to the friction coefficient from the propagating electromagnetic waves is given by

$$\gamma_{\perp}^{\mathrm{rad}} = \frac{\hbar}{4\pi^2}\int_0^{\infty} \mathrm{d}\omega \left(-\frac{\partial n}{\partial \omega}\right) \int_0^{\omega/c} \mathrm{d}q\, qp^2 \times \left[(1-|R_{1p}R_{2p}|^2)^2 + |(1-|R_{1p}|^2)R_{2p}\,\mathrm{e}^{ipd} + (1-|R_{2p}|^2)R_{1p}^{*}\,\mathrm{e}^{-ipd}|^2\right] \frac{1}{|1-\mathrm{e}^{2ipd}R_{1p}R_{2p}|^4} + [p \to s]\,, \quad (19.30)$$

and the contribution from the evanescent electromagnetic waves

$$\gamma_{\perp}^{\mathrm{evan}} = \frac{\hbar}{\pi^2}\int_0^{\infty} \mathrm{d}\omega \left(-\frac{\partial n}{\partial \omega}\right) \int_{\omega/c}^{\infty} \mathrm{d}q\, q|p|^2\,\mathrm{e}^{-2|p|d} \times \left[\left(\mathrm{Im}\, R_{1p} + \mathrm{e}^{-2|p|d}\,|R_{1p}|^2\,\mathrm{Im}\, R_{2p}\right)\left(\mathrm{Im}\, R_{2p} + \mathrm{e}^{-2|p|d}\,|R_{2p}|^2\,\mathrm{Im}\, R_{1p}\right) + \mathrm{e}^{-2|p|d}\left(\mathrm{Im}\left(R_{1p}R_{2p}\right)\right)^2\right] \frac{1}{\left|1-\mathrm{e}^{-2|p|d}R_{1p}R_{2p}\right|^4} + [p \to s]\,, \quad (19.31)$$

Eqs. (19.28)–(19.31) were first derived in [25, 38] using the semiclassical theory of the fluctuating electromagnetic theory.

There is a principal difference between the friction coefficient for normal and parallel relative motion, related to the denominator in the formulae for the friction coefficient. The resonant condition corresponds to the case when the denominator of the integrand in Eqs. (19.29)–(19.31), which is due to multiple scattering of the electromagnetic waves from the opposite surfaces, is small. For two identical surfaces and $R_i \ll 1 \leq R_r$, where R_i and R_r are the imaginary and real parts, respectively; for evanescent waves this corresponds to the resonant condition $R_r^2 \exp(-2kd) \approx 1$. At resonance, the integrand in Eq. (19.31) has a large factor of $\sim 1/R_i^2$, in sharp contrast to the case for parallel relative motion, where there is no such enhancement factor. The resonance condition can be fulfilled even for the case when $\exp(-2kd) \ll 1$, because for evanescent electromagnetic waves there is no restriction on the magnitude of the real part or the modulus of R. This opens up the possibility of resonant denominators for $R_r^2 \gg 1$.

For good metals ($k_{\mathrm{B}}T/4\pi\hbar\sigma \gg 1$), using (19.31) for $\lambda_T(k_{\mathrm{B}}T/4\pi\hbar\sigma)^{3/2} < d < \lambda_T(4\pi\hbar\sigma/k_{\mathrm{B}}T)^{1/2}$, we get

$$\gamma_{\perp p}^{\mathrm{evan}} \approx 0.13\frac{\hbar}{d^3\lambda_T}\left(\frac{k_{\mathrm{B}}T}{4\pi\hbar\sigma}\right)^{1/2} \quad (19.32)$$

and for $d < \lambda_T(k_{\mathrm{B}}T/4\pi\hbar\sigma)^{3/2}$ we get

$$\gamma_{\perp p}^{\mathrm{evan}} \approx \frac{\hbar}{d^4}\left(\frac{k_{\mathrm{B}}T}{4\pi\hbar\sigma}\right)^2\left(1+ln\frac{\hbar\sigma}{2k_{\mathrm{B}}T}\right)\,. \quad (19.33)$$

The last contribution will dominate for metals with conductivities that are not too high ($k_{\mathrm{B}}T/4\pi\hbar\sigma \simeq 1$).

For comparison, the p-wave contribution for parallel relative motion, for $d < \lambda_c = \lambda_T |\varepsilon(\omega = k_{\rm B}T/\hbar)|^{-1/2}$, is given by [3, 25]

$$\gamma_{\parallel p}^{\rm evan} \approx 0.3 \frac{\hbar}{d^4} \left(\frac{k_{\rm B}T}{4\pi\hbar\sigma} \right)^2 . \tag{19.34}$$

It is interesting to note that for normal relative motion, in contrast to parallel relative motion, for good conductors, then for practically all $d > 0$ the main contribution to friction comes from retardation effects, because Eq. (19.32), in contrast to Eq. (19.34), contains the light velocity in $\lambda_T = c\hbar/k_{\rm B}T$.

From Eq. (19.31), we get s-wave contributions to friction for $d < \lambda_c$:

$$\gamma_{\perp s}^{\rm evan} \approx 10^{-2} \frac{\hbar}{\lambda_c^4} (3 - 5\ln(2d/\lambda_c)) . \tag{19.35}$$

For parallel relative motion, the s-wave contribution is a factor of two smaller.

Figure 19.2 shows the contribution to the friction coefficient $\gamma_\parallel$ from the evanescent electromagnetic waves for two semi-infinite solids in parallel relative motion, with parameters chosen to correspond to copper ($\tau^{-1} = 2.5 \cdot 10^{13}\,{\rm s}^{-1}, \omega_p = 1.6 \cdot 10^{16}\,{\rm s}^{-1}$) at $T = 273\,{\rm K}$. The same result for normal relative motion is shown in Fig. 19.3. In both cases, we show the s- and p-wave contributions separately. The dashed lines show the result when the local (long-wavelength) dielectric function $\epsilon(\omega)$ is used, and full lines show the result obtained using the non-local optic dielectric formalism, which was

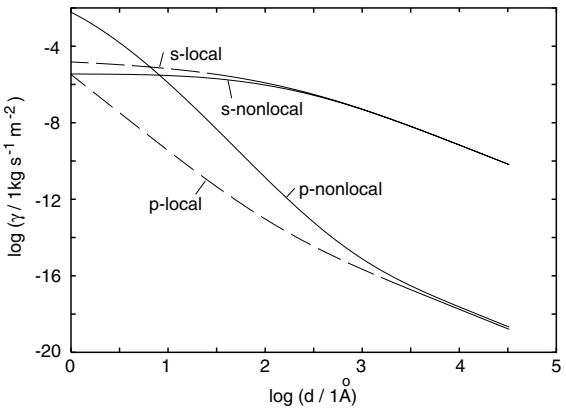


Fig. 19.2. The friction coefficient for two flat surfaces in parallel relative motion as a function of the separation d at $T = 273\,{\rm K}$ with parameters chosen to correspond to copper ($\tau^{-1} = 2.5 \cdot 10^{13}\,{\rm s}^{-1}, \omega_p = 1.6 \cdot 10^{16}\,{\rm s}^{-1}$). The contributions from the s- and p-polarized electromagnetic field are shown separately. The *full curves* represent the results obtained within the nonlocal optic dielectric formalism, and the *dashed curves* represent the result obtained within local optic approximation. (The base of the logarithm is 10)

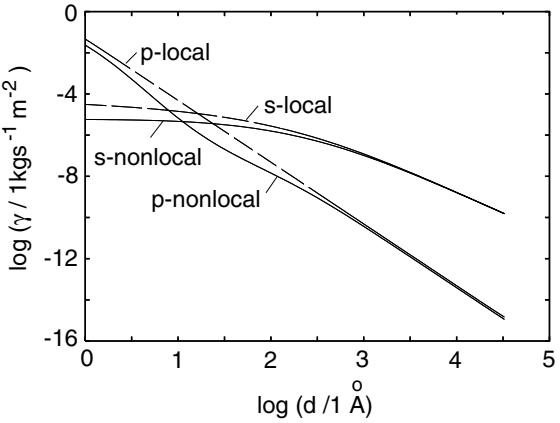


Fig. 19.3. The friction coefficient for two flat surfaces in normal relative motion as a function of separation d at $T = 273\,\mathrm{K}$ with parameters chosen to correspond to copper ($\tau^{-1} = 2.5 \cdot 10^{13}\,\mathrm{s}^{-1}, \omega_p = 1.6 \cdot 10^{16}\,\mathrm{s}^{-1}$). The contributions from the s- and p-polarized electromagnetic field are shown separately. The *full curves* represent the results obtained within the non-local optic dielectric formalism, and the *dashed curves* represent the result obtained within local optic approximation. (The base of the logarithm is 10)

proposed some years ago for studying anomalous skin effects [59]. This formalism takes into account the spatial dispersion of the dielectric function. Figure 19.2 shows that, for sufficiently small separations ($d < 1000\,\text{Å}$) and for parallel relative motion, nonlocal optic effects become important for the p-wave contribution. However, for the s-wave contribution, for both parallel and normal relative motion, the nonlocal optic effects are negligibly small for practically all separations. For the p-wave contribution the nonlocal optic effects are less important for normal relative motion, than for the parallel relative motion. The calculations presented take into account nonlocal effects in the bulk of the solid. There is also the contribution to the nonlocal optical response from the surface region, which was investigated in [3, 31]. For $d > 10\,\text{Å}$, the volume contribution from nonlocal effects is as important as the surface contribution.

For high-resistivity metals ($k_\mathrm{B}T/4\pi\hbar\sigma > 1$) and for $d < \lambda_c$ we get

$$\gamma_\perp \approx 0.48 \frac{\hbar}{d^4} \frac{k_\mathrm{B}T}{4\pi\hbar\sigma} \tag{19.36}$$

and $\gamma_\parallel \approx 0.1\gamma_\perp$. Thus, van der Waals friction diverges in the limit $\sigma \to 0$. Of course, in reality the friction must vanish in this limit because the conductivity is proportional to the concentration n of free electrons, and the friction must vanish when n vanishes. The origin of the discrepancy lies in the breakdown of the macroscopic theory, which is not valid at very low electron concentrations. The application of the macroscopic approach for the

electromagnetic properties of matter is valid only when the length scale of spatial variation of electromagnetic field is much larger than the average distance between the electrons. For evanescent waves, this length scale is determined by separation d between the bodies. From this condition, we can estimate the maximum friction that can be obtained for high-resistivity metals. The minimum conductivity can be estimated as

$$\sigma_{\min} \sim \frac{e^2\tau}{md^3}$$

and the maximum friction can be estimated as

$$\gamma_{\max} \sim \frac{\hbar}{d^4}\frac{k_{\rm B}T}{4\pi\hbar\sigma_{\min}} \sim \frac{mk_{\rm B}T}{4\pi e^2\tau d}\,.$$

To estimate the friction coefficient Γ for an atomic force microscope tip with a radius of curvature $R \gg d$, we can use the approximate relation $\Gamma \sim dR\gamma$. Using this approximation, we get a maximal friction coefficient for spherical tip of:

$$\Gamma^s_{\max} \sim \gamma_{\max} dR \sim \frac{mk_{\rm B}TR}{4\pi e^2\tau}\,. \tag{19.37}$$

For $\tau \sim 10^{-16}\,{\rm s}, R \sim 1\,\mu{\rm m}$ and $T = 300\,{\rm K}$, we get $\Gamma_{\max}10^{-14}\,{\rm kg/s}$. This friction is only one order of magnitude smaller than the friction observed in a recent experiment at $d = 10\,{\rm nm}$ [11]. Similarly, in the case of a cylindrical tip:

$$\Gamma^c_{\max} \sim \gamma_{\max}\sqrt{dR}w \sim \frac{mk_{\rm B}TR^{1/2}w}{4\pi e^2\tau d^{1/2}}\,, \tag{19.38}$$

where w is the width of the tip. For $w = 7\,\mu{\rm m}, d = 10\,{\rm nm}$ and with the other parameters as above, the friction is one order of magnitude larger than observed in the experiment. Thus, the van der Waals friction between high resistivity materials can be measured with present state-of-the-art equipment.

19.3.4 Resonant Photon Tunneling Enhancement of the van der Waals Friction

Surface Phonon Polariton Enhancement of the van der Waals Friction

The van der Waals friction can be greatly enhanced by resonant photon tunneling between localized surface states, e. g. surface plasmon polaritons and adsorbate vibration modes. Close to the resonance we can use the approximation

$$R_r + iR_i = \frac{\omega_a}{\omega - \omega_0 - i\eta}\,, \tag{19.39}$$

where ω_a is a constant. Then, from the resonance condition ($R_r = \pm\exp(qd)$), we get the positions of the resonances $\omega_\pm = \omega_0 \pm \omega_a\exp(-qd)$. For the resonance condition to be valid, the separation $\Delta\omega = |\omega_+ - \omega_-|$ between two

resonances in the transmission coefficient must be greater than the width of the resonance. From this condition, we get that the two poles approximation is valid only for $q \leq q_c \approx \ln(2\omega_a/\eta))/d$. For $\omega_0 > \omega_a$ and $q_c d > 1$, in the resonant approximation we get [38]:

$$\gamma_\perp = \frac{3}{128}\frac{\hbar^2\omega_a^2}{d^4 k_{\rm B}T\eta}\frac{1}{\sinh^2(\hbar\omega_0/2k_{\rm B}T)}\,, \tag{19.40}$$

and for parallel motion, we get:

$$\gamma_\parallel = \frac{\hbar^2\eta q_c^4}{128\pi k_{\rm B}T}\frac{1}{\sinh^2(\hbar\omega_0/2k_{\rm B}T)}\,. \tag{19.41}$$

Resonant photon tunneling enhancement of the van der Waals friction is possible for two semiconductor surfaces that can support low-frequency surface plasmon or surface polariton modes. As an example, we consider two clean surfaces of silicon carbide (SiC). The optical properties of this material can be described using an oscillator model [60]

$$\epsilon(\omega) = \epsilon_\infty\left(1 + \frac{\omega_L^2 - \omega_T^2}{\omega_T^2 - \omega^2 - i\Gamma\omega}\right) \tag{19.42}$$

with $\epsilon_\infty = 6.7, \omega_L = 1.8\cdot 10^{14}\,{\rm s}^{-1}, \omega_T = 1.49\cdot 10^{14}\,{\rm s}^{-1}$ and $\Gamma = 8.9\cdot 10^{11}\,{\rm s}^{-1}$. The frequency of surface plasmons is determined by the condition $\epsilon_r(\omega_p) = -1$, and from (19.42) we get $\omega_p = 1.78\cdot 10^{14}\,{\rm s}^{-1}$. The resonance parameters

$$\omega_a = \frac{\omega_L^2 - \omega_T^2}{\epsilon_\infty\omega_L} = 8.2\cdot 10^{12}\,{\rm s}^{-1}, \eta = \Gamma/2, q_c = 3.6/d\,, \quad \text{and} \quad \omega_0 \approx \omega_p\,.$$

Using these parameters in Eqs. (19.40)–(19.41), we get $\gamma_\perp = 4.2\cdot 10^3/d^4$ kg $\cdot$ s^{-1} $\cdot$ m^{-2} and $\gamma_\parallel = 2.2\cdot 10^2/d^4$ kg $\cdot$ s^{-1} $\cdot$ m^{-2}, where the distance d is in Å. Note that the friction between the two semiconductor surfaces is several orders of magnitude larger than between two clean good conductor surfaces (see Figs. 19.2 and 19.3).

Adsorbate Vibrational Mode Enhancement of the van der Waals Friction

Another enhancement mechanism for the van der Waals friction is connected with resonant photon tunneling between adsorbate vibrational modes localized on different surfaces. In [37, 38], we have shown that resonant photon tunneling between two surfaces separated by $d = 1$ nm and covered by a low concentration of potassium atoms at $d = 1$ nm gives rise to friction six orders of magnitude larger than that obtained for clean surfaces. The adsorbate-induced enhancement of the van der Waals friction is even larger for Cs adsorption onto Cu(100). In this case, even at very low coverage ($\theta \sim 0.1$),

the adsorbed layer exhibits an acoustic branch for vibrations parallel to the surface [61], which means that for such vibrations $\omega_{\parallel} = 0$. In this case, the reflection coefficient for p-polarized electromagnetic waves which takes into account the contribution from the adsorbed layer on the surface of the good conductor is given by [38]

$$R_p = 1 - \frac{2qa\omega_q^2}{\omega^2 - \omega_q^2 + i\omega\eta}\,, \tag{19.43}$$

where $\omega_q^2 = 4\pi n_a e^{*2} a q^2/M$; here e^* is the ionic charge and a is the separation between the ion and the image plane. Using Eq. (19.43) in Eq. (19.29) for

$$\frac{a}{\eta d}\sqrt{\frac{4\pi n_a e^{*2} a}{Md^2}} \ll 1$$

gives

$$\gamma_{\parallel} \approx 0.62\frac{k_{\rm B}Ta^2}{\eta d^6}\,. \tag{19.44}$$

It is interesting to note the vanishing dependence of $\gamma_{\parallel}$ on n_a, e^*, and M in Eq. (19.44). However, Eq. (19.43) is only valid when there are acoustic vibrations in the adsorbed layer. For Cs adsorbed onto a Cu(100) surface, the acoustic vibrations exist only for a coverage $\theta \geq 0.1$ [61]. The friction coefficient for a cylindrical atomic force microscope tip can be estimated using an approximation

$$\Gamma_{\parallel}^c \approx w\int_{-\infty}^{\infty} {\rm d}\rho\gamma_{\parallel}(d + \rho^2/2R) = 0.68\frac{k_{\rm B}Ta^2R^{0.5}w}{\eta d^{5.5}}\,, \tag{19.45}$$

where R is the radius of curvature of the tip and w is its width. For Cs adsorption on Cu(100), the damping parameter η was estimated in [44] as $\eta \approx 3\cdot 10^9\,{\rm s}^{-1}$. Using this value of η in Eq. (19.45) for $a = 2.94\,$Å [61], $R = 1\,\mu{\rm m}, w = 7\,\mu{\rm m}, T = 293\,{\rm K}$ at $d = 10\,{\rm nm}$, we get $\Gamma_{\parallel} = 0.5\cdot 10^{-13}\,{\rm kg/s}$, which is only three times smaller than the friction observed in [11] at the same distance. However, the van der Waals friction is characterized by a much stronger distance dependence ($\sim 1/d^{5.5}$) than observed in the experiment ($\sim 1/d^n$, where $n = 1.3 \pm 0.2$). Thus, at smaller distances the van der Waals friction will be much larger than the friction observed in [11], and can be measured experimentally. Figure 19.4 shows how the friction coefficient depends on the distance d between the copper tip and the copper substrate when the surfaces of the tip and the substrate are covered by a low concentration of Cs atoms, and for clean surfaces. In comparison, the friction between two clean surfaces at the separation $d = 1\,{\rm nm}$ is eleven orders of magnitude smaller. However, the friction between clean surfaces shown on Fig. 19.4 was calculated via the local optic approximation. For parallel relative motion, nonlocal optic effects are very important (see Fig. 19.2), and when they are taken into account at $d = 1\,{\rm nm}$, the friction between adsorbate-covered surfaces will be seven orders of magnitude greater than the friction between clean surfaces.

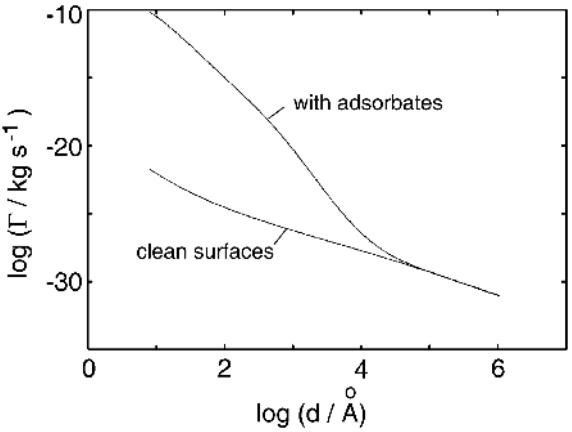


Fig. 19.4. The friction coefficient between a copper tip and a copper substrate, the surfaces of which are covered with a low concentration of cesium atoms, as a function of the separation d. The cylindrical tip is characterized by radius of curvature $r = 1\,\mu\text{m}$ and width $w = 7\,\mu\text{m}$. For other parameters corresponding to Cs adsorbed on a Cu(100) surface at coverage $\theta \approx 0.1$ and for Cu(100) [44, 61]: $e^* = 0.28e$, $\eta = 3 \cdot 10^9\,\text{s}^{-1}$, $a = 2.94\,\text{Å}$, $T = 293\,\text{K}$. (The base of the logarithm is 10)

19.3.5 Van der Waals Friction Between a Small Particle and a Plane Surface

The friction force acting on a small particle during motion parallel to a flat surface can be obtained from the friction between two semi-infinite bodies in the limit when one of the bodies is sufficiently rarefied. This means that dielectric function of this body, say body **2**, is close to unity, i. e., $\varepsilon_2 - 1 \to 4\pi n\alpha_2 \ll 1$, where n is the concentration of particles in body **2** and α_2 is their polarizability. To linear order in concentration n, the reflection coefficients

$$R_{2p} \approx 2\pi n \,\text{Im}\,\alpha_2\,, \quad R_s \approx \pi n \,\text{Im}\,\alpha_2 \left(\frac{\omega}{cq}\right)^2 .$$

The friction force acting on a particle moving parallel to a plane surface can be obtained as the ratio between the change in the frictional shear stress after the displacement of body **2** by a small distance dz, and the number of particles in a slab with thickness dz:

The friction force acting on a particle moving parallel to a plane surface with velocity v is given by

$$F_{\text{fric}} = \frac{d\sigma_{\parallel}(z)}{ndz}\Big|_{z=d}$$
$$== \frac{\hbar}{\pi^2}\int_{-\infty}^{\infty} \text{d}q_y \int_0^{\infty} \text{d}q_x q_x q \,\text{e}^{-2qd} \left\{ \int_0^{\infty} \text{d}\omega [n(\omega) - n(\omega + q_x v)] \right.$$

$$\times \Big[2\left(\mathrm{Im}\, R_{1p}(\omega+q_xv)\,\mathrm{Im}\,\alpha_2(\omega)+[\omega+q_xv\leftrightarrow\omega]\right)$$
$$+\left(\frac{1}{cq}\right)^2\left(\omega^2\,\mathrm{Im}\, R_{1s}(\omega+q_xv)\,\mathrm{Im}\,\alpha_2(\omega)+[\omega+q_xv\leftrightarrow\omega]\right)\Big]$$
$$-\int_0^{q_xv}\mathrm{d}\omega[n(\omega)+1/2]\Big[2\left(\mathrm{Im}\, R_{1p}(\omega-q_xv)\,\mathrm{Im}\,\alpha_2(\omega)+[\omega-q_xv\leftrightarrow\omega]\right)$$
$$+\left(\frac{1}{cq}\right)^2\left(\omega^2\,\mathrm{Im}\, R_{1s}(\omega-q_xv)\,\mathrm{Im}\,\alpha_2(\omega)+[\omega-q_xv\leftrightarrow\omega]\right)\Big]\Bigg\}, \tag{19.46}$$

where $[\omega\pm q_xv\leftrightarrow\omega]$ indicates the terms that can be obtained from the preceding terms by interchanging the positions of $\omega\pm q_xv$ and ω.

To linear order in the sliding velocity v, from (19.46) we get $F_{fric}=-\Gamma_\parallel v$, where

$$\Gamma_\parallel=\frac{\hbar}{\pi}\int_0^\infty\mathrm{d}\omega\left(-\frac{\partial n(\omega)}{\partial\omega}\right)\int_0^\infty\mathrm{d}qq^4\,\mathrm{e}^{-2qd}$$
$$\times\left[2\,\mathrm{Im}\, R_p(q,\omega)+\left(\frac{\omega}{cq}\right)^2\mathrm{Im}\, R_s(q,\omega)\right]\mathrm{Im}\,\alpha(\omega)\,. \tag{19.47}$$

In the nonretarded limit, this equation reduces to the formula obtained by Tomassone and Widom [40].

However Eq. (19.47) does not take into account the screening effects which become important at small separations between the particle and the surface. General formulae for friction coefficients which take into account screening effects were obtained in [31] using the semiclassical theory of the fluctuating electromagnetic field. Thus, for $d\gg R$, we can neglect screening effects, and in this case $\Gamma_\perp=2\Gamma_\parallel$.

For a spherical particle of radius R, the polarizability is given by

$$\alpha_0=\frac{\varepsilon_p-1}{\varepsilon_p+2}R^3\,, \tag{19.48}$$

where $\varepsilon_p=1+4\pi i\sigma/\omega$ is the particle dielectric function. For good conductors ($|\varepsilon|\gg1$), in the frequency range accessible by thermal radiation, the modulus of the reflection coefficient $R_{p(s)}\leq1$. In the limit $d<|\,\varepsilon(\omega=k_\mathrm{B}T/\hbar)\,|^{-1/2}\lambda_T$, the p- and s-wave contributions to the friction coefficient become [31]

$$\Gamma_{p\parallel}\approx3\frac{\hbar}{d^5}\left(\frac{k_\mathrm{B}T}{4\pi\hbar}\right)^2\sigma_s^{-1}\sigma_p^{-1}R^3\,, \tag{19.49}$$

$$\Gamma_{s\parallel}\approx\frac{\pi^3}{10}\frac{\hbar}{d\lambda_T^4}R^3\frac{\sigma_s}{\sigma_p}\,, \tag{19.50}$$

where σ_s and σ_p are the conductivities of the sample and the particle, respectively. For $d=2R=10\,\mathrm{nm}, \sigma_s=\sigma_p=4\cdot10^{17}\,\mathrm{s}^{-1}$ (corresponding to gold), we get very low friction: $\Gamma\sim10^{-29}\,\mathrm{kg\,s}^{-1}$.

However, the friction can be greatly enhanced for high-resistivity materials. Using Eq. (19.47) in the nonretarded limit (which can be formally obtained for the limit $c \to \infty$) and for high-resistivity material ($4\pi\sigma << k_{\mathrm{B}}T/\hbar$), we get:

$$\Gamma_{p\parallel} = 0.9 \frac{k_{\mathrm{B}}T}{4\pi\sigma} \frac{R^3}{d^5} \,, \tag{19.51}$$

where we have assumed that the particle and the substrate have the same dielectric function $\varepsilon = 1 + 4\pi i\sigma/\omega$. As we discussed above (see Sect. 19.3.3), the macroscopic theory used to obtain Eq. (19.51) is only valid when $\sigma \gg \sigma_{\min} \sim e^2\tau/md^3$. For $d = 2R = 10\,\mathrm{nm}$ and $\tau = 10^{-15}\,\mathrm{s}$, Eq. (19.51) gives $\Gamma_{p\parallel\,\max} \sim 10^{-18}\,\mathrm{kg/s}$.

If the particle and the substrate are made from the same material, which can support surface phonon polaritons, the friction is given by

$$\Gamma_{\parallel} = \frac{9k_{\mathrm{B}}T\eta R^3}{d^5} \left(\frac{1}{\omega_s^2} + \frac{1}{\omega_p^2} \right) \,, \tag{19.52}$$

where ω_s and ω_p are the frequencies of the surface phonon polaritons for the substrate and the particle, respectively. If the substrate and the particle are made from SiC, $\omega_s = 1.79 \cdot 10^{14}\,\mathrm{s}^{-1}$ and $\omega_p = 1.76 \cdot 10^{14}\,\mathrm{s}^{-1}$. Thus, for $d = 2R = 10\,\mathrm{nm}, \eta = 8.9 \cdot 10^{11}\,\mathrm{s}^{-1}$, and $T = 300\,\mathrm{K}$ we get $\Gamma \sim 10^{-21}\,\mathrm{kg\,s}^{-1}$. This friction coefficient is eight orders of magnitude greater than those obtained for good conductors.

19.3.6 Van der Waals Friction Mediated by Black Body Radiation

Van der Waals friction also occurs when a particle moves relative to black body radiation, such as particles moving between the walls of an oven at large distances from the walls or in the cosmic microwave background. This kind of friction has no position dependence, i. e., it is spatially homogeneous. The consequence is a universal dissipative drag acting on all matter in relative motion with respect to a thermalized photon gas. For the particle moving relative to black body radiation, the friction coefficient is given by [62]

$$\Gamma^{BB} = \frac{\beta\hbar^2}{3\pi c^5} \int_0^\infty \mathrm{d}\omega \frac{\omega^5 \,\mathrm{Im}\,\alpha(\omega)}{\sinh^2(\frac{1}{2}\beta\hbar\omega)} \,, \tag{19.53}$$

where $\beta^{-1} = k_{\mathrm{B}}T$. The electromagnetic field fluctuations exert a drag on any particle that moves with respect to the reference frame in which the electromagnetic field fluctuations are thermalized, and this drag is proportional to the relative velocity.

For a spherical particle, the polarizability α is given by Eq. (19.48). The dielectric and conducting condensed matter can be considered using the appropriate dielectric function, as was discussed above. The friction coefficient (19.53) is usually very small, corresponding to the general weakness of

this universal thermal drag. Two circumstances under which such low friction might be observable are in ovens and the cosmos.

Tungsten ovens can operate at temperatures as high as 3000 K. If a beam of atoms, ions or molecules passes through such an oven, it will be subject to drag due to the van der Waals friction mediated by the thermal radiation. For an atom or molecule, the polarizability α can be characterized by a single absorption line at ω_0. Because Γ^{BB} is proportional to $\mathrm{Im}\,\alpha(\omega)$, each absorption line for a molecule or a dielectric will contribute additively to Γ^{BB}. For the molecule $\alpha(\omega) = \alpha_0\delta(\omega/\omega_0 - 1)$, where α_0 is the static polarizability at $\omega = 0$.

Setting $m/\tau = \Gamma^{BB}$, where m is the mass of the molecule and τ is the relaxation time, yields the result that

$$\tau = \frac{3\pi mc^5\hbar^4}{2^6\alpha_0(k_{\mathrm{B}}T)^5}\frac{\sinh^2(x)}{x^6}\,, \tag{19.54}$$

where $x = \beta\hbar\omega_0/2$. The relaxation time has a minimum at a temperature-dependent frequency that coincides with the minimum of the function $f(x) = \sinh^2(x)/x^6$, at $x_m = 2.98$, where $f(x_m) = 0.137$. Ba^+ has a resonance near 500 nm, or 2 eV, which is about six times the thermal energy associated with a 3000 K oven. For this resonance, the relaxation time would be near the minimum, and for the ion polarizability $\alpha_0 \approx 1.0 \cdot 10^{-30}\,\mathrm{m}^3$, one obtains the relaxation time $\approx 10^5$ s, i. e., one day. This relaxation time can be measured using ion traps.

For the cosmos, it is believed that hydrogen atoms condensed from protons and electrons when the radiation cooled to about 3000 K, and that the coupling of the cosmic radiation to matter due to Compton scattering becomes ineffective below this condensation temperature [63]. However, atoms, ions and molecules with absorption in the appropriate frequency range should remain coupled to the cosmic radiation as its temperature drops from the 3000 K condensation temperature to perhaps 300 K or even a bit less. This coupling could influence the structure and anisotropies observed in recent experiments on the cosmic microwave background [64]. It could also influence the behavior of molecules formed from the residue of novae and supernovae, which are then subject to drag from a still-hot cosmic microwave (i. e., electromagnetic) background. At much lower temperatures, macroscopic bodies can coalesce, in which case geometrically determined resonances may become relevant.

19.3.7 Van der Waals Frictional Drag Force Between Quantum Wells

The van der Waals friction can be tested not only by measuring the friction force during relative motion of the two bodies, but a more elegant method consists of driving an electric current into one layer and registering the effect

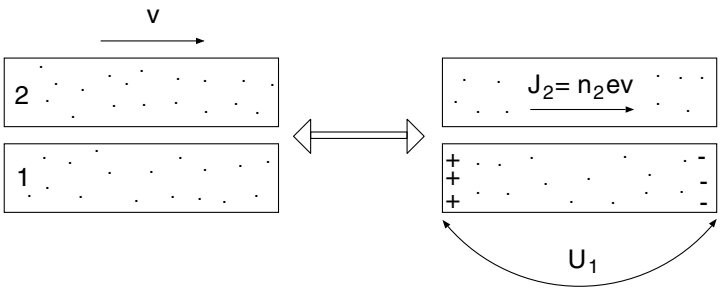


Fig. 19.5. *Left*: a metallic block is sliding relative to a metallic substrate with velocity v. An electronic frictional stress will act on the block (and on the substrate). *Right*: the shear stress σ can be measured if instead of sliding the upper block, a voltage U_2 is applied to the block, resulting in drift motion of the conduction electrons (velocity v). The resulting frictional stress σ on the substrate electrons will generate a voltage difference U_1 (proportional to σ), as indicated in the figure, which can be measured experimentally

of the frictional drag of the electrons in a second parallel layer (Fig. 19.5). Such experiments were predicted by Pogrebinskii [65] and Price [66], and were performed for 2 D quantum wells [15–17]. In these experiments, two quantum wells are separated by a dielectric layer thick enough to prevent electrons tunneling across it but allowing interlayer interaction between them. A current of density $J_2 = n_2 ev$ is driven through layer **2** (where n_2 is the carrier concentration per unit area in the second layer), as shown schematically in Fig. 19.5. Due to the proximity of the layers, the interlayer interactions will induce a current in layer **1** due to a friction stress $\sigma = \gamma v$ acting on the electrons in the layer **1** from layer **2**. If layer **1** is an open circuit, an electric field E_1 will develop in the layer, whose influence cancels the frictional stress σ between the layers. Thus, the frictional stress $\sigma = \gamma v$ must equal the induced stress $n_1 eE_1$ so that

$$\gamma = n_1 eE_1/v = n_1 n_2 e^2 E_1/J_2 = n_1 n_2 e^2 \rho_{12} \,, \tag{19.55}$$

where the *transresistivity* $\rho_{12} = E_1/J_2$ is defined as the ratio of the induced electric field in the first layer to the driving current density in the second layer. The transresistivity is often interpreted in terms of a drag rate which, in analogy with the Drude model, is defined by $\tau_D^{-1} = \rho_{12} n_2 e^2/m^* = \gamma/n_1 m^*$. The most widely used approach in the theory of drag effect is based on the Boltzmann equation [17,67–69] and the Kubo formalism [70,71]. In [72], a theory for the drag effect was developed based on the semiclassical theory of the fluctuating electromagnetic field. In this theory, when the separation between the quantum wells $d \ll \lambda_T$, the friction coefficient γ is given by Eq. (19.29), where $R_i(\omega)$ is the reflection coefficient for layer i. The retardation effects are automatically included in this approach.

For independent electrons and $d < v_{\rm F}\hbar/k_{\rm B}T$, the reflection coefficient for p-polarized electromagnetic waves is given by [36, 72]:

$$R_p = 1 + \frac{i\hbar\varepsilon\omega}{2k_{\rm F}e^2} \,, \tag{19.56}$$

where ε is the dielectric constant of the surrounded dielectric, $k_{\rm F} = \sqrt{2\pi n_s}$ is the Fermi wavevector for a 2-D electron gas, and n_s is the electron concentration for a 2-D electron layer. Using Eq. (19.56) in Eq. (19.29), we get the contribution to the frictional drag rate from p-polarized waves

$$\tau_{Dp} = \gamma_p/(nm^*) \approx 0.2360 \frac{(k_{\rm B}T)^2}{\hbar\epsilon_{\rm F}(q_{\rm TF}d)^2(k_{\rm F}d)^2} \,, \tag{19.57}$$

where $q_{\rm TF} = 2a_0^{-1}/\varepsilon$ is the single-layer Tomas-Fermi screening wavevector, $a_0 = \hbar/m^*e^2$, and $\epsilon_{\rm F}$ is the Fermi energy. Equation (19.57) was first derived by Gramila et al. using an approach based on the Boltzmann equation [16], and later by Persson and Zhang using the theory of van der Waals friction [36].

Figure 19.6 shows the friction coefficient γ as a function of the distance d between two quantum wells at $T = 3\,{\rm K}$ and with $n_s = 1.5 \cdot 10^{15}\,{\rm m}^{-2}, m^* = 0.067m_{\rm e}$, $v_{\rm F} = 1.6 \cdot 10^7\,{\rm cm/s}$, and for the electron mean free path $l = v_{\rm F}\tau = 1.21 \cdot 10^5\,{\rm Å}$. We have also assumed that $\varepsilon = 10$, which corresponds to the conditions for the experiment [15, 16]. In this case, the s-wave contribution is negligibly small in comparison with the p-wave contribution. For $d = 175\,{\rm Å}$,

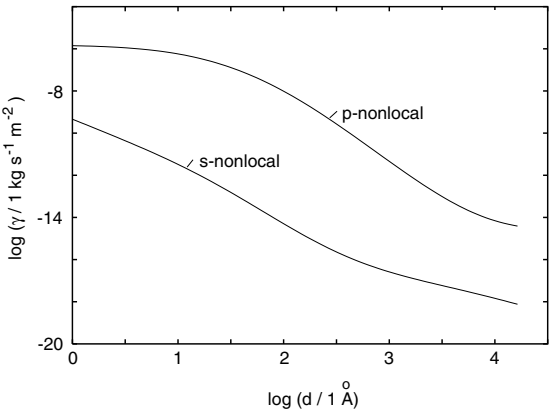


Fig. 19.6. The frictional drag coefficient for two quantum wells at $T = 3\,{\rm K}$ as a function of separation d. The s- and p-wave contributions are shown separately. The calculations were performed with a surface electron density $n_s = 1.5 \cdot 10^{15}\,{\rm m}^{-2}$, a damping constant $\eta = 1.3 \cdot 10^{10}\,{\rm s}^{-1}$, an effective electron mass $m^* = 0.067m_{\rm e}$, and a dielectric constant $\varepsilon = 10$. (The base of the log-function is 10)

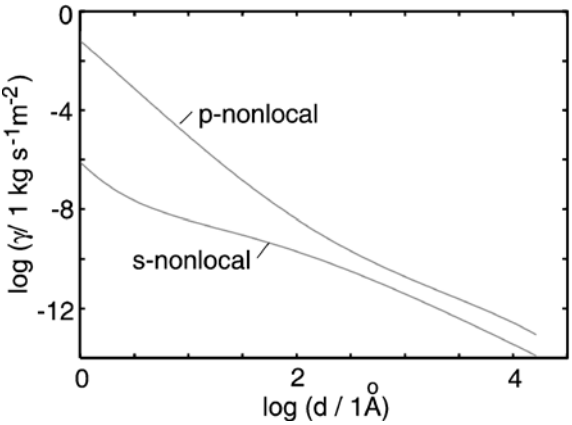


Fig. 19.7. The frictional drag coefficient for two quantum wells at $T = 273\,\mathrm{K}$ as a function of separation d. The s- and p-wave contributions are shown separately. The calculations were performed with a surface electron density $n_s = 1.05 \cdot 10^{19}\,\mathrm{m}^{-2}$, a damping constant $\eta = 2.5 \cdot 10^{13}\,\mathrm{s}^{-1}$, an effective electron mass $m^* = m_\mathrm{e}$, and a dielectric constant $\varepsilon = 1$. (The base of the log-function is 10)

we find $\gamma = 3.3 \cdot 10^{-9}\,\mathrm{kg} \cdot \mathrm{s}^{-1} \cdot \mathrm{m}^{-2}$, which corresponds to a drag rate $\tau_D^{-1} = 3.3 \cdot 10^7\,\mathrm{s}^{-1}$, which is close to the experimental value $(\tau_D^{-1})_\mathrm{exp} = 1.5 \cdot 10^7\,\mathrm{s}^{-1}$ [15,16]. Figure 19.7 shows the friction coefficient for 2-D quantum wells with high electron density $n_s = 10^{19}\,\mathrm{m}^{-2}, T = 273\,\mathrm{K}, \tau = 4 \cdot 10^{-14}$, and $\varepsilon = 1$, where the result for other ε can be obtained using the scaling $\tau_{Dp} \sim \varepsilon^2$ and τ_{Ds} is independent of ε. In Figs. 19.6 and 19.7, the p- and s-wave contributions are shown separately. The calculations show that p-waves provide a larger contribution to the friction, for both low-density and high-density 2-D quantum wells. Figure 19.8 shows the dependence of the friction coefficient on the electron density for the same parameters as in Fig. 19.6. In this case, the boundary between degenerate and nondegenerate electron density is determined by the Fermi density $n_\mathrm{F} = 3k_\mathrm{B}Tm^*/2\pi\hbar^2 = 1.09 \cdot 10^{14}\,\mathrm{m}^{-2}$. From the calculations, we find that the maximum of the frictional drag force for the electron density $n_\mathrm{max} \approx 1 \cdot 10^{15}\,\mathrm{m}^{-2}$; this means that the experiment [15,16] was performed near optimum conditions.

The frictional drag between quantum wells makes it possible to probe the interparticle interaction directly. Interparticle interactions form the cornerstone of many-body physics. Many-body effects are particularly important in low-dimensional systems. This leads to many intriguing phenomena, such as Luttinger liquid behavior in quantum wires, and the fractional quantum Hall effect and Wigner crystallization in two-dimensional electron gases in a magnetic field. As technology improves and semiconductor devices shrink further in size, interaction effects will become even more pronounced and it may become possible to probe these effects in novel experiments.

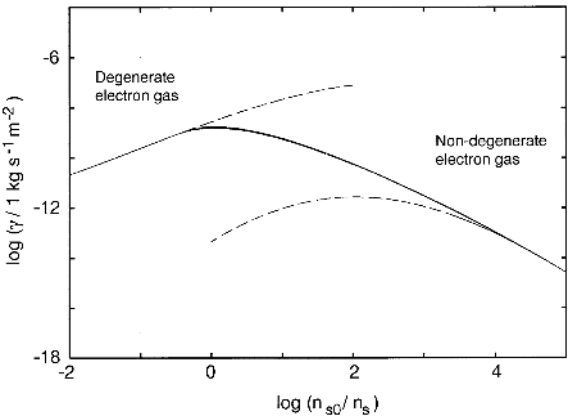


Fig. 19.8. The frictional drag coefficient for two quantum wells at $T = 3\,\mathrm{K}$ as a function of electron concentration n_s. The *full curve* was obtained by interpolation between the curves obtained with the nonlocal optic dielectric approach (*dashed lines*), with the dielectric functions corresponding to a degenerate electron gas for $n_s > n_\mathrm{F} \sim 10^{14}\,\mathrm{m}^{-2}$, and to a nondegenerate electron gas for $n_s < n_\mathrm{F}$. The electron density parameter $n_{s0} = 1.5 \cdot 10^{15}\,\mathrm{m}^{-2}$, the damping constant $\eta = 1.3\,\mathrm{s}^{-1}$, the effective electron mass $m^* = 0.067 m_\mathrm{e}$, the separation $d = 175\,\text{Å}$, and the dielectric function $\varepsilon = 10$

19.4 Electrostatic Friction

19.4.1 Effect of a Bias Voltage and the Spatial Variation of the Surface Potential

The electrostatic potential at the surface of a metal relative to its interior depends on the magnitude of the surface dipole moment per unit area which, in turn, depends on the separation of the lattice planes that are parallel to the surface [73]. Variations in the crystallographic direction at the surface of a clean polycrystalline metal result in a variation in the surface potential. This is referred to as the "patch effect". Patch potentials are also generated and influenced by surface contamination and, in the case of alloys, by variations in chemical composition. The surface potential can be easily changed by applying a voltage between the tip and the sample. The electrostatic forces between conducting surfaces due to spatial variation of the surface potential were studied in [74].

Observed variations in the surface potential: Patch potential variations are specific to the particular sample and depend on environmental factors. Spatial variations in surface potentials are expected to be related to the physical size of the surface crystallites, which in the case of bulk metal, are typically of the order of $2\,\mu\mathrm{m}$. Thin films deposited on substrates at temperatures much less than the melting point of the film are often amorphous, with nonuniform thickness, and the crystallite size is of the same order as the thickness of

the film [75]. Annealing of the film can produce grain structures that are substantially larger than the film thickness. The patch potential variations have been measured under various conditions using vibrating or rotating plate electrometers [76]. Notably, it was shown that large-scale variations in surface potential were caused by adsorption of contaminants [22, 77].

A General Theory

We begin by considering a model in which the tip of a metallic cantilever of length L is a section of a cylindrical surface with radius of curvature R (Fig. 19.9). The cantilever is perpendicular to a flat sample surface which occupies the xy plane, with the z-axis pointing outside the sample. The tip displacement $\boldsymbol{u}(t) = \hat{x}u_0\,\mathrm{e}^{-i\omega t}$ is assumed to be parallel to the surface (along the x axis), which will be a good approximation when the oscillation amplitude u_0 is sufficiently small. The cantilever width w (the size in the direction perpendicular to the xz plane) is taken to be much larger than the thickness c ($w \gg c$), and d is the separation between the tip and the sample surface. It is straightforward to obtain the static electric field distribution in the practically important case of small distances d such that the electrostatic field of the entire cylinder is effectively the same as that of its bottom part. (The criterion that d must satisfy for this to be the case is given by $\sqrt{d/R} \ll 1$.) The problem is then reduced to solving the two-dimension Laplace equation with the boundary conditions that the potential has constant values of V and 0 at the metallic surfaces of the tip and the substrate. The electric field distribution outside the conductors is equal to the field due to two charged wires

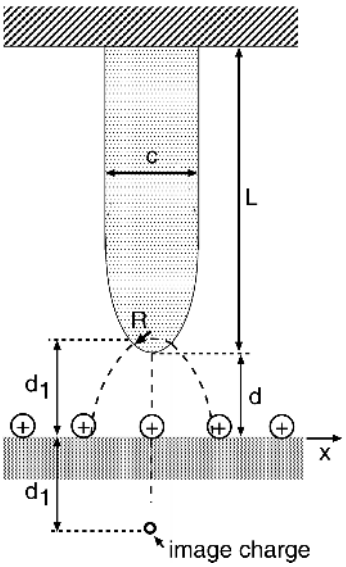


Fig. 19.9. Scheme of the tip–sample system. The tip shape is characterized by its length L and the tip radius of curvature r

passing through points at $z = \pm d_1 \pm \sqrt{(d+R)^2 - R^2}$ [79]. The wires have charges of $\pm Q$ per unit length, $Q = CV$, where $C^{-1} = 2\ln[(d+R+d_1)/R]$. The electric potential at a point $\boldsymbol{r}$ exterior to the tip and sample is given by

$$\begin{aligned}\varphi_0(\boldsymbol{r}) &= -2Q\left[\ln|\boldsymbol{r}-\boldsymbol{r}_+| - \ln|\boldsymbol{r}-\boldsymbol{r}_-|\right] \\ &= Q\int_{-\infty}^{\infty}\frac{\mathrm{d}q}{|q|}\mathrm{e}^{iqx}\left[\mathrm{e}^{-|q||z-z_+|} - \mathrm{e}^{-|q||z-z_-|}\right] , \end{aligned} \tag{19.58}$$

where $\boldsymbol{r}_\pm = \pm\hat{z}d_1$. The attractive cantilever surface force can be calculated in a straightforward way using Eq. (19.58) [43].

A somewhat different picture applies in the case of an oscillating charged tip. The cantilever charge is not changed when its tip moves parallel to the surface, while the sample charge varies with time at any fixed point. Thus, the electric field from the oscillating tip will be the same as that from the oscillating wire located at $d = d_1$. This electric field will induce an electric field outside the sample. The oscillating electric potential due to the tip oscillation at a point $\boldsymbol{r}$ exterior to the tip and sample is given by

$$\varphi_1(\boldsymbol{r},t) = \varphi_1(\boldsymbol{r})\,\mathrm{e}^{-i\omega t} + c.c.\,, \tag{19.59}$$

where *c.c.* stands for complex conjugate and where

$$\varphi_1(\boldsymbol{r}) = iQu_0\int_{-\infty}^{\infty}\frac{\mathrm{d}q\,q}{|q|}\mathrm{e}^{iqx}\left[\mathrm{e}^{-|q||z-z_+|} - \mathrm{e}^{-|q||z-z_-|}R_p(q,\omega)\right] , \tag{19.60}$$

and $R_p(q,\omega)$ is the reflection amplitude for the p-polarized electromagnetic waves. The electric field is given by $\boldsymbol{E}(\boldsymbol{r}) = -\boldsymbol{\nabla}\varphi(\boldsymbol{r})$. The energy dissipation per unit time induced by the electromagnetic field inside the metallic substrate is determined by integrating the Poynting vector over the surface of the metal, and is given by

$$\begin{aligned} P &= \frac{c}{4\pi}\int \mathrm{d}S\hat{z}\cdot[\boldsymbol{E}(\boldsymbol{r})\times\boldsymbol{B}^*(\boldsymbol{r})]_{z=+0} + c.c. \\ &= -\frac{i\omega}{4\pi}\int \mathrm{d}S\left(\varphi_1(\boldsymbol{r})\frac{\mathrm{d}}{\mathrm{d}z}\varphi_1^*(\boldsymbol{r})\right)_{z=+0} + c.c. \\ &= 4\omega Q^2|u_0|^2 w\int_0^\infty \mathrm{d}q\,q\,\mathrm{e}^{-2qd_1}\,\mathrm{Im}\,R_p(\omega,q)\,. \end{aligned} \tag{19.61}$$

Taking into account that the energy dissipation per unit time must be equal to $2\omega^2\Gamma|u_0|^2$, using (19.61) gives the friction coefficient:

$$\Gamma = \lim_{\omega\to 0} 2C^2V^2 w\int_0^\infty \mathrm{d}q\,q\,\mathrm{e}^{-2qd_1}\,\frac{\mathrm{Im}\,R_p(\omega,q)}{\omega}\,. \tag{19.62}$$

Now we assume that the electric potential on the surface of the tip is inhomogeneous, consisting of domains or "patches". Thus, the cylinder with

linear size w is "divided" into cylinder segments with linear sizes $w_i : w = \sum_i w_i \gg w_i \gg \sqrt{dR}$ and with surface potentials $V_{is} = V + V_i$, where V is the bias voltage and V_i is the randomly fluctuating surface potential for the domain i. In the case of a cylindrical tip geometry, all domains give independent contributions to friction, which can be obtained from Eq. (19.62) after the replacements $\to V + V_i$ and $w \to w_i$. The overall contribution from all of the domains to the friction is given by

$$\begin{aligned} \Gamma &= \sum_i \Gamma_i = \sum_i \lim_{\omega\to 0} 2C^2(V+V_i)^2 w_i \int_0^\infty \mathrm{d}q\, q\, \mathrm{e}^{-2qd_1} \frac{\operatorname{Im} R_p(\omega,q)}{\omega} \\ &= \lim_{\omega\to 0} 2C^2(V^2+V_0^2) w \int_0^\infty \mathrm{d}q\, q\, \mathrm{e}^{-2qd_1} \frac{\operatorname{Im} R_p(\omega,q)}{\omega}, \end{aligned} \tag{19.63}$$

where we consider that the average value of the fluctuating surface potential $\langle V_i \rangle = \sum_i w_i V_i = 0$ and $V_0^2 = \sum_i w_i V_i^2 / w$, so that V_0 is the root mean square variation of the surface potential. According to Eq. (19.63), bias voltage and patch contributions to the friction have the same dependence on d.

Many experiments use thermally evaporated thin films of gold [11]. The work functions of gold are 5.47, 5.37, and 5.31 eV for the $\langle 100 \rangle$, $\langle 110 \rangle$, and $\langle 111 \rangle$ directions, respectively [78]. If the surfaces are clean and amorphous, then we can assume that they consist of equal areas of these three crystallographic planes, and the root-mean-square $\langle \sigma_v^2 \rangle^{1/2}$ of the potential distribution becomes approximately 90 mV. When annealed, thin gold film forms mesa structures with their $\langle 111 \rangle$ crystallographic planes exposed. In this case, variations in the surface potential are presumably generated by the material lying between the mesas. The size of the mesa depends on the temperature of the substrate during the formation of the film.

Sukenik et al. studied the root mean square variation of the surface potential due to thermally evaporated gold using the Stark effect in sodium atoms [80]. The films were partially optically transparent with a thickness of 42 nm and heated at 120 °C for several hours in vacuum. They deduced that the magnitude of the fluctuating surface potential is $V_0 = 150\,\mathrm{mV}$, and showed that the scale of the lateral variation of the surface potential is on the order of the film thickness. The measurement of the noncontact friction between a gold tip and the gold sample gave $V_0 \sim 0.2\,\mathrm{V}$ [11], thus confirming the prediction of the theory that this parameter is determined by the root mean square variation in the surface potential.

Now, let us consider a spherical tip (radius R) with constant voltage surface domains of linear size R_i. If $R \gg R_i \gg \sqrt{dR}$, the domain on the apex of the tip will give the main contribution to the friction. In this case, we can neglect the spatial variation in the surface potential, and the electric field induced by the bias voltage is approximately the same as that which would be produced in the vacuum region between two point charges $\pm Q_i = \pm C(V + V_i)$ located at

$$z = \pm d_1 = \pm\sqrt{3Rd/2 + \sqrt{(3Rd/2)^2 + Rd^3 + d^4}}\,, \tag{19.64}$$

where

$$C = \frac{d_1^2 - d^2}{2d} \,. \tag{19.65}$$

It can been shown that the electrostatic force between the tip and the metal surface within this approximation agrees very well with the exact expression for a sphere above a metal surface [81]. The vibrations of the tip will produce an oscillating electromagnetic field, which, in the vacuum region, coincides with the electromagnetic field of an oscillating point charge. The friction coefficient for a point charge moving parallel to the surface due to the electromagnetic energy losses inside the sample is determined by [82]

$$\Gamma_{\parallel} = \lim_{\omega \to 0} \frac{Q_i^2}{2} \int_0^{\infty} \mathrm{d}q \, q^2 \, \mathrm{e}^{-2qd_1} \frac{\mathrm{Im}\, R_p(\omega, q)}{\omega} \,. \tag{19.66}$$

For motion normal to the surface, $\Gamma_{\perp} = 2\Gamma_{\parallel}$. Thus, just as for the cylindrical tip geometry, for a spherical tip the friction depends quadratically on the bias voltage. However, for a spherical tip, the parabola begins from zero in contrast to the parabola for the cylindrical tip, which begins from a finite positive value.

Clean Surface

For clean flat surfaces, the reflection coefficient is determined by the Fresnel formula (19.24). In this case, for a cylindrical tip radius with $R \gg d$, Eq. (19.63) gives:

$$\Gamma_{\mathrm{cl}}^{\mathrm{c}} = \frac{w(V^2 + V_0^2)}{2^6 \pi \sigma d^2} \,. \tag{19.67}$$

This formula, neglecting the contribution from the spatial variation of the surface potential, was first obtained recently in [43]. With $w = 7 \cdot 10^{-6}\,\mathrm{m}$ and $\sigma = 4 \cdot 10^{17}\,\mathrm{s}^{-1}$ (corresponds to gold at 300 K), and with $d = 20\,\mathrm{nm}$ and $V = 1\,\mathrm{V}$, Eq. (19.67) gives $\Gamma = 2.4 \cdot 10^{-20}\,\mathrm{kg/s}$, which is eight orders of magnitude smaller than the experimental value $3 \cdot 10^{-12}\,\mathrm{kg/s}$ [11].

Assuming $R \gg d$, then using (19.66) and (19.24) gives the friction between a spherical tip and a clean sample surface:

$$\Gamma_{\mathrm{cl}}^{s} = \frac{3^{1/2} R^{1/2} V^2}{2^7 d^{3/2} \pi \sigma} \,. \tag{19.68}$$

This expression is only a factor of 1.6 smaller than the result obtained independently in [43]. For the same parameters as shown above, and at $d = 20\,\mathrm{nm}$, the friction for a spherical tip is two orders of magnitude smaller than for the cylindrical tip. The friction determined by Eq. (19.68) has the same distance dependence as found in the experiment in [11]. However, the magnitude of the friction is too small to explain the experimental data.

To get an insight into possible enhancement mechanisms for noncontact friction, it is instructive to note that Eq. (19.67) can be obtained qualitatively

from the following simple geometrical arguments [83]. The vibrating tip will induce current in the sample in a volume of spatial dimensions L_x, L_y and L_z. The instantaneous dissipated power in the sample is given by $P \sim I^2 r$, where I is the current and r is the effective resistivity. The current I is proportional to the tip velocity v_x, and can be written as $I \sim v_x Q_\mathrm{t}/L_x$, where Q_t is the charge of the tip. The effective resistance r can be approximated by the macroscopic relation $r = \rho L_x/L_y L_z$, where ρ is the resistivity. Using these simple expressions for current I and resistance, and using the relation $Q_\mathrm{t} = C_\mathrm{t} V_s$ (where C_t is the tip–sample capacitance) for the induced charge, the instantaneous $P = I^2 r$ power dissipation is

$$P \sim \rho \frac{v_x^2 C_\mathrm{t}^2 V_s^2}{L_x L_y L_z} \,. \tag{19.69}$$

Comparing this expression with $P = \Gamma v_x^2$, we get

$$\Gamma \sim \rho \frac{C_\mathrm{t}^2 V_s^2}{L_x L_y L_z} \,. \tag{19.70}$$

For a cylindrical tip vibrating above the clean surface, $L_y \sim w$ and $L_x \sim L_z \sim d_1$. For $d \ll R$, the tip–sample capacitance $C_\mathrm{t} \sim w\sqrt{R/8d}$ and $d_1 \sim \sqrt{2dR}$. Substituting these expressions into Eq. (19.70) gives Eq. (19.67) to within a numerical factor of order unity. From Eq. (19.70), it follows that the friction will increase when the thickness L_z of the "dissipation volume" decreases. This is the reason for why 2-D systems may exhibit higher friction than 3-D systems.

Film On Top of a High-Resistivity Substrate

From the qualitative arguments given above, it follows that for a thin metal film on top of a high-resistivity substrate (a dielectric or a high-resistivity metal), the friction will be larger than for semi-infinite substrate with the same bulk resistivity as for the film. In this case, the thickness L_z of the volume where the dissipation occurs, will be determined by the thickness of the film, and according to Eq. (19.70), this will give rise to a strong enhancement of the friction.

For a planar film of thickness d_f and dielectric constant ϵ_2 on top of a substrate with dielectric constant ϵ_3, the reflection coefficient is determined by

$$R_p = \frac{R_{p21} - R_{p23} \exp(-2qd_\mathrm{f})}{1 - R_{p21} R_{p23} \exp(-2qd_\mathrm{f})} \,, \tag{19.71}$$

where

$$R_{pij} = \frac{\epsilon_i - \epsilon_j}{\epsilon_i + \epsilon_j} \,, \tag{19.72}$$

where index 1 is associated with the vacuum. For a metallic film on a dielectric substrate, or a metallic film on a metallic substrate with $\sigma_2 \gg \sigma_3$, and for $d_1 \gg d_\mathrm{f}$ and $R \gg d$, Eqs. (19.63) and (19.71) give

$$\Gamma_\mathrm{f}^c = \frac{w(V^2+V_0^2)R^{1/2}}{2^{9/2}\pi\sigma d_\mathrm{f} d^{3/2}} \,. \tag{19.73}$$

This is greater than the corresponding friction for the infinitely thick sample by a factor of $2d_1/d_\mathrm{f}$. For a thin film, the effective resistivity of the substrate is increased, giving rise to additional ohmic dissipation. In [43], Eq. (19.73) was obtained using a less general approach and by neglecting the spatial variation in the surface potential. The conditions necessary to make Eq. (19.73) valid could not be determined in this simplified approach.

2-D System On Top of a Dielectric or Metal Substrate

Let us now consider a 2-D system, e. g. electronic surface states or a quantum well, or an incommensurate layer of ions adsorbed on a metal surface. For example, for the Cs/Cu(100) system, experiment suggests the existence of an acoustic film mode even for a very dilute phase ($\theta \approx 0.1$). This implies that the Cs/Cu(100) adsorbate layer experiences a negligible surface pinning potential. The reflection coefficient for a 2-D system is given by Eq. (19.43) and

$$\mathrm{Im}\, R_p \approx \frac{2\omega\eta q a \omega_q^2}{(\omega^2-\omega_q^2)^2+\omega^2\eta^2} \,, \tag{19.74}$$

where $\omega_q^2 = 4\pi n_a e^{*2} a q^2/M$. In the case of a 2-D structure on top of a dielectric, the factor qa in Eq. (19.74) and in the expression for ω_q^2 must be replaced by $1/\epsilon$, where ϵ is the dielectric function of the substrate. Using (19.74) in (19.63) for $R \gg d$, we get

$$\Gamma_\mathrm{ad}^c = \frac{w\eta M R^{1/2}(V^2+V_0^2)}{2^{9/2} d^{3/2}\pi n_a e^{*2}} \,. \tag{19.75}$$

This friction exhibits the same distance dependence as observed experimentally [11]. The same expression for the friction is valid for a 2-D structure on top of a dielectric. Comparing Eqs. (19.67) and (19.75), we find that a 2-D structure on top of a substrate gives the same friction as for a clean surface with effective conductivity $\sigma_\mathrm{eff} = n_a e^{*2}/M\eta 2d_1$. We obtain agreement with experiment at $d = 20\,\mathrm{nm}$ if $\sigma_\mathrm{eff} \approx 4 \cdot 10^9\,\mathrm{s}^{-1}$. In the case of a 2 D electron system, such an effective conductivity is obtained for $R = 1\,\mu\mathrm{m}$ if $\eta = 10^{14}\,\mathrm{s}^{-1}$ and $n_a = 10^{15}\,\mathrm{m}^{-2}$. For Cs/Cu(100) and for $n_a = 10^{18}\,\mathrm{m}^{-2} (\theta \approx 0.1)$, the electric charge of the Cs ions $e^* = 0.28e$ [61]. Due to the similarities of Cu and Au surfaces, a similar effective charge can be expected for the Cs/Au surface. For such a 2-D system, agreement with experiment is obtained for $n_a = 10^{18}\,\mathrm{m}^{-2}$ and $\eta = 10^{11}\,\mathrm{s}^{-1}$. In [44], we estimated the damping parameter for Cs atoms associated with the covalent bond $\eta_{\parallel\mathrm{cov}} = 3 \cdot 10^9\,\mathrm{s}^{-1}$.

However, the collisions between the ions, and between the ions and other surface defects, will also contribute to η. In this case, $\eta_{\rm col} \sim v_T/l$, where $v_T \sim \sqrt{k_{\rm B}T/M}$, and l is the ion mean free path. For $T = 293\,{\rm K}$ and $l \sim 1\,{\rm nm}$ we get $\eta_{\rm col} = 10^{11}\,{\rm s}^{-1}$.

For a spherical tip with a 2-D system on top of the substrate, from Eqs. (19.74) and (19.66) and for $R \gg d$ we get the contribution to the friction from the 2-D system

$$\Gamma^s_{\rm ad} = \frac{3RM\eta V^2}{2^6 d\pi n_a e^{*2}} \,. \tag{19.76}$$

At $d = 20\,{\rm nm}$, this friction is approximately two orders of magnitude smaller than that for the cylindrical tip.

19.4.2 Friction Due to Spatial Fluctuations of Static Charge in the Bulk of the Sample

In this section we consider a dielectric substrate with a stationary, inhomogeneous distribution of charged defects. Such a situation was investigated experimentally [11] by employing a fused silica sample irradiated with γ-rays. Positively charged centers (Si dangling bonds) are generated during irradiation. Randomly distributed positive charges are compensated by randomly distributed negative charges, meaning that, on average, the sample is electrically neutral. We model the sample as consisting of microscopically small volume elements ΔV_i. Each element is chosen to be small enough that no more than one charge center is present in it. The electric charge q_i of each element is equal to $\pm e$ or 0, in such a way that the average $\langle q_i \rangle = 0$. We will consider the charge fluctuations in different volume elements i, j to be statistically independent, so that $\langle q_i q_j \rangle = 0$ for $i \neq j$. The mean square of charge fluctuations within a given element $\langle q_i q_i \rangle \approx 2ne^2$, where n is the average number of positive charges in one volume element. In the absence of cross terms, the average tip–sample friction coefficient is determined by adding the friction coefficients from all charges q_i. According to Eq. (19.66), the contribution to the friction coefficient from charge q_i in element ΔV_i is given by

$$\Delta\Gamma_{i\|} = \lim_{\omega\to 0} ne^2 \int_0^\infty {\rm d}q\, q^2\, {\rm e}^{-2qd_i}\, \frac{{\rm Im}\, R_p(\omega, q)}{\omega} \,, \tag{19.77}$$

where $d_i = D(x_i, y_i) - z_i$. Here the coordinates x_i, y_i, z_i give the position of the i-th volume element in the substrate, and $D(x_i, y_i)$ is the distance between the substrate and points x_i, y_i located on the surface of the tip. The total friction coefficient is obtained by summing over all elements. Replacing the sum by an integral ($n\sum \to c\int {\rm d}^3 r$, where c is the number of positive charge centers per unit volume), and integrating over z gives

$$\Gamma_\| = \lim_{\omega\to 0} \frac{ce^2}{2} \int_0^\infty {\rm d}q\, q \int {\rm d}x \int {\rm d}y\, {\rm e}^{-2qD(x,y)}\, \frac{{\rm Im}\, R_p(\omega, q)}{\omega} \,. \tag{19.78}$$

For a cylindrical tip $D(x,y) = d + x^2/2R$, and so we get

$$\Gamma_{\parallel}^{c} = \lim_{\omega\to 0} \frac{\sqrt{\pi R} c e^2 w}{2} \int_0^\infty \mathrm{d}q\, q^{1/2}\, \mathrm{e}^{-2qd} \frac{\mathrm{Im}\, R_p(\omega, q)}{\omega} \,. \tag{19.79}$$

Using the same parameters as used in Sect. 2.2, for a gold tip separated by $d = 10$ nm from a dielectric sample with $c = 7 \cdot 10^{17}\,\mathrm{cm}^{-3}$, we get $\Gamma_{\parallel} = 4.4 \cdot 10^{-20}\,\mathrm{kgs}^{-1}$.

For the tip surface with a 2-D structure on it, using Eq. (19.74) we get

$$\Gamma_{2D\parallel}^{c} = \frac{1}{2^{5/2}} \left(\frac{e}{e^*}\right)^2 \sqrt{\frac{R}{d}} \frac{cw}{n_a} M\eta = \frac{e^2 cw}{16\sigma_{eff} d} \,. \tag{19.80}$$

With $\sigma_{\mathrm{eff}} = n_a e^{*2}/2M\eta d_1 = 4 \cdot 10^9, c = 7 \cdot 10^{17}\,\mathrm{cm}^{-3}$, and with the other parameters kept the same as before, for $d = 10\,\mathrm{nm}$ we get $\Gamma_{2D\parallel}^{c} = 3.5 \cdot 10^{-12}\,\mathrm{kg\,s}^{-1}$, which is nearly the same as was observed experimentally [11]. Thus our theory of friction between a gold tip and silica substrate with an inhomogeneous distribution of charged defects is consistent with the theory of friction between a gold tip and gold substrate (see Sect. 2.4). In both theories we have assumed that the gold surfaces are covered by a 2-D structure.

The above analysis has ignored the screening of the electric field in the dielectric substrate. This can be justified in the case of very small tip– sample separations (substantially smaller than screening length), as only defects in the surface layer of thickness d contribute to the integral in Eq. (19.78). When the screening is important, the effective electric field outside the sample will be decreased by a factor of $(\varepsilon + 1)/2$ [79], and so the friction coefficient will be decreased by a factor of $((\varepsilon + 1)/2)^2$, which is equal to ≈ 6.25 in the case of silica. However, the inhomogeneity of the surface of the tip may be larger than that of the sample surface, so that the damping parameter η may be larger for the 2-D structure on the surface of the tip. This increase in η and screening effects will compensate each other.

19.5 Phonon and Internal Noncontact Friction

19.5.1 Noncontact Friction Due to Excitation of Substrate Phonons

Consider a tip which performs harmonic oscillation, $u = u_0 \exp(-\mathrm{i}\omega t) + c.c.$, above an elastic body with a flat surface. This will result in a fluctuating stress acting on the surface of the solid which excites acoustic waves with parallel wavenumbers $q < \omega/c_{\mathrm{s}}$, where c_{s} is the sound velocity. The stress σ_{iz} that acts on the surface of the elastic solid can be represented through the

Fourier integral

$$\sigma_{iz}(\boldsymbol{x},t)=\int\frac{\mathrm{d}^2q}{(2\pi)^2}\sigma_i(\boldsymbol{q})u_0\,\mathrm{e}^{i\boldsymbol{q}\boldsymbol{x}-i\omega t}+c.c. \tag{19.81}$$

Using the theory of elasticity (assuming an isotropic elastic medium for simplicity), one can calculate the displacement field u_i on the surface $z=0$ in response to the surface stress distribution σ_{iz}:

$$u_i(\boldsymbol{x},t)=\int\frac{\mathrm{d}^2q}{(2\pi)^2}M_{ij}(\boldsymbol{q},\omega)\sigma_j(\boldsymbol{q})u_0\,\mathrm{e}^{i\boldsymbol{q}\boldsymbol{x}-i\omega t}+c.c. \tag{19.82}$$

The explicit form of the stress tensor in the model of the elastic continuum is given in [84]. The energy dissipation per unit time equals

$$\begin{aligned}P&=\int\mathrm{d}^2x\langle\dot{u}_i(\boldsymbol{x},t)\sigma_{iz}((\boldsymbol{x},t)\rangle\\&=2\omega\int\frac{\mathrm{d}^2q}{(2\pi)^2}\,\mathrm{Im}\,M_{ij}(\boldsymbol{q},\omega)\sigma_i(\boldsymbol{q})\sigma_j^*(\boldsymbol{q})|u_0|^2\end{aligned} \tag{19.83}$$

where $\langle\ldots\rangle$ stands for a time average. The energy dissipation per unit time must be equal to $\Gamma<\dot{u}(t)^2>=\Gamma 2\omega^2\left|u_0\right|^2$. Comparing this expression with (19.83) gives

$$\Gamma=\int\frac{\mathrm{d}^2q}{(2\pi)^2}\frac{\mathrm{Im}\,M_{ij}(\boldsymbol{q},\omega)}{\omega}\sigma_i(\boldsymbol{q})\sigma_j^*(\boldsymbol{q})\,. \tag{19.84}$$

At typical experimental conditions, we have $\omega\sim10^3-10^6\,\mathrm{s}^{-1}$ and $qr^*<\omega r^*/c_\mathrm{s}<10^{-3}\ll1$, where the effective radius of the interaction $r^*\approx\sqrt{dR}$, and where d is the separation between the tip and the sample, and R is the radius of curvature of the tip. In this case, the contribution from the excitation of acoustic waves to the friction can be determined by calculating the energy dissipation due to the oscillating point force applied to the surface of the semi-infinite elastic continuum. These calculations were done in connection with the vibrational energy relaxation of adsorbates [85]. According to this theory, the friction coefficient for vibration of the tip normal to the surface is given by

$$\Gamma_\perp=\frac{\xi_\perp}{4\pi}\frac{K^2}{\rho c_\mathrm{t}^3}\,, \tag{19.85}$$

where $\xi_\perp\approx1.65$, c_t is the transverse sound velocity of the solid, ρ is the mass density of the sample, $K=\partial F_z/\partial d$, where $F_z(d)$ is the force acting on the tip due to interaction with the sample.

For vibration of the tip parallel to the flat surface, the friction coefficient due to excitation of the acoustic waves is given by

$$\Gamma_\parallel=\frac{\xi_\parallel}{4\pi}\frac{\omega^2}{\rho c_\mathrm{t}^5}F_z^2(d)\,, \tag{19.86}$$

where $\xi_\parallel \approx 1.50$. From a comparison of Eqs. (19.85) and (19.86), we see that $\Gamma_\parallel/\Gamma_\perp \sim (\omega d/c_{\rm t})^2 \ll 1$. We now consider two different contributions to the tip–sample interaction.

Van der Waals Interaction

Accordingly to the Lifshitz theory [23, 27], the stress $\sigma_{zz}(d)$ acting on the surfaces of two identical semi-infinite bodies due to van der Waals interactions at small separation $d \ll c/\omega_p$ (where ω_p is the plasma frequency) and $d \ll \lambda_T$ is given by:

$$\sigma_{zz}(d) = \frac{\hbar}{8\pi^2 d^3} \int_0^\infty {\rm d}\xi \frac{[\varepsilon(i\xi)-1]^2}{[\varepsilon(i\xi)+1]^2} \,. \tag{19.87}$$

In the Drude model the explicit form of ε is

$$\varepsilon(i\xi) = 1 + \frac{\omega_p^2}{\xi(\xi+\eta)} \,. \tag{19.88}$$

For a typical metal, the damping constant $\eta \ll \omega_p$ and this can be neglected when integrating Eq. (19.87). It follows from Eqs. (19.87) and (19.88) that

$$\sigma_{zz} = \frac{\hbar\omega_p}{32\sqrt{2}\pi d^3} \,. \tag{19.89}$$

For a spherical tip of radius R, we get

$$F_z(d) \approx 2\pi \int_0^\infty {\rm d}\rho\rho\sigma_{zz}(d + \frac{\rho^2}{2R}) = \frac{R\hbar\omega_p}{32\sqrt{2}d^2} \tag{19.90}$$

and

$$K^s = \frac{R\hbar\omega_p}{16\sqrt{2}d^3} \,. \tag{19.91}$$

Similarly, in the case of a cylindrical tip, we have

$$F_z^c(d) = \frac{3wR^{1/2}\hbar\omega_p}{2^8 d^{5/2}} \tag{19.92}$$

and

$$K^c = \frac{15wR^{1/2}\hbar\omega_p}{2^9 d^{7/2}} \,. \tag{19.93}$$

For a copper tip separated from a copper substrate by $d = 10\,{\rm nm}$, and with $= 1\,\mu{\rm m}, w = 7\,\mu{\rm m}$, we get for a spherical tip $\Gamma_\perp^s = 6.3 \cdot 10^{-18}\,{\rm kg\,s^{-1}}$ and for a cylindrical tip $\Gamma_\perp^c = 1.3 \cdot 10^{-14}\,{\rm kg\,s^{-1}}$. The phononic friction decreases as d^{-6} and d^{-7} for spherical and cylindrical tips, respectively.

Electrostatic Interaction Due to a Bias Voltage

In the presence of the bias voltage V, the attractive force between the tip and the sample at $d \ll R$ is given by

$$F^c(d) = \frac{wV^2R^{1/2}}{2^{7/2}d^{3/2}} \tag{19.94}$$

for a cylindrical tip, and

$$F^s(d) = \frac{RV^2}{4d} \tag{19.95}$$

for a spherical tip. For a bias voltage $V = 1\,\mathrm{V}$, and with the other parameters the same as above, we get $\Gamma_\perp^s = 8.8 \cdot 10^{-17}\,\mathrm{kg\,s^{-1}}$ and $\Gamma_\perp^c = 1.2 \cdot 10^{-13}\,\mathrm{kg\,s^{-1}}$ for the spherical and cylindrical tips, respectively. Note that in this case the friction depends on the bias voltage as V^4.

For vibrations of the tip parallel to the sample surface, the expression for the friction coefficient contains the additional small factor $(\omega d/c_\mathrm{s})^2 \ll 1$. Thus, the friction coefficient for parallel vibrations of the tip will be many orders of magnitude smaller than for normal vibrations.

19.5.2 Noncontact Friction Due to Internal Friction of the Substrate

When studying the phononic friction in the preceding section, it was assumed that deformations of the solids are purely elastic. However, the deformation will be purely elastic or adiabatic only for an infinitesimally small velocity, so that at every moment in time the system stays in the equilibrium state. However, real motion always occurs with finite velocity, and the body does not stay in equilibrium, so "flow processes" occur which tend to bring it back to equilibrium. This leads to nonadiabatic deformations, resulting in dissipation of the mechanical energy.

The energy dissipation is determined by two kind of processes. First, in the presence of a temperature gradient in the body, heat flow occurs. Secondly, if some kind of internal motion occurs in the body, then nonadiabatic processes occur, related to the finite velocity of the motion; these processes of energy dissipation can be denoted, as in liquids, as internal friction or viscosity.

The friction coefficient due to the internal friction is determined by Eq. (19.84). However, in contrast to phononic friction, large values of $q \gg \omega/c_\mathrm{t}$ play the most important role in internal friction. For $q \gg \omega/c_\mathrm{s}$, the tensor component M_{zz} is given by [84]

$$M_{zz} = \frac{2(1-\nu^2)}{Eq}\,, \tag{19.96}$$

where $E(\omega)$ is the complex elastic modulus and ν is the Poisson ratio.

Van der Waals Interaction

For $R \gg d$, only the σ_{zz} component of the stress tensor due to the van der Waals interaction is important. In this case, for vibrations of the cylindrical tip parallel to the sample surface, we get

$$\sigma_z(\boldsymbol{q}) = \int \mathrm{d}^2x\, \mathrm{e}^{i\boldsymbol{q}\boldsymbol{x}} \frac{\partial}{\partial x}\sigma_{zz}(\boldsymbol{x}) = -\frac{iq_x R^{1/2}}{2^7 d^{5/2}} \frac{\sin(q_y w/2)}{q_y}(3+\xi^2+3\xi)\,\mathrm{e}^{-\xi}\,, \tag{19.97}$$

where $\xi = \sqrt{2dR}q_x$. Using (19.97) and (19.96) in (19.84), we get for a cylindrical tip

$$\Gamma_{\parallel}^{c} = \frac{75\pi}{2^{16}} \frac{w\hbar^2\omega_p^2}{d^6} \frac{\mathrm{Im}(E/(1-\nu^2))}{\omega|E/(1-\nu^2)|^2}\,. \tag{19.98}$$

For the spherical tip, similar calculations give

$$\Gamma_{\parallel}^{s} = \frac{0.25}{2^9\sqrt{2}\pi} \frac{R^{1/2}\hbar^2\omega_p^2}{d^{11/2}} \frac{\mathrm{Im}(E/(1-\nu^2))}{\omega|E/(1-\nu^2)|^2}\,. \tag{19.99}$$

In general, $\mathrm{Im}[E(\omega)/(1-\nu^2)]$ has many resonance peaks, corresponding to different thermally activated relaxation processes. One important source of internal friction at high frequencies is related to thermal currents: elastic compression of a material is commonly associated with heating effects. If the compression takes place sufficiently rapidly, there is no opportunity for heat to be conducted away, while for very slow compression, temperature gradients are eliminated by thermal conduction. In both of these cases, the process of compression will be reversible. In the former case it will be adiabatic and in the latter one isothermal. In both of these limiting cases the contribution from thermal current to the internal friction will be negligible. However, in the intermediate frequency regime we expect dissipation of mechanical energy into heat. The characteristic frequency for the maximum dissipation will be of order $\omega_t = 1/\tau$, where, from dimensional arguments, we expect the relaxation time $\tau \sim l^2/D$, where l is the linear size of the compression region and D the thermal diffusibility $D = \kappa/\rho C_p$ (where C_p is the specific heat and κ the heat conductivity). For $l \sim 10^3$ Å, this gives $\omega_t \approx 10^{11}\,\mathrm{s}^{-1}$ for gold, which is much higher than the resonance frequency of the cantilever of the atomic force microscope. Another very important contribution to the internal friction is point-defect flipping. This involves thermally activated transitions of point defects or loose sites in crystalline and amorphous networks. A special case is the vibrational motion of adsorbates on the surface of the substrate and/or on the tip, as was treated separately above. Another contribution to the internal friction comes from grain boundary slip [86]. For a copper cylindrical tip and a copper substrate with $d = 10\,\mathrm{nm}, w = 7\,\mu\mathrm{m}, R = 1\,\mu\mathrm{m}, \omega = 10^4\,\mathrm{s}^{-1}$, and, as

is typical for metals [87], $\mathrm{Im}\, E(\omega)/\,|E(\omega)| \approx 10^{-5}$ and $E \approx 10^{11}\,\mathrm{N/m^2}$, then $\Gamma_{\parallel}^{c} \approx 10^{-16}\,\mathrm{kg \cdot s^{-1}}$. Thus, at this separation the internal friction provides a much smaller contribution to the friction coefficient than electrostatic friction due to bias voltage or spatial variation of the surface potential. However, internal friction can provide the dominant contribution for small separations $d \leq 1\,\mathrm{nm}$. For a spherical tip with $R = 1\,\mu\mathrm{m}$, the friction coefficient is two orders of magnitude smaller. Finally, we note, as a curiosity, that the internal friction of solids makes a very important contribution to the rolling resistance of most solids [88], and is the main contribution to the friction from rubber on rough substrates (e. g. road surfaces [88]), where $\mathrm{Im}\, E(\omega)/\,|E(\omega)| \approx 1$ in the transition region between the rubbery and glassy regions of the viscoelastic spectrum of rubber.

19.6 Summary and Outlook

All material bodies are surrounded by a fluctuating electromagnetic field due to thermal and quantum fluctuations of the current density inside the bodies. This fluctuating electromagnetic field can be studied using the semiclassical theory due to Rytov, or via quantum field theory. In the near-field zone of any body, the electromagnetic field is greatly enhanced in comparison with the far-field region due to the existence of evanescent waves. This enhancement is especially large if the surface of the body can support surface modes, e. g., surface plasmon polaritons, surface phonon polaritons or adsorbate vibrational modes of adsorbed ions. The thermal radiation emitted by surface modes is spatially and temporally coherent, which can be used to develop tunable infrared emitters with high power in a narrow spectral band, which may in turn be used for sensing, spectroscopy, and thermophotovoltaic applications.

The fluctuating electromagnetic field is responsible for many important phenomena, e. g., radiative heat transfer between bodies, van der Waals interactions and van der Waals friction. The heat transfer between two bodies in a vacuum is strongly modified when the surfaces become closer than the characteristic wavelength of the the thermal radiation, $\lambda_T = c\hbar/k_{\mathrm{B}}T$. For a separation $d < \lambda_T$, the heat transfer is greatly enhanced in comparison with black body radiation, due to the contribution from evanescent waves. The latter contribution is particularly large if the conductivity of the materials is tuned to maximize the heat flow due to photon tunneling. At room temperature, this implies conductivities typical of semimetals, such as carbon, or of metal–insulator composites. The radiative heat transfer can also be greatly enhanced when resonant photon tunneling between surface modes, like surface plasmon polaritons, surface phonon polaritons, or adsorbate vibrational modes can occur. In the case of resonant photon tunneling, the heat flux occurs within a very narrow frequency band, which can find application

in energy conversion technologies, such as thermophotovoltaic energy conversion. Adsorbates can generate evanescent waves that are very localized in space and fall within a very narrow frequency band. These phenomena can be used in scanning probe microscopy for local heating and modification of surfaces.

Similar to radiative heat transfer, van der Waals friction can also be greatly enhanced in the case of resonant photon tunneling between surface modes. This enhancement is especially large between surfaces which can support a 2-D acoustic branch in the spectrum of the elementary excitations, e. g. 2-D electron systems on dielectric surfaces or adsorbed layers of ions with acoustic vibrations parallel to the surface. For such systems the electrostatic friction can be also important. In particular, the friction observed in [11] can be explained by electrostatic friction when the electromagnetic field from a moving atomic force microscope tip is mediated by bias voltage or spatial fluctuations of the charge in the bulk or on the surface. Van der Waals friction is responsible for frictional drag between quantum wells and can also be measured in noncontact friction experiments using present state-of-the-art equipment. Noncontact friction is technologically important for ultrasensitive force detection and from a basic scientific point of view.

Van der Waals friction will be even more important for 1-D structures like carbon nanotubes. The one-dimensional nature of carbon nanotubes may have profound consequences for the basic physical phenomenology used to describe them: single-wall carbon nanotubes have been predicted to be Luttinger liquids [89,90], and some experimental evidence for this exists [91,92], even though other interpretations has been suggested [93]. The question of whether multiwall carbon nanotubes are Fermi or Luttinger liquids has been investigated extensively experimentally [94–96] and theoretically [97] and seems to depend on the situation. The situation is also not clear for ropes constructed from nanotubes yet [98]. It may be possible to get additional information about strongly correlated low-dimensional systems from studying the noncontact friction between them.

In a recent, remarkable experiment [99,100], it was observed that the flow of an ion-rich liquid such as water through bubbles of single-walled carbon nanotubes induces a voltage in the nanotube along the direction of the flow. Several mechanisms were proposed to explain the frictionally induced voltage in the carbon nanotube [99–101]. In one proposed mechanism, the induced voltage is related to the fluctuating Coulomb field created by charge density fluctuations in the moving liquid. Thus the problem of induced voltage can be related to the van der Waals friction. However, the situaion is not clear yet.

Acknowledgement. A.I.V acknowledges financial support from the Russian Foundation for Basic Research (Grant N 04-02-17606). B.N.J.P. and A.I.V. acknowledge support from the European Union's Smart Quasicrystals project.

References

1. K. Joulain, J.P. Mulet, F. Marquier, R. Carminati, and J.J. Greffet, Surf. Sci. Rep. **57,** 59 (2005)
2. J.B. Pendry, J. Phys.: Condens. Mat. **11**, 6621 (1999).
3. A.I. Volokitin and B.N.J. Persson, Phys. Rev. B **63**, 205404 (2001); Phys. Low-Dim. Struct. **5/6**, 151 (2001)
4. A.I. Volokitin and B.N.J. Persson, Phys. Rev. B **69**, 045417 (2004)
5. A.I. Volokitin and B.N.J. Persson, JETP Lett. **78**, 457 (2003)
6. J.P. Mulet, K. Joulin, R. Carminati, and J.J. Greffet, Appl. Phys. Lett. **78**, 2931 (2001)
7. A. Majumdar, Ann. Rev: Mater. Sci. **29**, 505 (1999).
8. A. Kittel, W. Müller-Hirch, J. Parisi, S.-A. Biehs, D. Reddig, and M. Holthaus, Phys. Rev. Lett. **87,** 096801 (2001).
9. I. Dorofeyev, H. Fuchs, G. Wenning, and B. Gotsmann, Phys. Rev. Lett. **95,** 224301 (2005).
10. B. Gotsmann and H. Fuchs, Phys. Rev. Lett. **86,** 2597 (2001).
11. B.C. Stipe, H.J. Mamin, T.D. Stowe, T.W. Kenny, and D. Rugar, Phys. Rev. Lett. **87,** 096801 (2001).
12. H.J. Mamin and D. Rugar, Appl. Phys. Lett. **79**, 3358 (2001)
13. P.M. Hoffmann, S. Jeffery, J.B. Pethica, H. Özgür Özer and A. Oral, Phys. Rev. Lett. **87,** 265502 (2001).
14. J.J. Greffet, R. Carminati, K. Joulain, J.P. Mulet, S. Mainguy, and Y. Chen, Nature **416**, 61 (2002)
15. T.J. Gramila, J.P. Eisenstein, A.H. MacDonald, L.N. Pfeiffer, and K.W. West, Phys. Rev. Lett. **66**, 1216 (1991); Surf. Sci. **263,** 446 (1992).
16. T.J. Gramila, J.P. Eisenstein, A.H. MacDonald, L.N. Pfeiffer, and K.W. West, Phys. Rev. B **47**, 957 (1993); Physica B **197,** 442 (1994).
17. U. Sivan, P.M. Solomon, and H. Shtrikman, Phys. Rev. Lett. **68,** 1196 (1992).
18. D. Rugar, R. Budakian, H.J. Mamin and B.W. Chui, Nature **430**, 329 (2004)
19. J.A. Sidles, J.L. Carbini, K.J. Bruland, D. Rugar, O. Zuger, S. Hoen, and C.S. Yannoni, Rev. Mod. Phys. **67**, 249 (1995)
20. G.P. Berman, G.D. Doolen, P.C. Hammel, and V.I. Tsifrinovich, Phys. Rev. B **61**, 14694 (2000).
21. N. Arkani-Hamed, S. Dimopoulos, and G. Dvali, Phys. Lett. B **429**, 263 (1998); Sci. Am. **283**, 62 (2000)
22. U. Mohideen and A. Roy, Phys. Rev. Lett. **81**, 4549 (1998)
23. I.E. Dzyaloshinskii, E.M. Lifshitz and L.P. Pitaevskii, Adv. Phys. **10**, 165 (1961)
24. J.B. Pendry, J. Phys. C **9**, 10301 (1997).
25. A.I. Volokitin and B.N.J. Persson, J. Phys.: Condens. Mat. **11**, 345 (1999); Phys. Low-Dim. Struct. **7/8**, 17 (1998)
26. E.V. Teodorovitch, Proc. Roy. Soc. A **362**, 71 (1978).
27. E.M. Lifshitz, Zh. Eksp. Teor. Fiz. **29,** 94 (1955) [Sov. Phys.-JETP **2**, 73 (1956)]
28. J. Mahanty, J. Phys. B: Atom. Molec. Phys. **13**, 4391 (1980)
29. G.V. Dedkov and A.A. Kyasov, Phys. Lett. A **259**, 38, (1999).
30. A.A. Kyasov and G.V. Dedkov, Surf. Sci. **463**, 11 (2000)

31. A.I. Volokitin and B.N.J. Persson, Phys. Rev. B **65**, 115419 (2002)
32. W.L. Schaich and J. Harris, J. Phys. F **11**, 65 (1981).
33. L.S. Levitov, Europhys. Lett. **8**, 499 (1989)
34. V.G. Polevoi, Zh. Eksp. Teor. Fiz. **98**, 1990 (1990). [Sov. Phys. JETP **71**, 1119 (1990)].
35. V.E. Mkrtchian, Phys. Lett. **207**, 299 (1995).
36. B.N.J. Persson and Z. Zhang, Phys. Rev. B **57**, 7327 (1998)
37. A.I. Volokitin and B.N.J. Persson, Phys. Rev. Lett. **91**, 106101 (2003).
38. A.I. Volokitin and B.N.J. Persson, Phys. Rev. B **68**, 155420 (2003).
39. A. Krishnan, T. Thio, T.J. Kim, H.J. Lezec, T.W. Ebbesen, P.A. Wolf, J. Pendry, L. Martin-Moreno, and F.J. Garcia-Vidal, Opt. Commun. **200**, 1 (2001)
40. M.S. Tomassone and A. Widom, Phys. Rev. B **56**, 4938 (1997)
41. B.N.J. Persson and A.I. Volokitin, Phys. Rev. Lett. **84**, 3504 (2000).
42. J.R. Zurita-Sánchez, J.J. Greffet, L. Novotny, Phys. Rev. A **69**, 022902 (2004).
43. A.A. Chumak, P.W. Milonni, and G.P. Berman, Phys. Rev. B **70**, 085407 (2004).
44. A.I. Volokitin and B.N.J. Persson, Phys. Rev. Lett. **94**, 86104 (2005)
45. S.M. Rytov, Yu.A. Kravtsov, and V.I. Tatarskii, *Principles of Statistical Radiophysics* (Springer, New York, 1989), Vol. 3
46. D. Polder and M. van Hove, Phys. Rev. B **4**, 3303 (1971)
47. J.J. Loomis and H.J. Maris, Phys. Rev. B **50**, 18517 (1994)
48. A.A. Abrikosov, L.P. Gor'kov and I.Ye. Dzyaloshinskii, *Quantum Field Theoretical Methods in Statistical Physics* (Pergamon Press, Oxford, 1965)
49. L.D. Landau and E.M. Lifshitz, *Statistical Physics* (Pergamon, Oxford, 1970).
50. R.C. Carminati and J.J. Greffet, Phys. Rev. Lett. **82**, 1660 (1999).
51. A.V. Shcheglov, K. Joulain, R. Carminati, and J.J. Greffet, Phys. Rev. Lett. **85**, 1548 (2000).
52. F. Marquier, K. Joulain, J.P. Mulet, R. Carminati, and J.J. Greffet, Opt. Commun. **237**, 379 (2004)
53. M.U. Pralle et al., Appl. Phys. Lett. textbf81, 4685 (2002)
54. M. Kreiter, J. Oster, R. Sambles, S. Herminghaus, S. Mittler- Neher, W. Knoll, Opt. Commun. **168**, 117 (1999)
55. W.L. Schaich, J. Chem. Phys. **60**, 1087 (1874); Solid. St. Commun. **15**, 357 (1974)
56. E.G. d'Agliano, P. Kumar, W. Schaich and H. Suhl, Phys. Rev. B **11**, 2122(1975)
57. A. Nourtier, J. Physique, **38**, 479 (1977)
58. G. Mahan, *Many-Particle Physics*, 2nd edn. (Plenum Press, New York, 1990)
59. K.L. Kliewer and R. Fuchs, Phys. Rev. **172**, 607 (1968); R. Fuchs and K.L. Kliewer, Phys. Rev. **186**, 905 (1969)
60. E.D. Palik, *Handbook of Optical Constants of Solids* (Academic, San Diego, CA, 1985)
61. P. Senet, J.P. Toennis and G. Witte, Chem. Phys. Lett. **299**, 389 (1999)
62. V. Mkrtchian, V.A. Parsegian, R. Podgornik, and W.M. Maslow, Phys. Rev. Lett. **91**, 220801 (2003)
63. P.J.E. Peebles, *Principles of Physical Cosmology* (Princeton University Press, Princeton, NY, 1993)

64. J.M. Kovac, E.M. Leitch, C. Pryke, J.E. Carlstom, N.W. Halverson, and W.L. Holzapfel, Nature (London) **420**, 772 (2002)
65. M.B. Pogrebinskii, Fiz. Tekh. Poluprov. **11**, 637 (1977) [Sov. Phys. Semicond. **11**, 372 (1977)
66. B.J. Price, Physica B **117**, 750 (1983)
67. L. Zheng and A.H. MacDonald, Phys. Rev. B **48**, 8203 (1993)
68. H.C. Tso and P. Vasilopoulos, Phys. Rev. B **45**, 1333 (1992)
69. A.-P. Jauho and H. Smith, Phys. Rev. B **47**, 4420 (1993)
70. K. Flensberg, B.Y.-K. Hu, A.-P. Jauho, and J. Kinaret, Phys. Rev. B **52**, 14761 (1995)
71. A. Kamenev and Y. Oreg, Phys. Rev. B **52**, 7516 (1995)
72. A.I. Volokitin and B.N.J. Persson, J. Phys.: Condens. Mat. **13**, 859 (2001)
73. N.D. Lang and W. Kohn, Phys. Rev. B **3**, 1215 (1971)
74. C.C. Speake and C. Trenkel, Phys. Rev. Lett. **90**, 160403 (2003)
75. Z.H. Liu, N.M.D. Brown, and A. McKinley, J. Phys.: Condens. Mat. **9**, 59 (1997)
76. F. Rossi and G.I. Opat, J. Phys. D **25**, 1349 (1992)
77. T.E. Ederth, Phys. Rev. A **62**, 062104 (2000)
78. C.R. Lide (ed), *CRC Handbook of Chemistry and Physics*, 82nd edn (CRC Press, BOCA Raton, FL, 2001).
79. L.D. Landau and E.M. Lifshitz, *Electrodynamics of Continuous Media* (Pergamon, Oxford, 1960).
80. C.I. Sukenik, et al., Phys. Rev. Lett. **70**, 560 (1993)
81. S. Hudlet, M.S. Jean, C. Guthmann, and J. Berger, Eur. Phys. J. B **2**, 5 (1998)
82. B.N.J. Persson, Phys. Rev. B **44**, 3277 (1991)
83. T.D. Stowe, W. Kenn, D.J. Thomson, and D. Rugar, Appl. Phys. Lett. **75**, 2785 (1999)
84. B.N.J. Persson, J. Chem. Phys. **115**, 3840 (2001)
85. B.N.J. Persson and R. Ryberg, Phys. Rev. B **32**, 3586 (1985)
86. F.A. McClintock and A.S. Argon, *Mechanical Behavior of Material* (Addison-Wesley, Reading, MA, 1966), pp. 475–483.
87. B.N.J. Persson, *Sliding Friction: Physical Principles and Applications*, 2nd edn (Springer, Heidelberg, 2000).
88. B.N.J. Persson, Surf. Sci. **401**, 445 (1998)
89. R. Egger and A.O. Gogolin, Phys. Rev. Lett. **79**, 5082 (1997)
90. R. Egger and A.O. Gogolin, Eur. Phys. J. B **3**, 281 (1998)
91. M. Bockrath, D.H. Gobden, J. Lu, A.G. Rinzler, R.E. Smalley, L. Balents, and P.L. McEuen, Nature (London) **397**, 598 (1999)
92. H. Ishii et al., Nature (London) **426**, 540 (2003)
93. A. Kasumov, M. Kociak, M. Ferrier, R. Deblock, S. Guéron, B. Reulet, I. Khodos, O. Stéphan, and H. Bouchiat, Phys. Rev. B **68**, 214521 (2003)
94. R. Tarkiainen, M. Ahlskog, J. Penttitä, L. Roschier, P. Hakonen, M. Paalanen, and E. Sonon, Phys. Rev. B **64**, 195412 (2001) and references therein
95. V. Krstić, S. Blumentritt, J. Muster, S. Roth, and A. Rubio, Phys. Rev. B **67**, 041401(R) (2003)
96. N. Kang, L. Lu, W.J. Kong, J.S. Hu, W. Yi, Y.P. Wang, D.L. Zhang, Z.W. Pan, and S.S. Xie, Phys. Rev. B **67**, 033404 (2003)
97. R. Egger, Phys. Rev. Lett. **83**, 5547 (1999)

98. T. Hunger, B. Lengeler, and J. Appenzeller, Phys. Rev. B **69**, 195406 (2004)
99. S. Ghost, A.K. Sood, and N. Kumar, Science **299**, 1042 (2003)
100. S. Ghost, A.K. Sood, S. Ramaswamy, and N. Kumar, Phys. Rev. B **70**, 205423 (2004)
101. B.N.J. Persson, U. Tartaglino, E. Tosatti, and H. Ueba, Phys. Rev. B **69**, 235410 (2004)
102. N.D. Mermin, Phys. Rev. B **1**, 2362 (1970)

20 Dissipation at large Separations

S. Rast[1], U. Gysin[1], E. Meyer[1], and D.W. Lee[2]

[1] Institute of Physics, University of Basel, Klingelbergstrasse 82, CH-4056 Basel, Switzerland
[2] Chonnam National University, South Korea

20.1 Introduction

In a closed system, friction forces transfer work to heat. Energy dissipation is a non-reversible process, which is well known from everyday life. The mechanisms, which lead to these irreversible processes, are complex and still poorly understood. The vibration amplitude of a damped oscillator decays in time that is equivalent to the fact that the kinetic energy is converted to heat. The energy transfer lasts until the cantilever system reaches its thermodynamic equilibrium. In this, state fluctuations from the mean value $\langle x \rangle$ are observed. Both decay time and equilibrium fluctuations $x(t)$ contain information about the dissipative process. The dynamic of a linear damped harmonic oscillator can be described by the knowledge of the spring constant k, the eigenfrequency ω_0 and the friction coefficient Γ, the effective mass m_{eff} and the external force f_{ext}:

$$m_{\text{eff}} \frac{\mathrm{d}^2 x}{\mathrm{d}t^2} + \Gamma \frac{\mathrm{d}x}{\mathrm{d}t} + m_{\text{eff}} \omega_0^2 x = f_{\text{ext}}(t) \tag{20.1}$$

The external force f_{ext} can be regarded as sum of a non-stochastic force and a stochastic force. Since Eq. (20.1) is a linear differential equation, both contributions can be treated separately. In both cases the same parameter set (ω_0, Γ, m_{eff}) is required to describe the system. All quantities (fluctuating and non-fluctuating) derived from the cantilever movement depend on the same parameter set. An experimentally accessible quantity to describe the dissipation process of a vibrating cantilever at its resonance frequency is its decay time τ. The quality factor Q is given by:

$$Q = \frac{\tau \omega_0}{2} \tag{20.2}$$

The knowledge of the spring constant k and decay time τ allow us to calculate the friction coefficient Γ which is the constant of proportionality between the non-conservative friction force F and the velocity v of the oscillator.

$$F = -\Gamma \cdot v \tag{20.3}$$

where the friction coefficient is given by

$$\Gamma = \frac{k}{\omega_0 Q} \tag{20.4}$$

For a soft free cantilever with a resonance frequency of 2.7 kHz and a vibration amplitude of $x0 = 20\,\text{nm}$ a Q of 500,000 and a $\Gamma = 10^{-14}\,\text{kg/s}$ we obtain friction force $F = \Gamma\omega_0 x_0 = 3.4 \cdot 10^{-18}\,\text{aN}$. The dissipated power $P = \Gamma \cdot v^2 = 1.15 \cdot 10^{-21}$ Watt ($7.17 \cdot 10^{-3}\,\text{eV/s}$). For non-contact force sensors with a resonance frequency of 300 kHz, a $\Gamma = 10^{-11}\,\text{kg/s}$ a typical amplitude of 1 nm the friction force is $F = 1.8 \cdot 10^{-14}\,\text{N}$ which corresponds to a dissipated power of $3.5 \cdot 10^{-17}$ Watt (221 eV/s). The experimentally determined friction coefficient Γ can be a superposition of different friction coefficients Γ_i

$$\Gamma = \Gamma_0 + \Gamma_1 + \Gamma_2 + \cdots + \Gamma_N \tag{20.5}$$

$\Gamma(T, p, x,)$ represents a quantity which depends on the temperature T, the material properties, the pressure p and external fields (e. g. magnetic or electrical fields). By approaching the cantilever close to a surface the electromagnetic fields between tip and surface have an impact on the friction coefficient. This leads to a distance dependent friction coefficient at small tip sample separations.

This contribution is structured as follows: First we are investigating the internal friction of the cantilever Γ_0. This is the friction coefficient which describes the friction losses which occur by simply bending the cantilever. Internal friction has to be measured under vacuum conditions to minimize the viscous damping due to inelastic scattering between the vibrating cantilever and gas molecules. At pressures below $10^{-6}\,\text{mbar}$ viscous damping can be neglected [2]. The internal friction depends on the material properties and the sensor geometry. There are two possibilities to decrease the kinetic energy of a vibrating cantilever dissipation and sound wave scattering. The cantilever can be regarded as a one side clamped bar. In reality the cantilever is a micromechanically etched silicon bar which is connected to a support.

If the junction between bar and support is not ideally constructed, there exists the possibility that sound waves are scattered from the cantilever to the support, which is reducing the kinetic energy of the cantilever. This is not a dissipative process! On the other hand kinetic energy is transferred into heat. It is experimentally a proven fact that rectangular bar cantilever scatters a minimum of stored energy into the support. The dissipative energy losses can be induced due to stress and strain acting on the cantilever. Experimentally it is hard to distinguish between the two mechanisms. The internal friction Γ_0 is omnipresent and determines if an additional friction coefficient $\Gamma_\text{R} = \Gamma_1 + \Gamma_2 + \Gamma_3 + \cdots + \Gamma_N$ is detectable at all. Assuming the errors $\delta\Gamma$ and $\delta\Gamma_0$ are Gaussian distributed we find for Γ_R:

$$\Gamma_\text{R} = \Gamma - \Gamma_0 \pm (\delta\Gamma + \delta\Gamma_0) \tag{20.6}$$

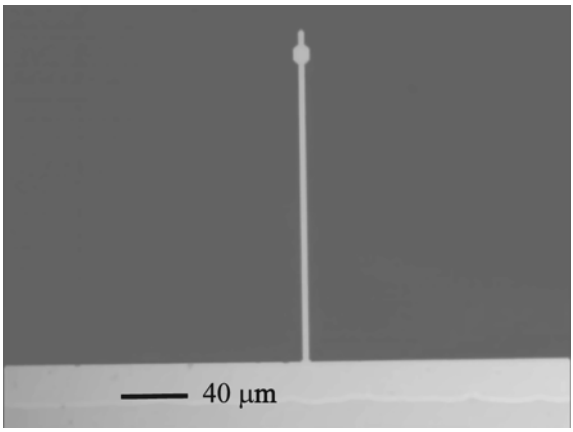


Fig. 20.1. Ultrasensitive single crystalline silicon force sensor with a length of 454 μm, a width of 4 μm and a thickness of 400 nm. The spring constant is $k = 0.00015$ N/m and the eigenfrequency $f_0 = 2.7$ kHz. A quality factors up to 500,000 can be obtained after annealing under UHV conditions for several hours. An internal friction coefficient Γ_0 in the order of $1.7684 \cdot 10^{-14}$ kg/s is achieved. With a vibration amplitude of 20nm the dissipated power is: $2 \cdot 10^{-21}$ Watt ($1.27 \cdot 10^{-2}$ eV/s)

Γ_{R} is limited by the sum of the statistical errors $\delta\Gamma + \delta\Gamma_0$. For an ultrasensitve cantilever (cf. Fig. 20.1) an internal friction coefficient of $\Gamma_0 = 1.7684 \cdot 10^{-14}$ kg/s can be obtained under ambient conditions.

Approaching the free cantilever close to a flat surface without bringing it into contact which opens new dissipative channels. An important limitation is that soft cantilevers jump into contact, when the attractive force gradient is larger than the spring constant, which happens on almost all surfaces because of van der Waals surfaces. The only exception is the operation under liquid environment, where attractive forces are greatly reduced [7]. An alternative way has been recently introduced by the use of the so-called pendulum geometry [4]. Spring constants of the order of mN/m are used, which greatly improves the force sensitivity to the level of atto-Newtons close to the surface. The cantilever is oriented perpendicular to the surface, which avoids a jump-into contact (Fig. 20.2).

To measure the friction force between tip and sample, ultrasensitive force detection is required (Fig. 20.1). There are several methods to determine the friction coefficient Γ: Ring-down measurements or the measurement of the power spectral density $\mathrm{S}(\omega)$ of the cantilever fluctuations [11]. The dimensions of the sensor and the temperature dependent Young's modulus [15] determine the spring constant k. Smallest possible friction coefficients Γ (Eq. (20.4)) can be sensed by using soft cantilever with a smallest possible constant k, a highest possible eigenfrequency and quality factor Q. Cantilevers with these features are very force sensitive. The minimal detectable force F_{min} is given

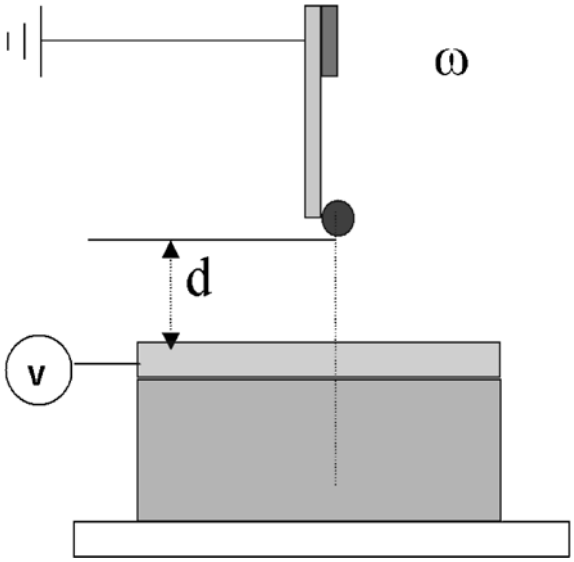


Fig. 20.2. Experimental setup: The cantilever is mounted perpendicular to the surface to avoid a jump into contact. The distance d between force sensor and surface can be varied. A bias voltage V between cantilever and surface can be applied

by :

$$F_{\min} = \sqrt{\frac{2k_{\mathrm{B}}T\Gamma\Delta\omega}{\pi}} \tag{20.7}$$

where k_{B} is the Boltzmann constant, T the temperature and the bandwidth of the measurement. A change of the friction coefficient is always accompanied by a change of the minimal detectable force (Eq. (20.7)).

Quantities like the frequency noise $\delta\omega$ of a cantilever which oscillates with an amplitude x_0 at its resonance frequency ω_0 depend on the friction coefficient Γ.

$$\delta\Gamma = \frac{\omega}{x_0 k}\sqrt{\frac{2k_{\mathrm{B}}T\Gamma\Delta\omega}{\pi^5}} \tag{20.8}$$

To obtain the best possible frequency resolution of a self-driven cantilever the frequency noise $\delta\omega$ has to be minimized. Therefore, the measurement has to be performed at low temperatures, where small thermal fluctuations and low Γ are found. The oscillator frequency noise is influenced by Γ and the temperature T. With Eq. (20.7) the power spectral noise density $S(\omega)^2)\Delta\omega = \omega^2$ can be estimated in a narrow bandwidth

$$S(\omega)^2) = \left(\frac{\omega}{x_0 k}\right)^2 \left(\frac{2k_{\mathrm{B}}T\Gamma\Delta\omega}{\pi^5}\right) \tag{20.9}$$

20.2 Internal Friction of the Cantilever

The temperature dependence of resonance frequency of cantilevers is rather well understood [15]. Geometry changes due to thermal expansion can be

neglected. However, the temperature dependence of the Young's modulus, $E(T)$, is given by the Wachter-formula:

$$E(T) = E_0 - B \cdot T \exp -\frac{T_0}{T} \tag{20.10}$$

where T_0 is related to the Debye-temperature of the sensor material. With Eq. (20.10) the temperature dependent resonance frequency can be calculated:

$$\omega_n = \alpha_n^2 \frac{t}{L^2} \sqrt{\frac{E}{12\rho}} \tag{20.11}$$

where $\alpha_1 = 1.875$ for the first resonance mode. t is the thickness, L the length and ρ the mass density. The experimental frequency vs. temperature data are well fitted with $T_0 = 317\,\mathrm{K}$ [15]. According to $D = T_0/2$ a Debye temperature of $D = 634\,\mathrm{K}$ is determined, which is in good agreement with literature values of $D = 645\,\mathrm{K}$ for silicon.

In contrast, the damping of cantilevers is still rather poorly understood. Several contributions have to be distinguished:

1. Damping due to thermoelastic damping
2. Damping due bulk losses
3. Damping due to surface losses
4. Damping due to acoustic emission into the bulk
5. Losses due to the clamping
6. Viscous damping due to the presence of gases or liquids

As far as ultra-sensitive measurements under ultrahigh vacuum conditions are concerned, the influence of viscous damping at pressures below $10^{-6}\,\mathrm{mbar}$ can be neglected. The influence of clamping can be optimized by rigid holders and the exclusion of glues with high damping rates. Damping due to acoustic emission is also found to be negligible in most practical cases. Therefore, the first three mechanisms are the most important ones.

20.2.1 Thermo-Elastic Damping

The conduction of heat is an important energy loss mechanism. Periodical compression and expansion of oscillating micromechanical elements is associated with heat flow between compressed and expanded areas. The Zener-model is a continuum model of this thermo-elastic damping mechanism [13]. The internal friction is given by

$$Q^{-1} = \frac{\alpha^2 T E}{\rho c_P} \frac{\omega\tau}{1 + (\omega\tau)^2} \tag{20.12}$$

where α is the thermal expansion coefficient, c_p the specific heat capacity and ρ is the mass density. The relaxation time τ is given by

$$\tau = \frac{t^2}{\pi^2} \frac{\rho c_p}{\kappa} \tag{20.13}$$

where κ is the thermal conductivity. Typical parameters for silicon at room temperature are $E = 1.68\,\mathrm{GPa}, \alpha = 2.54 \cdot 10^{-6}\,\mathrm{K}^{-1}, \rho = 2.33 \cdot 10^3\,\mathrm{kgm}^{-3}$, $c_p = 711\,\mathrm{J\,kg}^{-1}\mathrm{K}^{-1}$ and $\kappa = 150\,\mathrm{Wm}^{-1}\mathrm{K}^{-1}$. The temperature dependence of the Young's-modulus is small compared to the strong variations of thermal expansion (zero crossings at 20 K and 125 K). The thermal conductivity in the bulk varies between $100 - 5000\,\mathrm{Wm}^{-1}\mathrm{K}^{-1}$. One should also take into account that the thermal conductivity is reduced due to phonon-boundary scattering for thickness of the order of microns below $100\,\mathrm{Wm}^{-1}\,\mathrm{K}^{-1}$ at temperatures below 30 K [21].

At present, many experimental data indicate that thermo-elastic damping is the dominant loss mechanism at room temperature. At temperature below 200K other channels start to dominate, which may be related to bulk or surface losses.

20.2.2 Bulk and Surface Losses

The scattering of elastic waves with defects on the surface or in the bulk is an important loss mechanism. The oscillation of the cantilever leads to a time dependent local stress field. The energy landscape of the defects is changed by this stress field. Instabilities of these defects may occur, where atoms jump from one equilibrium position to another position. The energy difference between equilibrium positions is the activation energy. Therefore, damping vs. temperature curves show activation peaks, also called Debye peaks. So far, most of the experimental work is limited to silicon cantilevers which exhibit the highest Q-factor of available cantilevers. Typical Q-factors are between 10,000 up to 500,000. Comparable cantilevers made of Si_3N_4 or SiO_2 show much smaller Q-factors of 100 to 1000. Therefore, we conclude that bulk losses are dominant for these amorphous structures. In the case of silicon, bulk or surface losses may become dominant at temperatures below 200 K. At 160 K a peak is observed, which may be related to such an activation peak with an activation energy of 0.25 eV. Unfortunately, the nature of these defects in silicon is still poorly understood. Simple defects, such as vacancies or interstitials are ruled out because of their high activation barriers [15]. Recently, it has been observed that the 160 K peak can be reduced strongly by annealing under vacuum conditions [22]. It is also observed that the peak does not shift with the resonance frequency, which is not in agreement with the simple activation energy model. The authors suggest that the 160 K peak is related to an adsorbate layer. Another peak at 30 K shifts with the resonance frequency and seems to be in better agreement with a Debye peak [22].

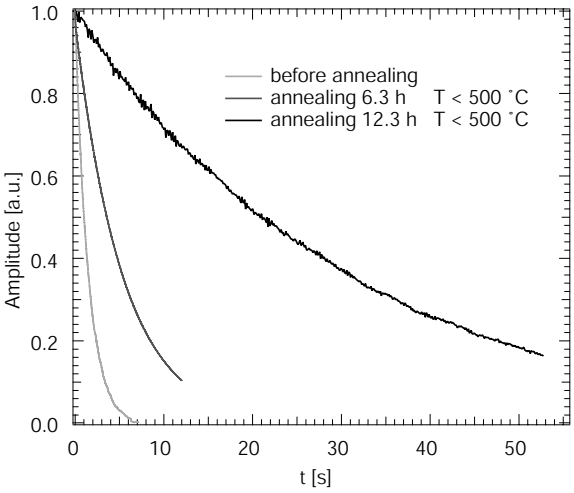


Fig. 20.3. Ring down measurement of a cantilever before and after annealing under UHV-conditions. The initial quality factor of 62,000 could be improved by an order of magnitude after 6 h annealing. Further annealing improved the quality factor to 1,239,500. The annealing temperature was too low to remove the oxide layer. Thus, the removal of weakly bound molecules, such as H_2O, OH or hydrocarbons, improves the quality factor

Coating of cantilevers leads to a strong increase of dissipation. The polycrystalline nature of these metallic films implies grain boundaries, where increased phonon scattering leads to an increase of damping losses. Other surface coatings, such as silicon oxide, or adsorbates, such as H_2O or hydrocarbons, lead to rather large damping losses. Yang et al. annealed extremely small cantilevers (length $< 80\,\mu m$) [12]. They used rather high annealing temperatures (1000 °C), which was sufficient to remove the oxide layers. Recently, it has been shown that annealing at temperatures below 600 °C of rather large silicon cantilevers (length of 400 or 500 µm and thickness of 0.5 to 1.5 µm) under ultrahigh vacuum (UHV) conditions can also lead to a reduction of dissipation [26]. In the case of Fig. 20.3 the quality factor of 62,000 could be improved by an order of magnitude after 6h annealing at temperatures below 600 °C. Further annealing improved the quality factor to 1,239,500. The annealing temperature was too low to remove the oxide layer. Thus, the removal of weakly bound molecules, such as H_2O, OH or hydrocarbons, improves the quality factor. Alternatively, defects on the surface or in the bulk of the cantilever may be reduced by the annealing procedure. Figure 20.4 shows the temperature dependence of the minimal detectable force.

It is evident that the internal friction coefficient reduces in a rather continuous way down to 9 K, which is in qualitative agreement to previous studies of internal friction of silicon [13–15]. Below 9 K a small plateau of constant

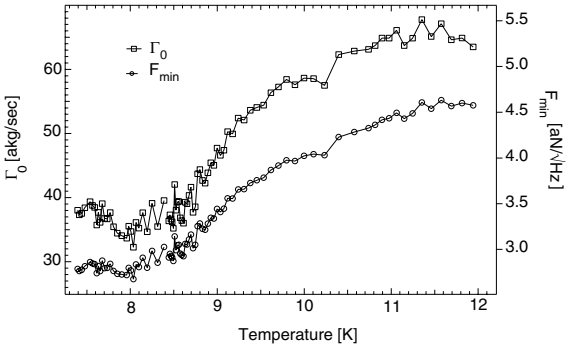


Fig. 20.4. Minimum detectable force of a cantilever. (*red datapoints*) For a rectangular, ultrahigh-sensitive cantilever ($L = 454\,\mu\text{m}, w = 4\,\mu\text{m}, t = 0.44\,\mu\text{m}$, and spring constant $0.1310\,\text{mN/m}$) (*black datapoints*) ($L = 450\,\mu\text{m}, w = 45\,\mu\text{m}, t =$ µm, and spring constant $D = 0.176\,\text{N/m}$)

dissipation is observed. In order to investigate the influence of geometries and unknown micro fabrication processes a variety of silicon cantilevers from different manufacturers were investigated. As shown in Fig. 20.4 the increase of the quality factor and corresponding force sensitivity is observed in a systematic way. One to two orders of magnitude of improvement are found. Some of the annealed cantilevers are found to have aN-force sensitivity even at room temperature due to Q-value enhancement, which opens new possibilities for experiments, such as magnetic resonance force microscopy at room temperature or cantilever mass spectroscopy. The force sensitivity of commercial silicon sensors with a spring constant of $0.176\,\text{N/m}$ changes an order of magnitude after annealing. Quality factors higher than 6,000,000 can be reached after heat treatment. Cantilevers with mN/m spring constant were fabricated by the use of a dry etching method [23]. For these cantilevers with a spring constant of $1\,\text{mN/m}$, which have a thickness of $200\,\text{nm}$, the Q factor can be improved by a factor of 100.

20.3 Dissipation at Large Separations

The dissipation between two moving bodies separated by a distance d is due to an electromagnetic interaction. The range of the interacting force determines the distance dependency of the friction coefficient Γ. Therefore, it is possible to distinguish between long-range and short-range friction forces. The mechanism of the long-range, electrostatic dissipation known also as Joule Dissipation is well understood. Static electric fields between two different surfaces can exist without any externally applied voltage due to different work functions of different orientations of the crystallites of a polycrystalline surface. The work function measured by applying a bias voltage between tip and sample is distance dependent [26]. This is due to the fact that the cantilever senses at large separations an average patch force resulting of several different oriented crystallites.

This dissipative force can be minimized compensating the local electric field. Electrostatic dissipation was previously observed by Denk et al. [24]. Later Stipe et al. [16] observed electrostatic dissipation at separations of 1 – 200 nm by using ultrasensitve force sensors. A gold tip was attached to an ultra-sensitive cantilever in the pendulum geometry. Friction coefficients of the order of 10^{-13} kg/s between tip and metal substrate were observed. An increase of dissipation with increasing temperature was observed. The distance dependence of the friction coefficient was fitted by a power law $\Gamma \propto d^{-n}$ with an exponent $n = 1.3 \pm 0.2$. The measurements of Stipe et al. were performed under high vacuum conditions. Volokitin et al. [5] have calculated theoretically that dissipation at these large separations is mainly due to charge fluctuations. Rast et al performed dissipation measurements under ultra high vacuum conditions at 7 K of an ultrasensitive cantilever with a Co-Sm magnetic tip ($1.3\,\mu\text{m} \times 1.2\,\mu\text{m} \times 2.8\,\mu\text{m}$) . The front edge of the tip is flat and has an area of $1.56\,\mu\text{m}^2$. The tip was glued with an non-conductive glue to the cantilever. As a substrate irradiated quartz was used (Suprasil 300). 20nm gold film (deposited at room temperature and transferred through air and stored for about 3 days under UHV-conditions). As shown in Fig. 20.5 both conservative forces and dissipative forces increase with applied bias voltage and are is observed at separations larger than 200 nm. The conservative force obeys to a quadratic power law. The friction coefficient has its minimum, where the conservative force is minimal.

The friction coefficients fits to a quadratic power law $\Gamma \propto (V_{\text{bias}}^2)$. The friction coefficient of the force sensor changes approximately $2 \cdot 10^{-13}$ kg/s by applying 1 V. Chumak et al. [25] calculated the friction coefficient between a metallic tip and a flat metallic surface. Due to the cantilever oscillations the

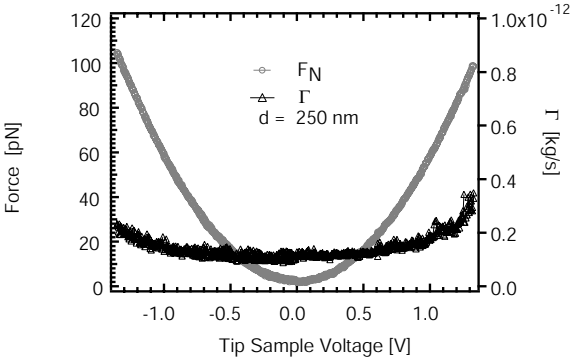


Fig. 20.5. Voltage-dependence of the normal force and friction coefficient at a separation of $d = 250$ nm of a 20 nm thick gold film. The parabolic force dependence is related to capacitive forces. The friction coefficient increases with applied voltage. The quadratic behavior of the normal force and the quadratic behavior of friction coefficient in respect to bias voltage indicate that an electrical field is involved in the process

electrical field which is perpendicular to the cantilever surface varies in time and induces a current in the cantilever, which is responsible for dissipation.

Compensating the contact potential allows us to investigate the friction force which is not dominated by Joule losses. Figure 20.6 shows the distance dependence of the friction coefficient as a function of distance with compensated contact potential. Absolute values are between $1 \cdot 10^{-14}$ kg/s to $5 \cdot 10^{-14}$ kg/s (The friction coefficient of the free cantilever was $5 \cdot 10^{-14}$). At tip sample separations which are larger than 30 nm the friction force is dominated by the internal friction of the force sensor. At this separation the friction coefficient is smaller than $\ll 10^{-14}$ kg/s, the internal friction of the force sensor. The strength of the electromagnetic field between tip and sample is too small to create a measurable friction force. separations below 30 nm the friction coefficient is larger than the internal friction of the cantilever and an non-reversible energy flow between tip and sample is observed. Several contributions can enhance the friction between tip and sample. Stipe [16] observed that artificial electric fields generated by defects in quartz have an influence on the friction coefficient. The higher the defect concentration the higher the value of distance dependent friction coefficient. Volokitin (see chapter in this book) calculated the friction coefficient of a tip separated by a distance d generated by a van der Waals friction force. For a cylindrical tip with radius R and width w the calculated friction coefficient which scales with $d^{-1/2}$ which fits to our data.

Both Joule and non-Joule dissipation have an influence on the minimum detectable force of a cantilever close to the surface. The electrostatic friction between tip and sample can be minimized by applying a bias voltage between tip and surface. The minimal detectable force of the cantilever is given by (Eq. (20.7)). Force sensitivities in the order of $10^{-18}\,\mathrm{N}/\sqrt{\mathrm{Hz}}$ which correspond to the force sensitivity of the free cantilever can be obtained at separations

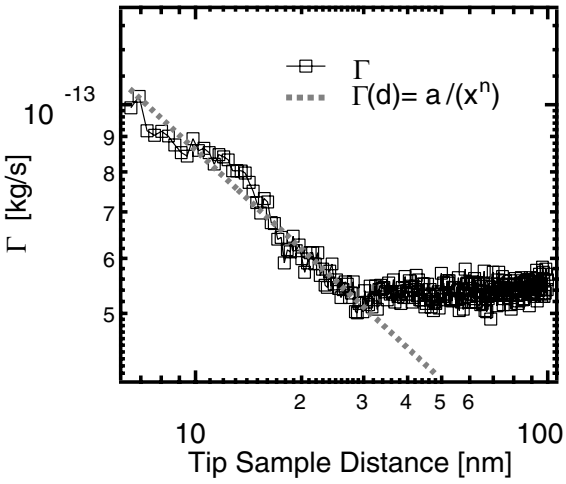


Fig. 20.6. Illustration of the friction coefficient with compensated contact potential at separations between $1-100$ nm. At a separation larger than 30 nm the friction is governed by the internal friction 0 of the cantilever. Therefore, the friction coefficient of the long-range interaction has to be smaller then $\Gamma \ll 10^{-14}$ kg/s

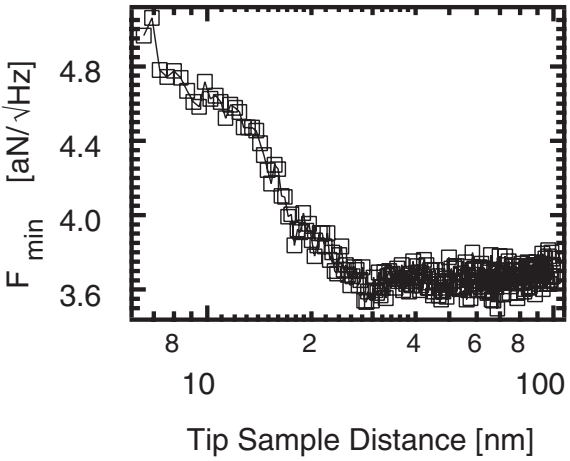


Fig. 20.7. The corresponding force sensitivity is calculated by Eq. (20.7) between 2 to 4 aN at separations from 6 to 100 nm. Above 30 nm the force sensitivity of the cantilever is limited by the force sensitivity of the free cantilever

$>$ 30 nm at 7 K. Below 30 nm the minimum detectable force increases but remains still in the attonewton regime.

Most AFM- measurements close to the surface are not made in the dynamic mode where the frequency ω of a self oscillating is measured. Phase Locked Loops are used to measure frequency change $\delta\omega$ caused e. g. by the tip sample interaction. The frequency noise determines frequency resolution of the measurement (Eq. (20.8)). Figure 20.8 illustrates the power spectral density $S(\omega)^2$ of the frequency noise of a self-oscillating cantilever with an amplitude of 10 nm close to a CaF_2 substrate at room temperature under high vacuum conditions. The power spectral density is measured by sampling the frequency noise of the FM-demodulated signal of the cantilever vibration amplitude. At smaller tip sample separations the noise floor is increasing several orders of magnitude due to the distance dependent change

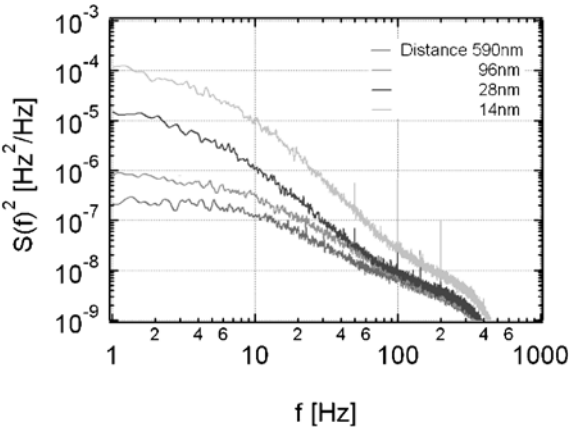


Fig. 20.8. Illustration of the power spectral density $S(\omega)^2$ of the frequency noise of a self-oscillating cantilever with an amplitude of 10 nm close to a CaF_2 substrate under ambient conditions. At large tip sample separations the frequency noise corresponds to the frequency noise of the free cantilever

of the friction coefficient Γ. The noise floor $S(\omega)^2$ is changing more at low frequencies (1 – 10 Hz) than at frequencies above 10 Hz.

20.4 Summary and Conclusions

Thermal fluctuations and fluctuating electromagnetic fields are the limiting factors for ultrasensitive measurements close to the surface. It could be demonstrated that the force sensitivity of the free cantilever is limited by thermal fluctuations material properties. Annealing the cantilever under UHV-conditions removes the adsorbates on the cantilever and increases the force sensitivity about an order of magnitude. Cooling to cryogenic temperatures reduces the thermal fluctuations and is lowering the internal friction coefficient due to the change of thermodynamic properties. The temperature dependent quality factor can be described by Zener's theory of thermo-elastic damping. Varying the distance between the tip of the cantilever and substrate opens new dissipative channels. Up to tip sample separations of 250 nm the long-range dissipation can be observed. Joule dissipation is relatively well understood, but the quantitative analysis indicates that both the probing tip and the cantilever contribute to dissipation. By compensating the local electric field Joule dissipation can be minimized and attonewton sensitivity can be achieved at separations down to 30 nm. Below 30 nm fluctuations are observed, which are not related to Joule dissipation. These fluctuations are lowering the force sensitivity close to the surface. The losses may be caused by short-range forces, such as van der Waals forces. Since the experimental access to this ultra-low dissipation and force regime is relatively new, the exact mechanisms are still to be explored.

Acknowledgement. This work was supported by the Swiss National Science Foundation, the "Kommission zur Förderung von Technologie und Innovation", and the National Center of Competence in Research on Nanoscale Science.

References

1. B. Gottsmann and H. Fuchs, Phys. Rev. Lett. 86, 2597 (2001).
2. Sensoren mit geringer Dissipation zur Messung kleiner Kräfte, Diss. Phil.-naturwiss. Fak. Basel, 1999.
3. I. Dorofeyef, H. Fuchs, G. Wenning and B. Gotsmann, Phys. Rev. Lett 83, 2402 (1999)
4. D. Rugar, R. Budakian, H.J. Mamin and B.W. Chui, Nature 43, 329 (2004).
5. A. Volokitin and B. Persson, Phys. Rev. Lett 94, 086104. (2005).
6. G. Binnig, C. Quate and Ch. Gerber, Phys. Rev. Lett. 56, 930 (1986).
7. F. Ohnesorge and G. Binnig, Science 260, 1451 (1993).
8. S. Morita, R. Wiesendanger, and E. Meyer, Noncontact Atomic Force Microscopy (Springer, Berlin, Germany, 2002).

9. T. Albrecht et al., Albrecht, T. R., P. Grutter, H. K. Horne, and D. Rugar, J. Appl. Phys. 69, (1991).
10. F.J. Giessibl, Rev. Mod. Phys. 75, 949 (2003).
11. S. Rast, C. Wattinger, U. Gysin and E. Meyer, Nanotechnology 11, 169 (2000).
12. J. Yang et. al, Appl. Phys. Lett. 77, 3860 (2000).
13. C. Zener, Phys. Rev. 53, 100 (1938).
14. R. Lifshitz and M. Roukes, Phys. Rev. B 61, 5600 (2000).
15. U. Gysin, S. Rast, P. Ruff, D.W. Lee, P. Vettiger and C. Gerber, Phys. Rev. B 69, 045403 (2004).
16. B.C. Stipe, H.J. Mamin, T.D. Stowe, T.W. Kenny and D. Rugar, Phys. Rev. Lett. 86, 2874 (2001)
17. M. Roukes et al., J. Appl. Phys. 95 2682 (2004),
18. A.H. Morrish and S.P. Yu, Phys. Rev. 102, 670 (1956).
19. M.J. Naughton et al., Rev. Sci. Instr. 68, 4062 (1997).
20. B.C. Stipe, H.J. Mamin, T.D. Stowe, T.W. Kenny and D. Rugar, Phys. Rev. Lett. 87, 096801-1 (2001).
21. M. Ashegi, Y.K. Leung, S.S.Wong and K.E. Goodson, Appl. Phys. Lett. 71, 1798 (1997).
22. H. Haucke, Xiao Liu, J.F. Vignola, b.H. Houston, M.H. Marcus, and J.W. Baldwin, Appl. Phys. Lett. 86, 191903 (2005).
23. D.W. Lee et al., JMMM (2006).
24. W. Denk and D. W. Pohl, Appl Phys. Lett., 59, 2173 (1991)
25. A. A. Chumak, P. W. Milonni, and G. P. Berman Phys Rev. B 70, 085407 (2004).
26. S. Rast, U. Gysin, P. Ruff, C. Gerber, E. Meyer, and D.W. Lee, Nanotechnology 17, 189 (2006).

21 Surface-Damage Mechanisms: from Nano- and Microcontacts to Wear of Materials

R. Colaço

Mat. Eng. Department, Instituto Superior Técnico Av. Rovisco Pais 1, 1049-001, Lisboa, Portugal rogerio.colaco@ist.utl.pt

21.1 Introduction

Most certainly life on Earth would be quite different from what we know without (or with different) tribologically related phenomena. For a few examples, one can think of the Earth's landscape, formed by a billion years of water and air erosion, the friction-controlled metastable equilibrium of geological faults and cracks, or simply by the way we (and all the things) move on the Earth's surface. Nevertheless, until the mid-1960s, not much attention had been paid to this science. Of course, some important exceptions are worth mentioning, but even the word "tribology" did not exist forty years ago. However, in 1966 a report made by a UK government committee presided by H.P. Jost, known as the *Jost Report*, found that a large waste of resources – estimated in 515 million pounds sterling per year (approximately 4% of the 1965 UK GNP) – occurred because of ignorance of surface-interaction phenomena [1]. Although friction, lubrication and wear had been studied for many years before (Leonardo da Vinci was among those who paid attention to tribological phenomena), the fact is that the Jost Report launched several R&D and education programs in tribology. The introduction in the lexicon of the word "tribology" – the science of rubbing – was one of the first consequences of this report.

Initially it was widely felt that the Jost Report greatly exaggerated the savings that might result from improved tribological expertise. Nevertheless, later it became clear that, on the contrary, the report underestimated the economical importance of tribology, since it paid small attention to wear, which happens to be, from an economical point of view, the most significant tribological phenomenon [2, 3].

In the last decades the experimental and theoretical research in tribology and, particularly, in the wear mechanisms of materials, have led to considerable advances. The generic "wear of materials" has been classified in different modes, in accordance with its type of occurrence and with its phenomenology (abrasion, adhesion, erosion, fretting, etc.); the damage mechanisms of rubbing surfaces have been identified (cutting, plowing, cracking, delamination, fatigue, oxidation, etc.) and mathematical models for wear, with more or less wider application ranges, have been developed.

These advances have certainly brought wear studies closer to a more predictive science, and important technological benefits have resulted from it. Some examples of technological advances closely related with the advances in tribological knowledge, and particularly in wear knowledge, are the hard disk drive technology and the development of digital light processing micromirrors (DLPMs) [4], the development of ceramic bearings [5] and the development of composite hardfacing materials for protective coatings [6]. Nevertheless, recent studies [7–9] still point to a non-negligible economical impact of tribologically related phenomena. In the US alone, recent estimations lead to values around US$ 8 billion per year spent in wear-related problems [10] and it is widely accpeted that at least 1% of the GNP of an industrialized country might be saved with minimal further investment in research [11, 12].

How to understand the maintenance of such high economical wastes in wear- and friction-related problems, in spite of the unquestionable increase in tribological knowledge in the last decades?

Most probably, many causes are the origin of this persistency over the years of the economical impact of tribological phenomena, particularly of wear. Perhaps the first factor to be considered is the fact that tribologically related problems are still not yet a main topic of concern for the common industrial engineer, who faces wear as an inevitability. This surely can be overcome with a stronger emphasis on tribology in undergraduate engineering programs. But, certainly, it also contributes to this persistency technologically related causes. On the one hand, in parallel with the growing development of tribological and surface engineering knowledge, there is also a growing demand to increase the production rates and efficiency of machines, tools and engineering components in general, which leads to higher working speeds and loads that result in more aggressive wear conditions. On the other hand, new systems and applications, in which tribological performance is a critical issue, have arisen in recent years. Some examples of this are the increasingly growing demands for higher storage densities and speeds in the magnetic disk drive technology, the growing interest in micro-/nano-electromechanical systems (MEMS and NEMS) and the demands for more wear-resistant biomedical implant materials, with increased *in-vivo* lifetimes and satisfying biocompatibility and bioactivity requirements. In these examples, new challenges for the tribologists arise, either because the variability of the contact conditions is rather complex (biomedical implant materials), or because of the reduction of the contact scales to the nanometric or even atomic ranges (electromagnetic microdevices).

In what concerns wear, a large number of recent studies have pointed to the need to understand deeper the atomic origins of wear [12] as well as to establish relations between macro-, micro- and nanowear phenomena [5], i. e. bridging the gap between observation scales.

From the application point of view, nanowear studies are not only important because of the growing importance of disk storage and of micro- and

nanotechnologies, but also for the optimization of more common and conventional tribological systems in engineering components. In fact, since the pioneering work of Greenwood and Williamson [13], it is well known that the contact between nominally flat surfaces (the most frequent in engineering components) occurs between small surface asperities, whose contact areas can be only of a few tens of square nanometers or even smaller. The point being made is that, even in conventional engineering systems, the interactions at nanometric scales cannot be neglected.

With the advent of the atomic force microscope (AFM) [14], it became possible to study wear phenomena at very small scales and loads. The AFM tips can be used to simulate a sharp single asperity travelling over a surface [15, 16]. By using a stiff steel cantilever with a sharp diamond tip (or diamond-coated tip) mounted on its end, nanowear experiments at high loads in hard materials can also be made [17–21]. AFMs can be used to measure or monitor, at very fine scales, the damage and/or topographical changes in surfaces after local rubbing (e. g. [22–24]) or to measure the work of adhesion between the surface and the single asperity simulated by the tip (e. g. [25,26]). As a consequence, the appearance of this instrument has opened a wide new field of research for tribologists, enabling entrance to the study of wear and damage mechanisms at the scales that were not acessible before.

This possibility of going to smaller and smaller scales in wear studies has raised a number of new questions, such as:

- what are the (or, are there) phenomenological thresholds between nanowear, microwear and macrowear?
- are the classical wear equations valid for rubbing at nanoscale, or are new ones needed for nanowear?

Generally speaking, these questions can be summarized in the general form "is the *scale* a characteristic of the tribological system that influences its response (such as speed, load, atmosphere, etc.) or, on the contrary, is the phenomenology kept if the tribological system is (homothetically) scaled down?"

Results obtained up to now are still scarce and progress still needs to be made, both from the experimental and theoretical points of view. There are not only scale gaps between "conventional" wear and nanowear, but also gaps between theory and experimental results on going to micro- and nanomechanical phenomena in general. As an example, there is clearly a need for the development of a micrometer level continuum theory, enabling a link from the macroscopic fracture process to the atomistic one to be made [27]. In fact, according to conventional plasticity theories, which are length-scale independent, the maximum stress level that can be achieved at a crack tip is not larger than 4–5 times the tensile yield stress of the material [28]. However, experimental studies show that the stress level needed to produce atomic decohesion in Nb-sapphire crystals can be up to 10 times its tensile yield stress [29]. This incompatibility between macroscopic fracture theories and

atomistic fracture results is apparently due to factors that become relevant at atomistic scales but that are not included in conventional plasticity theories.

This chapter aims to give a brief overview of the state-of-the-art in atomic, nano- and microwear knowledge.

21.2 The Nature of Solid Surfaces

The term "surface" can be understood as the transition region of a material to its environment [30], and this term will be used from now on with the more strict meaning of "solid surface".

From the tribological point of view, surface properties are a key factor for the performance of any moving component since, in contacting rubbing pieces, the work is dissipated in the surface and subsurface region. An "engineering surface" can be quite different from an ideal atomically smooth surface in vacuum conditions, whose equilibrium structure can be described, for instance, by the TLK (terrace, ledge, kink) model [31]. Two important differences should be considered between an engineering surface and a TLK surface. The first is the fact that, in the former case, the transition between the bulk properties and the environment is not as sharp as in the latter. The second is the fact that engineering surfaces have topographical features that can be several orders of magnitude larger than the atomic terrace and ledge scales. The tribological performance of an engineering component is therefore strongly dependent on the physical, chemical and mechanical properties of the surface (and by the gradients of these properties) and on the surface topography. Some aspects of these two topics will be briefly discussed next.

21.2.1 Surface Constitution

Solid surfaces in general can present several layers, with properties that can be quite different from the bulk ones. As a consequence, a solid surface is a portion of material that can present a rather complex structure and properties, which, in part depend on the surface preparation method, the nature of the solid and the interactions between the surface and environment (e. g. [11]).

Usually, in mechanically machined, ground or polished surfaces, there is a layer of deformed material (formed during the surface-preparation process). In metals and metallic alloys, this deformed zone usually is work hardened and can present a hardness higher than that of the bulk (see point 4). Bowden, Moore and Tabor were the first to observe clearly this deformed zone, by using taper-section optical microscopy [32].

On the top of this deformed layer sometimes is referred the presence of a smeared layer with an amorphous structure, whose thickness is a few nanometers, the so-called Beilby layer [33]. The Beilby layer is supposed to be produced by localized melting and surface flow during the mechanical surface-preparation process. However, its existence is indeed quite controversial [34],

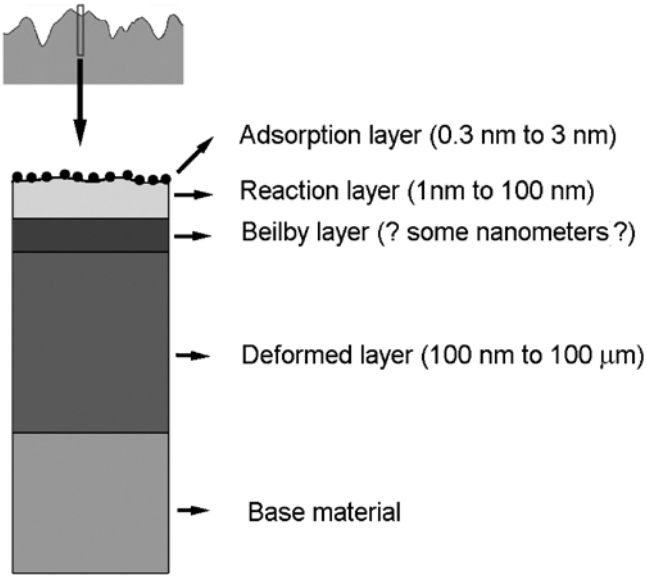


Fig. 21.1. Schematic view of a cross section in an engineering surface

and the attempts to characterize this layer by fast electron scattering [35] or by transmission electron microscopy [36] have never been conclusive.

Most surfaces are chemically reactive and form native surface oxide layers in air, or other reaction layers depending on the environment (nitrides, sulfides, chlorides, etc.). Besides these native films, adsorbed layers of molecules present in the environment (water, gaseous molecules, organic molecules, etc.) can also appear at the surface.

The presence of all these regions (schematically represented in Fig. 21.1), with their different properties, influences the tribological response of the material. Moreover, if the interaction scale of the damage during contact and rubbing changes between these regions, alterations in the wear response of the material can occur.

21.2.2 Surface Topography

Surfaces are rough. Even the most highly polished component has irregularities (asperities) significantly larger than the atomic scale. Figure 21.2 shows an AFM topographic profile (tapping mode) of a finely polished metallic surface (mirror-like finishing). The irregularities, or deviations from the nominal ideal flat surface, form the surface topography. The surface topography can show different features (such as flaws, pores, waviness, etc.). However the contact mechanics of solids depends essentially on the shorter-wavelength fluctuations [13, 37–39], the so-called roughness. Roughness also influences the dynamics (and vibrational modes) of the contact region of rubbing (or

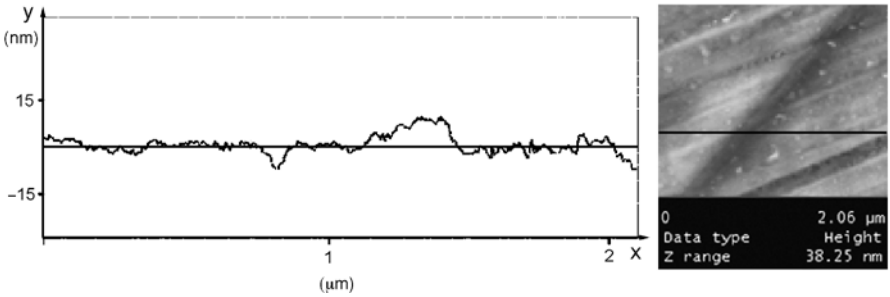


Fig. 21.2. AFM topographic profile of a mirror-like metallic surface, obtained by fine mechanical polishing with a suspension of 1 μm diamond particles

near-rubbing) sliding bodies (such as hard disk drive sliders) [40] and, of course, the wear response of the materials [41].

The surface roughness has a fractal nature, in the sense that it is formed of a large number of length scales superimposed on each other [42]. A wide number of roughness parameters and parametric functions can be used to describe and quantify surface roughness (a rather exhaustive presentation can be found in [39]). The most commonly used are roughness amplitude parameters (vertical descriptors), such as the average roughness, R_a, and the average maximum roughness, R_z, defined as:

$$R_\mathrm{a} = \frac{\int_0^{l_\mathrm{m}} |y| \, \mathrm{d}x}{l_\mathrm{m}} \tag{21.1}$$

in which l_m is the profile length, and $y(x)$ is the profile function in an axis whose origin is at the centerline of the profile (Fig. 21.2), and:

$$R_\mathrm{z} = \frac{1}{5} \sum_{i=1}^{5} y_i \tag{21.2}$$

with y_i being the maximum distance peak to valley in the i^{th} interval of the profile.

21.2.3 Topographic Mechanisms of Wear

Experience shows that, in general, smoother surfaces in rubbing contact exhibit lower wear than rougher ones, either in dry or lubricated contact conditions, and that the running-in period also decreases with increasing smoothness of surfaces [41, 43, 44].

An explanation for this behavior can be found in the early work of Greenwood and Williamson [13], on the contact mechanics of rough surfaces. The

basic idea of the GW model is that, since contact of rough surfaces occurs between the small asperities of opposing surfaces, the real area of contact, A_{R}, is different from the apparent area of contact, A_{N}. According to the GW model, A_{R} is independent of A_{N} and the ratio $A_{\mathrm{R}}/A_{\mathrm{N}} = f$ can be rather small (typically between 10^{-5} and 10^{-1}).

Since the height of the asperities is not uniform (see Fig. 21.2), when the surface is submitted to a normal load, F_{N}, the touching asperities are under different stresses, but it can be shown that the average stress in the touching asperities, $\langle\sigma\rangle$, is given by [45]:

$$\langle\sigma\rangle = \frac{\sigma_{\mathrm{N}}}{f} \tag{21.3}$$

in which $\sigma_{\mathrm{N}} = F_{\mathrm{N}}/A_{\mathrm{N}}$ is the normal (compressive) stress in the bulk material, in a plane parallel to the surface.

Equation (21.3) shows that even small elastic compressive stresses in the bulk can lead to high average stresses in the contact spots of the surfaces, since f can be rather small. Moreover, some of the asperities (those of higher amplitude) may deform plastically, while the others (the smaller amplitude ones) can deform in the elastic regime [45].

Although Greenwood and Williamson did not discuss the mechanism of creation of a wear particle, they pointed out that wear is much more probable in touching plastic asperities than in elastic ones. They introduced a plasticity index, Ψ, given by:

$$\Psi = \frac{E^*}{H}\sqrt{\frac{R_{\mathrm{p}}}{\rho}} \tag{21.4}$$

R_{p} and ρ are topographical parameters: the standard deviation of the asperities height distribution function and the curvature tip radius of the asperities, respectively ($R_{\mathrm{p}} \approx 1.25R_{\mathrm{a}}$, for a Gaussian height distribution [39]). E^* is the reduced Young modulus given by:

$$E^* = \left(\frac{1-\nu_1^2}{E_1} + \frac{1-\nu_2^2}{E_2}\right)^{-1} \tag{21.5}$$

with ν_i and E_i the Poisson's and Young's moduli, respectively, of each of the surfaces. H is the hardness of the softer surface (see Sects. 21.3 and 21.4 of this chapter for a more detailed discussion on the concept of hardness and hardness at different scales, respectively).

For $\Psi < 0.6$ the contacts are predominantly elastic and for $\Psi > 1$ the contacts are predominantly plastic. For intermediate values of the plasticity index, part of the asperities in contact are in the elastic regime and part in the plastic regime. The most interesting thing about Ψ is its independence on load: it only depends on mechanical properties of the surfaces (the ratio E^*/H is inversely proportional to a yield extension) and of their topographic characteristics.

Only very finely polished surfaces have plasticity indexes smaller than 0.6 [46]. As a consequence, most of the surfaces in contact rubbing engineering components have a "mirror-like" finishing, i. e. they are finely polished, in order to decrease Ψ, bringing the contacts as much as possible to the elastic regime. The idea is that this will result in a shorter running-in period and in a lower wear rate. It should be pointed out that, in a mirror-like surface, such as the one whose topographical profile is shown in Fig. 21.2, the maximum peak-to-valley amplitude is typically smaller than 50 nm and R_a is typically smaller than 10 nm. In such surfaces, very common in dry or lubricated moving engineering parts, the contacts will occur in nanometric scales, as schematically shown in Fig. 21.3. Even in 3-body abrasive wear situations, with free hard micrometric particles moving in-between well-polished surfaces, the wear damage and debris can be in the nanometric range. Figure 21.4 shows a scanning electron microscopy (SEM) image of the initial debris formed in a finely polished tempered tool steel surface, submitted to

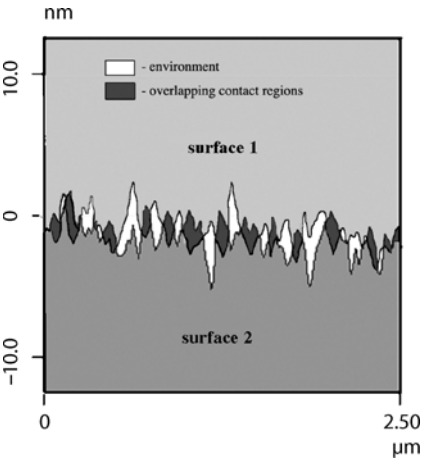


Fig. 21.3. Simulation of the contact regions of two surfaces with a roughness similar to that in Fig. 21.2

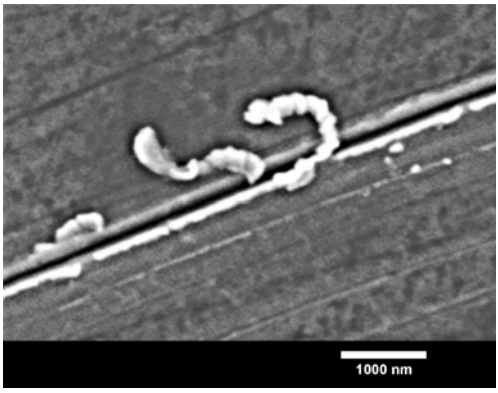


Fig. 21.4. Formation of the wear debris in a Cr tool steel when submitted to abrasion by hard 6 μm diamond particles. Typical wear coefficients obtained in such experiments are between 10^{-12} and 10^{-14} N/m^2 [47]

a common laboratorial ball-cratering abrasive wear test, with 6-μm diamond abrasive particles [47]. It can be observed that the typical dimensions of the scratch caused by the abrasive diamond particles (as well as that of the debris formed) are smaller than 150 nm.

The point being made is that, although up to now the studies concerning wear at nanometric scales have been performed mainly in electronically relevant materials (see Sect. 21.5 of this chapter), the wear mechanisms at this scale are certainly of a much wider relevance than that restricted to the wear of electronic and ultraprecision components.

21.3 Wear Theories

21.3.1 Classical Wear Theories

Probably, Holm [48] was the first to discuss the mechanisms of wear in his analysis on the relations between contact area and conductance. Later, Archard developed, based on some of Holm's ideas, a simple mathematical model for wear caused by adhesion between two opposing asperities [49]. The Archard wear equation for plastic contacts (usually known simply as the "Archard equation") is similar to that proposed by Holm, but Archard replaced the Holm's concept of "removal of atoms" by the concept of "removal of wear particles". So, in fact, the first model for wear was an atomic wear model.

In spite of its simplicity, the Archard model for "adhesive" wear is still widely used (sometimes outside of its developing context). The model considers two opposing asperities of materials with similar mechanical properties contacting during sliding. At some moment, the location of the two areas forming the contact is fully established (as shown in Fig. 21.5). A short time later the contact area is reduced to zero but, in addition, it is assumed that, at this moment, a new similar contact area has just been fully established somewhere in the surface. In each asperity contact, chemical interbonding and diffusion can lead to local sticking. The continuous relative motion can, thereof, result in a separation of the asperities at a surface different from

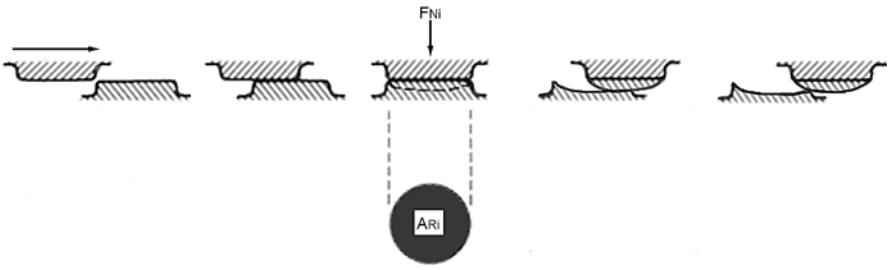


Fig. 21.5. Scheme of the Archard model for adhesive wear

the original one (Fig. 21.5), leading to surface damage and wear by material transfer.

The Archard model considers that the wear rate, Q_i, i. e. the volume of material transferred from one asperity to another per unit of sliding length, is simply proportional to the contact area when the contact is fully established, $A_{\mathrm{R}i}$:

$$Q_i = \frac{V_i}{L} = k_{\mathrm{ad}} A_{\mathrm{R}i} \tag{21.6}$$

Although it can be interpreted in various ways (e. g. see [50] for a detailed discussion), the factor, k_{ad}, is proportional to the probability that touching asperities adhere and originate mass transfer from one surface to another.

As previously discussed (point 2), the most probable wear mechanism is that the contacting asperities are plasticized and wear occurs by the removal of material lumps. If the asperities are fully plasticized, when submitted to a force $F_{\mathrm{N}i}$, than $A_{\mathrm{R}i}$ results directly from Tabor's definition of hardness [51]:

$$A_{\mathrm{R}i} = \frac{F_{\mathrm{N}i}}{H} \tag{21.7}$$

with H the hardness of the material. From Eqs. (21.6) and (21.7) one can obtain:

$$Q_i = k_{\mathrm{ad}} \frac{F_{\mathrm{N}i}}{H} \tag{21.8}$$

Assuming that all the contacts are fully plasticized ($\Psi > 1$), the total wear rate is simply:

$$Q = \sum Q_i = \frac{k_{\mathrm{ad}}}{H} \sum F_{\mathrm{N}i} = k_{\mathrm{ad}} \frac{F_{\mathrm{N}}}{H} \tag{21.9}$$

Eq. (21.9) was deduced for fully plastic contact conditions but Archard proposed the more general relation:

$$Q_i \propto F_{\mathrm{N}}^n \tag{21.10}$$

with $n = 0.6, 0.75, 0.8$ or 1, depending on whether the contacts are elastic or plastic and on whether the wear particle transferred is layer-like or lumpy-like (e. g. Eq. (21.9) is the particular case of Eq. (21.10) with $n = 1$, which stands for plastic contacts and lumpy-like transfer).

However, the subsequent work of Greenwood and Williamson [13] showed that, for multiasperity contacts, A_{R} is always directly proportional to F_{N}, independently of the contacts being elastic or plastic. Assuming this, the direct proportionality between wear rate and load stands not only for fully plastic contact conditions. In this way, a more general form of Equation (9) is simply:

$$Q = k_{\mathrm{ad}} \frac{F_{\mathrm{N}}}{\phi} \tag{21.11}$$

ϕ is the hardness, H, if the contacts are fully plasticized, or an "elastic contact hardness" if the contacts are in the elastic regime. For Gaussian surface

profiles this elastic contact hardness is given by [13]:

$$\phi = 0.25E^* \sqrt{\frac{R_\mathrm{d}}{\rho}} \tag{21.12}$$

A different situation is that of contacting surfaces with very different chemical and mechanical properties. In particular, if the hardness of one material is higher than that of the other (typically $H_1 > 1.2H_2$ as a rule of thumb [10]), the harder asperities will indent the softer material. The relative motion will lead to the formation of a groove (as shown in Fig. 21.6), resulting in "abrasive" wear of the softer material. This type of wear damage is not caused by adhesion, but rather by mechanical deformation and cutting. Therefore, the Archard model is not valid for "abrasive" wear.

Rabinowicz [52] proposed a different approach for abrasive wear. Assuming a conical asperity of the harder surface, infinitely rigid, that, when submitted to a load F_N indents the softer surface. After the softer surface become fully plasticized (Fig. 21.6), the equilibrium condition in the load application axis is given by:

$$F_\mathrm{N} = HA_\mathrm{R} = H\frac{\pi r^2}{2} = H\frac{\pi}{2}tg^2\theta h^2 \tag{21.13}$$

with r, h and θ defined in Fig. 21.6.

If all the material displaced from the groove is removed, than the worn volume, V, is given by:

$$V = Lrh = Lh^2tg\theta \tag{21.14}$$

From Eqs. (21.13) and (21.14) one can obtain,

$$Q = \frac{V}{L} = k_\mathrm{ab}\frac{F_\mathrm{N}}{H} \tag{21.15}$$

with $k_\mathrm{ab} = 2/(\pi tg\theta)$.

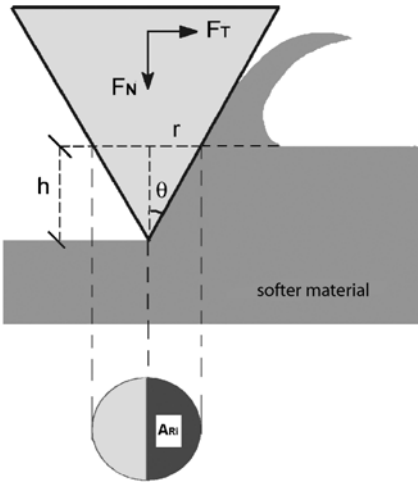


Fig. 21.6. Scheme of the Rabinowicz model for abrasive wear

By comparing Eqs. (21.9), (21.11) and (21.15), it can be concluded that the Rabinowicz model for abrasive wear is formally identical to Archard's model for adhesive wear. Moreover, the equations are identical to those obtained by Holm for "atomic wear" [48].

Apparently, this can be a quite surprising result since the phenomenology of each type of wear is different. But, in fact, all these approaches are based on the same basic idea: the worn volume is directly proportional to the real contact area between the tribological pair. This real contact area is proportional to the ratio F_N/H, for fully plastic contacts, or to the ratio F_N/ϕ (with ϕ defined in Eq. (21.12)), for GW Gaussian multiple elastic contacts.

Therefore, it seams that a more general equation for wear due to the relative motion of contacting surfaces can simply be:

$$Q = k_{tot} A_R \tag{21.16}$$

with k_{tot} an additive wear coefficient, with contributions of the active wear mechanisms in a given situation. Equation (21.16) applies in all the cases for which the wear rate is proportional to the contact area (there are some exceptions, e. g., the case of wear of brittle materials [53]).

21.3.2 Atomic Wear Theories

From the simple formulation given by Eq. (21.16), there is no discontinuity between wear occurring at atomic/nanometric scales and at larger ones: if the real contact area decreases to nanometric scales the wear rate decreases proportionally (as already mentioned, Holm's "removal of atoms" theory [48] led to an equivalent equation). Of course, from the predictive point of view, Eq. (21.16) is of limited helpfulness, since there is an uncertainty related with k_{tot}. The wear coefficient, k_{tot}, by depending on the active wear mechanisms, will surely also be scale dependent.

As the contact scale decreases to near-atomic lengths, the wear mechanisms involving extensive plastic deformation and crack propagation, such as grooving, delamination or fatigue, cannot be predominant if only a few atomic layers are involved in the contact. Molecular dynamics simulations, in contact scales involving a few atomic layers, show that at this small length scale, wear is essentially controlled by the dragging of atoms from their initial positions [54, 55].

Bassani and D'Acunto [56] developed a theoretical approach to the atomic dragging between a flat surface and an AFM tip under adhesive contact conditions. The approach used is based in a double-well potential model, in which the wear volume is quantified in terms of the atomic transition between the two well minima. According to this model, the total wear volume per unit time, V_t (i. e. the total volume of the atoms jumping from the flat surface to the tip surface in unit time) is given by:

$$V_t = n\omega V_{at} \tag{21.17}$$

where n is the initial atomic population, ω is the transition rate, and V_{at}, the atomic debris volume (note that this atomic debris is not necessarily an atom, since it can be an ionic pair or a polyatomic molecule, for instance). The transition rate between wells is given by an Arrhenius law:

$$\omega = \omega_{\mathrm{o}} \exp\left(-\frac{\Delta U}{kT}\right) \tag{21.18}$$

in which ΔU is the height of the energy barrier between the two well minima.

Later, D'Acunto [57,58] extended the model to the quantification of wear mechanisms at the nanoscale. It was proposed that the atomic wear rate (i. e. wear in which the active mechanisms are dragging of atoms (or ionic pairs) from their original positions) could be separated into two basic mechanisms: adhesion and abrasion. At this point it should be emphasized that the concept of "adhesive" and "abrasive" wear at the atomic scale does not have exactly the same meaning as defined previously, when the Archard and Rabinowicz models were presented. At atomic scales, adhesive wear was used for the cases in which the vertical (Van der Waals) forces are predominant and the atoms transfer from one surface to another. Conversely, atomic abrasive wear was used for the case in which shear forces are predominant, leading to the dragging of atoms to further positions from the primitive ones (in this case, the atoms do not jump from one surface to another). Figure 21.7a and b illustrates schematically the atomic adhesive and abrasive wear concepts.

The wear volume per unit of time defined in Eq. (21.17) can be transformed in a wear rate by dividing by the scanning speed of the asperity / tip, V_{s}. By doing this, D'Acunto's model leads to the following expression for the

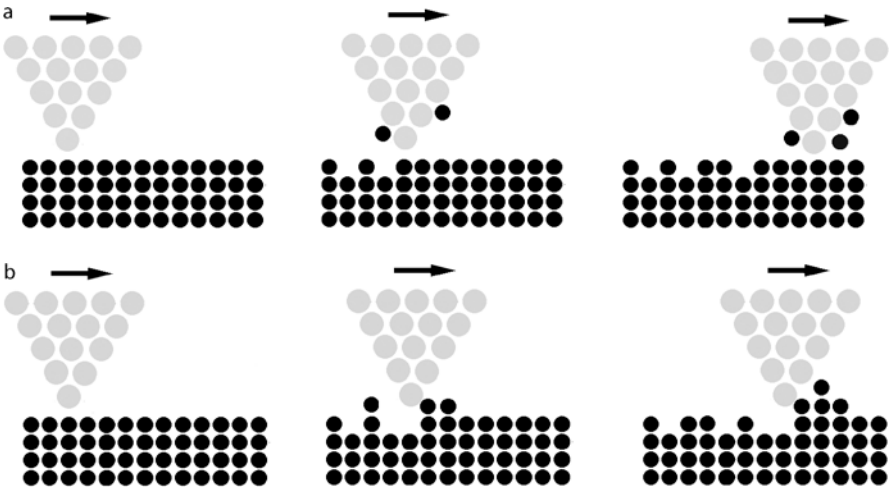


Fig. 21.7. Scheme of atomic adhesive wear (**a**) and atomic abrasive wear (**b**)

wear rate at the atomic scale:

$$Q = \frac{n\omega V_{\mathrm{at}}}{V_{\mathrm{s}}} \tag{21.19}$$

By noting that n can be given by $A_{\mathrm{R}}/A_{\mathrm{R}/n}$, with $A_{\mathrm{R}/n}$ the specific area per contact atom, Eq. (21.9) can be rewritten in the equivalent form:

$$Q = k_{\mathrm{at}} A_{\mathrm{R}} \tag{21.20}$$

with,

$$k_{\mathrm{at}} = \frac{\omega V_{\mathrm{at}}}{A_{\mathrm{R/n}} V_{\mathrm{s}}} \tag{21.21}$$

In spite of the formal similarity between Eq. (21.20) and Eq. (21.16), this result is different from what is obtained for bulk wear. According to theories for abrasive and adhesive wear at bulk scales (Eqs. (21.9) and (21.15), respectively), the wear rate is independent of the scanning speed, if the wear mechanism does not change within a certain range of speeds (e. g., see Lim and Ashby approach to sliding wear [59]). Therefore, an important result arising from this model, is that, atomic-scale wear is inversely proportional to the travelling speed of the counterbody. This is caused by the fact that, if speed increases, the interaction time for atom jumping decreases. Experimental observations for atomic wear of ionic crystals [23] do not show a significant velocity dependence of atomic wear as expected from the model. However, the experimental data and the velocity ranges tested are still scarce and further work is certainly necessary (in point 5 a more detailed overview of the experimental results, obtained up to now in nanoscale and atomic wear experiments, will be presented).

21.4 Hardness at Different Scales: is Smaller Harder?

Hardness is a measure of the yield stress of the material [51]. The appearance of this property in Archard and Rabinowicz wear models (Eqs. (21.9) and (21.15), respectively) is related to the fact that hardness measures the resistance of the asperities to the start of the dissipative plastic deformation processes, which can result in the formation of wear particles. Plastic deformation mechanisms are essentially related to the creation and movement of dislocations in metals and ceramics and chain reptation in polymers. As a consequence, if the interaction between body and slider is only of a few atomic layers, the deformation processes are inactive, and the nature of wear must be distinct, as discussed in point 3.2.

However, there is a length-scale range, from a few atomic layers to submicrometric lengths, in which, although the deformation mechanisms are already still active, the mechanical response of the material can be different

from its "bulk" mechanical response. In this (poorly defined) length range, the transition from "bulk wear" mechanisms to "atomic wear" mechanisms occurs. Changes in hardness with the indentation depth are particularly useful to identify this transition region, with the advantage that this property can be directly related to the wear behavior of the material.

In recent years there has been a growing interest in the use of molecular dynamics simulations (MD) methods to investigate nanoindentation [12]. Landman et al. [54], carried out one of the first MD simulations of a Ni tip indentating a gold surface. Subsequently, MD has been used in several studies of nanoindentation in different materials (e. g. [60–63]). Although, at present, atomistic methods allow for simulations of up to millions of atoms in time scales of microseconds, and provide information about the single-atom role in the contact [12], the length and time scales are still insufficient to provide more extensive links to experimental data [5]. Particularly, in the scale lengths of a few hundreds of nanometers there is a gap of understanding, since it is one too large scale for atomistic simulations and a too small scale for continuum mechanics theories to be still valid [64].

The fact is that, in this scale range of a few hundreds of nanometers, a large number of experimental works have shown that, in general, an increase of the hardness of the material occurs as compared with its bulk hardness, the so-called *indentation size effect* (ISE). ISE has been observed in a significant variety of metals, metallic alloys and ceramics: Cu [65, 66], Mo and W [67], Ag [68], Al and brass [69], Ni and Co [70], sapphire and MgO [71], TiO_2 and SnO_2 [72], LiF and NaCl [73], among others.

There are various explanations for the observed indentation size effect, including the sample preparation and test methods and the increasing perfection of the materials as the volume is reduced [71]. Of course, as mentioned in point 2.1, the constitution of the material changes on approaching the surface (Fig. 21.1). This, most certainly, will lead to different material properties, in particular to different hardness results, as the indenter crosses the different layers, schematically represented in Fig. 21.1. In particular, it has been observed that the deformed layer caused by the mechanical polishing of surfaces may contribute to the increase the hardness of metallic surfaces [66, 71]. Nevertheless, the ISE is still observed in electropolished samples and, in fact, the results strongly indicate that the ISE is an intrinsic material characteristic [74].

Conventional plasticity theories, which are based in continuous mechanics and, thus, do not involve a length-scale dependency, do not provide an explanation for the size dependency of hardness. Up to now, there are two theories that seem to offer the best explanation for the ISE: geometrically necessary dislocations (strain-gradient plasticity) [75] and surface free-energy effects [76, 77]. These two explanations for the indentation size effect will be briefly reviewed.

When a material is plastically deformed dislocations are generated, moved and stored. This storage causes the material to work-harden [78]. Disloca-

tions can be randomly distributed and stored, forming a network of defects in the crystalline structure of the material and, in this case, they are called statistically stored dislocations (STDs) [79]. In nonuniform deformation, dislocations are necessary for compatibility reasons and, in this case, they are called geometrically necessary dislocations (GNDs) [79, 80].

In the case of indentation experiments, the hypothesis is that plastic strain gradients surrounding the indentation are accommodated by the formation of these geometrically necessary dislocations. Since, in general, strain gradients are inversely proportional to the length scale over which plastic deformation occurs as the indentation depth decreases the density of GNDs must increase. This will lead to an increase of the hardness of the material [81, 82]. Based on these assumptions, Nix and Gao [75], proposed a model that relates the indentation size effect with the density of GNDs. The Nix–Gao model leads to a simple relation between the variation of the materials hardness caused by the GNDs, ΔH_{GND}, and the indentation depth, h [75]:

$$\Delta H_{\mathrm{GND}} = H_0 \left(\sqrt{1 + \frac{h^*}{h}} - 1 \right) \tag{21.22}$$

H_0 is the material hardness at an infinite depth (the "bulk hardness"). h^* is a characteristic length that depends on the shape of the indenter and on the tested material, given by:

$$h^* = \frac{3}{2} \frac{\tan^2 \theta}{b \rho_{\mathrm{s}}} \tag{21.23}$$

where b is the Burgers vector of the dislocations and ρ_{s} is the density of statistically stored dislocations. θ is the angle between the top surface of the material and the indenter.

Surface free-energy-related effects can also contribute to the increase of the hardness with decreasing scale, since the area to volume ratio increases as the scale decreases [76, 77]. According to Jager [77], surface free energy can have a non-negligible contribution to hardness if the three following conditions are simultaneously met: a) soft ductile samples with high SFE; b) sharp indenter geometries; c) shallow indentations.

Recently developed models accounting for the contribution of SFE to hardness lead to a relation of the type [76, 77]:

$$\Delta H_{\mathrm{SFE}} = \kappa \frac{E_{\mathrm{s}}}{h} \tag{21.24}$$

where ΔH_{SFE}, is the hardness variation with depth caused by SFE effects, E_{s} is the surface free energy, h the indentation depth, and κ a constant that depends on the geometry of the indenter (e. g., for a Veeco DNISP diamond AFM tip, $\kappa \approx 3.43$ [70]).

Assuming that both these effects (GNDs and SFE) are independent, and thus additive, it results from Eqs. (21.23) and (21.24), that a general equation

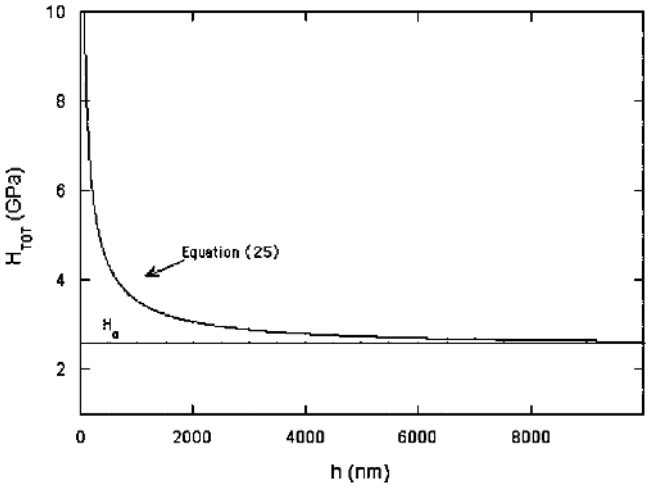


Fig. 21.8. Hardness variation with indentation depth according to Eq. (21.25). The plot was calculated with a set of values typical of a metallic alloy, e. g. a plain carbon steel: $H_0 = 2.5\,\mathrm{GPa}, E_\mathrm{s} = 2\,\mathrm{J/m^2}$ and $h^* = 1000\,\mathrm{nm}$ [75]). κ was taken as 3.43 [70]

for hardness as a function of indentation depth, accounting for both effects, has the form:

$$H_\mathrm{TOT} = H_0 + \Delta H_\mathrm{GND} + \Delta H_\mathrm{SFE} = H_0\sqrt{1 + \frac{h^*}{h}} + \kappa\frac{E_\mathrm{s}}{h} \qquad (21.25)$$

A typical plot of Eq. (21.25) is shown in Fig. 21.8. The plot was calculated with a set of values typical of a metallic alloy, e. g. a steel: $H_0 = 2.5\,\mathrm{GPa}, E_\mathrm{s} = 2\,\mathrm{J/m^2}$. h^* was taken as 1000 nm [75] and κ as 3.43 [70].

As can be observed, for indentation depths smaller than some thousands of nanometers, the combined effect of geometrically necessary dislocations and surface free energy can result in a non-negligible increase of the hardness. This increase can be particularly relevant in the submicrometric ranges. Therefore, both experimental and theoretical results obtained up to now seem to lead to the same conclusion: smaller tends to be harder, i. e. as the contact scale decreases to submicrometric ranges the resistance of the material to plastic deformation increases.

Our understanding of the theory of nanoindentation has certainly improved in recent years. However, many questions remain, such as strain rate and temperature dependence, dislocation nucleation and glide mechanisms [12]. Nanoindentation mechanics must certainly have an impact on the wear behavior of materials as the contact scale is reduced. There is certainly a need for comparative wear studies at micro- and (supra-atomic) nanometric scales. This will be an interesting research topic that, surely, will grow in the near future.

21.5 Nanoscale Wear Experiments

The development of atomic force microscopes [14], twenty years ago, has provided tribologists with a new tool that enables experiments from atomic scales to (almost) bulk ones. Up to now, an increasing number of studies has dealt with this topic but, as pointed out some years ago by Carpick and Salmeron [83], since this field is still in its infancy (may be in its preadolescence now), results so far are generally sporadic in scope and direction. Yet, a trend of growth is clearly perceptible, which will certainly continue in the future, due to the undeniable scientific and technological relevance of the field [84].

AFM experiments have shown that observable tip-induced wear in the sample (and also wear of the tip itself [20,85,86]) can occur when the AFM tip slides in contact with the sample, above some threshold load [83]. Tip-induced nanowear experiments have been performed (in UHV and environmental conditions) in a relatively large range of materials, such as thin films of AgBr and C_{60} deposited on NaCl [87], NaCl [88], KBr [23,89,90], PZTs [20], polymeric magnetic tapes [84], silicon [91, 92], SiO_2 [92], mica [93, 94], thermoplastic and thermoset polymers [95–97], among other materials. Fewer experiments have been performed with metals, possibly because, as pointed out by Gnecco et al. [89], metals are not the best candidates to study wear mechanisms by AFM, since the debris tend to accumulate on the tip, leading to a poor reproducibility of measurements. Still, some results on gold [98], MnZn ferrites [99], Cu [91] and Cr thin films [100] can be found in the literature.

One of the first in situ observations of the transition threshold to detectable wear in AFM experiments was made by Hu et al. [93]. The experiments were performed in mica, with silicon nitride probes, in air, water and ethanol environments. While monitoring the frictional forces, these authors observed that, once a certain critical load threshold was reached, the wear of mica (remotion of an atomic layer) was responsible for a transition from a linear relation between friction and load to a stochastic behavior (Fig. 21.9). The experiments showed that the wear of this layer of mica occurs both at high loads, in a single scan, or at lower loads, in multiple scans. This result strongly indicates that the load onset for atomic wear of mica has somehow a memory effect to the number of scans of the counterface slider. Hu et al. [93] suggested that, in the lower-load regime, point defects were accumulating during each scan. When the number of scans reaches a critical value, this accumulation will result in the formation of a small cluster (a nanodebris), resulting from the removal of one atomic layer. Recently, Helt and Bateas [101] were able to observe the nucleation of defects in muscovite mica under aqueous environments, prior to gross wear, confirming the nanowear mechanism proposed by Hu et al.

Gnecco et al. [23] have performed abrasive wear experiments at nanoscale on (001) KBr. They have shown that, under UHV conditions, the atomic-scale wear of KBr is due to the removal of single ion pairs. They have also shown

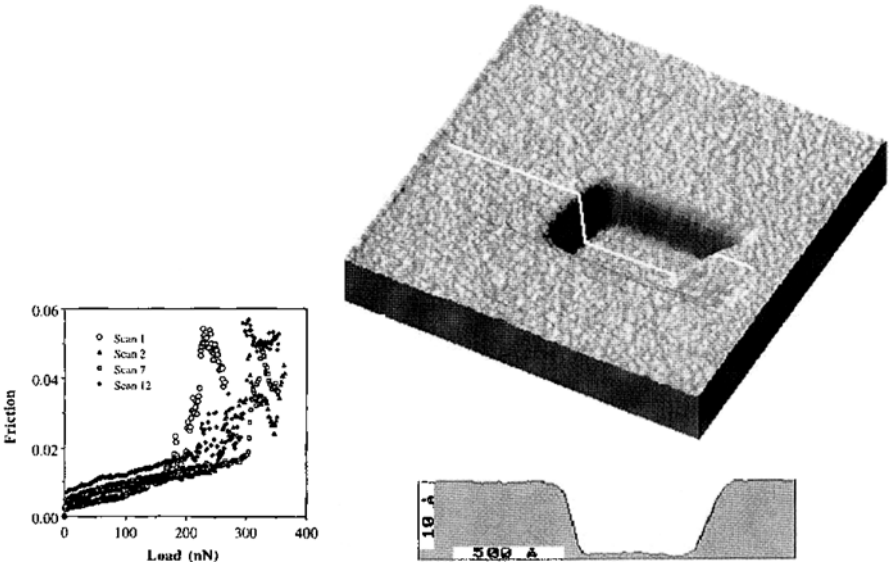


Fig. 21.9. Friction force (au) versus load in the high-load regime for multiple scans of a silicon nitride tip in contact with mica. Friction increases smoothly with load until wear occurs. At this time abnormal changes in the friction behavior are observed [93]

that the wear debris is reorganized in regular terraces with similar structure and orientation as in the unscratched surface. More recent experiments performed by the same group [90], also confirmed the cumulative nature of wear at atomic scales. In fact, they observe that, even at very low loads (between 1.7 and 30.1 nN) scratching of KBr with silicon probes in UHV conditions, always results in groove formation, after a sufficient number of scans is reached. Concomitantly, an irregular ripple formation in the periphery of the grooves was observed. This ripple-like structure appears after the onset of wear and has a periodicity that matches well with the scale length of the tip radii. In this way, the authors attributed the KBr wear mechanism at atomic scale to an accumulation in front of the tip of ionic pairs or small ion clusters. Ripple arises when the material transported by the tip increases friction and the tip jumps over it as the process restarts.

The experimental results obtained up to now seem to indicate that, in general, the precursor mechanism for atomic-scale wear is the formation of point defects that lead to the formation of small clusters. It is interesting to note that these observations match quite well with the theoretical double-well approach presented in Sect. 21.3.2. In fact, according to this model, in each interaction between surface and counterface atoms, there is a probability of atomic jumps to occur to the opposite surface (adhesion) or to another position in the same surface (abrasion). These jumps will result in the formation of point defects. Thus, both experimental and theoretical results seem to con-

firm that, in general, the formation of point defects in sliding contacts is the precursor mechanism for atomic wear. This mechanism is predominant in situations in which the contact loads are below the onset of plastic deformation or crack nucleation.

In a small, but relatively larger contact scale, Wang and Kato [102, 103] made a set of interesting wear experiments inside the chamber of an E-SEM (environmental scanning electron microscope). The experiments were performed with a pin-on-disk geometry in carbon nitride coatings and in bare Si(111). A diamond pin with a 10-μm curvature radius and a load range between 10 and 250 mN were used. One of the aims of the experiments was to observe the onset of wear from "no observable wear particles" to "wear-particle formation". This transition was defined when the formation of a cluster of worn particles, larger than 0.25 μm, could be confirmed by in situ E-SEM observation. The worn particles were described as "feather-like", for the mild wear regime, and "plate-like", for the higher wear regime.

Figure 21.10 summarizes the results obtained for the carbon nitride coatings. It can be observed that the transition from "no observable wear particles" to "wear-particle formation" depends on the load (as expected). However, the transition also occurs, in certain load conditions, when a critical number of cycles is reached. The authors have attributed this wear transition dependence of the number of cycles to a low-cycle fatigue wear mechanism [103]. However, and in spite of the higher loads used in Wang and Kato's experiments, it is worth mentioning the similarities with the observations for the nanowear onset of mica (Fig 21.9) [93].

As previously mentioned, when the load increases, resulting in higher depth interactions, other wear mechanisms become active. Also, the chemical

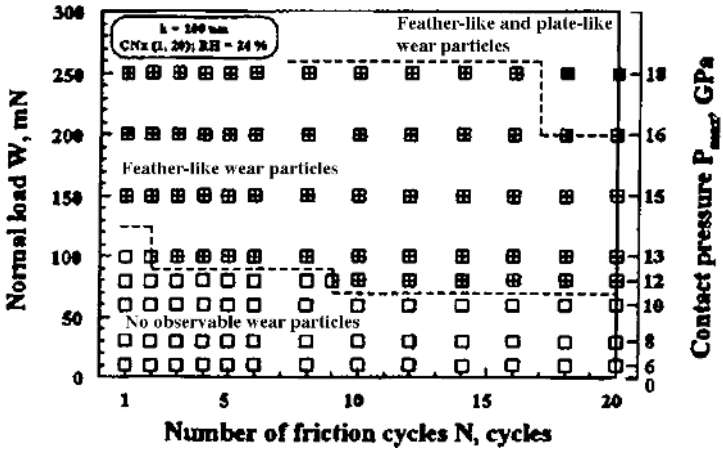


Fig. 21.10. Occurrence of wear particles of a carbon nitride coating, in terms of load and friction cycles, based on E-SEM observations [102]

Table 21.1. Wear coefficients for Si obtained in tests performed at different loads (after [91])

Material	Load	Test	Wear coefficient (as defined by Eq. (21.9) or (21.15))	Ref.
Si (100)	10 – 30 μN	Specimen-on-disk	10^{-5}	[104]
Si (100)	100 μN	AFM	10^{-1}	[105]
Si (111)	100 mN	Pin-on-disk in SEM	10^{-4}	[102]
Si	100 mN	Pin-on-disk	10^{-3}	[106]

and structural composition of the surface interlayers can change (Fig. 21.1). As a consequence, the wear response of the material can be quite different and, moreover, unpredictable, depending on the contact scale. For now, it seems that the only choice for tribologists is to carefully choose representative test conditions, in view of the application. The results obtained by Miyake et al. [19] for diamond and N^+-implanted diamond films clearly illustrate this point. These authors have observed that nanowear AFM tests, with loads in the range of μN, lead to the lowest wear resistance in N^+-implanted diamond films, as compared with diamond films. On the contrary, N^+-implanted films present the highest wear resistance in reciprocating wear tests, with loads in the range 0.098 to 4.9 N. This difference was attributed by the authors to the formation of a thin amorphous layer in the N^+-implanted samples (as observed by Raman spectroscopy), whose properties are relevant for AFM wear tests but not for the wear tests at higher loads.

Chung and Kim [91] presented a summary of the results obtained for low-load wear experiments, in materials for MEMS and ultraprecision electronic applications. The data was obtained by different authors, using low-load pin-on-disk and AFM tests in Si-based materials, carbon films and DLCs. The load ranges used changes from μN to mN. Table 21.1 shows the results for silicon, as reviewed by Chung and Kim. These results show that, although a general perception of the basic small-scale wear mechanisms starts to exist, extreme difficulties in quantifying wear still remain.

Clearly, advances can only be made if precise and reproducible experiments are available to test theoretical models [5]. Due to the complex nature of mass-dissipation phenomena, this goal is still to be achieved and, certainly, will continue to be a target for future endeavors.

21.6 Conclusions

In conclusion, one can say that there has been a great interest in submicrometric wear studies within the last 10 – 15 years. These studies cover different types of materials, different length scales (going from quasiatomic to quasi-

bulk ones) and different research objectives (from generic theoretical studies to well-focused application ones). Although the natural dispersion of results, caused by the great complexity of the phenomena involved, and to the youthful nature of the field, significant advances have been made in recent years.

The models and results obtained up to now suggest that, on approaching atomic scales, a division of mechanisms can be made between "atomic wear" and "bulk wear", i. e. wear involving atomic jumps and point-defect formation (the former) and wear involving plastic deformation and crack propagation (the latter).

For atomic wear, experimental results strongly suggest that there is threshold for detectable wear. This threshold is a load threshold but can also be a number of contact cycles threshold. This suggests a cumulative nature for atomic wear, which is in agreement with theoretical models.

For higher loads and higher contact interaction depths, plastic deformation and crack propagation mechanisms will be activated, and the wear can, in principle, be described by classical continuum-mechanics-based theories. However, it should be noted that there is a length scale interval (from some nanometers to some hundreds of nanometers) in which, although the deformation and crack-propagation mechanisms can be active, important differences as compared with the same mechanisms occurring at larger scales can occur. These differences can be caused by surface-topography-related factors, changes in the structure and chemical composition of the material on approaching the surface and intrinsic effects related with increasing surface to volume ratio (surface free energy, strain gradient plasticity phenomena). What happens in this poorly known transition region from bulk wear to atomic wear is still far from being understood. Certainly, in the near future, experimental results, MD simulations and the development of scale-dependent plasticity theories will bring new insights to this topic, reducing the gap between the comprehension of macro-, micro- and nanowear mechanisms.

Acknowledgement. The author would like to thank S. Graça, M. A. Fortes and M. D'Acunto for the useful discussions and suggestions, and R. Vilar for creating the conditions that result in this work. The author also thanks the Portuguese Foundation for Science and Technology (FCT – Project Nanonico, POCTI/CTM/59376/2004), and also the European Science Foundation (ESF) scientific program "NATRIBO", for the financial support of the nanotribology research line at IST.

References

1. H.P. Jost, Lubrication (tribology) - A report of the present position and industry's needs, Dept. of Science and Education, H. M. Stationery Office, London, U.K. (1966).
2. H.P. Jost, Economic impact of tribology, Mechanical Engineering, August, 26–33 (1975).

3. E. Rabinowicz, Friction and Wear of Materials. 2nd edn, John Wiley and Sons (1995).
4. H. Liu and B. Bhushan, Nanotribological characterization of digital micromirror devices using an atomic force microscope, Ultramicroscopy, 100, 391–412 (2004).
5. S.S. Perry and W.T. Tysoe, Frontiers of fundamental tribological research, Tribology Letters, 19 (3), 151–161 (2005).
6. R. Colaço and R. Vilar, A model for the abrasive wear of metallic matrix particle-reinforced materials, Wear, 254 (7–8), 625–634 (2003).
7. A Strategy for Tribology in Canada, National Research Council of Canada, Canada (1986).
8. An investigation on the application of tribology in China, Tribology Institute of the Chinese Mechanical Engineering Society, China (1986).
9. J. Krim, Surface science and the atomic scale origins of friction: what once was old is new again, Surface Science, 500, 741–758 (2002).
10. J.B. Adams, L.G. Hector, D.J. Siegel, H. Yu and J. Zhong, Adhesion, lubrication and wear on the atomic scale, Surface and Interface Analysis, 31, 619–626 (2001).
11. I.M. Hutchings, Tribology: friction and wear of engineering materials, Edward Arnold (1992).
12. B. Bushan, An introduction to tribology, John Wiley and Sons, Inc. (2002).
13. J.A. Greenwood and J.B. Williamson, Contact of nominally flat surfaces, Proceedings of the Royal Society of London, A295, 300–319 (1966).
14. G. Binnig, C.F. Quate and C. Gerber, Atomic Force Microscope, Physical Review Letters, 56 (9), 930–933 (1986).
15. S. Sundararajan and B. Bhushan, Micro/nanotribology of ultra-thin hard amorphous carbon coatings using atomic force friction force microscopy, Wear, 225–229, 678–689 (1999).
16. B. Bhushan, Nano to microscale wear and mechanical characterization using scanning probe microscopy, Wear, 251, 1105–1123 (2001).
17. R. Kaneko, K. Nonaka and K. Yasuda, Scanning Tunneling Microscopy and Atomic Force Microscopy for Nanotribology, Journal of Vacuum Science & Technology A - vacuum surfaces and films, 6 (2), 291–292 (1988).
18. Z. Jiang, C.-J. Lu, D.B. Bogy and T. Miyamoto, An investigation of the experimental conditions and characteristics of a nanowear test, Wear, 181–183 777–783 (1995).
19. S. Miyake, T. Miyamoto and R. Kaneko, Increase of nanometer-scale wear of polished chemical-vapor-deposited diamond films due to nitrogen ion implantation, Nuclear Instruments and Methods in Physics Research B - beam interactions with materials & atoms, 108, 70–74 (1996).
20. K.-H. Chung, Y.-H. Lee, D.-E. Kim, J. Yoo and S. Hong, Tribological characteristics of probe tip and PZT media for AFM-based recording technology, IEEE Transactions on Magnetics, 41 (2), 849–854 (2005).
21. S. Graça, R. Colaço and R. Vilar, Using atomic force microscopy to retrieve nanomechanical surface properties of materials, Materials Science Forum, 514–516, 1598–1602 (2006).
22. A.R. Machcha, M.H. Azarian and F.E. Talke, An investigation of nano-wear during contact recording, Wear, 197, 211–220 (1996).
23. E. Gnecco, R. Bennewitz and E. Meyer, Abrasive wear on the atomic scale, Physical Review Letters, 88 (21), 215501/1–215501/4 (2002).

24. W. Gulbinski, T. Suszko and D. Pailharey, High load AFM friction and wear experiments on V2O5 thin films, Wear, 254, 988–993 (2003).
25. J.Y. Park, D.F. Ogletree, M. Salmeron, C.J. Jenks and P.A. Thiel, Friction and adhesion properties of clean and oxidized Al-Ni-Co decagonal quasicrystals: a UHV atomic force microscopy/scanning tunneling microscopy study, Tribology Letters, 17 (3), 629–636 (2004).
26. J. Drelich, G.W. Tormoen and E.R. Beach, Determination of solid surface tension from particle-substrate pull-off forces measured with the atomic force microscope, Journal of Colloid and Interface Science, 280, 484–497 (2004).
27. H. Gao, Y. Huang, W.D. Nix and J.W. Hutchinson, Mechanism-based strain gradient plasticity-I. Theory, Journal of the Mechanics and Physics of Solids, 47, 1239–1263 (1999).
28. Y.G. Wey and J.W. Hutchinson, Steady-state crack growth and work of fracture for solids characterized by strain gradient plasticity, Journal of the Mechanics and Physics of Solids, 45 (8), 1253–1273 (1997).
29. G. Elssner, D. Korn, R.M. Cannon and M. Rühle, Fracture properties of interfacially doped Nb-Al_2O_3 bicrystals: I, fracture characteristics, Acta Materialia, 50 (15), 3881–3901 (2002).
30. K.-H. Zum Gahr, Microstructure and Wear of Materials, Elsevier Scientific Publishing Company (1987).
31. H. Luth, Surfaces and interfaces of solids, Springer-Verlag (1993).
32. F.P. Bowden, A.J.W. Moore and D. Tabor, The plowing and adhesion of sliding metals, Journal of Applied Physics, 14, 80–91 (1943).
33. G.I. Finch, A.G. Quarrekk and J.S. Roebuck, The Beilby Layer, Proc. of the Royal Society of London, 145 (855), 676–681 (1934).
34. B.W.E. Avient and H. Wilman, New features of the abrasion process shown by soft metals: the nature of mechanical polishing, British Journal of Applied Physics, 13, 521–526 (1962).
35. L.H. Gerner, Diffraction of electrons by metal surfaces, Physical Review, 43 (9), 724–726 (1933).
36. D.M. Turley and L.E. Samuels, The nature of mechanically polished surfaces of copper, Materials Characterization, 39 (2–5), 399–418 (1997).
37. K.L. Johnson, K. Kendall and D. Roberts, Surface Energy and the Contact of Elastic Solids, Proceedings of the Royal Society of London, A 324, 301–313 (1971).
38. K.N.G. Fuller and D. Tabor, The effect of surface roughness on the adhesion of elastastic solids, Proceedings of the Royal Society of London, A 345, 327–342 (1975).
39. T.R. Thomas, Rough surfaces, Longman, London and New York (1982).
40. L. Zhou, M. Beck, H. Gatzen, K. Altshuler and F. Talke, Slider vibration reduction using slider surface texture, Microsystem Technologies - micro- and nanosystems - information storage and processing systems, 11 (8–10), 857–866 (2005).
41. T. Hisakado and T. Tsukisoe, Effect of surface roughness on transient wear, Journal of the Japan Society of Lubrication Engineers, 21 (4), 228–235 (1976).
42. S. Gantiand B. Bhushan, Generalized fractal analysis and its application to engineering surfaces, Wear, 180, 17–34 (1995).
43. H.H. Gatzen and M. Beck, Wear of single crystal silicon as a function of surface roughness, Wear, 254, 907–910 (2003).

44. X. Wang, K. Kato and K. Adachi, Running-in effect on the load-carrying capacity of a water-lubricated SiC thrust bearing, Proceedings of the Institution of Mechanical Engineers Part J - Journal of Engineering Tribology, 219 ((J2)), 117–124 (2005).
45. M.A. Fortes, R. Colaço and M.F. Vaz, Contact mechanics of cellular solids, Wear, 230, 1–10 (1999).
46. D. Tabor, Lubrication and Wear, in Surface and Colloid Science, E. Matijevic, Ed. John Wiley. 245–312 (1972).
47. R. Colaço and R. Vilar, On the influence of retained austenite in the abrasive wear behavior of a laser surface melted tool steel, Wear, 258 (1–4), 225–231 (2005).
48. R. Holm, Electrical contacts, H. Gerber Pub, Stockholm (1946).
49. J.F. Archard, Contact and rubbing of flat surfaces, Journal of Applied Physics, 24 (8), 981–988 (1953).
50. D.A. Rigney, Some thoughts on sliding wear, Wear, 152 187–192 (1992).
51. D. Tabor, The hardness of metals. Oxford Classic Texts, Clarendon Press (1951).
52. E. Rabinowicz, Friction and wear of materials, John Wiley and Sons (1965).
53. A.G. Evans, Abrasive wear of ceramics, American Ceramic Society Bulletin 56 (3): 292–292 1977, 56 (3), 292–292 (1977).
54. U. Landman, W.D. Luedtke and E.M. Ringer, Atomistic mechanisms of adhesive contact formation and interfacial processes, Wear, 153 (1), 3–30 (1992).
55. J.A. Harrison, R.J. Colton, C.T. White and D.W. Brenner, Effect of atomic-scale surface roughness on friction - a molecular-dynamics study of diamond surfaces, Wear, 168 (1–2), 127–133 (1993).
56. R. Bassani and M. D'Acunto, Nanotribology: tip-sample wear under adhesive contact, Tribology international, 33, 443–452 (2000).
57. M. D'Acunto, Wear and diffusive processes, Tribology international, 36, 553–558 (2003).
58. M. D'Acunto, Theoretical approach for the quantification of wear mechanisms on the nanoscale, Nanotechnology, 15, 795–801 (2004).
59. S.C. Lim and M.F. Ashby, Wear-mechanism maps, Acta Metallurgica, 35 (1), 1–24 (1987).
60. C.L. Kelchner, S.J. Plimpton and J.C. Hamilton, Dislocation nucleation and defect structure during surface indentation, Physical Review B, 58 (17), 11085–11088 (1998).
61. J.A. Zimmerman, C.L. Kelchner, P.A. Klein, J.C. Hamilton and S.M. Foiles, Surface step effects on nanoindentation Physical Review Letters, 87 (16), paper 165507 (2001).
62. I. Szlufarska, R.K. Kalia, A. Nakano and P. Vashishta, Atomistic mechanisms of amorphization during nanoindentation of SiC: A molecular dynamics study, Physical Review B, 71 (17), paper 174113 (2005).
63. E.T. Lilleodden, J.A. Zimmerman, S.M. Foiles and W.D. Nix, Atomistic simulations of elastic deformation and dislocation nucleation during nanoindentation, Journal of the Mechanics and Physics of Solids, 51 (5), 901–920 (2003).
64. N.A. Fleck and J.W. Hutchinson, A phenomenological theory to for strain gradient effects in plasticity, Journal of the Mechanics and Physics of Solids, 41 (12), 1825–1857 (1993).

65. K.W. McElhaney, J.J. Vlassak and W.D. Nix, Determination of indenter tip geometry and indentation contact area for depth-sensing indentation experiments, Journal of Materials Research, 13 (5), 1300–1306 (1998).
66. Y. Liu and A.H.W. Ngan, Depth dependence of hardness in copper single crystals measured by nanoindentation, Scripta Materialia, 44, 237–241 (2001).
67. N.A. Stelmashenko, M.G. Walls, L.M. Brown and Y.V. Milan, Microindentations on W and Mo oriented single crystals - an STM study, Acta Metallurgica et Materialia, 41 (10), 2855–2865 (1993).
68. Q. Ma and D.R. Clarke, Size dependent hardness of silver single crystals, Journal of Materials Research, 10 (4), 853–863 (1995).
69. A.A. Elmoustafa and D.S. Stone, Indentation size effect in polycrystalline FCC metals, Acta Materialia, 50 (14), 3641–3650 (2002).
70. S. Graça, R. Colaço and R. Vilar, Indentation size effect in laser clad nickel and cobalt, (to be published).
71. S.J. Bull, On the origins and mechanisms of the indentation size effect, Z. Metallkd, 94 (7), 787–792 (2003).
72. H. Li, A. Ghosh, Y.H. Han and R.C. Bradt, The frictional component of the Indentation size effect in low load microhardness testing, Journal of Materials Research, 8 (5), 1028–1032 (1993).
73. J.G. Swadener, A. Misra, R.G. Hoagland and A. Nastasi, A mechanistic description of combined hardening and size effects, Scripta Materialia, 47 (5), 343–348 (2002).
74. H. Gao and Y. Huang, Geometrically necessary dislocation and size-dependent plasticity, Scripta Materialia, 48 113–118 (2003).
75. W.D. Nix and H. Gao, Indentation size effects in crystalline materials: a law for strain gradient plasticity, Journal of Mechanics and Physics of Solids, 46 (3), 411–425 (1998).
76. T.-Y. Zhang and W.-H. Xu, Surface effects on nanoindentation, Journal of Material Research, 17 (7), 1715–1720 (2002).
77. I.L. Jager, Surface free energy - a possible source of error in nanohardness?, Surface Science, 565 (2–3), 173–179 (2004).
78. F.R.N. Nabarro and J.P. Hirth, Dislocations in solids, ed. F.R.N. Nabarro and J.P. Hirth, Vol. Volume 11, Elsevier (2002).
79. M.F. Ashby, The deformation of plastically non-homogeneous alloys, Philosphical Magazine, 21 399–424 (1970).
80. J.F. Nye, Some geometrical relations in dislocated crystals, Acta Metallurgica, 1 (2), 153–162 (1953).
81. N.A. Fleck, G.M. Muller, M.F. Ashby and J.W. Hutchinson, Strain gradient plasticity: theory and experiment, Acta Metallurgica et Materialia, 42 (2), 475–487 (1994).
82. N.A. Fleck and J.W. Hutchinson, Strain gradient plasticity, Advances in Applied Mechanics, 33 295–361 (1997).
83. R.W. Carpick and M. Salmeron, Scratching the surface: fundamental investigations of tribology with atomic force microscopy, Chemical Review, 97, 1163–1194 (1997).
84. B. Bhushan, Nanotribology and nanomechanics, Wear, 259, 1507–1531 (2005).
85. A.G. Khursudov, K. Kato and H. Koide, Wear of the AFM diamond tip sliding against silicon, Wear, 203–204, 22–27 (1997).

86. K.-H. Chung, Y.H. Lee and D.-E. Kim, Characteristics of fracture during the approach process and wear mechanism of a silicon AFM tip, Ultramicroscopy, 102, 161–171 (2005).
87. R. Lüthi, E. Meyer, H. Haefke, L. Howald, W. Gutmannsbauer, M. Guggisberg, M. Bammerlin and H.-J. Güntherodt, Nanotribology: an UHV-SFM study on thin films of C60 and AgBr, Surface Science, 338 (1–3), 247–260 (1995).
88. P.E. Sheehan, The wear kinetics of NaCl under dry nitrogen and at low humidities, Chemical Physics Letters, 410 (1–3), 151–155 (2005).
89. E. Gnecco, R. Bennewitz, A. Socoliuc and E. Meyer, Friction and wear on the atomic scale, Wear, 254, 859–862 (2003).
90. A. Socoliuc, E. Gnecco, R. Bennewitz and E. Meyer, Ripple formation induced in localized abrasion, Physical Review B, 68, paper 115416 (2003).
91. K.-H. Chung and D.-E. Kim, Fundamental investigation of micro wear rate using an atomic force microscope, Tribology Letters, 15 (2), 135–144 (2003).
92. B. Bhushan and A.V. Kulkarni, Effect of normal load on microscale friction measurements, Thin Solid Films, 278 (1–2), 49–56 (1996).
93. J. Hu, X.-D. Xiao, D.F. Ogletree and M. Salmeron, Atomic scale friction and wear of mica, Surface Science, 327 358–370 (1995).
94. S. Miyake, 1 nm deep mechanical processing of muscovite mica by atomic force microscopy, Applied Physics Letters, 67 (20), 2925–2927 (1995).
95. D.D. Woodland and W.N. Unertl, Initial wear in nanometer-scale contacts on polystyrene, Wear, 203–204, 685–691 (1997).
96. S.P. Ho, R.W. Carpick, T. Boland and M. LaBerge, Nanotribology of CoCr-UHMWPE TJR prosthesis using atomic force microscopy, Wear, 253, 1145–1155 (2002).
97. J.S.S. Wong, H.-J. Sue, K.-Y. Zeng, R.K.Y. Li and Y.-W. Mai, Scratch damage of polymers in nanoscale, Acta Materialia, 52 (2), 431–443 (2004).
98. Z.G. Jiang, C.J. Lu, D.B. Bogy and T. Miyamoto, An investigation of the experimental conditions and characteristics of a nano-wear test, Wear, 181, 777–783 (1995).
99. Z.G. Jiang, C.J. Lu, D.B. Bogy and T. Miyamoto, Dependence of nano-friction and nano-wear on loading force for sharp diamond tips sliding on Si, Mn-Zn ferrite and Au, Journal of Tribology - Transactions of the ASME, 117 (2), 328–333 (1995).
100. W. Lu and K. Komvopoulos, Nanomechanical and nanotribological properties of carbon, chromium, and titanium carbide ultrathin films, Journal of Tribology - Transactions of the ASME, 123 (4), 717–724 (2001).
101. J.M. Helt and J.D. Batteas, Wear of mica under aqueous environments: direct observation of defect nucleation by AFM, Langmuir, 21, 633–639 (2005).
102. D.F. Wang and K. Kato, Nano-scale fatigue wear of carbon nitride coatings: part I-wear properties, Journal of Tribology: Transactions of the ASME, 125 430–436 (2003).
103. D.F. Wang and K. Kato, Nano-scale fatigue wear of carbon nitride coatings: part II-wear mechanisms, Journal of Tribology - Transactions of the ASME, 125 437–444 (2003).
104. U. Beerschwinger, T. Albrecht, D. Mathieson, R.L. Reuben, S.J. Yang and M. Taghizadeh, Wear at microscope scales and light loads for MEMS applications, Wear, 181 426–435 (1995).

105. S. Sundararajan and B. Bhushan, Micro/nanotribological studies of polysilicon and SiC films for MEMS applications, Wear, 217 251–261 (1998).
106. A.R. Krauss, O. Auciello, D.M. Gruen, A. Jayatissa, A. Sumant, J. Tucek, D.C. Mancini, N. Moldovan, A. Erdemir, D. Ersoy, M.N. Gardos, H.G. Busmann, E.M. Meyer, M.Q. Ding, Ultrananocrystalline diamond thin films for MEMS and moving mechanical assembly devices, Diamond and Related Materials, 10 (11), 1952–1961 (2001).

22 Single Asperity Nanometer-Scale Studies of Tribochemistry

J.Thomas Dickinson

Department of Physics/Astronomy and Materials Science Program Washington State University, Pullman, WA 99164-2814

22.1 Introduction

There are a number of examples in nature of synergisms (where the net gain is greater than the sum of the parts), e. g., beneficial interactions between two species and favorable combinations of dietary components in living creatures. In everyday life, we appreciate that combining soap and water is a synergism, where water or dry soap alone is ineffective. In our laboratory, we have a long history of examining synergisms involving the combination of radiation and mechanical or thermal treatment, [1–5], radiation and chemical exposure [4,6] and (the topic of this review) mechanical stress and chemical exposure. In relation to desirable (e. g., polishing) or undesirable (wear) removing of material, simultaneous application of tribological and chemical stimuli are often much more effective than either stimulus alone and relate to the general areas of tribo- and mechano-chemistry. These effects can be exploited to produce important technological surfaces as in polishing (usually implying large quantities of material removed – e. g., polishing a lens or initial polishing of Si wafers) or extremely flat surfaces via planarization (usually a process that removes only surface topologies, tending to smooth and flatten the surface – e. g., the high tolerance planarization at several intermediate steps in the fabrication of semiconductor chips). These important processes are generally referred to as "CMP" which we combine to mean: chemical-mechanical polishing/planarization. Conversely, corrosive wear, environmentally induced crack growth, and related phenomena can dramatically shorten useful device lifetimes. Atomic force microscopy (AFM) is a particularly valuable tool for the study of these mechanical/chemical synergisms, being able to (i) localize the tribological stimulation near and at a surface and (ii) to image the resulting wear with nm-scale resolution and in situ [7–9]. In many respects, the AFM tip can simulate a single asperity or abrasive particle interacting with nearly ideal substrates.

Material removal is often a relatively inefficient means of obtaining a smooth surface. Energetically, filling in small, shallow pits is much more effective. An AFM tip can also be used to nucleate, accelerate, and control deposition along step edges on certain single crystal surfaces exposed simultaneously to solutions favoring crystal growth, i. e., in supersaturated aqueous solutions. We present strong evidence that the tip sweeps adsorbed ions to

nucleation sites [10, 11]. The resulting deposition can be localized, again on the nanometer size scale, providing a unique means of generating nanometer scale structures, growing atomically flat surfaces, and new methods of controlling biomineralization.

The removal of submicron particles poses a severe challenge in the production of optical components (mirrors and lenses) and high density integrated circuits. Adhesion of metal particles to semiconductor substrates is a significant issue, for instance [12–14]. Whole technologies for particle removal have been developed, including laser assisted particle removal [15,16]. In integrated circuit manufacture, CMP often serves this function [17]. These processes typically employ both a liquid phase to reduce adhesion and a mechanical or thermal shock stimulus to actually remove the particle. However, the mechanisms of adhesive force reduction and the mechanical details of removal are not well understood. By applying measurable stresses to the particle in situ, a much more detailed picture of the release mechanism can be deduced [18].

An interesting question arises when we consider polymer substrates exposed to solvents and mechanical stimulation. Initial studies were aimed at stimulating local dissolution of the polymer. For instance, the response of the polymer (e. g., PMMA) is strongly dependent on the substrate molecular weight, often resulting in enhanced local swelling of the polymer rather than dissolution [19, 20].

In the above work, the AFM tip is a tool to modify various surfaces. Clearly one expects the tip to be modified in a similar fashion under appropriate conditions. In the role of a model asperity, the tip experiences "asperity wear". We show that the wear of silicon nitride AFM tips is dominated by chemical-mechanical effects involving not only the solution but also the chemical nature of the substrate.

Combined chemical and mechanical attack is most effective for material removal and other forms of materials modification in a wide variety of contexts. Together, they provide an especially effective "**one-two punch**" to surfaces that can be exploited to produce desired structural features or, conversely, atomically flat surfaces. Applications to the formation of sub-micron structures for micro/nano-fluidic, sensor, MEMS (Micro-Electro-Mechanical Systems) and NEMS (Nano-Electro-Mechanical Systems) devices are high interest areas of possible application.

22.2 Underlying Mechanisms for Substrate Material Removal

In the case of tribologically loaded surfaces interacting with one another in an active chemical solution, the interactions leading to material removal involve asperities on both surfaces and/or carefully selected abrasive particles (usually embedded at least temporarily in a soft polymer pad) and the substrate. If stress is a critical component of the process, the asperities and/or particles

are clearly the origin of the highest stresses and are therefore expected to play a critical role. We wish to distinguish and partition substrate materials and chemical conditions into two extreme cases:

(1) Substrates that tend to dissolve in the surrounding solvent;
(2) Substrates that tend to form reaction layers (e. g., softer and/or chemically protective).

Examples of (1) are basically all crystals that can be *grown* from solution (e. g., aqueous based solutions: halides, nitrates, carbonates, phosphates, sulfates, borates, oxalates, chromates, and possibly molybdates, tungstates, vanadates, niobates, and tantalates; organic solvent based solutions: these are so extensive that we simply list organic salts, organometallic complexes, biomolecules (including proteins, viruses, and nucleic acids, and some crystalline polymers). When such crystals are near equilibrium, i. e., the solution chemistry is near saturation, they will tend to have a susceptibility to the presence local stress (this is a general statement that we have *not* proven except for a few cases!). For the few cases studied we will try to show how this occurs and under what conditions.

Examples of (2) are also fairly numerous, particularly in technologically interesting materials. Consider the polishing, wear, and corrosion of metals such as Al and Cu. Generally, a pH is chosen which promotes the formation of a chemical layer (e. g., an oxide or hydroxide) that is removed through tribological stimulus. In a number of cases (2a), the reaction layer is softer than the wear particle (or asperity) and is therefore plastically deformed and readily removed [21]. Another variation related to (2), in terms of material removal mechanisms (2b) follows when a more aggressive attack of the substrate by the solution occurs when a protective layer is abrasively removed. This (2b) mechanism will be discussed in more detail when AFM studies of the etching of Si (100) are examined in high pH NaOH solutions are described below.

22.3 Examples of Single Asperity Tribochemical Wear

In the next few sections we present evidence of these types (22.3.1 and 22.3.2) of material removal mechanisms and their consequences when nanometer scale stimulation is employed.

22.3.1 Dissolution along Monolayer Steps in Calcite

Calcite is readily imaged by AFM in aqueous solutions [22]. The dissolution and growth of calcite crystals in aqueous solution has been previously studied by AFM [23–30] and by other means [31–35], due to its importance in mineral formation, global CO_2 exchange, and strong surface interactions with heavy

metals in the environment. We have shown that dramatic tribochemical wear can be induced by scanning the AFM tip back and forth across the edge of a naturally occurring etch pit at high contact forces [8]. Indenting the surface with an AFM tip near the edge of a etch pit also locally enhances dissolution along steps near the tip. We attribute these effects to increased rates of double kink nucleation in the strain field of the AFM tip [8].

Etch pits on calcite typically form parallelograms bounded by two pairs of crystallographically distinct steps. During steady state dissolution, one pair moves much more rapidly than the other [27]. We designate these steps "fast steps" and "slow steps", respectively. Physically, the difference between fast and slow steps is related to the inclination of the plane of the carbonate ions relative to the sample surface – a steric effect. The rate limiting step in pit growth is believed to be double kink nucleation along the pit edges [27]; the weakly bound ions at the resulting kinks are readily incorporated into solution, resulting in rapid kink motion to the corners of the pit. Fast steps are more vulnerable to dissolution than slow steps because the CO_3^{2-} ions along fast steps are more exposed to the surrounding water. This lowers the activation energy for double kink nucleation (ion removal) on fast steps. A similar model was employed by Hirth and Pound to describe evaporation from crystal surfaces [36].

Drawing the AFM tip back and forth across the edge of a monolayer deep etch pit creates a wear pattern that can be directly interpreted in terms of double kink nucleation. The geometry of the experiment and resulting features are shown in Fig. 22.1. The path of the AFM tip during wear is marked by the white line. The maximum step displacement in the scanned region, Δx, was measured from the end of the "wear track" to the line defined by the unperturbed portions of the original step (as far as possible from the wear track). Similar treatments on flat terraces (away from steps) had no detectable effect on the surface. Likewise, repeated linear scans across cleavage steps on dry calcite surfaces (in ambient air) had no effect when imaged on this scale. This synergism between the chemical environment *and* mechanical loading clearly marks this as an example of tribochemical wear. Essentially no tip wear is seen in these experiments, most likely because the chemistry involved does not lead to attack of the Si_3N_4 tip material. This is in sharp contrast with silica based substrates (discussed below) that lead to significant tip wear.

The geometry of the "wear track" in Fig. 22.1 illustrates two important features of molecular scale dissolution. The jogs along the edges of the wear track reflect the tendency of atomic scale kinks to aggregate into larger structures. Further, the two edges of the wear track show significantly different patterns of jogs, reflecting the different propagation behavior of crystallographically distinct kinks along these steps. These differences are also responsible for the marked contrast in dissolution along the two halves of the original step. Dissolution along the portion of the original step to the left of the wear track (accomplished by kinks moving in the $+\langle 100 \rangle$ direction) has

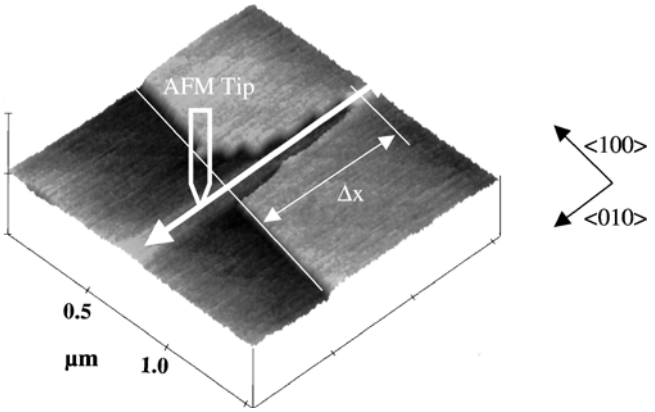


Fig. 22.1. AFM image of a slow step on calcite after 3800 linear scans at a normal force of $F_{\mathrm{N}} = 160\,\mathrm{nN}$

been much more rapid than dissolution along the portion of the original step to the right of the wear track (accomplished by kinks moving in the $-\langle 100\rangle$ direction). Analysis of material dissolution after wear track formation suggests that the jogs produced by wear are especially vulnerable to dissolution, and play an important role in the planarization of stepped material.

This experimental geometry was chosen because the length of the wear track (Δx in Fig. 22.1) is quantitatively related to the rate of double kink nucleation where the AFM tip crosses the step. Further, the wear track growth rate ($\Delta x/\Delta t$) is a strong function of contact force. Plots of growth rate vs. contact force for two solution concentrations are shown in Fig. 22.2. These measurements employed $500 \times 500\,\mathrm{nm}^2$ scans acquired at a scan rate of 24 Hz (tip velocity of $12\,\mu\mathrm{m/s}$). The dark lines in Fig. 22.2 represent a least squares fit of the data to a model function described below (an exponential function of the surface stress). Growth rate measurements taken with different cantilevers show a high degree of consistency, as indicated by the data points represented by different symbols in Fig. 22.2a. Each symbol represents measurements made on different days, with different calcite samples (from the same block) and cantilevers (from the same wafer).

Wear tracks across fast steps grow much faster than wear tracks across slow steps. At the low solution concentration employed in Fig. 22.2a ($60\,\mu\mathrm{M}$ solution flowing at $\sim 10\,\mu\mathrm{L/s}$), wear tracks across fast steps grew too quickly for practical measurements. We therefore reduced the growth velocity by providing a more saturated solution, where a certain fraction of the nucleated double kinks are annihilated by redeposition of material from solution. Even at these higher solution concentrations, the AFM tip effectively mixes the nearby solution, preventing the development of concentration gradients that would complicate analysis.

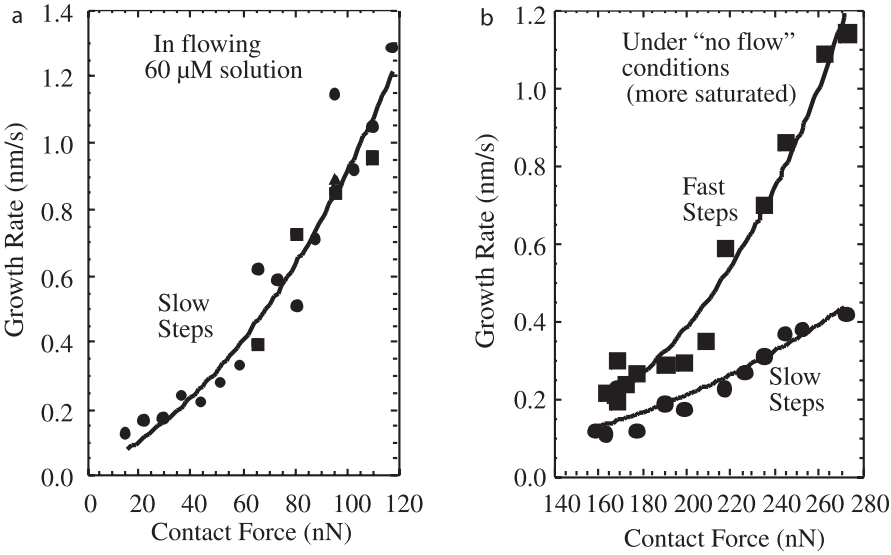


Fig. 22.2. Wear track growth rate versus contact force **a** under flowing, 60 μM solution and **b** in more saturated solution (obtained by turning off flow). The data points represented by different symbols were made on different calcite samples with different cantilevers. The dark line is a least squares fit of Eq. (22.2) (described below) to the data

The resulting wear track growth rates are shown in Fig. 22.2b, where a high degree of saturation was provided by stopping the flow of solution for about 30 minutes. Under these conditions, a contact force of 270 nN is required to produce growth rates on slow steps comparable to 70 nN in flowing solution. Nevertheless, the strong dependence of growth rate on contact force confirms that reaction limited conditions prevail – that is, wear track growth is not limited by concentration gradients. Over most of the range of contact forces probed in Fig. 22.2b, wear tracks across fast steps grew at least twice as fast as wear tracks across slow steps. Thus fast steps are considerably more vulnerable to tribologically enhanced dissolution than slow steps.

We have observed similar, strong enhancements in dissolution during 2D AFM imaging [8]. However, 2D scanning nucleates kinks at many points along the step; mutual annihilation of kinks nucleated at different points along the step make it difficult to infer the rate of kink nucleation from the step velocity. In contrast, kinks nucleated by linear scanning are formed along a narrow portion of the step and propagate away from this point in opposite directions. At the highest contact forces employed in this work, the average number of double kinks nucleated per scan is about 0.2, which corresponds to about 5 per second. This allows sufficient time for kinks nucleated along one row of ions to propagate away from the line scan before the next kink

is nucleated, so that the next kink is nucleated along a new row. Then the total number of nucleated kinks can be estimated by dividing the length of the wear track, Δx, by the distance between ionic rows (3.2 Å).

Despite the high stresses applied by the AFM tip, we see no evidence for plastic deformation in this work or in AFM tip indentation experiments [8]. (We cannot rule out deformation-related features that do not survive long enough to be imaged – at least a few seconds.) The lowest contact forces employed in this work (15 nN) correspond to average compressive stresses of about 2 GPa; Vickers indentation (which employs mm-sized tips) at these stresses would produce an indent some tens of microns across [37]. We attribute the absence of AFM-induced deformation to the small size of the AFM tip (tip radius $\sim$ 40 nm). Deformation is strongly hindered when the indentor is smaller than a typical slip band (usually $\sim 1\,\mu$m) [38]. Deformation is also hindered at high strain rates [39], such as those associated with the motion of the tip across the surface. The absence of a threshold stress for the onset of enhanced dissolution also argues against dislocation emission and twinning as possible sources of double kinks. We conclude that plastic deformation does not contribute to mechanically enhanced dissolution under these conditions.

In terms of the available mechanical energy, the operation of mechanical effects is striking. At the highest contact forces employed (assuming elastic interactions only), the total work done by the tip on the substrate is less than 200 eV – and this mechanical energy is distributed over thousands of bonds. Further, double kink nucleation is not likely under the center of the tip, where the stresses are highest, because the compressive stresses there hinder the escape of solvated ions. The (albeit weaker) tensile stresses along the surface surrounding the AFM tip are much more likely to promote dissolution. Stress effects are strongly enhanced along steps, where surface ions are less constrained by surrounding ions. Ion displacement can play a key role in volume-activated processes, where the work done (force $\times$ displacement) can directly reduce the binding energy.

The magnitude of the stresses adjacent to the AFM tip can be estimated for the case of isotropic, elastic behavior. The maximum tensile stress involves the radial component, σ_r, along the circle where the AFM tip contacts the substrate. Double kink nucleation will be enhanced over a modest range of distances from edge of tip contact (a few tip radii, from Saint Venant's principle), so that the relevant average stress will be somewhat lower. This suggests that mechanical enhancements are insensitive to variations in the geometry of the AFM tip itself, but also introduces some uncertainty in the value of the (average) stress responsible for enhanced dissolution. The peak tensile stress given by the Hertz relation for an infinitely stiff, spherical tip is: [40]

$$\sigma_{\mathrm{r}} = \frac{(1-2\nu)}{\pi}\left(\frac{2F_{\mathrm{N}}E^2}{9(1-\nu^2)^2 r^2}\right)^{1/3}, \tag{22.1}$$

where E and ν are Young's modulus and Poisson's ratio, respectively, for the substrate; F_{N} is normal force between the sphere and substrate; and r is the sphere radius. σ_r is relatively insensitive to errors in the contact force ($F_{\mathrm{N}}^{1/3}$ dependence), and somewhat more sensitive to errors in the tip radius ($r^{-2/3}$). To avoid the complications of material anisotropy, we use technical moduli (directionally averaged values) appropriate for $CaCO_3$: $E = 81\,\mathrm{GPa}, \nu = 0.32$ [41]. For example, a tip radius of curvature $r = 40\,\mathrm{nm}$, and an applied normal force of 100 nN yields a maximum radial stress of about 560 MPa.

The dependence of the wear track growth rate (expressed as a velocity, V) on contact force is readily modeled with a Zhurkov-Arrhenius expression [42]

$$V = V_o \exp\left[-\left(\frac{E_{\mathrm{act}} - v^*\sigma}{kT}\right)\right] = V_o' \exp\left(\frac{v^*\sigma}{kT}\right) , \tag{22.2}$$

where V_o is the appropriate pre-exponential, E_{act} is the zero stress activation energy for double kink nucleation and v^* is an activation volume. The best fit curve of Eq. (22.2) to the data of Fig. 22.2a (flowing solution), using the stresses given by Eq. (22.1), is shown by the dark line in Fig. 22.2. The best fit values of the parameters correspond to $V_o' \approx 6 \pm 3\,\mathrm{pm/s}$ [where $V_o' = V_o \exp(-E_{\mathrm{act}}/kT)$] and $v^* \approx 3.7 \pm 0.3 \cdot 10^{-29}\,\mathrm{m}^3$, respectively. This activation volume is slightly larger than the average volume per ion in the $CaCO_3$ lattice ($3.3 \cdot 10^{-29}\,\mathrm{m}^3$), making it reasonable to suppose that this activation volume corresponds to the displacement of one or perhaps two ions from a step site. (The CO_3^{2-} ion is considerably larger than the Ca^{2+} ion.) We note that the displacement of ions in flat terrace sites is limited by the surrounding ions; this may explain why scratching does not nucleate new etch pits over the range of contact forces used here. The large size of the CO_3^{2-} ion and the fact that a component C–O bond is directed into the solution would make it especially vulnerable to combined mechanical and chemical effects.

A similar analysis of the wear track growth rates at slow steps under "no flow" conditions (Fig. 22.2b) yields $v^* \approx 3.9 \pm 0.5 \cdot 10^{-29}\,\mathrm{m}^3$. The agreement with v^* under flowing solution is well within the numerical uncertainty of the curve fitting procedure. This agreement is consistent with the expectation that the degree of saturation controls the lifetime of nucleated kinks and not the (stress-enhanced) nucleation rate. The best-fit value for v^* for dissolution along fast steps is somewhat larger, $v^* \approx 6.0 \pm 0.5 \cdot 10^{-29}\,\mathrm{m}^3$. A higher activation volume is consistent with the reduced steric constraint experienced by CO_3^{2-} ions along fast steps, which would allow for larger displacements at a given stress and render them more vulnerable to water attack.

An estimate of the activation energy for double kink nucleation at zero stress can be derived from Eq. (22.2) assuming that V_o corresponds to a typical "attempt frequency", f_o, multiplied by the step displacement per successful attempt (one lattice spacing: 3.2 Å). Setting f_o equal to typical vibrational frequencies ($10^{13}\,\mathrm{s}^{-1}$), the measured pre-exponential (V_o' from Fig. 22.2a) at

$T = 293\,\mathrm{K}$ corresponds to an activation energy of $0.8 \pm 0.2\,\mathrm{eV}$ – a plausible value for reaction limited nucleation. This analysis neglects the effects of possible redeposition reactions due to finite kink lifetimes. [Redeposition under "no flow" conditions is strongly affected by kink lifetimes. Therefore, this calculation uses the data taken in flowing solution (Fig. 22.2a).] At the highest stresses employed in this work (corresponding to a normal force of 270 nN), this activation energy for double kink nucleation is reduced by about 0.2 eV.

Thus, it is reasonable to interpret the observed step growth in terms of a thermally activated process. In addition, the Zhurkov-Arrhenius description of the observed stress dependence of the wear track growth also appears to be valid. We point out that crack velocities in environmental crack growth [43] often display a similar dependence on stress intensity, although in this case the relation between stress intensity and stress at the crack tip can be very difficult to deduce.

22.3.2 Dissolution along Monolayer Steps in Brushite

The calcium phosphates are important biological minerals, occurring in both normal (e. g., enamel, dentine, bone) and pathological (e. g., dental cavities, kidney stones) calcifications. In weak acid solution (pH 4–5), equilibrium conditions favor the formation of brushite (monoclinic calcium hydrogen phosphate dihydrate – $CaHPO_4 \cdot 2H_2O$). Since urine often displays a pH in this range, kidney stones are often formed of this material. Brushite is also used as an abrasive in toothpaste, and is an intermediate in phosphate fertilizer production. Stress-induced surface modifications in brushite therefore have potentially important medical and commercial implications. It also provides another model inorganic material for the study of planarization mechanisms, where we can generate atomically smooth surfaces, of interest to the important technological area of CMP.

Solution grown single crystal brushite forms crystalline [44] plates with broad surfaces [45]. In undersaturated solutions, these surfaces develop triangular etch pits with three crystallographically distinct steps bounded by steps along the [101], [201], and $\sim$ [001] directions [45]. These directions correspond to rows of strong ionic bonds with contrasting bonding environments. This variability makes brushite a valuable system for testing models of corrosion and tribochemical wear.

Step velocity measurements made at very low contact forces reflect the step stabilities against dissolution [46]. These measurements indicate that the [201] steps are most vulnerable to dissolution and that the [001] steps are significantly less vulnerable. No dissolution along [101] steps was observed on times scales of tens of minutes. The [101] steps are extremely resistant to dissolution at low contact forces. Like calcite, essentially no tip wear was observed for any of the experiments described here.

AFM images obtained immediately after high contact force linear scans across two different monolayer etch pits on brushite are shown in Fig. 22.3.

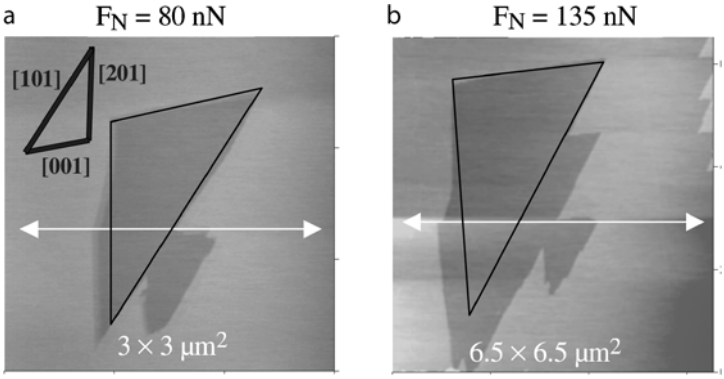


Fig. 22.3. AFM images of triangular etch pits on brushite immediately after linear scanning along the white line to form wear tracks. The orientation of the steps along the edges of the pit are indicated in the inset of **a**. The black outline shows approximate location of the original etch pit. Note the contrasting sizes of the images in **a** and **b**

As in the case of calcite, crystallographically distinct steps show contrasting resistance to wear. From Fig. 22.3, wear across [101] steps is much faster than wear across [201] steps. Similar measurements on [001] steps show much less wear. In brushite, the steps which are most vulnerable to dissolution ([201]) are not the most vulnerable to wear. This contrasts with calcite, where fast steps are more vulnerable than slow steps in both dissolution and wear experiments.

Like calcite, the two edges of the wear tracks on brushite dissolve at different rates. The situation in brushite is especially severe, however. Wear tracks aligned along [001] directions (as in Fig. 22.3) show especially dramatic contrasts: one edge of the wear track (designated the +[001]) is quite stable, while the other literally falls apart into segments made up of the other two stable steps. Significant differences in dissolution are also observed along the original steps adjacent to the wear track. For instance, much more material has been removed from the [201] steps in Fig. 22.3 below the wear track than from this same step above the wear track. We attribute this to the different propagation velocities of crystallographically distinct kinks along these steps.

Wear track growth for brushite can be analyzed in much the same fashion as for calcite. The wear track growth rates for all three distinct steps are plotted as a function of contact force in Fig. 22.4a. At all contact forces yielding measurable wear tracks, the wear track growth rates were highest along the [101] and lowest along the [001] steps. As noted above, [201] steps are more vulnerable to zero force dissolution than [101] steps. Indeed, when the model curves in Fig. 22.4a are extrapolated to zero, the curve for [101] steps drops below the curve for [201] steps. In Eq. (22.2), this corresponds to a lower prefactor for [101] steps; as the contact force is raised, the higher

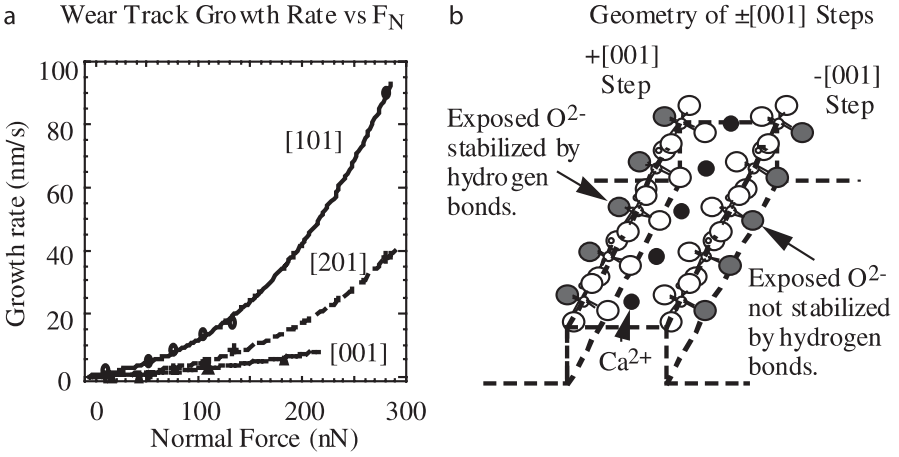


Fig. 22.4. a Wear track growth rate vs. contact force in undersaturated solution for [101], [201] and [001] steps on single crystal brushite. **b** Geometry of ±[001] steps, showing how the +[001] steps can be stabilized by hydrogen bonds

v* for the [101] steps ensures that the [101] growth rate overtakes the [201] growth rate.

The remarkable contrast between the stabilities of +[001] and –[001] steps may account for the presence of triangular etch pits in this material. Most ionic materials (including gypsum, $CaSO_4 \cdot 2H_2O$, which has an analogous structure yet lacks the HPO_4^{2-} hydrogen bonds), display four-sided etch pits, composed of the two most stable steps. The triangular steps in brushite would be accounted for if the step stabilities were ranked according to

$$+[001] \gg \pm[201]\,, \quad \pm[101] \gg -[001]\,. \tag{22.3}$$

We attribute the exceptional stability of +[001] steps relative to −[001] steps to the effect of the HPO_4^{2-} hydrogen bond. The [001] steps run parallel to the HPO_4^{2-} rows that make up the anion sublattice. Assuming that the HPO_4^{2-} ions lie along the step edges (which is physically reasonable), +/−[001] steps would have the structure shown in Fig. 22.4b. This orientation was confirmed by high resolution AFM images, which exploit the relative positions of the uppermost HPO_4^{2-} oxygen ions to positively orient the lattice relative to the observed steps. The hydrogen bonds are localized on the +[001] side of the HPO_4^{2-} ions, reducing the effective negative charge on the $+c$ side of these ions. This reduced charge would weaken the water-anion interaction. Since the $+c$ sides of the HPO_4^{2-} ions face the solvent on +[001] steps, hydrogen bonding would stabilize these steps. Conversely, the −[001] side of the HPO_4^{2-} ions lacks the hydrogen bond, so that −[001] steps present greater negative charge to the solvent and thus interact more strongly with water. Neutron diffraction studies confirm the presence of long range order in

Table 22.1. Step Stability in the $CaHPO_4$ Sheets of Brushite [44] Planes

Step	Bond Strength Bond/Å^2	Missing Bonds Per Ion	V_{step} (zero stress) nm/s	Activation Volume (Å^3)
[101]	0.100	1.0	< 0.1	62
[201]	0.121	1.5	1.5	48
[001]	0.067	1.5	0.5	41

the hydrogen bonding in brushite [47], as required if hydrogen bonding were to account for a consistent difference in step stability.

A semi-quantitative comparison between the expected and observed step stabilities is presented in Table 22.1. The effective bond strength was estimated by assuming that all ionic bonds in the lattice have the same strength, and counting the number of bonds per unit area along each of the observed step directions. This parameter correlates with the step dissolution rates on many ionic surfaces. For brushite, the correlation between bond strength and the zero-stress (low contact force) dissolution rates is poor. A much better correlation is obtained with the number of "missing bonds" per ion – that is, the number of ionic bonds per ion which have been disrupted by creating the step. Missing bonds reflect uncompensated charge (similar to dangling bonds on covalent materials), which renders a step vulnerable to attack by water. This parameter accounts for the observed step stability if we also factor in the stabilizing effect of hydrogen bonds on the observed +[001] steps. On more symmetric ionic crystals, each of the observed steps usually display the same number of missing bonds per ion. On such materials, second-order effects, such as effective bond strength, should dominate the stability ranking.

The effective bond strength also correlates poorly with the activation volumes determined by fitting the wear track growth rate data of Fig. 22.4a to Eq. (22.2). (Steps with high activation volumes are more sensitive to stress effects.) In the case of calcite, the difference in activation volume appears to be principally steric in nature: the CO_3^{2-} ions along one pair of steps are freer to move in response to an applied stress. In brushite, however, the most vulnerable, [101], steps are also the most constrained, with five strong ionic bonds holding each HPO_4^{2-} ion in place.

The most likely explanation for the variation of activation volume among the three brushite steps relates to the elastic anisotropy of the lattice. Because the AFM tip imposes a roughly uniform displacement on the surface ions, the mechanical energy delivered to the various bonds will vary in proportion to the bond stiffness. In this sense, the surface is analogous to a 2D array of springs (bed springs), with stiff springs along one direction and soft springs in the other. Deforming the surface stretches both sets of springs, but the stiff springs resist the deformation more, resulting higher stresses along the stiff direction. Although Eq. (22.2) assumes isotropic material behavior, this

difference can be numerically accounted for by assigning higher activation volumes to steps lying along stiff directions.

The brushite lattice displays rows of strong, ionic, double bonds along the [101] direction, suggesting that the lattice stiffness is indeed maximum along the [101] direction ($v^* \sim 62\,\text{Å}^3$). The activation volumes for the other two steps are lower in rough proportion to their inclination to the [101] direction. The [201] steps intersect the [101] steps at a 31° angle ($v^* \sim 48\,\text{Å}^3$), while the [001] steps intersect the [101] steps at a 51° angle ($v^* \sim 41\,\text{Å}^3$). Ignoring steric effects, a 35% variation in principal tensile stress as one moves from the [101] to the [001] direction would account for the observed variation in activation volume. This effect would not apply to calcite, where the lattice stiffness along the two observed steps (fast and slow) are equal.

The case of brushite illuminates two factors that strongly affect the material response to combined chemical and mechanical attack: mechanical stiffness and the details of bonding, including possible hydrogen bonds. The importance of these factors is easily overlooked in more homogenous materials.

22.3.3 Si_3N_4 Tip Wear During Scanning on Reactive Surfaces

Silicon nitride ceramics are hard, inert, and stable at high temperatures, making them attractive for use in extreme environments. However, conventional polishing and grinding operations often produce surface defects and cracks that lead to premature component failure. Tribochemical or chemical-mechanical polishing [48] can produce very smooth, defect-free, silicon nitride surfaces, but this process is not well understood. Hydrodynamic lubrication, which results in low friction and almost no wear, is observed with sufficiently smooth silicon nitride surfaces [49–51] and has potential applications in high performance ceramic bearings. Wear of silicon nitride and silicon oxide is also of great interest in chemical-mechanical polishing of semiconductors for integrated circuits.

Geometrically, AFM wear experiments are analogous to the traditional pin-on-disk or ball-on-disk wear experiments [52–54]. Wear of silicon nitride AFM tips has been previously studied [55,56] and attributed to adhesive wear, plastic deformation, and low cycle fatigue fracture. Wear of silicon AFM tips can depend on the ambient solution [57, 58], suggesting that tribochemical effects can be important.

In these experiments, tip wear was induced by scanning $3 \times 3\,\mu\text{m}^2$ patterns in raster mode at a tip velocity of $16\,\mu\text{m/s}$ on a wide variety of substrates. These scanning conditions correspond to a sliding distance of 1.5 mm per scan; the largest sliding distance reported here is 67.5 mm. Due to the important role of the substrate in tip wear, and because it is nearly impossible to reposition a tip on exactly the same portion of the surface after a wear measurement, each wear measurement required a new AFM tip and a previously unscanned portion of the substrate. The tip shape before and after each wear experiment was characterized by scanning sharp silicon spikes $\sim 600\,\text{nm}$ tall

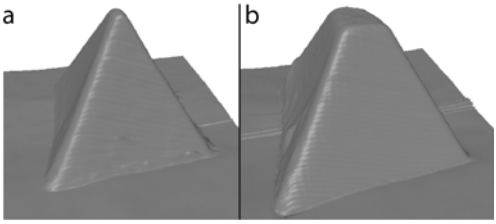


Fig. 22.5. AFM images of **a** a fresh AFM tip, and **b** the same tip after wear. The AFM tip was scanned 45 scans times across a sodium trisilicate glass substrate in ammonium hydroxide solution (pH $\sim$ 11) at an applied force of 125 nN. Each image has been slightly cropped from the original $1500 \times 1500\,\mathrm{nm}^2$ image. The vertical (z) axis has the same scale as the x- and y-axes

with a tip radius of curvature of $< 10\,\mathrm{nm}$ (MikroMasch, TGT01). Because the spikes are significantly sharper than the AFM tip, imaging the spikes produces an image of the AFM tip. Typical images before and after a wear experiment are shown in Fig. 22.5. These tip images could then be analyzed to obtain wear height and area of the end of the worn tip.

The effects of scan duration (T) (which is proportional to the sliding distance) and contact or normal force (F_N) were explored using sodium trisilicate glass ($Na_2O \cdot 3SiO_2$) substrates and ammonium hydroxide solution (pH $\sim$ 11). The progression of wear was observed by measuring the amount of material removed from a series of AFM tips, where each tip was treated with a different number of scans at the same contact force ($\sim 120\,\mathrm{nN}$). The scan size and tip velocity were chosen so that each scan involved 1.5 mm of total tip travel distance along the substrate in 50 sec. The total change in tip length, h, for each new tip is plotted as a function of the number of scans in Fig. 22.6a. In terms of instantaneous depth of tip wear, tip wear is initially rapid and gradually slows. Because of the pyramidal tip shape, the area of the tip is increasing with time as wear occurs, thus the applied stress depends on time (or scan distance). The dark line shows a simultaneous least squares fit of a power law-expression (described below).

Another set of measurements were undertaken as a function of contact force. Again using a fresh tip for each measurement, 15 scans were performed on a sodium trisilicate glass substrate at the contact force of interest, followed by depth of wear characterization. The resulting change in tip height as a function of normal force is displayed in Fig. 22.6b. The dark lines in Fig. 22.6a and b show the result of a least squares fit of the form $(F_\mathrm{N} \times \mathrm{time})^M$ to both data sets simultaneously. The best fit yielded a functional dependence of $(F_\mathrm{N} \times \mathrm{time})^{0.51}$. No significant changes in tip wear with pH were observed.

When compared to macroscopic pin-on-disk wear of silicon nitride in similar solutions, the volume removal rates represented in Fig. 22.6 are high. For comparison with conventional wear measurements, we convert the height data of Fig. 22.6a to a volume removal rate per unit normal force per unit

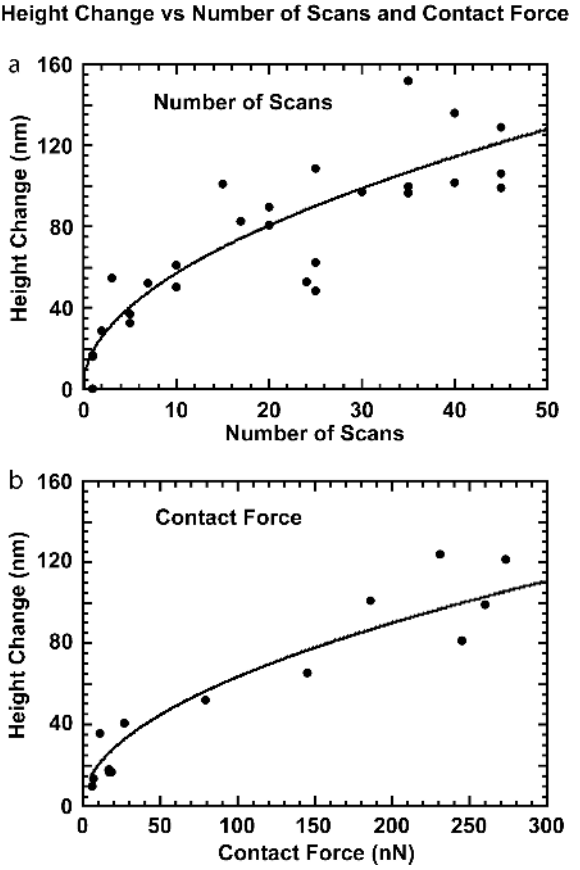


Fig. 22.6. Height change in AFM tip **a** as a function of the number of scans on trisilicate glass in ammonium hydroxide (pH $\sim$ 11) at a contact force of $\sim$ 120 nN, and **b** as a function of contact force over 15 scans. Each scan was $3000 \times 3000\,\mathrm{nm}^2$ at 5.2 Hz. Each tip wear measurement was undertaken with a fresh AFM tip. *Circles* – experimental data. *Line* – least squares fit of Eq. (22.1) to *both* sets of data simultaneously

sliding distance. The slope of the wear rate data in Fig. 22.6a between 5 and 20 scans corresponds to a volume removal rate of $1 \cdot 10^{-13}\,\mathrm{m}^3/\mathrm{Nm}$. This is an order of magnitude higher than the highest volume removal rates observed by Chen et al. [59] and Fischer et al. [60,61]. This may in part be due to the exceptionally high stresses at the AFM tip contact (initially > 1 GPa) and the single asperity nature of contact, as discussed below.

The tip wear produced during 25 scans across a variety of substrates in distilled water (pH = 7) at a contact force of $\sim$ 120 nN is shown in Fig. 22.7. Three tip wear measurements were made for each sample, starting with a fresh tip each time. Tip wear for five of the substrates was also measured in NH_4OH solution (pH $\sim$ 11, data not shown); in all cases tip wear at pH = 11 was almost identical to tip wear at pH = 7.

All of the substrates producing significant tip wear (i. e., quartz, fused silica, zirconia, titania, sodium trisilicate glass, polished soda lime glass, and silicon nitride) form modified hydroxide layers in aqueous solution. Twenty-five scans on these materials at a contact force of 120 nN typically removed

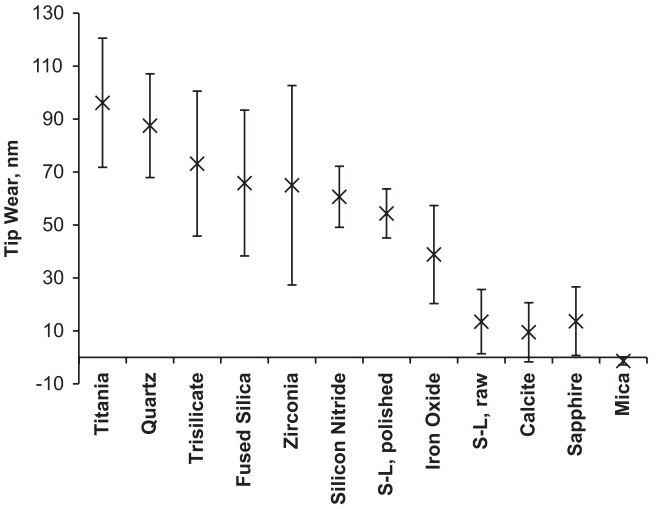


Fig. 22.7. Amount of tip wear on various substrates in water produced by $25,3000 \times 3000\,\mathrm{nm}^2$ scans at 5.2 Hz and a contact force of $\sim 120\,\mathrm{nN}$. A fresh tip was employed for each set of scans. Plot shows average and standard deviation for three measurements on each substrate material. "S-L" refers to soda-lime glass

$50-100\,\mathrm{nm}$ of material from the tip. This wear is sufficient to produce a flat area of about $120 \times 120\,\mathrm{nm}^2$ at the end of the tip. Images of these tips showed that for all of these substrates this flat area was quite smooth, as in Fig. 22.5b, with no detectable signs of roughening.

Conversely, substrates which do not form significant hydroxide layers in aqueous solution (calcite and mica) produced little tip wear. Typical wear during 25 scans of these substrates was $0-15\,\mathrm{nm}$, with typical uncertainties less than $\pm 2\,\mathrm{nm}$. Although the mica structure possesses Si–OH bonds, these bonds are not exposed to the surface by cleavage. The Si–OH bonds on adjacent layers are linked by polyvalent cations, and cleavage occurs between surfaces lacking these bonds [62, 63]. The lack of significant tip wear during scanning on these surfaces is evidence that hydroxide bonds are needed. Note that hardness of the substrate has no influence on tip wear.

Given the role of water in hydroxide formation, the effect of surface hydroxide formation was confirmed by performing wear measurements under ethyl acetate, which eliminates the presence of water. Neither the tip nor the sodium trisilicate glass substrates were given any special treatment to eliminate hydroxyl groups. The resulting tip wear was insignificant ($10 \pm 8\,\mathrm{nm}$ – not shown on plot) at $F_\mathrm{N} = 120\,\mathrm{nN}$; this wear is similar to the (lack of) tip wear produced by scanning calcite in water. Similarly, no tip wear is seen under ethyl acetate on a fused silica substrate. The exclusion of water prevents the replacement of hydroxide bonds consumed in initial wear-related reactions with the AFM tip. It is well known that glass substrate wear rates during

polishing operations drop dramatically when water is excluded, presumably for similar reasons [64, 65].

We also have noticed a very interesting result, namely that lateral force measurements (relative values only) such as those shown in Fig. 22.8, indicate a small but very reproducible reduction in friction ($\sim 10\%$) during the course of a typical wear experiment. Since the contact area, A, is increasing, this shows a frictional decrease as A increases. Several AFM studies have shown the contact area dependence to be of the form:

$$F_{\mathrm{L}} = \tau A \tag{22.4}$$

where F_L is the lateral force, τ is the interfacial shear strength (frictional force/area) [66–68]. The small magnitude in the decrease of F_L and the significant wear rates that prevail throughout the duration of scanning in this work argue strongly against any transition to hydrodynamic lubrication. The high stresses under the tip, even at our largest contact areas ($\sim 2.5 \cdot 10^{-14}\,\mathrm{m}^2$), should prevent the formation of a continuous fluid film between the tip and the substrate.

It is possible that in spite of the increase in A, the drop in F_L may be due to the significant decrease in roughness of the substrate (see Sect. 22.3.4 below). Roughness is known to play a role in macroscopic friction as discussed by Bowden and Tabor as well as others [69–71]. This mechanism deals with the irreversible work done by asperities climbing over one another and is sometimes referred to as a "ratchet" mechanism. A similar mechanism was suggested to explain friction in an AFM experiment performed on surfaces

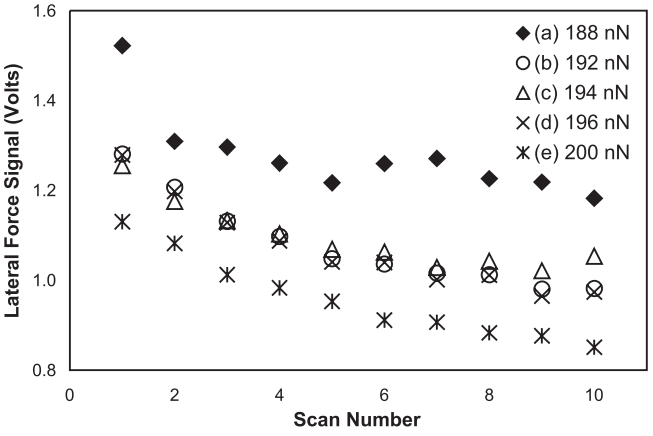


Fig. 22.8. Lateral force signal during sequential scanning at five different locations on sodium trisilicate glass in ammonium hydroxide solution (pH $\sim$ 11). Ten 750 $\times$750 nm^2 scans at 7.8 Hz were taken at the first location with a new tip (a), then the tip was moved to an unscanned area and another ten scans were taken (b), then (c) etc.

with controlled asperities, where the tip is considered to be a single asperity [72]. With commercial tips such as most of us use this may be a rash assumption (e. g., note our chosen chapter title). We also point out that macroscopic CMP studies have frequently reported decreases in friction as wear progresses [73–75].

To interpret the observed dependence of tip wear on time and F_{N}, it is useful to consider the *rate* of tip wear, $\mathrm{d}h_{\mathrm{tip}}(t)/\mathrm{d}t$, which is related to the experimentally measured net change in tip height, $H_{\mathrm{tip}}(F_{\mathrm{N}}, t)$ by:

$$\frac{\mathrm{d}h_{\mathrm{tip}}}{\mathrm{d}t} = \frac{\partial}{\partial t} H_{\mathrm{tip}}(F_{\mathrm{N}}, t_{\mathrm{F}}) = \left(\frac{kF_{\mathrm{N}}}{4t}\right)^{1/2} , \tag{22.5}$$

where the experimental result for$H_{\mathrm{tip}}(F_{\mathrm{N}}, t) = (kF_{\mathrm{N}}t)^{1/2}$ has been employed on the right; k is the constant of proportionality. Normally, one is interested in the volume removal rate, $\mathrm{d}V_{\mathrm{tip}}/\mathrm{d}t$, which equals product of the tip area, A_{tip}, and $\mathrm{d}h/\mathrm{d}t$.

$$\frac{\mathrm{d}V_{\mathrm{tip}}}{\mathrm{d}t} = A_{\mathrm{tip}} \frac{\mathrm{d}h_{\mathrm{tip}}}{\mathrm{d}t} = A_{\mathrm{tip}} \left(\frac{kF_{\mathrm{N}}}{4t}\right)^{1/2} . \tag{22.6}$$

For the purposes of interpretation, we remove the explicit time dependence in Eq. (22.6) by exploiting the time dependence of the tip dimensions. This is consistent with the expectation that the physical properties of the tip/substrate system are not time dependent. For a square pyramidal tip with a tip angle of 70°, the basal area $A_{\mathrm{tip}} = \alpha H^2$, where $\alpha = 1.99$. The experimental value of $H(F_{\mathrm{N}}, t)$ gives $A_{\mathrm{tip}} = \alpha k F_{\mathrm{N}} t$. (The tip area increases with time.) Then

$$\frac{\mathrm{d}V_{\mathrm{tip}}}{\mathrm{d}t} = A_{\mathrm{tip}} \left(\frac{kF_{\mathrm{N}}}{2t}\right)^{1/2} \left(\frac{\alpha k F_{\mathrm{N}} t}{A_{\mathrm{tip}}}\right)^{1/2} = \left(\frac{\alpha A_{\mathrm{tip}}}{4}\right)^{1/2} kF_{\mathrm{N}} \tag{22.7}$$

Finally, it is convenient to cast the dependence on normal force, F_{N}, in terms of the applied stress $(F_{\mathrm{N}}/A_{\mathrm{tip}})$. This reflects our desire to employ microscopic variables where possible. Further, macroscopic volume removal rates are often proportional to stress (as in the Preston Equation) [76]. Then

$$\frac{\mathrm{d}V_{\mathrm{tip}}}{\mathrm{d}t} = A_{\mathrm{tip}}^{3/2} \left(\frac{\alpha}{4}\right)^{1/2} \left(\frac{kF_{\mathrm{N}}}{A_{\mathrm{tip}}}\right) \tag{22.8}$$

Normally, one expects the local volume removal rate to vary linearly with the contact area $(= A_{\mathrm{tip}})$, as in Eq. (22.6). The extra factor of $(A_{\mathrm{tip}})^{1/2}$ in Eq. (22.8) suggests that the volume wear rate is "increased" by a factor proportional to the length of base of the tip, $l = (A_{\mathrm{tip}})^{1/2}$.

One possible source of this unusual dependence on A_{tip} is a gradual, stress-activated production of a chemical precursor state on the substrate as the tip passes over. If the time constant for precursor formation is small

relative to the duration of the applied stress, the precursor concentration will be approximately uniform and constant under the AFM tip. However, if the time constant for precursor formation is long relative to the duration of applied stress, the concentration of states under the tip will increase in a linear fashion from the leading edge of the tip to the trailing edge. Further, the maximum concentration will be proportional to the time required for the tip to pass over any given point on the surface. This time is proportional to the length of the flat portion of the tip, $l = (A_{\text{tip}})^{1/2}$. The stress at a given substrate location under tips with large flat areas is sustained for longer times than tips with small flat areas, producing higher concentrations of precursor states and yielding faster wear. Given the scan speeds and tip dimensions involved in this work, the relevant time scale for this effect is on the order of milliseconds.

Numerous studies have shown that on exposure to moist air or water, the silicon nitride surface rapidly oxidizes to form a thin layer of silicon oxides. Angle resolved XPS measurements by Hah et al. [77] show 2-nm thick oxide layers on silicon nitride surfaces formed by cleaving in air. Tribochemically polished surfaces showed a similar oxide layer, with approximate composition corresponding to $0.2-0.5\,\text{nm}$ SiO_2 and $1.0-1.5\,\text{nm}$ SiO_xN_y. Abrasively polished surfaces showed thicker oxide layers. Adhesion between silicon nitride AFM tips and a variety of substrates in water was inconsistent with predictions assuming that the tip was composed of Si_3N_4, but did match predictions for tips composed of SiO_2 [78]. In a separate study, XPS of Si_3N_4 AFM tips found considerable amounts of oxygen on the surface, which was attributed to an oxide layer [79]. Traces of surface oxide are also found in wear tracks on silicon nitride [60]. The ubiquity of oxide films on silicon nitride suggests that the wear of silicon nitride AFM tips is dominated by removal of the oxide, which is subsequently regenerated by further oxidation.

Significantly, AFM tip wear in this work is strongly affected by the chemical, as opposed to mechanical, properties of the substrate. Relatively soft substrates, such as the sodium trisilicate glass, often produced much more tip wear than much harder substrates, such as sapphire. All of the substrates yielding extensive tip wear form surface metal hydroxide species. Quartz, fused silica, zirconia, titania, sodium trisilicate glass, polished soda lime glass, and silicon nitride substrates in aqueous solution all have high densities of surface M–OH bonds and showed significant wear. Iron (II–III) oxide in aqueous solution is expected to display a high density of hydroxide bonds and produced intermediate levels of tip wear. Only one substrate with a strong potential for M–OH bonds (alumina – note: a relatively *hard* material) failed to produce significant tip wear, possibly due to kinetic effects. In contrast, all substrates expected to lack surface M–OH in aqueous solutions (calcite and mica) produced little if any tip wear. Therefore, we propose that tip wear results from reactions between hydroxide bonds on the tip with hydroxide bonds on the substrate.

On close contact, silicon hydroxyls on the AFM tip can form chemical bonds to hydroxyl groups on the substrate surface [57, 64, 80]. A schematic diagram of such tip-substrate bond formation is shown in Fig. 22.9a and b. As the tip moves, these tip-surface bonds would be stretched (Fig. 22.9c), and eventually rebreak (Fig. 22.9d). Simple bond-breaking will not result in tip wear (requiring removal of Si). Breaking any connecting bond in Fig. 22.9c would leave the silicon atom attached to one of the surfaces, and breaking two bonds simultaneously is statistically unlikely. However, when a tip-surface bond breaks, the energy of the stretched bond may transfer to kinetic energy of the atoms on each side of the bond (Fig. 22.9d) which weakens the other Si–O (or M–O) bonds toward reaction with water. We propose that this stretching and breaking process plays a critical role in tip wear.

Recoil energy as the stretched bonds break provides the mechanism to break the second bond after the first bond breaks, and it is the breaking of the second bond that leads to wear. The importance of this second step can also be observed in the lack of wear in the absence of water. As long as surface hydroxyls are present, tip-substrate bonds can form in the absence of water,

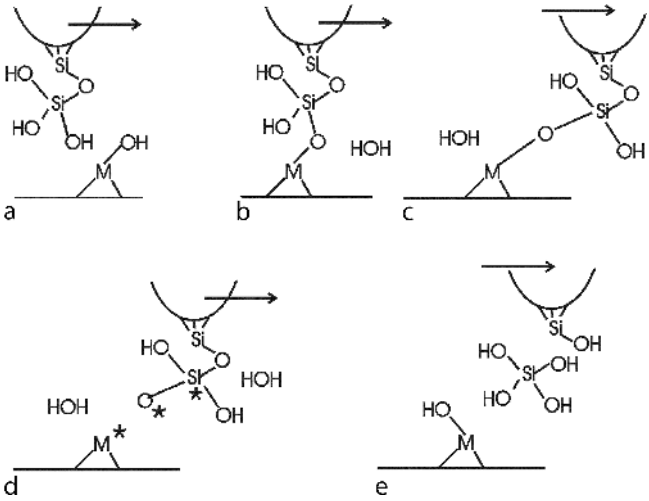


Fig. 22.9. Chemical reaction model for tribochemical tip wear. In water, the silicon nitride AFM tip is coated with a layer of silicon oxide. **a** During scanning, a hydroxyl group on the tip (Si–OH) encounters a hydroxyl group on the surface (M–OH). **b** The two groups react to form a bond, releasing water. **c** As the tip continues its motion across the surface, the bridging bond is stretched. **d** The bond breaks. Energy that was formerly stored in the stretched bridging bond is deposited in the atoms on either side (*asterisks*). **e** These "activated" atoms react with water, which may also break additional bonds. In this case, a bond between the silicon atom and the AFM tip breaks, releasing the silicon atom from the tip and producing aqueous $Si(OH)_4$. Further tip motion yields additional reactions and more broken Si–O bonds on the tip surface, as well as M–O bonds on the substrate

and if bond breaking (Fig. 22.9c,d) was all that was required for wear, then water would not be required for wear to occur. The fact that no wear occurs when ethyl acetate solution is used to exclude water strongly suggests that wear occurs not during bond breaking, but during subsequent reactions with water (Fig. 22.9d,e).

Tip-substrate bond formation plays a critical role in the proposed model of tip wear. One might expect formation of such bonds would depend on the ratio of protonated (M–OH) to deprotonated (M-O^-) hydroxyls on both surfaces, and thus vary with pH. Katuski et al. [57] observed a weak dependence of silicon nitride AFM tip wear on pH, while Hah et al. [53] observed no significant pH dependence of the macroscopic wear of silicon nitride in the presence of various chemicals. No significant pH dependence was observed in our work. A weak or absent pH dependence may reflect the heterogeneity of the surface. With SiO_xN_y and SiO_2 components, the tip surface may not have a well-defined isoelectric point. It is also possible that the local pH at the tip-substrate interface becomes nearly independent of the solution pH when the two surfaces are pressed tightly together and most of the solution is excluded.

In some models of chemical mechanical polishing, material removal is attributed to the mechanical abrasion of a soft surface reaction layer. If, as we propose, this layer is formed between asperities, material removal will involve competition between the rate of the chemical reaction (to form the modified layer) and the rate of asperity wear (which limits the time for surface modification). In the context of commercial chemical mechanical polishing operations, asperity contact is ideally confined to colloidal particles on the polishing pad, while chemical surface modification is largely confined to the substrate being polished. However, it is difficult to account for the effectiveness of certain abrasive materials, such as cerium oxide, in terms of mechanical effects alone. The effectiveness of some polishing materials has been attributed to a "chemical tooth", wherein beneficial chemical bonds are formed between the particulates and the substrate [65]. In our case, it is clear that the chemical reactivity of the substrate plays a key role in tip wear, which we feel is due to the formation of bridging oxide bonds between the tip and the substrate. The effectiveness of colloidal silica particles in chemical mechanical polishing (although not as effective as some other oxides) suggests that bridging oxide bonds may play a similar role in many chemical-mechanical polishing systems.

22.3.4 Single Asperity Tribochemical Interactions with Silicate Glasses

Given the results in Sect. 22.3.3, we can quickly understand why preliminary experiments involving wear of silicates in solution showed a striking slowdown of material removal rate with time – the tip was wearing and so applied

stresses were decreasing. Armed with the above quantitative knowledge and model, we now can address single asperity silicate wear kinetics.

Silicate wear plays an important role in the polishing of optics, the planarization of silicon wafers, and the machining of ceramics. Polishing typically employs hard particles, such as aluminum oxide, silica, or cerium oxide, where chemical interactions involving the substrate, mechanical stress, and ambient fluid promote controlled material removal. To produce nanometer-scale smoothness, submicron particles are often employed, and the nanotribology of the particle-substrate interaction becomes important. In the case of single asperity "polishing" of sodium containing glasses, we have found that over micron scale square raster scanning with a Si_3N_4 tip in 1 M NaOH aqueous solutions we can readily obtain several nm deep square pits with roughness ~ 0.1 nm; clearly, effective tribochemical polishing is occurring.

Considerable work has been performed on the wear of oxidized silicon surfaces by AFM tips [57,58,81–83], often with silicon AFM tips at relatively high contact forces ($1-100\,\mu$N). The effect of solution pH, scanning time, and scanning force are relatively well understood [83] but little has been done to address the effect of tip (asperity) wear on substrate wear rates.

Our choice of substrate is sodium trisilicate glass (allowing us to use our tip-wear data and model presented above) and we will scan in basic solution (normal conditions for CMP of glass). We have studied the consequences of two types of tip motion: $1 \times 1\,\mu m^2$ square raster scans which produced flat, smooth glass surfaces, and linear scans along a 500 nm line which produced grooves in the glass substrate (square and linear scanning provide models of nanometer-scale material removal representing polishing and "nano"-machining, respectively) [84]. Here we present results only on the square raster scans. After diamond polishing, the RMS roughness of these surfaces was typically 1.9 ± 0.6 nm. Wear experiments were performed under ammonium hydroxide solution (pH = 11). Wear was induced by repeatedly scanning the tip over a $1000 \times 1000\,nm^2$ square at 3.1 Hz. The worn tip was used to image (at low normal force) the worn region on the substrate. A new tip was used for each measurement and the tip shape was measured before and after as described above.

The AFM sensitivity was determined for each sample by noting the slope of the contact region of the force curve. Frictional forces were estimated from lateral force measurements using a numerically derived relationship between the normal and lateral force constants, $k_{normal}/k_{lateral} \sim 0.0032$ [85]. The lateral detector sensitivity was measured by taking the slope of the friction loop before the tip begins to slide [86]. In this work, the estimated lateral forces on a given surface were a linear function of applied normal force. Although our calibration is not adequate for the accurate determination of friction coefficients, changes in the ratio of the estimated lateral force to the measured normal force are significant and reflect changes in the tip-surface interface.

Our first observation was that tip wear continued to follow predictions from the model in Sect. 22.3.3. After an initial break in period, the final tip area is proportional to the total number of scans (and therefore proportional to time, t). The pyramidal geometry of the tip requires that the change in tip height be proportional to the square root of the number of scans (or $t^{1/2}$) [74]. Second, in terms of material removal during each wear test, we note that the change in tip height due to the flattening of the pyramidal tip during square raster scanning is greater than the depth of the resulting square pit in the substrate. For instance, square scans that produce 20-nm deep pits typically shorten the AFM tip by about 80 nm. Nevertheless, the total volume removed from the substrate typically exceeds the volume of material removed from the tip by an order of magnitude, consistent with the large area of the square raster scans. Third, substrate wear was found be independent of tip velocity over a range from $\sim 1 - 30\,\mu$m/s.

The initial surface roughness of the polished sodium trisilicate substrates was ~ 2 nm; high contact force scanning in high pH solutions produced surfaces with roughness between 0.1 and 0.2 nm, approaching the roughness due to the expected atomic fluctuations in bond angles for amorphous silicate surfaces [87]. A Fourier analysis of these surfaces and discussion can be found in ref [84].

Substrate wear depth versus number of square scans (or time) is shown in Fig. 22.10a. Despite the scatter, the wear depth clearly increases with time, while the wear rate (the slope of the curve) decreases. The dashed line in Fig. 22.10a is a least squares fit to a function of the form $kt^{1/2}$, where k is a constant and t is the time in units of number of scans (note this is the same functionality as the fit to the tip wear data in Fig. 22.6a). The measured wear volume is an important parameter and is typically a simple function of wear depth. Since the wear depth increases with the square root of time, one expects the volume to increase linearly with time. The observed relation is $V \sim t^{1.01}$. Again, we point out that the volume of material removed from the substrate was typically ten times the volume removed from the tip.

The substrate wear versus applied normal force, F_{N}, is shown in Fig. 22.10b. Despite the scatter, it is clear that the wear depth is not a linear function of contact force; the dashed line shows a least squares fit of the function $k(F_{\mathrm{N}})^{1/2}$ to the data, again similar to tip wear. As we shall see, the departure from linear behavior in both the time and F_{N} is due to the blunting of the AFM tip during scanning. Note that during a square raster scan, the tip is displaced about 2 nm each cycle; thus, for a typical contact diameter of 80 nm, each point in the scanned area would see the tip pass over it some 1200 times during a thirty scan treatment. This exposure would be in form of bursts of about 40 passes at a rate of about eight passes per second, separated by intervals of about 80 s in which the point in question is not exposed to the tip.

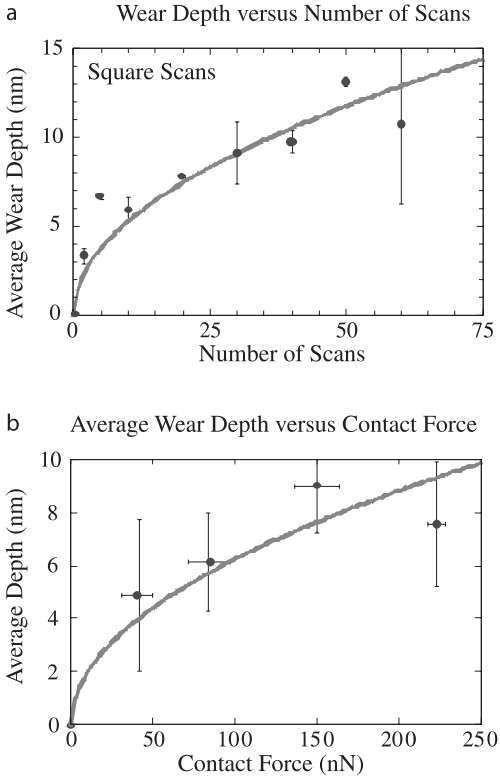


Fig. 22.10. **a** Wear depth as a function of number of scans for a $1000 \times 1000\,\mathrm{nm}^2$ square raster scans with 512 passes per scan at a contact force of 150 nN. **b** Wear depth as a function of contact forced produced by **a** thirty square raster scans

During wear we see again a drop in F_L with time and again we attribute this drop to the decrease in roughness of the substrate (in line with a ratcheting friction mechanism) [72].

The wear rate of the glass substrate can be interpreted in much the same fashion as the tip wear rate above. However, the volume removal rate is now equal to the product of the scan area A_{scan} and the rate at which the pit deepens, $\mathrm{d}h_{\mathrm{sub}}/\mathrm{d}t$, which is given by the time derivative of the experimental result, $H_{\mathrm{sub}}(F_{\mathrm{N}}, t) = (kF_{\mathrm{N}}t)^{1/2}$.

$$\frac{\mathrm{d}V_{\mathrm{sub}}}{\mathrm{d}t} = A_{\mathrm{scan}}\frac{\mathrm{d}h_{\mathrm{sub}}}{\mathrm{d}t} = A_{\mathrm{scan}}\left(\frac{kF_{\mathrm{N}}}{4t}\right)^{1/2} . \tag{22.9}$$

The explicit time dependence of the volume removal rate can again be eliminated, using the inferred time dependence of the tip geometry ($A_{\mathrm{tip}} = \alpha k F_{\mathrm{N}} t$).

$$\frac{\mathrm{d}V_{\mathrm{sub}}}{\mathrm{d}t} = A_{\mathrm{scan}}\left(\frac{kF_{\mathrm{N}}}{4t}\right)^{1/2}\left(\frac{\alpha k F_{\mathrm{N}} t}{A_{\mathrm{tip}}}\right)^{1/2} = A_{\mathrm{scan}}\left(\frac{\alpha}{4A_{\mathrm{tip}}}\right)^{1/2} kF_{\mathrm{N}} . \tag{22.10}$$

Finally, we wish to express the normal force dependence of the volume removal rate in terms of the corresponding microscopic (single asperity) stress

(F_N/A_{tip}). Then

$$\frac{dV_{sub}}{dt} = kA_{scan}\left(\frac{\alpha}{4A_{tip}}\right)^{1/2}\left(\frac{F_N}{A_{tip}}\right)A_{tip} = kA_{scan}\left(\frac{\alpha A_{tip}}{4}\right)^{1/2}\left(\frac{F_N}{A_{tip}}\right). \tag{22.11}$$

As noted above, one generally expects the volume removal rate to be proportional to the area of tip-substrate contact, A_{tip}. In contrast, the removal rate in Eq. (22.11) is proportional to the length of the edge of the tip base, $l = (A_{tip})^{1/2}$. One possible explanation is that the AFM tip removes a thin strip of vulnerable material, whose width is proportional to the width of the tip (the snow shovel approximation). The depth of this layer would be proportional to the stress applied by the tip. We note that if the associated plowing forces contribute to the measured lateral force, the decrease in lateral force during wear could in part be due to the decreased penetration of the tip as A_{tip} increases.

Somewhat by coincidence both the tip wear rate, $H_{tip}(F_N, t)$, and the substrate pit depth $H_{sub}(F_N, t)$ vary as $(F_N t)^{1/2}$. However, due to the contrasting geometries of the tip and substrate (which fix the relation between the volume removal rate and the rate of height change) we cannot interpret these two wear rates in the same way. Tip wear is proportional to stress and $(A_{tip})^{3/2}$, consistent with the removal of an intermediate produced by a stress activated chemical reaction involving the tip and the substrate. In contrast, substrate wear is proportional to stress and $(A_{tip})^{1/2}$, consistent with a removal of a thin strip of material whose width is proportional to the tip width and whose thickness is proportional to the applied stress.

Although the proposed mechanism for glass wear does not explicitly involve chemistry, the presence of the basic solution is essential. Significantly, the velocity dependence of substrate wear does not change over the experimentally accessible range of tip velocities. Thus any chemical reactions that promote substrate wear must occur on time scales shorter than the time between successive passes of the AFM tip. Since subsequent scan lines overlap, any reactions would have to take place during the time between scan lines, i. e., in less than 30 ms at the at the highest tip velocities we examined (30 μm/s). Rapid hydration/hydroxylation is consistent with the high alkali content of this glass, which is associated with a high density of non-bridging oxygens (NBOs). These NBOs are charge compensated by sodium ions, which react with water to form surface hydroxyls and sodium hydroxide, which can promote further dissolution by reacting with bridging oxygens [88]. At the pH employed in this work, any hydroxides on the surface are quickly dissociated to form dangling Si–O bonds [65]. These dangling bonds are available for reaction with the silicon nitride tip, which would play an important role in tip wear. The weakened bonds would also render the near surface glass vulnerable to attack by the AFM tip.

22.4 Response of Crystals to Single Asperity Tribological Loading in Saturated Solutions

Returning to slightly soluble crystalline surfaces exposed to aqueous solutions, one has to ask what happens to these surfaces under conditions of supersaturation? Are their any tip induced effects? The answer is surprisingly simple: Yes. Here we present this highly controllable atomic layer growth induced at step edges simply by scanning over the step edge. We begin with a study on $CaHPO_4 \cdot 2H_2O$ (brushite).

22.4.1 Deposition along Monolayer Steps in Single Crystal Brushite [89]

Figure 22.11 shows several etch pits on a cleaved surface of brushite in supersaturated solution. The degree of supersaturation, σ, is in multiples of the concentration of nominally saturated solution; $\sigma = 5$ corresponds to Ca^{2+} and HPO_4^{2-} concentrations five times that of a nominally saturated solution. Even at these supersaturations, etch pits tend to grow (material locally dissolves) unless they are manipulated with sufficient force by the AFM tip. In more dilute solutions ($\sigma \sim 1$), dissolution produces etch pits with very

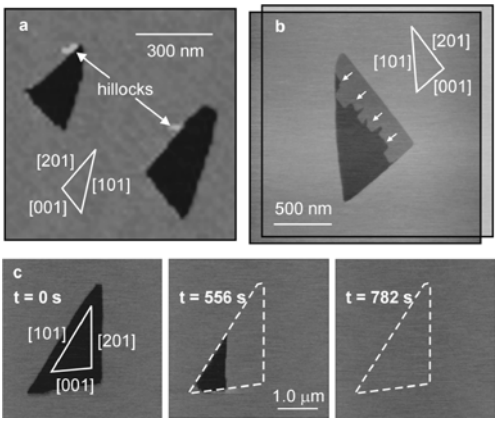


Fig. 22.11. a AFM image of two spontaneously dissolving etch pits in a (0$\underline{1}$0) brushite surface exposed to supersaturated solution. 3D hillock growth nucleates along the top edges of [201] steps. These hillocks are 0.8 nm high and are easily re-dissolved by local scanning with the AFM tip. **b** Fingering growth normal to the [201] step of a single atomic layer deep etch pit during 37 scans at a normal force of 5 nN and scan speed of 70 μm/s. **c** A sequence of AFM images of a one atomic layer deep etch pit in a brushite (010) surface scanned a total of 23 times at a tip speed of 70 μm/s and normal force of 5 nN. We have superimposed images acquired before and after growth, where the darker triangle is the original pit, and new single atomic layer growth is shown in a lighter shade

sharp corners. However at higher supersaturations, the corner joining [201] and [101] steps becomes rounded, as shown in Figs. 22.11a and b at $\sigma = 4$. This rounding reflects the influence of solution chemistry on the dynamics of ion removal from these two steps. Despite the fact that material is removed from the interior of the etch pits (material is dissolving), localized nucleation and growth of 3D hillocks is often observed along the top terrace edges of [201] steps, as shown in Fig. 22.11a. Since the most favorable sites for ion attachment are actually on the lower terrace due to higher coordination, on-top nucleation indicates the presence of an Ehrlich–Schwoebel type barrier [10,11] that hinders ion diffusion from the top terrace down to the lower terrace.

Although spontaneous pit dissolution is generally observed at these supersaturations, localized, directional growth is readily induced by continuously scanning over an etch pit at low normal forces ($F_N < 50\,nN$). Growth produced by 23 scans at $F_N = 5\,nN$ and $\sigma = 5.3$ is shown the sequence of images in Fig. 22.11c, eventually filling in the pit and producing an atomically flat surface. Growth is predominately along the [201] step and propagates normal to this step. Larger images taken before and after the images in Fig. 22.11c show that pits outside the continuously scanned area continue to dissolve even as the etch pit inside the continuously scanned area grew smaller. Thus we conclude that the observed deposition is *tip-induced*. This behavior (dissolution and tip induced growth at the same time) was observed over a wide range of supersaturations ($1 \ll 6$). Scanning at $F_N > 50\,nN$ immediately produces localized wear (dissolution) rather than growth as previously described [46].

At high saturations, growth along the [201] step induced by continuous scanning often shows a fingering instability. This is seen in the composite of two images in Fig. 22.11b, one acquired before and one after 37 scans at $F_N = 5\,nN$ and a tip speed of $70\,\mu m/s$. In spite of this instability, continued scanning fills in the entire pit (two complete layers) and leaves no visible defects in either the topographic (constant contact force) image nor in the lateral force images, although point defects such as vacancies would not be detectable in this mode of imaging.

More localized growth can be induced by drawing the AFM tip back and forth along a line normal to the [201]. Figure 22.12a shows a "single finger" growth feature generated by repeated linear scanning along the white line at a frequency of 2 Hz. The width of the resulting deposit is $\sim 120\,nm$, several times wider than the region actually contacted by the tip. Assuming elastic contact, the width of the strip contacted by the moving tip is only $\sim 20\,nm$ wide. Figure 22.12b shows "single finger" growth formed by scanning across the [201] edge of a two atomic layer deep pit. The finger on the lower atomic layer is slightly wider and grew at approximately twice the speed of the finger on the upper layer. Growth rates vary with pit size (smaller pits shrink faster) and the degree of supersaturation, with higher σ yielding higher growth rates. At $\sigma = 5$, we have measured growth rates during linear scanning as high as 7 nm/s, corresponding to the deposition of ~ 3 rows of calcium phosphate ions per pass of the AFM tip over the step. We were unable to induce growth

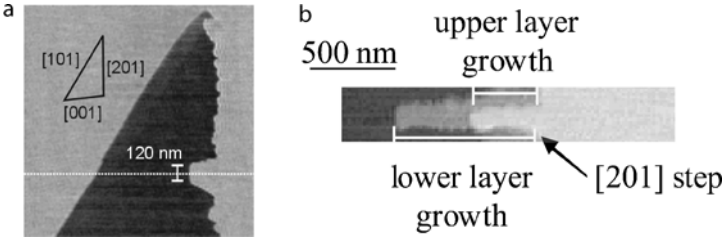


Fig. 22.12. a A single atomic layer deposit ~ 120 nm wide growing normal to the [201] step produced by repeated linear scanning normal to the step. **b** Deposition in a two atomic layer deep pit showing that the lower layer grows faster than the upper layer. The arrow on the right shows the position of the linear scan over the [201] step

at [001] steps; [101] steps showed slight growth at rates $< 5\%$ of the growth rate on the [201] step.

In the dissolution work described above [9, 46], the material removal rate increases continuously with increasing stress, even at very low stresses. Therefore, stress applied at or near the step edge might be expected to hinder rather than favor deposition. However, the presence of an Ehrlich–Schwoebel barrier for motion of ions down the step suggests a possible mechanism for tip induced deposition. In supersaturated solutions, one expects the nucleation and disappearance of transient, sub-critical, 3D clusters on the terraces [90]. If the tip can detach and sweep ions from these clusters over the step, this would increase the concentration of adsorbed ions on the bottom terrace at the step-edge and promote deposition. The growth pattern in Fig. 22.12b suggests that most of the swept material accumulates on the lower of the two terraces. Moving ions to the lowest terrace is of highest probability and would favor faster growth of the lowest finger. Diffusion along the bottom terrace would account for the broad (120 nm wide relative to the ~ 20 nm tip contact diameter) patches of deposited material in the pits near the linear scans in Fig. 22.12a and b. The high growth rates along the [201] steps may be due to the zig-zag rows of alternating Ca^{2+} and HPO_4^{2-} ions expected along these steps in order to maintain charge neutrality. This zig-zag structure would provide high binding energy sites for the nucleation of new ion rows. New ion rows are much more difficult to nucleate along the other two steps.

To examine the possible role of adsorbed material swept from the upper terraces, we performed back and forth linear scans of exactly the same length (1.5 µm), normal force (10 nN), scan frequency (1.0 Hz), and number of scans (32) at the same supersaturation ($\sigma = 4.3$), but changed the fraction of the scan taking place on the top terrace. The length of the growing [201] deposit was recorded during each linear scan, and the growth rate was determined by differentiation. Figure 22.13 shows the measured growth rates for ten experiments along a single [201] step, where we alternated between a 10 : 1 ratio

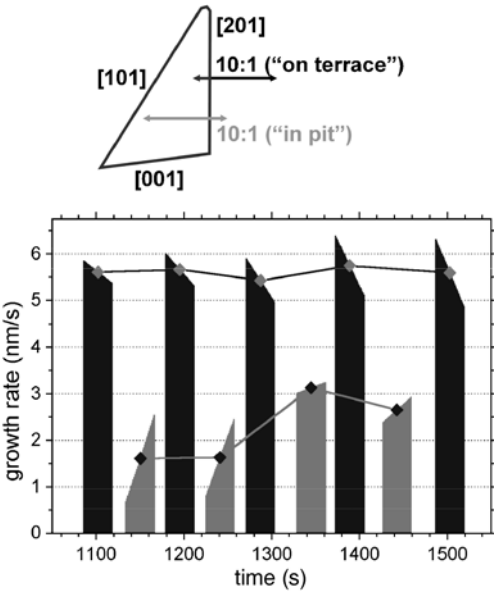


Fig. 22.13. Comparison of growth rates along the [201] edge of a single etch pit during scanning primarily on the top terrace (10 : 1) (*dark bars*) versus primarily on the bottom terrace (1 : 10) (*lighter bars*). The top of each bar shows the growth velocity vs. time during that particular set of scans. The points joined by lines join the average growth rates for each set of scans

of scan length on the top terrace to scan length on the bottom terrace (dark bars) and a 1 : 10 ratio (light bars). When the majority of the scan took place on the top terrace, the growth rates were 2–3 times higher than when the majority of the scan took place on the lower terrace. Very little material is swept into the pit when most of the scan is confined to the lower terrace, but scanning on the upper terrace sweeps more adsorbed material over the Ehrlich–Schwoebel barrier into the pit, where it becomes available for deposition. Note that the growth rates for the 10 : 1 scans consistently decrease with time (the top of each bar slopes down), consistent with the depletion of sub-critical, adsorbed clusters along the upper terrace. In contrast, the growth rates for the 1 : 10 scans consistently increase with time (the top of each bar slopes up). This suggests that on the bottom terrace near the step the adion concentration is initially low; as the 1 : 10 scans move ions into this depleted region the growth rate increases. (A suggested and plausible ion movement mechanism was suggested by a reviewer, namely tip-enhanced diffusion of ions to and over the step edge.)

Although transient clusters of sub-critical radii were not observed in any AFM images, the fact that at high σ and long exposure times we eventually observe 3-D nucleation on atomically flat terraces and away from step edges, strongly supports their existence. We have acquired indirect evidence for their presence by examining small fluctuations ("noise") in the lateral twist of the AFM cantilever [91]. (The lateral twist of the cantilever is highly sensitive to surface roughness.) These fluctuations are much stronger during scanning in highly supersaturated solution ($\sigma = 5$) than during scanning

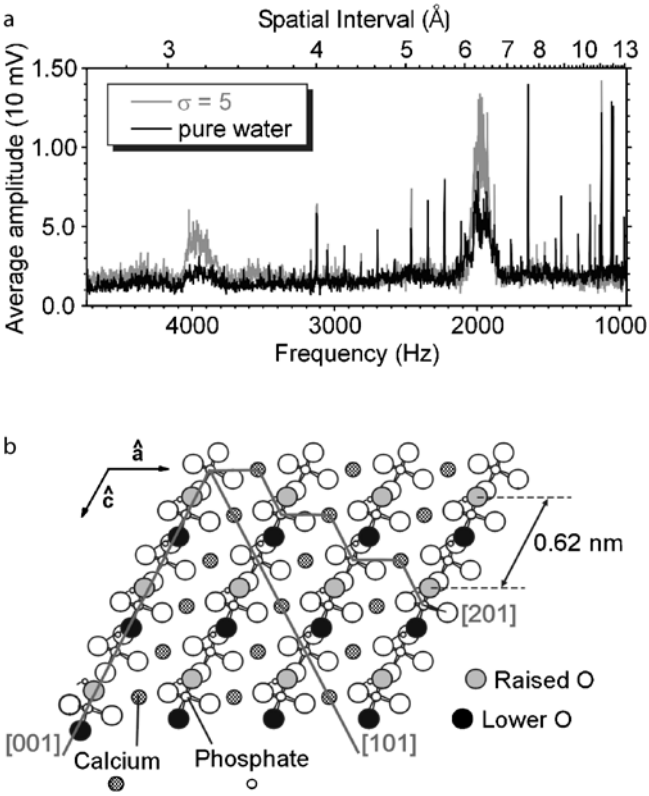


Fig. 22.14. Characterization of fluctuations during scanning in pure water and in supersaturated solution on the (010) surface of brushite. **a** Power spectra of fluctuations in the lateral displacement of the AFM cantilever during scanning along the crystal *c*-axis. We attribute the higher amplitudes during scanning in supersaturated solution to increased periodic stimulation of the tip in the presence of small crystallites on the terrace. **b** Schematic of the (010) surface of brushite

in pure water (no cluster formation). Figure 22.14a shows power spectra of the fluctuations averaged over ten 1-μm linear scans along the crystal's *c*-axis at 1.25 μm/s and $F_N = 5$ nN in supersaturated solution and in pure water. Both power spectra in Fig. 22.14a show peaks near ~ 2 and 4 kHz, but their amplitudes in supersaturated solution are higher by a factor of 2. The spatial frequency corresponding to the larger 2 kHz peak (given by the scan speed/frequency) is 0.63 nm. This corresponds well with the distance between the uppermost phosphate oxygen ions (which interact most strongly with tip asperities) along the *c*-axis of the unreconstructed (010) surface of 0.62 nm [47,92]. A schematic of an unrelaxed brushite cleavage plane (010) is shown in Fig. 22.14b. We propose that this peak is due to a periodic rocking motion induced in the tip as asperities interact collectively with successive

rows of oxygen ions. The second harmonic arises due to a non-linearity in the cantilever response.

A related mechanism accounts for many images showing atomic-scale periodic structures (not true atomic resolution) with relatively large radii AFM tips acquired in contact mode on graphite [93], brushite [46], calcite (another crystal with protruding oxygen ions) [22, 24, 29] and other single crystals. The resonant frequency of the lateral twist of our cantilevers (estimated from the properties of Si_3N_4 and the dimensions of our cantilevers) is $\sim 40\,\mathrm{kHz}$, high above the observed spectral features. We propose that the increased amplitude of cantilever rocking at high saturations is caused by tip-cluster collisions and therefore serves as evidence of transient clusters on the terraces. These collisions may increase normal mode motion which enhances the lateral deflection due to stronger interactions with the lattice. Numerous noise measurements in both pure water and saturated solutions at various scanning directions all show comparable increases in amplitude for the saturated solutions, supporting our hypothesis. Similarly, scanning atomically flat brushite terraces vs. regions with high densities of single atomic layer steps show precisely the same response: an increase in noise amplitude for the latter surfaces.

22.4.2 Tip Induced Growth at Step Edges on Single Crystal Calcite

Brushite is not the only material for which we see tip induced crystal growth at step edges. Figure 22.15 shows a series of images acquired during low contact force imaging of a three-monolayer deep etch pit on single crystal calcite. The first image was acquired at a relatively low supersaturation, $\sigma = 1.5$. This solution was then replaced with a solution of $\sigma = 2.2$. In this series, the innermost (bottom) pit filled in very quickly, before the next scan was completed. The middle pit required seven more scans to fill in completely, and the outermost pit required thirteen more scans.

Again, deposition is much more rapid along some steps (the fast steps) than others. As noted above, stress-enhanced dissolution is also enhanced

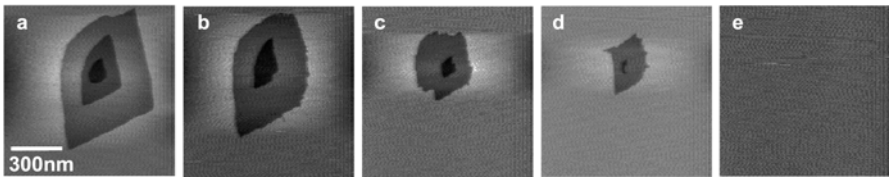


Fig. 22.15. AFM images of an etch pit on single crystal calcite, three monolayers deep: **a** imaged at $\sigma = 1.5$; **b** imaged immediately after replacing fluid with solution at $\sigma = 2.2$; **c** after two more scans; **d** after four more scans; and **e** after thirteen more scans, completely filling in the pit. Images were acquired at $F_N = 15\,\mathrm{nN}$ and a scan speed of 700 nm/s

along the fast steps. The resulting surface is free of defects observable by AFM. Quantitative measurements of the pit area as a function of time show that, initially, the pit area falls exponentially with time. As the pit size decreases, the kinetics change to a super-exponential regime, where the pit fills in quickly. Visually, small pits can appear to fill in much faster than large pits.

Linear scanning on calcite produces localized deposits, much as with brushite above. When linear scanning is performed across noncrystallographic cleavage steps, the deposits often form prominent fast steps – the steps where deposition is strongly favored. Figure 22.16 shows images of a cleavage step before and after 2560 back-and-forth linear scans along the white line in (a) at a supersaturation $\sigma = 3.7$. Subsequent scans were taken in undersaturated solution to verify that the edges of the new deposit in (b) are indeed fast steps.

Importantly, the formation of the deposit in (b) requires significant material transport along the growing, new fast step. Mobile material can even turn the convex corner (joining the two fast steps) at the upper right end of the deposit. The relatively straight sides of the main deposit suggest that this transport is efficient over distances of at least 2 µm. The visual evidence for the diffusion of material upward along the fast step in Fig. 22.16b is compelling.

Diffusion of material along pre-existing steps corresponds to ledge diffusion in terrace-ledge-kink (TLK) models of crystal growth [36]. Easy ledge diffusion in calcite is not necessarily expected, because diffusion on terraces is apparently slow [23, 94]. One normally expects ledge diffusion to be slower than terrace diffusion. However, if the relevant deposits along terraces were

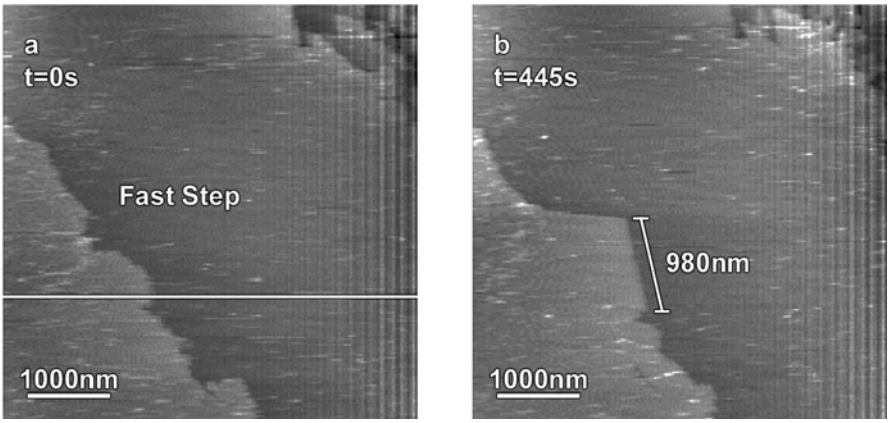


Fig. 22.16. Localized growth induced by 2560 passes of the AFM tip across a cleavage step along the white line in **a** at supersaturation $\sigma = 3.7$, $F_N = 22$ nN, and tip speed 31 µm/s

composed of many ions, they would be relatively immobile. Rapid diffusion of ion pairs along ledges would help explain why calcite etch pits are relatively straight over a wide range of solution conditions [28].

The total number of ion pairs incorporated into the new growth under these conditions can be quite large. In Fig. 22.16, the deposit area corresponds to at least 1×10^7 ion pairs, or about 4000 per pass of the tip across the step. Transport of this magnitude is probably best visualized in terms of the sweeping of clusters or mats of material, rather than individual ion pairs. Detailed comparison of the images in Fig. 22.16 suggests that almost 1000 ion rows have been nucleated at the point where the scan crosses the step. This corresponds to about one kink nucleation even per 2.5 passes of the tip. At least in the case of calcite, our measurements indicate that the total amount of material deposited along steps depends much more strongly on solution supersaturation than the number of kinks nucleated where the tip intersects the surface.

22.5 Conclusions

The tip of an AFM probe provides a useful surrogate for the study of nanoscale, single asperity wear. The synergism between mechanical (asperity) wear and chemical dissolution plays an important role in many commercial processes, including chemical mechanical polishing. Much of this synergism can be attributed to the vulnerability of defect structures such as steps and kinks to combined-mechanical chemical attack. This vulnerability is readily observed on single-crystal surfaces of materials such as calcite and brushite. These surfaces display remarkable anisotropies in wear track growth and evolution during and after linear scanning across molecular steps. These anisotropies reflect the important role of local geometry and bonding in determining step stabilities in aqueous environments, with and without applied stress.

In principle, knowledge of E_{act}, v^*, and the relevant prefactors for the key dissolution processes (e. g., pit nucleation, double kink nucleation, and kink propagation) would form the basis for a complete predictive model for tribochemical wear in these materials. Where deformation processes play an important role, dislocation emission and similar processes would also require characterization. Our understanding how E_{act}, v^*, and other material parameters are affected by the local structure would also be enhanced by molecular dynamics simulations of water-surface interactions. The rather low rate of dissolution (one molecule removed per kink site in millions of vibrational periods) may complicate these calculations.

Low normal force scanning of the surface of an ionic crystal (brushite) with an AFM tip in supersaturated solution can induce and control atomic layer re-growth at step edges. This process can be exploited to produce atomically smooth surfaces by “filling” rather than “polishing”. The chief role of the

tip appears to be the controlled transport of sorbed ion clusters from upper terraces over the step edge, thereby raising concentration of adsorbed ions near the step inside the pit. Secondly, 3D nucleation is totally suppressed by scanning, thereby maintaining flat surfaces. The use of fluctuations ("noise") in the lateral rocking of the cantilever to probe dynamic processes during deposition, such as the formation of transient sorbed clusters, is an exciting prospect.

In work not shown here we were able to characterize forces necessary for an AFM tip to detach nanoparticles strongly adhering to inorganic substrates with the aid of exposure to water vapor [18]. Again, we see that the interaction between mechanical and chemical effects plays an important role. The dramatic effect of particle size and humidity on the removal of salt particles from a soda lime glass substrates is mirrored in practical CMP operations, where chemically active of components of the solution lower the surface energy of the surfaces formed by removal. It is particularly significant that the humidity dependence in the salt-glass system can be quantitatively described in terms of the change in interfacial energy due to water absorption. This knowledge, in conjunction with the relevant particle size dependencies, would facilitate intelligent design and improvement of CMP processes in new materials systems. *Combined* chemical and mechanical attack is most effective for particle removal as well as for eliminating monolayer surface roughness.

Human-scale machining operations are often limited by wear of the tool, and it is no surprise that this is often true at the nanometer scale as well. Over a large range of contact forces and tip velocities, the wear of silicon nitride AFM tips is dominated by chemical-mechanical effects involving not only the solution but also the chemical nature of the substrate. In the present work, the only substrates producing significant tip wear are characterized by high densities of metal-hydroxide bonds. Wear also requires an aqueous environment. Solution pH has little influence on tip wear rates. The role of water is to replenish hydroxide bonds on the substrate surface that are consumed in mechanically mediated reactions with the tip.

Measurements of AFM tip wear provide a powerful probe of single asperity wear – both in the tip itself and in the affected substrate. Ultimately, we hope to understand macroscopic wear dominated by asperity contacts in terms of the wear of individual asperities from a given initial size distribution as a function of applied stress and chemical environment. Together, stress and corrosion are much more effective than either alone, and provide many variables for the control and localization of material wear or deposition.

Acknowledgement. We wish to thank Steve Langford, Louis Scudiero, Forrest Stevens, Ann McEvoy, and Rizal Hariadi of Washington State University, and Nam-Seok Park, of Chungbuk University, South Korea, for performing portions of the AFM work and analysis discussed above. We thank Ryan Leach, Washington State University, for help with computer programming, and Larry Pederson, Pacific Northwest National Laboratories, for supplying sample materials.

This work was supported by the US National Science Foundation under Grants CHE-02-34726 CMS-0409861 and by NIH under Contract No. HG-002647-01A1.

References

1. J. T. Dickinson, L. C. Jensen, R. L. Webb, M. L. Dawes, and S. C. Langford, "Interactions of wide band gap single crystals with 248 nm excimer laser radiation. III. The role of cleavage-induced defects in MgO," J. Appl. Phys. **74**, 3758–3767 (1993).
2. J. T. Dickinson, L. C. Jensen, R. L. Webb, M. L. Dawes, and S. C. Langford, in *Laser Ablation in Materials Processing: Fundamentals and Applications*; *Vol. 285*, edited by B. Braren, J. J. Dubowski, and D. P. Norton (Materials Research Society, Pittsburgh, Pennsylvania, 1993), p. 131–143.
3. M. L. Dawes, W. Hess, Y. Kawaguchi, S. C. Langford, and J. T. Dickinson, "Laser-induced positive ion and neutral atom/molecule emission from single crystal $CaHPO_4 \cdot 2H_2O$: The role of electron beam induced defects," Appl. Phys. A **69**, S547-S522 (1999).
4. J. T. Dickinson, S. C. Langford, C. Bandis, M. L. Dawes, and Y. Kawaguchi, "Consequences of combining laser irradiation with other stimuli on laser desorption and ablation from wide bandgap insulators," Appl. Surf. Sci. **154–155**, 291–304 (2000).
5. J. T. Dickinson, Y. Kawaguchi, M. L. Dawes, and S. C. Langford, in *Laser Applications in Microelectronic and Optoelectronic Manufacturing V*, SPIE Proceedings Vol. 3935, edited by R.F. Haglund and R.F. Wood (SPIE, Bellingham, WA, 2000), p. 38–36.
6. J. T. Dickinson, M. L. Dawes, K. H. Nwe, and S. C. Langford, in *Photon Processing in Microelectronics and Photonics*; *Vol. 4637*, edited by K. Sugioka, R. F. Haglund, A. Pique, F. Traeger, J. J. Dubowski, and W. Hoving (SPIE – The International Society for Optical Engineering, Bellingham, WA, 2002), p 1–12.
7. S. Nakahara, S. C. Langford, and J. T. Dickinson, "Surface force microscope observations of corrosive tribological wear on single crystal $NaNO_3$ exposed to moist air," Tribology Lett. **1**, 277–300 (1995).
8. N.-S. Park, M.-W. Kim, S. C. Langford, and J. T. Dickinson, "Tribological enhancement of $CaCO_3$ dissolution during scanning force microscopy," Langmuir **12**, 4599–4604 (1996).
9. N.-S. Park, M.-W. Kim, S. C. Langford, and J. T. Dickinson, "Atomic layer wear of single-crystal calcite in aqueous solution using scanning force microscopy," J. Appl. Phys. **80**, 2680–2686 (1996).
10. G. Ehrlich and F. G. Hudda, "Atomic view of self-diffusion: Tungsten on tungsten," J. Chem. Phys. **44**, 1039–1049 (1966).
11. R. L. Schwoebel and E. J. Shipsey, "Step motion on crystal surfaces," J. Appl. Phys. **37**, 3682–3686 (1966).
12. E. Meyer, R. Luthi, L. Howald, and M. Bammerlin, in *Micro/Nanotribology and its Applications*, edited by B. Bhusan (Kluwer Academic, Dordrecht, 1997), p. 193–215.

13. T. Junno, K. Deppert, L. Montelius, and L. Samuelson, "Controlled manipulation of nanoparticles with an atomic force microscope," Appl. Phys. Lett. **66**, 3627–3629 (1995).
14. C. Lebreton and Z. Z. Wang, "Critical humidity for removal of atoms from the gold surface with scanning tunneling microscopy," J. Vac. Sci. Technol. B **14**, 1356–1359 (1996).
15. A. C. Tam, W. P. Leung, W. Zapka, and W. Ziemlich, "Laser-cleaning techniques for removal of surface particulates," J. Appl. Phys. **71**, 3515–3523 (1992).
16. A. C. Tam, H. K. Park, and C. P. Grigoropoulos, "Laser cleaning of surface contaminants," Appl. Surf. Sci. **127–129**, 721–725 (1998).
17. G. Nanz and L. E. Camilletti, "Modeling of chemical-mechanical polishing: A review," IEEE Trans. Semiconductor Manufacturing **8**, 382–389 (1995).
18. R. F. Hariadi, S. C. Langford, and J. T. Dickinson, "Atomic force microscope observations of particle detachment from substrates: the role of water vapor in tribological debonding," J. Appl. Phys. **86**, 4885–4891 (1999).
19. R. N. Leach, F. Stevens, C. Seiler, S. C. Langford, and J. T. Dickinson, "Nanometer-scale solvent-assisted modification of polymer surfaces using the atomic force microscope," Langmuir **19**, 10225–10232 (2003).
20. F. Stevens, R. N. Leach, S. C. Langford, and J. T. Dickinson, "Influence of molecular weight on nano-scale modification of poly(methyl methacrylate) due to simultaneous mechanical and chemical stimulation," Langmuir **22**, 3320–3325 (2006).
21. G. Fu, A. Chandra, S. Guha, and G. Subhash, "A Plasticity-Based Model of Material Removal in Chemical–Mechanical Polishing (CMP)," IEEE Trans. Semi. Manufact. **14**, 406–417 (2001).
22. F. Ohnesorge and G. Binnig, "True atomic resolution by atomic force microscopy through repulsive and attractive forces," Sci. **260**, 1451–1456 (1993).
23. P. E. Hillner, A. J. Gratz, S. Manne, and P. K. Hansma, "Atomic scale imaging of calcite growth and dissolution in real time," Geology **20**, 359–362 (1992).
24. P. E. Hillner, S. Manne, A. J. Gratz, and P. K. Hansma, "AFM images of dissolution and growth on a calcite crystal," Ultramicroscopy **42–44**, 1387–1393 (1992).
25. S. L. S. Stipp, C. M. Eggleston, and B. S. Nielsen, "Calcite surface structure observed at microtopographic and molecular scales with atomic force microscopy (AFM)," Geochim. Cosmochim. Acta **58**, 3023–3033 (1994).
26. S. L. S. Stipp, W. Gutmannsbauer, and T. Lehmann, "The dynamic nature of calcite surfaces in air," Am. Mineralogist **81**, 1–8 (1996).
27. Y. Liang, D. R. Baer, and A. S. Lea, in *Evolution of Thin-Film and Surface Structure and Morphology*; *Vol. 355*, edited by B. G. Demczyk, E. D. Williams, E. Garfunkel, B. M. Clemens, and J. J. Cuomo (Materials Research Society, Pittsburgh, PA, USA, 1995), p. 409.
28. Y. Liang, D. R. Baer, J. M. McCoy, and J. P. LaFemina, "Interplay between step velocity and morphology during the dissolution of $CaCO_3$ surface," J. Vac. Sci. Technol. A **14**, 1368–1375 (1996).
29. Y. Liang, A. S. Lea, D. R. Baer, and M. H. Engelhard, "Structure of the cleaved $(10\underline{1}4)$ surface in an aqueous environment," Surf. Sci. **351**, 172–182 (1996).
30. Y. Liang and D. R. Baer, "Anisotropic dissolution at the $CaCO_3(10\underline{1}4)$-water interface," Surf. Sci. **373**, 275–287 (1997).

31. J. Schott, S. Brantley, D. Crerar, C. Guy, M. Borcsik, and C. Willaime, "Dissolution kinetics of strained calcite," Geochim. Cosmochim. Acta **53**, 373–382 (1989).
32. R. G. Compton, P. J. Daly, and W. A. House, "The dissolution of Iceland spar crystals: The effect of surface morphology," J. Colloid Interface Sci. **113**, 12–20 (1986).
33. J. Paquette and R. J. Reeder, "New type of compositional zoning in calcite: Insights in to crystal-growth mechanisms," Geology **18**, 1244–1247 (1990).
34. I. N. MacInnis and S. L. Brantley, "The role of dislocations and surface morphology in calcite dissolution," Geochim. Cosmochim. Acta **56**, 1113–1126 (1992).
35. I. N. MacInnis and S. L. Brantley, "Development of etch pit size distributions on dissolving minerals," Chem. Geology **105**, 31–49 (1993).
36. J. P. Hirth and G. M. Pound, "Evaporation of metal crystals," J. Chem. Phys. **26**, 1216–1224 (1957).
37. P. Anuradha and I. V. K. B. Raju, "Study of plastic flow in $CaCO_3$ and $NaNO_3$ crystals under concentrated loads," Physica Status Solidi (a) **95**, 113–119 (1986).
38. S. J. Bull, T. F. Page, and E. H. Yoffe, "An explanation of the indentation size effect in ceramics," Philos. Mag. Lett. **59**, 281–288 (1989).
39. W. W. Walker and L. J. Demer, "Effect of loading duration on indentation hardness," Trans. Metall. Soc. AIME **230**, 613–614 (1964).
40. S. P. Timoshenko and J. N. Goodier, *Theory of Elasticity*, 3rd ed. (McGraw-Hill, New York, 1970).
41. D. H. Chung and W. R. Buessem, in *Anisotropy in Single-Crystal Refractory Compounds*, edited by F. W. Vahldiek and S. A. Mersol (Plenum, New York, 1968), p. 217–245.
42. W. B. Hillig and R. J. Charles, in *High Strength Materials*, edited by V. F. Zackay (John Wiley, New York, 1965), p. 682–705.
43. S. M. Wiederhorn, "Influence of water vapor on crack propagation in soda-lime glass," J. Am. Ceram. Soc. **50**, 407–414 (1967).
44. Y. Ben-Eliahu, Y. Haas, and S. Welner, "Laser initiation of the decomposition of energetic polymers: Shock wave formation," J. Phys. Chem. **99**, 6010 (1995).
45. M. Ohta, M. Tsutsumi, and S. Ueno, "Observations of etch pits on as-grown faces of brushite crystals," J. Crystal Growth **47**, 135–136 (1979).
46. L. Scudiero, S. C. Langford, and J. T. Dickinson, "Scanning force microscope observations of corrosive wear on single-crystal Brushite ($CaHPO_4{\cdot}2H_2O$) in aqueous solution," Tribology Lett. **6**, 41–55 (1999).
47. N. A. Curry and S. W. Jones, "Crystal structure of brushite, calcium hydrogen orthophospate dihydrate: A neutron diffraction investigation," Chem. Soc. (London) A **1971**, 3725–3729 (1971).
48. V. A. Muratov and T. E. Fischer, "Tribochemical polishing," Annu. Rev. Mater. Sci. **30**, 27–51 (2000).
49. F. Honda and T. Saito, "Tribochemical characterization of the lubrication film at the Si_3N_4/ Si_3N_4 interface sliding in aqueous solutions," Appl. Surf. Sci. **92**, 651–955 (1996).
50. T. Saito, T. Hoseo, and F. Honda, "Chemical wear of sintered Si_3N_4, hBN and Si_3N_4-hBN composites by water lubrication," Wear **247**, 223–230 (2001).
51. J. Xu and K. Kato, "Formation of tribochemical layer of ceramics sliding in water and its role for low friction," Wear **245**, 61–75 (2000).

52. J. Xu, K. Kato, and T. Hirayama, "The transition of wear mode during the running-in process of silicon nitride sliding in water," Wear **205**, 55–63 (1997).
53. S. R. Hah and T. E. Fischer, "Tribochemical polishing of silicon nitride," J. Electrochem. Soc. **145**, 1708–1714 (1998).
54. H. Tomizawa and T. E. Fischer, "Friction and wear of silicon nitride at 150 °C to 800 °C," ASLE Transactions **29**, 481–488 (1986).
55. A. Khurshudov and K. Kato, "Wear of the atomic force microscope tip under light load, studied by atomic force microscopy," Ultramicroscopy **60**, 11–16 (1995).
56. M. L. Bloo, H. Haitjema, and W. O. Pril, "Deformation and wear of pyramidal, silicon-nitride AFM tips scanning micrometer-size features in contact mode," Measurement **25**, 203–211 (1999).
57. F. Katsuki, K. Kamei, A. Saguchi, W. Takahashi, and J. Watanabe, "AFM studies on the difference in wear behavior between Si and SiO_2 in KOH solution," J. Electrochem. Soc. **147**, 2328–2331 (2000).
58. S. Seta, T. Nishioka, Y. Tateyama, and N. Miyashita, in *Chemical Mechanical Planarization IV*, edited by R. L. Opila, C. Reidsema-Simpson, K. B. Sundaram, and S. Seal (Electrochemical Society, Pennington, NJ, USA, 2001), p. 28–33.
59. M. Chen, K. Kato, and K. Adachi, "Friction and wear of self-mated SiC and Si_3N_4 sliding in water," Wear **250**, 246–255 (2001).
60. S. Jahanmir and T. E. Fischer, "Friction and wear of silicon nitride lubricated by humid air, water, hexadecane and hexadecane + 0.5 percent stearic acid," STLE Transactions **31**, 32–43 (1988).
61. V. A. Muratov, T. Luangvaranunt, and T. E. Fischer, "The tribochemisty of silicon nitride: Effects of friction, temperature and sliding velocity," Tribology Int. **31**, 601–611 (1998).
62. S. W. Bailey, in *Rev. Mineral.*; *Vol. 13* (BookCrafters, Inc., Chelsea, MI, 1984).
63. R. V. Gaines, H. C. W. Skinner, E. E. Foord, B. Mason, A. Rosenzweig, and V. T. King, *Dana's New Mineralogy* (John Wiley & Sons, Inc., New York, 1997).
64. N. B. Kirk and J. V. Wood, "Glass polishing," Br. Ceram. Trans. **93**, 25–30 (1994).
65. L. M. Cook, "Chemical processes in glass polishing," J. Non-Cryst. Solids **120**, 152–171 (1990).
66. R. W. Carpick and M. Salmeron, "Scratching the surface: Fundamental investigations of tribology with atomic force microscopy," Chem. Rev. **97**, 1163–1194 (1997).
67. O. Piétrement and M. Troyon, "Study of the interfacial shear strength pressure dependence by modulated lateral force microscopy," Langmuir **17**, 6540–6546 (2001).
68. M. A. Lantz, S. J. O'Shea, M. E. Welland, and K. L. Johnson, "An Atomic Force Microscope Study of Contact Area and Friction on NbSe2," Phys. Rev. B. **55**, 10776–10785 (1997).
69. F. P. Bowden and D. Tabor, *The Friction and Lubrication of Solids vol 1* (Oxford, Clarendon, 1950).
70. E. Rabinowicz, *Friction and Wear of Materials* (John Wiley and Sons, Inc., New York, 1965).
71. M. Casey and J. Wilks, "The friction of diamond sliding on polished cube faces of diamond," J. Phys. D **6**, 1772–1782 (1973).

72. S. Sundararajan and B. Bhushan, "Topography-induced contributions to friction forces measured using an atomic force/friction force microscope," J. Appl. Phys. **8**, 4825–4831 (2000).
73. U. Mahajan, M. Bielmann, and R. K. Singh, "Dynamic lateral force measurements during chemical mechanical polishing of silica," Electrochemical and Solid-State Letters **2**, 80–82 (1999).
74. W. Maw, F. Stevens, S. C. Langford, and J. T. Dickinson, "Single asperity tribochemical wear of silicon nitride studied by atomic force microscopy," J. Appl. Phys. **92**, 5103–5109 (2002).
75. A. Opitz, S. I.-U. Ahmed, J. A. Schaefer, and M. Scherge, "Nanofriction of silicon oxide surfaces covered with thin water films," Wear **254**, 924–929 (2003).
76. F. W. Preston, "The theory and design of plate glass polishing machines," J. Soc. Glass Technol. **11**, 214–256 (1927).
77. S. R. Hah, C. B. Burk, and T. E. Fischer, "Surface quality of tribochemically polished silicon nitride," J. Electrochem. Soc. **146**, 1505–1509 (1999).
78. C. Jacquot and J. Takadoum, "A study of adhesion forces by atomic force microscopy," J. Adhesion Sci. Technol. **15**, 681–687 (2001).
79. G. U. Lee, L. A. Chrisey, C. E. O'Ferrall, D. E. Pilloff, N. H. Turner, and R. J. Colton, "Chemically-specific probes for the atomic force microscope," Israel J. Chem. **36**, 81–87 (1996).
80. T. Hoshino, Y. Kurata, Y. Terasaki, and K. Susa, "Mechanism of polishing of SiO_2 films by CeO_2 particles," J. Non-Cryst. Solids **283**, 129–136 (2001).
81. W. C. Moon, T. Yoshinobu, and H. Iwasaki, "Nanotribology of Si oxide layers on Si by atomic force microsocopy," Ultramicroscopy **86**, 49–53 (2001).
82. W.-C. Moon, S. A. Contera, T. Yoshinobu, and H. Iwasaki, "Nanotribology of clean and oxide-covered silicon surfaces using atomic force microscopy," Jpn. J. Appl. Phys. Part 1 **39**, 272–274 (2000).
83. F. Katsuki, A. Saguchi, W. Takahashi, and J. Watanabe, "The atomic-scale removal mechanism during Si tip scratching on Si and SiO_2 surfaces in aqueous KOH with an atomic force microscope," Jpn. J. Appl. Phys. Part 1 **41**, 4919–4923 (2002).
84. F. Stevens, S. C. Langford, and J. T. Dickinson, "Tribochemical wear of sodium trisilicate glass at the nanometer size scale," J. Appl. Phys. (2006).
85. J. M. Neumeister and W. A. Ducker, "Lateral, normal, and longitudinal spring constants of atomic force microscopy cantilevers," Rev. Sci. Instrum. **65**, 2527–2531 (1994).
86. R. G. Cain, S. Biggs, and N. W. Page, "Force calibration in lateral force microscopy," J. Colloid Interface Sci. **227**, 55–65 (2000).
87. J.-F. Poggemann, G. Heide, and G. H. Frischat, "Direct view of the structure of different glass fracture surfaces by atomic force microscopy," J. Non-Cryst. Solids **326-327**, 15–20 (2003).
88. C. G. Pantano, in *Strength of Inorganic Glass*, edited by C. R. Kurkjian (Plenum, New York, 1985), p. 37–66.
89. R. Hariadi, S. C. Langford, and J. T. Dickinson, "Controlling nanometer-scale crystal growth on a model biomaterial with a scanning force microscope," Langmuir **18**, 7773–7776 (2002).
90. Z. Toroczkai and E. D. Williams, "Nanoscale fluctuations at solid surfaces," Phys. Today **52**, 24–28 (1999).
91. R. Hariadi and J. T. Dickinson, "in preparation".

92. R. W. G. Wyckoff, *Crystal Structures: Inorganic Compounds* $R_x(MX_4)_y$, $R_x(M_nX_p)_y$, *Hydrates and Ammoniates*, Vol. 3, 2nd ed. (Interscience, New York, 1965).
93. J.-A. Ruan and B. Bhushan, Atomic-scale and microscale friction studies of graphite and diamond using friction force microscopy, J. Appl. Phys. **76**, 5022–5035 (1994).
94. A. J. Gratz, P. E. Hillner, and P. K. Hansma, Step dynamics and spiral growth on calcite, Geochim. Cosmochim. Acta **57**, 491–495 (1993).

23 Nanotribology of MEMS/NEMS

Satish Achanta and Jean-Pierre Celis

Katholieke Universiteit Leuven, Dept. MTM, Kasteelpark Arenberg 44, 3001 Leuven, Belgium

List of symbols

Symbol	Definition	Units
r	Radius of a meniscus	m
R	Radius of a ball	m
γ_l	Surface tension of a fluid	J/m^2
F_{cap}	Capillary force between ball and counterbody	N
z	Distance of separation between two surfaces	m
d_{cap}	Characteristic distance for capillary condensation	m
θ	Contact angle	–
$e_{\mathrm{cap}}(z)$	Interaction energy function	J/m^2
A_{Ham}	Hamaker constant	J
d_{ret}	Maximum distance between molecules for Van der Waals forces	m
d_{co}	Cutoff distance between molecules for repulsive interactions	m
e_{vdW}	van der Waals interaction function	J/m^2
F_{total}	Total normal force	N
F_{applied}	External applied normal force	N
F_{adhesion}	Adhesion force	N
F_a	Pull-off force	N
γ	Work of adhesion	J/m^2
Γ	Surface-interaction energy	J/m^2
t	Thickness of beam	m
h	Initial separation between substrate and cantilever	m
s	Deflection length of a cantilever	m
F_{f}	Friction force	N
a	Modified Hertzian contact length	m
μ	Coefficient of friction	–
L	Applied normal force	N
K	Effective elastic modulus of a material couple	N/m^2

23.1 MEMS/NEMS Devices, Applications, and Their Reliability Issues

The miniaturization of devices continues to fuel the integrated circuit industry. The fabrication of micro-/nanosized complex structures and devices have been made possible thanks to the advancement in lithographic techniques. Micro-/nanoelectromechanical systems (MEMS/NEMS) form an integral part of the microelectronic industry. In 2000, the MEMS/NEMS industry was worth approximately $15 billion and with a projected 10–20% annual growth rate, it is expected to be worth more than $100 billion by the end of this decade [1]. MEMS/NEMS devices are a group of products ranging in size from a micrometer to a centimeter that combine mechanical and electrical structures. They may also consist of micromechanical components such as comb drives, microgears, microlevers, etc., which move to perform certain tasks, and microelectronic components to control motion or to obtain information from that motion [2]. To give an idea of their size and complex structure, a spider mite next to a MEMS device with a series of comb drives is shown in Fig. 23.1 [3]. MEMS technology is a general term used for materials and processes required to make MEMS components, the integration of such components to make devices (e. g., sensors, actuators), and their applications. MEMS/NEMS find a wide range of applications in engineering fields ranging from electronic devices, space technology to biological sciences due to their existence in many forms and the ability to engineer these devices as necessary for given applications [4]. MEMS are commercially used in inkjet printer heads, microwave switches, accelerometers, and sensors.

The future of MEMS/NEMS looks bright as new types of microsystems emerge in drug-delivery systems, optical switches, chemical lab-on-a-chip systems, gas-turbine engines, microgears, chemical sensors, infrared im-

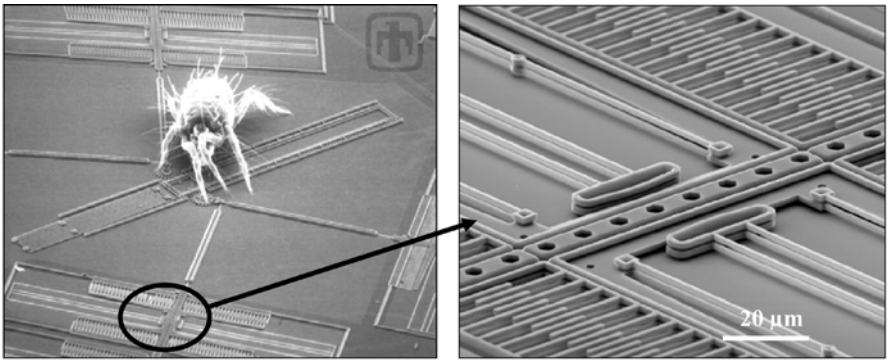


Fig. 23.1. MEMS device along with a spider mite indicating the size of these microcomponents, **b** zoom-in of the picture showing comb drives (series of aligned cantilevers) used in MEMS devices [3]

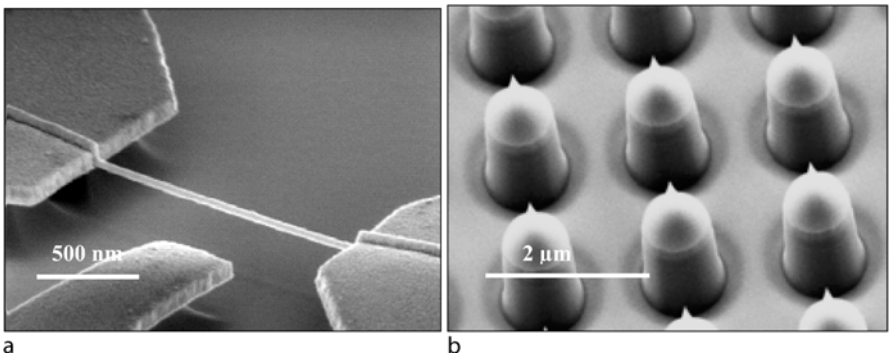


Fig. 23.2. a NEMS-based ultrasensitive sensor [6] and **b** protein motors attached to a substrate with precision [7]

agers, etc. [5]. Nanoelectromechanical systems (NEMS) are future devices that are further miniatures of MEMS in which typical dimensions of the device structures range in nanometers. Recent examples of NEMS devices are shown in Fig. 23.2. Sensitive sensors (Fig. 23.2a) are being developed, like force and mass detection down to the molecular level, high-frequency resonators (GHz range), and ultrasensitive low-power switches [6]. Molecular gears (Fig. 23.2b), motors and nanobearings and engines are future devices that will be used in medical treatments inside the human body [7]. Even though MEMS/NEMS offer a wide spectrum of applications, their poor reliability is a major hindrance to their commercialization. Most of these devices suffer from a range of reliability issues depending on the conditions under which they are manufactured, and operated. In recent years, the reliability of MEMS has gained importance and researchers around the world are putting a lot of effort into unraveling these failure issues. MEMS technology is still in its infancy, and not much has been done concerning their reliability [8].

Hereafter, tribological issues like adhesion, friction and wear that are hampering the commercialization of MEMS/NEMS devices are addressed. Recent advancements made in evaluating MEMS reliability and ways to mitigate these disastrous effects are reviewed. On miniaturizing any device or system, it is critical to have a good understanding of the scaling properties of the overall design, materials, and the fabrication processes involved. The scaling properties related to any one of these components could present a formidable barrier to achieving adequate performance or economic feasibility [9]. MEMS are designed with some basic parts that appear repeatedly in the structure like cantilever beams, membranes, springs, hinges, gears, etc. A variety of problems arise from the environment and the contact or design under which these devices function. Hence, material-related problems and their failure modes depending on factors like design and operating conditions have always been important for evaluating reliability and performance. MEMS/NEMS devices can be broadly classified into 4 types depending on factors like design

Table 23.1. Types of MEMS devices and their description and operating conditions [10]

Type	Description	Examples
I	Devices with no moving parts	Accelerometers, pressure sensors, ink-jet heads, etc.
II	Devices with moving parts without rubbing or impacting surfaces	Gyros, comb drives, resonators, RF MEMS
III	Devices with impacting surfaces	Relays, valve pump
IV	Moving parts with impacting and rubbing surfaces	Shutters, scanners, microgears, electrostatic rotors, optical switches

and contact mode, as shown in Table 23.1 [10]. Based on this classification, material failure mechanisms can be fracture, creep, tribological issues like stiction, friction, and wear, delamination, and other factors like electromigration, corrosion, pitting, and contamination [8]. In the following section only components operating in intermittent and continuous contact conditions are addressed (Types III and IV) where tribological issues like stiction, friction, and wear are of prime concern.

23.2 Tribological Problems in MEMS/NEMS

A major challenge for MEMS designers is to overcome the effects of stiction. As the name suggests, stiction is the effect where microscopic structures tend to adhere to each other when they come into contact [11]. MEMS engineers employing surface micromachining frequently encounter a fatal stiction effect when they attempt to release the structures in the final step of processing. A practical example of stiction faced in comb drives is shown in Fig. 23.3a: a free-standing cantilever sticks to the substrate due to stiction in the final release step [12]. With the increasing complexity of devices, an in-depth investigation of this phenomenon is essential to counteract it. Stiction is also a leading cause of failure in many MEMS/NEMS applications including accelerometers used in air-bag devices in automobiles [13] and digital micromirror devices (DMDs) used in commercial digital light processing (DLP) equipment [14].

The friction force is still the limiting factor to a successful operation and the missing reliability of MEMS having parts in relative motion to each other. Micromotors, microgears, and microturbines are examples of MEMS that operate in contact mode. For example, a typical microgear unit may rotate at a very high speed up to 250,000 rpm. The damage resulting from friction and wear between contacting surfaces at various locations of such a microgear after its use is illustrated in Fig. 23.3b. Friction and wear are the dominant degradation mechanisms noticed in microgears [17]. These devices are not

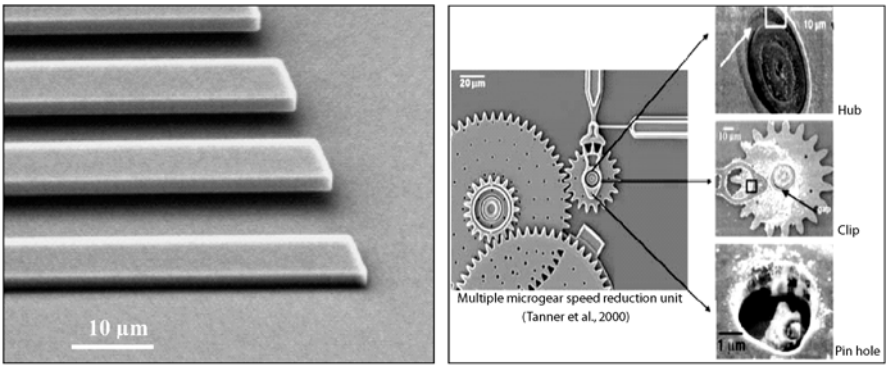


Fig. 23.3. Tribological problems encountered in MEMS: **a** stiction problem resulting in the collapse of free-standing cantilever structure in a comb drive [15] and **b** friction and wear problems resulting in severe damage of microgears that operate at 250,000 rpm [16]

yet commercialized due to their poor reliability as a result of friction and wear. Since the advent of the first micromotors in the late 1980s, much work has been dedicated to the reliability of electrostatic motors (Fig. 23.4) but the understanding of friction and wear in such devices is still insufficient. A failure is said to occur when a micromotor or a system no longer performs the required functions under the special conditions within the stated period of time. Electrostatic motors suffer from a variety of tribological problems. The intermittent contact at the rotor/stator interface and physical contact at the hub/flange interface result in friction and wear. In a micromotor, the rotor is driven electrostatically in the stator. Adhesive wear and abrasive wear often occur between the rotor and the ground plane. Stiction is a problem often occurring between the rotor and stator, rotor and the substrate. Moreover,

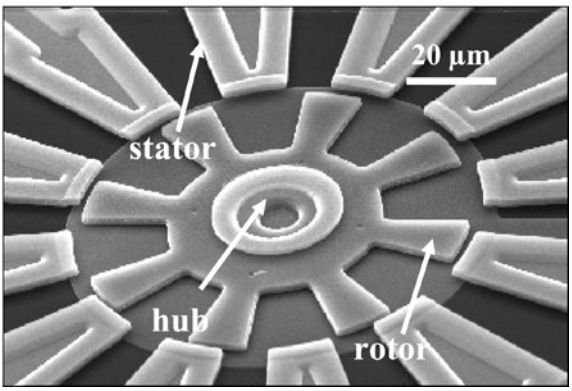


Fig. 23.4. Electrostatic motor [19]

a coefficient of friction as high as 0.5 – 1.1 has been recorded in linear stepper motors [18].

To date, most of the microelectromechanical system (MEMS) devices are based on silicon technology. This is due to the large technological view accumulated on manipulating, machining, and manufacturing of silicon. Silicon has a poor fracture toughness ($0.8\,\mathrm{MPa/m^{-1/2}}$) and undergoes a severe wear and degradation in contact MEMS [20]. This has opened a new area of research in tribology namely special coatings that can be used to protect silicon. As the devices range in micrometers, the coatings or the protective layers on silicon should be only a few monolayers or nanometers thick. Notwithstanding this, the tribological characterization of coatings and materials using special techniques and the fundamental investigation of tribological mechanisms in such materials have gained importance. In the following chapter, the various available techniques are discussed, and materials that were developed successfully to mitigate various tribological issues of MEMS/NEMS devices are reviewed.

23.3 Tribological Evaluation of Materials for MEMS/NEMS

23.3.1 Background on Adhesion, Friction and Wear at Nano-/Microscales

Friction and wear are the two system properties that significantly influence material losses in contacting materials. In earlier years, friction was considered to be a dissipation process associated with large machinery and moving components. With the advancements in technology it was realized that friction could wreak havoc at the micro/nanoscale. Wear resistance depends mostly on the mechanical properties of materials, whereas friction is a complex phenomenon that depends on parameters belonging to the material and to the system. Surface, physical, mechanical, and chemical properties influence friction. Friction at high normal forces is dominated by mechanical aspects like plastic deformation, plowing and chemical aspects like wear-debris chemistry, triboreactions, etc. At micro-/nano-Newton normal forces, these properties are of different relevance because the contact pressures and the size of the contact are small. Adhesion forces arise from van der Waals forces, and capillary forces acting between two contacting surfaces. Adhesion forces depend on surface reactivity, and the affinity to water (hydrophobic/hydrophilic behavior) of the interacting surfaces [21]. Adhesion forces in turn affect friction at micro/nanoscales. As the size of component decreases to micrometers and nanometers, surface-related properties like roughness, adhesion, capillary, and electrostatic forces play a major role in affecting the friction force over mechanical properties, unlike in large components [22]. Conventional tribological tests, such as pin-on-disk or reciprocating fretting equipment,

are associated with large quantities of wear that are unrealistic for real engineering applications. They are no more suitable to evaluate materials in microelectronic applications where the contact areas involved are hundreds of nm^2 and contact loads are in the μN or mN range [23].

Mechanisms and dynamics of interactions between contacting solids in relative motion, ranging from atomic- to micro- load and length scales must be understood to develop fundamental understanding of adhesion, friction, wear, indentation, and lubrication processes occurring in miniature devices. Another way to understand tribological phenomena is by analytical models that help in predicting the sequence of events. Analytical modeling in tribology is still in its infancy but it is promising. Efforts are being made to model tribological phenomena by molecular dynamics and finite elements. The biggest challenge in atom-based computational techniques is to simulate friction and wear in traditional engineering applications with respect to both size and time scales. Explicit atomistic simulations on rough surfaces are still intractable, and investigation carried out with interfacial velocities below 1 m/s requires prohibitively long run times. The simulation cell size in analytical models is barely large enough to adequately model a single asperity contact [24]. Because of these constraints, experimental methods and appropriate techniques are still of prime interest to evaluate materials at low normal forces and small contact areas as prevailing in MEMS/NEMS components.

23.3.2 Techniques for Tribological Characterization of Materials

Scanning probe microscopy (SPM) has opened a new era of science called nanotechnology. High-resolution microscopes enable atoms and molecules to be seen. Scanning tunneling microscope (STM) and atomic force microscopes (AFM) are broadly used microscopes belonging to this branch of SPMs. With the introduction of the first friction measurement technique using an AFM by Mate et al. [25] in 1985, known as friction force microscopy (FFM) or lateral force microscopy (LFM) using an atomic force microscope (AFM), a new branch of science known as nanotribology emerged. This branch aims at studying tribological properties like friction, wear, adhesion, and lubrication at nanocales. AFM simulates a single asperity contact and thereby enables fundamental understanding of tribological phenomena at nanoscales. AFM is also being used for quantifying adhesion, friction, scratch resistance, wear resistance, and nanohardness. Its ability to unveil topographical information with subnanometer resolution makes it perfect surface characterization equipment for materials used in microdevices. Apart from AFM, surface-force apparatus (SFA), developed in 1968, is commonly employed to study both static and dynamic properties of molecularly thin films sandwiched between two molecularly smooth surfaces [26]. Recent developments on pin-on-disk and fretting equipments are also used for the tribological characterization of surfaces and coatings that are operated at low normal force and contact

pressures simulating contact pressures prevailing in real applications. Apart from laboratory-level tribological tools, characterization tools like cantilever beam array (CBA), inchworm, side-wall friction measurement devices, etc., are also used to evaluate tribological properties of devices [27].

23.3.3 Tribological Evaluation of Materials

Quantitative Measurement of Adhesion/Stiction

Adhesion is an important cause of failure in MEMS/NEMS with compliant structural members. When the surface to volume ratios of structures become large, surface forces dominate inertial forces and adhesion becomes important. Adhesion can be classified into "in-process adhesion" and "in-use adhesion". The "in-process adhesion" refers to the sticking of movable elements typically due to drying after wet chemical etching of the sacrificial layers. The "in-use adhesion" refers to the collapse of initially freely movable structures during handling or in operation [8]. The adhesion force is the source of stiction. Therefore, adhesion is by far the most important parameter that must be minimized to improve the reliability of MEMS/NEMS operating in intermittent or continuous contact modes. A systematic study on adhesion between two surfaces was first done by Bowden and Tabor in 1950 [28]. In a dry environment no adhesive force was measured, but in a humid environment a measurable force holds the two surfaces together. Water vapor has a tendency to condense in small cracks and pores to form a thin layer due to capillary condensation. This water generates a force that one encounters when trying to separate two macroscopic wet glass plates. Hence, hydrophilic surfaces suffer badly from capillary forces in humid environments. Also, high free-surface energies lead to high adhesive forces. When a wet ball is pressed against a flat surface and the radius of the meniscus "r" is smaller than the radius of the ball "R", the pressure inside the liquid is smaller than the atmospheric pressure by the ratio γ_l/R, where γ_l is surface tension of the water. The adhesive capillary force, F_{cap}, is then:

$$F_{\text{cap}} = 4R\pi\gamma_l \tag{23.1}$$

A roughening of the ball surface reduces the adhesive force. However, this is applicable only up to a certain level of relative humidity after which adhesion remains the same for both rough and smooth balls due to attainment of saturation. Water at atmospheric pressure is highly responsible for stiction and silicon as well as its oxide are hydrophilic in nature and hence small traces of water on the surface after etch release can lead to stiction. The capillary condensation of water vapor occurs when the surfaces come closer than a characteristic distance $z = d_{\text{cap}}z$ being the separation distance between the surfaces. Maboudian and Howe [12] calculated the energy required to pull the two surfaces apart against the capillary forces between them. The

surface interaction energy due to capillary forces, $e_{\mathrm{cap}}(z)$, that equals work of adhesion W, can be expressed as:

$$\langle e_{\mathrm{cap}}(z) = 2\gamma_l \cos\theta|_{z\leq d_{\mathrm{cap}}} \tag{23.2}$$

$$\langle e_{\mathrm{cap}}(z) = 0|_{z>d_{\mathrm{cap}}} \tag{23.3}$$

The function $e_{\mathrm{cap}}(z)$ does not depend on the exact distance as long as the surfaces are closer than d_{cap} and the contact angle θ is a constant for a given surface. Both γ_l and d_{cap} are functions of temperature and d_{cap} in turn is also a function of relative humidity. The surface interaction energy can thus be manipulated by changing temperature and relative humidity. The above expressions were deduced assuming surfaces to be flat. Hence calculated forces from these equations may vary for real situations. The van der Waals interaction forces are short-range attractive forces that operate when two surfaces are brought extremely close to each other. The molecular van der Waals forces are caused by mutually induced dipoles in molecules. The interaction energy due to van der Waals forces depends on the medium between the surfaces and the distance between them. The force description becomes complex in a media other than air and vacuum. The energy due to van der Waals, e_{vdW}, forces is expressed as:

$$\langle e_{\mathrm{vdW}}(z) = \frac{A_{\mathrm{Ham}}}{12\pi z^2}|_{d_{\mathrm{co}}<z<d_{\mathrm{ret}}} \tag{23.4}$$

$$\langle e_{\mathrm{vdW}}(z) = 0|_{z>d_{\mathrm{ret}},z<d_{\mathrm{co}}} \tag{23.5}$$

with A_{Ham} the Hamaker constant of the molecule. d_{ret} the retarded distance after which force does not act, and d_{co} the cutoff distance close to the molecule below which repulsive forces start to operate. The Hamaker constant depends on the molecular surface termination of the approaching surfaces. The Hamaker constant for most nonpolar molecules lies in the range of $0.4-4\cdot 10^{-19}$ J [29]. Stifter et al. [30] studied the distance dependence of the meniscus and van der Waals forces to determine their relative importance at various operating conditions. At larger distances, the meniscus force is always stronger than the van der Waals force. The adhesion contributes to friction between surfaces when operated under low normal forces as in many MEMS. Adhesion force acts as an additional force to the applied external normal force and thereby results in higher friction [31]. This can be mathematically expressed as

$$F_{\mathrm{total}} = F_{\mathrm{applied}} + F_{\mathrm{adhesion}} \tag{23.6}$$

with F_{total} the total normal force, F_{applied} the external applied normal force and F_{adhesion} the adhesion force.

AFM is a commonly used tool to determine adhesion force on surfaces at nanoscale with a force resolution of in the range of pico-N. The sensing probe has a nanometer-dimension tip usually made of silicon nitride. The adhesion force is calculated from an approach-retraction cycle between tip and

a surface, known as "force-calibration curve". The adhesion force is calculated by multiplying the maximum vertical deflection of the cantilever with the cantilever stiffness. The measured adhesion force is also known as the pull-off force. Precise knowledge of the cantilever stiffness is thereby essential to obtain reliable results. According to the Johnson–Kendall–Roberts (JKR) theory [32], the pull-off force, F_a, is related to the work of adhesion as:

$$F_a = \frac{3}{2}\pi\gamma R \tag{23.7}$$

with γ the work of adhesion expressed in $\mathrm{J/m^2}$, and R the radius of the AFM tip. Adhesion measurements with an AFM are a versatile technique because the tip can be coated with virtually any material of interest. Biomolecules, self-assembled monolayers were coated on AFM tips to investigate the interaction forces between various molecules [33]. The adhesion force in general decreases with increasing surface roughness because the true contact area of interaction between the surfaces decreases [34]. Using colloidal probes of different radii and roughness, Yang et al. [35] proved that the adhesion force does not vary linearly with the probe radius, contrary to the JKR theory due to surface roughness. This suggests that a quantitative adhesion calculated from AFM does not take into account surface roughness. It is unlikely that microcomponents possess a single asperity contact (i. e. no surface roughness). De Boer et al. [36] estimated that for planar polysilicon surfaces, the actual contact area is approximately eight orders of magnitude smaller than the apparent area, and individual contacts are tens of micrometers apart. Therefore adhesion measurements by AFM are limited to the fundamental understanding and comparison of different materials. Alternative measurement tools have been developed to take into account surface roughness in quantifying adhesion in real microsystems as MEMS.

Mastrangelo and Hsu [37] developed the cantilever beam array (CBA) technique to measure adhesion taking into account surface roughness. The setup consists of a 20-µm wide array of cantilever beams that extend from 150 – 1700 µm in length, and are 2 µm above the surface. The beam lengths are incremented by 50 µm. These beams are brought into contact with the underlying surface by electrostatic actuation or by pushing it down with a probe needle. From a measurement of the length over which the beam is not adhering to the surface, the interaction energy can be calculated. Two modes of cantilever deflection were described in the literature, arc-shaped or S-shaped. S-shaped beam deflections give adhesion measurements that are statistically more reliable than arc-shaped ones due to a substantial contact area. The contact length of the beams that adhere to the substrate is determined by differential interference contrast microscopy. The surface energy of the S-shaped beam is calculated from Eq. (23.8):

$$\Gamma = \frac{3}{8}E\frac{t^3h^2}{s^4} \tag{23.8}$$

Table 23.2. Adhesion measurements on various materials measured using AFM and cantilever beam array method [39–41]

AFM technique			CBA technique	
Material couple	Adhesion force (nN)	Work of adhesion (mJ/m^2)	Material tested	Work of adhesion (mJ/m^2)
Si_3N_4 tip / Si	50, 80, 52, 33	530	Silicon oxide (SiO_2)	> 8
Si_3N_4 tip / SiO_2	35	370		
Si_3N_4 tip / carbon	33	350	Octyltrichorosilane (OTS)	0.012
Si_3N_4 tip / Au	14	148	Octadimethylsilane (ODMS)	0.045
diamond tip / Si	17.4	36.9	Perfluorodecyltrichlorosilane	0.005
diamond /carbon	10.9	23.1	(FDTS)	

with Γ the surface interaction energy, t thickness of the beam, h the initial separation between the substrate and the cantilever, and s the deflection length of the cantilever. Using the CBA technique, De Boer et al. [38] proved that theoretically calculated surface interaction energies are much higher than experimental values. This was attributed to surface roughness, not included in most adhesion models. Recently, a comprehensive stiction model was proposed by Van Spengen et al. [8] taking into account surface roughness. The theory assumes surface asperities with a Gaussian distribution. However, the applicability of theoretical models to adhesion is still limited and much still has to be done to formulate a unified model because surface asperities are rarely Gaussian, and the contact mode can be elastic, plastic, or, as in most cases, a combined elastic and plastic contact mode depending on the geometry, applied load, and distribution of asperities. Adhesion force and work of adhesion of some MEMS/NEMS materials calculated from AFM pull-off curves and CBA method are summarized in Table 23.2.

Four different adhesion values reported in the literature on Si (001) are mentioned in Table 23.2. This scatter is understandable because most surfaces are nonhomogeneous at the nanoscale, and since surface preparation can vary from one laboratory to another. Such a scatter in adhesion data obtained with AFM is associated to differences in surface condition, calibration technique used for cantilever stiffness and piezo, etc.

Quantification of Friction at Micro-/Nanoscales

Friction is hard to characterize due to its dependence on chemical, mechanical, environmental and physical aspects of a material couple. Control of friction is, however, essential for the development of contact MEMS. For example, in situ friction measurements on a microengine running at 18,300 rpm

revealed a coefficient of friction of 0.5 for normal forces less than 4 μN [42]. Such a high coefficient of friction must be minimized to increase the reliability. In nanoscopic contacts under low normal forces, adhesion dominates the friction force. Santner et al. [43] conducted FFM measurements on a Au (001) surface with a Si tip and measured the nanoscopic frictional forces as a function of the applied normal force. A nonzero friction force was noticed even in the absence of any external applied normal force. This nonzero frictional force (finite coefficient of friction) in the contact arises from adhesion. These observations are in contradiction with the empirical Amontons' law that predicts an infinite coefficient of friction. Carpick et al. [44] observed a finite friction force for negative applied load between a Pt-coated AFM tip and a mica surface. The friction force between a Pt-coated AFM tip and mica in ultrahigh vacuum (UHV) varies with load in proportion to the contact area, as predicted by the JKR theory. The above observations revealed that the adhesion force in the contact acts as an additional normal force. The contributions to the contact load from adhesion are described in detail by theories like the JKR [32] for short-range adhesion forces between compliant materials, and Derjaguin, Muller, and Toporov (DMT) [45] for long-range forces acting on stiffer materials. The JKR theory was found to fit with the measured friction versus normal force curves at nanoscales. According to the JKR theory, the total friction force, F_{f}, in the contact and the modified Hertzian contact radius, a, are:

$$F_{\mathrm{f}} = \mu\left(L + 3\pi R\gamma + \sqrt{6\pi RL\gamma + (3\pi R\gamma)^2}\right) \tag{23.9}$$

$$a^3 = \frac{R}{K}\left\{L + 3\pi R\gamma + \sqrt{6\pi RL\gamma + (3\pi R\gamma)^2}\right\} \tag{23.10}$$

$$\left\langle \begin{matrix} F_{\mathrm{f}} = \mu\,(6\pi R\gamma) \\ a = \left(\frac{6\pi R^2\gamma}{K}\right)^{1/3} \end{matrix} \right|_{L=0} \tag{23.11}$$

with μ the coefficient of friction, L the applied normal force, R the radius of the contacting body, K the effective elastic modulus of the material couple, and γ the adhesion energy per unit area also known as the surface interaction energy. This γ is equivalent to Dupre's energy of adhesion [8]: $\gamma = \gamma_1 + \gamma_2 - \gamma_{12}$. At $\gamma = 0$, Eq. (23.9) reduces to the Hertzian expression. At an external load equal to zero, Eq. (23.9) and (23.10) reduce to Eq. (23.11). The existence of adhesion was also confirmed at micro-Newton (μN) normal forces by Ando et al. for steel against steel at normal loads ranging from 0.8 μN to 3000 μN [46]. The coefficient of friction increased with decreasing normal force but, when the same friction force is divided by the applied normal force plus the adhesion force, a constant coefficient of friction was obtained. Therefore, at low normal forces, the coefficient of friction can be apparent (calculated from the applied normal force alone) and real (calculated from adhesion + applied normal force). Thus, adhesion is a significant factor that must be controlled to reduce both stiction and friction force in mi-

crocomponents. The true contact area between surfaces affects both adhesion and friction. The true contact area can be defined as the sum of individual microscopic areas at contact points between surfaces. The friction force increases with increasing true contact area. As the true contact area decreases, the interaction zone between two surfaces becomes small and the adhesion reduces. At the macroscale, surface asperities are leveled after the first few initial contact events due to high contact pressures. On the other hand, at low contact pressures, as in the case of microcomponents, the roughness greatly affects adhesion and therefore friction.

The measurement of friction with an AFM is known as lateral/friction force microscopy (LFM/FFM). The force sensor in FFM is a microfabricated cantilever ending with a sharp tip. The tip is brought into repulsive contact with the counterpart. The load is derived from the normal deflection of the cantilever. On sliding, the lateral force acting on the tip is derived from the twist on the cantilever. Bending and twisting are usually measured by sensing with a laser beam reflected on the backside of the cantilever. The applied normal force is derived from the cantilever stiffness. Hence, accurate cantilever stiffness values are desired for reliable friction data. Apart from AFM, a surface-force apparatus (SFA) is used for characterizing surface interactions at atomic scale. A quartz crystal microbalance (QCM) is another technique used to measure frictional forces between surfaces and thin adsorbed layers [47]. SFA and QCM are suitable for fundamental investigation of tribological phenomena and hence seldom used for materials characterization on a regular basis. Sundararajan and Bhushan [48] successfully determined, by AFM, the static friction forces encountered in a surface micromachined micromotor. The AFM tip was pushed against a rotor arm of the micromotor so as to generate a lateral deflection of the tip measured by the AFM lateral deflection signal. The maximum lateral deflection obtained prior to the rotor movement is a measure of the static friction of the micromotor. The same technique was used to investigate the effect of humidity and rest time on the static friction force of polysilicon motors using polyfluoropolyether as a solid lubricant.

Using FFM, Liu and Bhushan [49] successfully characterized the stiction and friction with a digital micromirror device (DMD) used in lightweight projection displays. The DMD chip is an array of two million independently controlled reflective aluminum alloy micromirrors (14 μm square and 15 μm pitch) that switch forward and backward at a frequency of 5 to 7 kHz integrated onto a silicon CMOS static random access memory (SRAM) array. The micromirror rotates as a result of electrostatic attraction between the micromirror structure and the underlying electrodes. A schematic view of two pixels of a DMD is shown in Fig. 23.5. The micromirror is attracted to the side towards the landing site with the largest electrostatic field differential. To release the micromirror from the landing site, a short reset pulse is applied and the bias voltage is removed. The lifetime of micromirrors is a challenge for the commercial success of DMDs. At an operating frequency

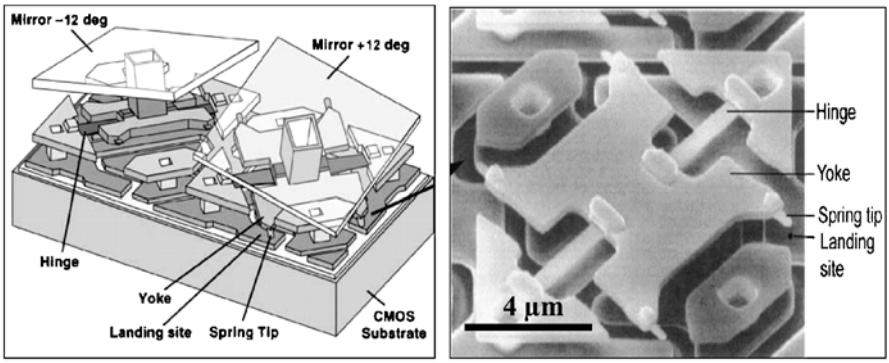


Fig. 23.5. Schematic picture of **a** 2 micromirror pixels and **b** SEM image of a pixel in digital micromirror device (DMD) [49]

of 7 kHz, each micromirror element has to switch about 2.5 trillion cycles. Stiction and wear between the spring tips and the landing sites are important issues that affect its reliability. Liu and Bhushan [49] presented a method to analyze and compare the tribological properties of a defective mirror with normally functioning mirrors. Using AFM adhesion and friction maps, they concluded that the adhesion and friction on the landing site of stuck mirrors (defective) are higher than of properly functioning mirrors. They found a significant influence of capillary forces on stiction between mirrors and the landing sites.

Limitations of Tribological Labscale Tests

The quantitative data obtained from FFM measurements do not represent data during real applications even though FFM is a good technique for a fundamental study and a comparison of materials. The contact pressures in FFM are high, namely in the range of GPa, whereas MEMS have maximum operating pressure of 400 MPa. Single-asperity contact in FFM eliminates the effect of roughness and oversimplifies the real contact situation in MEMS. LFM/FFM measurements are extremely sensitive, and a meticulous calibration of the cantilever stiffness and force-conversion factor is essential. It was proved by Karuppiah and Sundararajam [50] that the same material tested with different lateral force calibration techniques gives rise to different results. Moreover, nanotribology techniques like AFM, QCM and SFA operate at sliding speeds that are too slow compared to the typical micromachine operating speeds. From Fig. 23.6 it appears that the conventional tribological equipments can better simulate the contact pressures and velocities existing in MEMS, but this is not the case. Indeed, large contact size and nonisothermal heat generation, unlike in MEMS, limit the usage of conventional equipment. Coming to analytical models and atomistic simulations, there is no means to

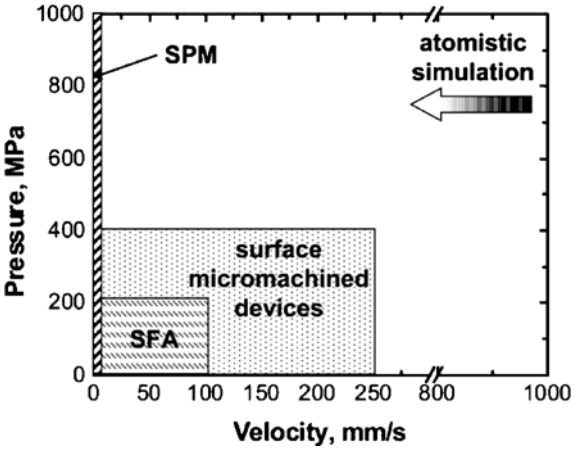


Fig. 23.6. Contact pressure and sliding speed of nanotribological techniques or atomistic simulations compared to surface micromachined devices [27]

extrapolate AFM data and to simulate tribological events at the dimensions and operating conditions of real devices [27]. On the other hand, atomistic simulations are normally performed at sliding speeds above the operating speeds of MEMS devices in the range of thousands of mm/s [51, 52].

Notwithstanding this, new tribological tools are developed using microfabricated devices with a bulk microstructure, surface topography, and a surface chemistry that closely represent real contact situations in MEMS. Lim et al. [53] developed a planar surface friction apparatus for measuring friction at microscales. The device uses an electrostatic comb drive to move a shuttle laterally. The contact occurs between a planar surface and dimples on the bottom of the shuttle. They obtained a very high coefficient of friction of 4.9 ± 1.0 between polysilicon surfaces, and 2.5 ± 1.0 between polysilicon and silicon nitride tested in ambient air. A continuous sidewall tribometer was developed by Senft and Dugger [54], which allows in situ measurement of the kinetic friction between MEMS sidewall contacts while operating at realistic speeds for millions of sliding cycles. The latest addition to characterize friction between micromachined surfaces is an inchworm tribometer developed by de Boer et al. [55]. The device consists of planar polysilicon rails that make contact with a planar track. Parallel-plate capacitors with large area can be used to tailor the size of the contact. This instrument can operate under loads in the range of μN, and therefore can simulate contact conditions prevailing in microdevices.

Evaluation of Wear Resistance by Commonly Used and In Situ Measurements

Materials or coatings used in contact MEMS/NEMS applications must have a good wear resistance apart from low friction. Wear-resistant materials form a basis for achieving reliable contact MEMS like microgears, micromotors,

etc. In macroscale tribological tests, it is customary to quantify wear loss by measuring the wear volume. A novel way of representing the wear rate is by expressing the wear volume generated per unit dissipated energy that predicts the wear rate of any given material couple [56]. However, at micro-/nanonormal loads, it is hard to accurately determine the wear volume because wear occurs only on a few atomic layers. Hence, wear is characterized and quantified using topographical changes. The atomic force microscope (AFM) is undisputedly the best available equipment for an in-depth investigation of wear mechanisms. There are various methods to characterize the wear resistance of a material, like measuring the critical load, recording dimensional changes of a reference point with time, measuring the maximum wear depth after certain duration of a test, and time required for complete removal of a coating. AFM is highly used for characterizing wear at micro-/nanoscales. AFM is operated under high normal forces using stiff cantilevers with a diamond tip. The wear tests conducted using an AFM are thus accelerated ones due to high contact pressures. Presently, a new generation of tribometers, like, e. g., the MUST microtribometer, (producer Falex Tribology N.V. Belgium) that can operate in milli-/micro-N normal forces are also used for evaluating frictional wear of thin films [57].

In wear tests performed with AFM, the diamond tip of the AFM cantilever is brought into contact with the counter material at a high normal force and the surface is scratched for a given period of time. Afterwards, the profile of the scratch mark is scanned with the same tip but at a light load. The wear depth is then determined using a 2-D profile of the wear scar (Fig. 23.7 [58]). There are two scratching methods: line-scratching (scanning repeatedly over a line) and scanning scratching (scanning over an area) [59]. Using this technique, Miyake and Kaneko [60] showed that a fluorinated Si–

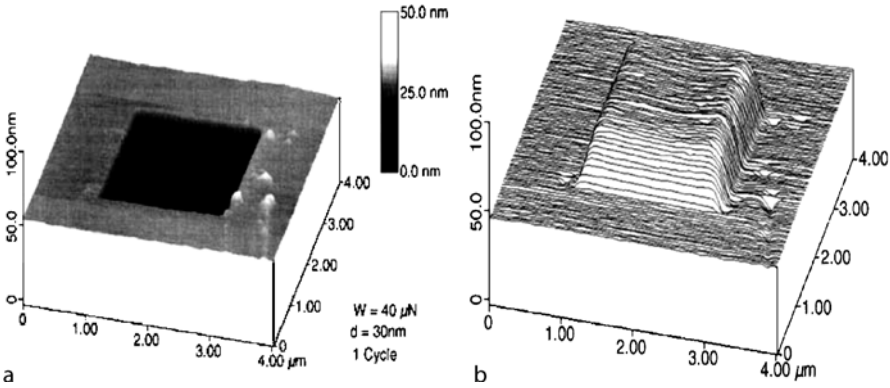


Fig. 23.7. a AFM image of wear scar on Si (001) surface after scanning wear test with 30-nm diameter diamond tip at 40 µN normal force after one scan, i. e. 1 cycle **b** inverted image of the wear track [58]

C film has a 30-times better wear durability than conventional amorphous carbon films. Sundararajan and Bhushan [61] used AFM for studying the dependence of wear resistance of DLC films on thickness. They found that the wear resistance of DLC films is directly proportional to their thickness. AFM is now used to characterize the wear resistance of SAM monolayers for MEMS/NEMS applications. Using this technique, it is possible to quantitatively determine the critical load at which a SAM coating fails.

Machcha [62] proposed another method for accurately measuring the wear rate at nanoscale. This method relies on 20 – 30 nm deep, micrometer-sized indentations made on the sliding surface that act as reference points. Changes in the size and depth of these indentations as a function of sliding time are monitored using an AFM. This method was used to quantify wear on a contact recording head consisting of diamond-like carbon pads used in magnetic hard disks. Wear rates of 4.5 nm per week were reported for constant speed test at $7.84\,\mathrm{m\,s^{-1}}$ with $350\,\mu\mathrm{N}$ load and the wear coefficient using Archard's law was found to be 0.2×10^{-10}.

The wear resistance of coatings can be measured in terms of the coating lifetime by observing the evolution of friction force with time. When a coated substrate is tested, a sudden change in the friction force indicates coating wear-through. The coating durability is reported in terms of the sliding distance or by total sliding time before wear-through [63]. This method is appropriate for comparing thin films. Conventional sliding equipments like pin-on-disk and fretting equipment are used for this method and operated under low contact pressures and sliding speeds as existing in microdevices. Tao and Bhushan [64] successfully used this technique for determining the wear resistance of polyfluoropolyethylene (PFPE) solid lubricants in different environments for MEMS/NEMS applications. Beerschwinger et al. [65] showed the possibility of using conventional pin-on-disk equipment for simulating wear in microdevices. They used specially designed triangular macroscopic specimens with 1.5 mm side length on which circular-shaped microstructures were micromachined and coated with different coatings. The contact pressures can be varied from 10^{-6} MPa to 2 MPa depending on the contact area, as in MEMS. The wear rate was determined by measuring the decrease in the length of circular projections. Achanta et al. [66] elaborated a method for determining nanowear by first performing a wear test on coatings at low normal forces using a MUST microtribometer and then characterizing wear by observing changes in surface topography using an AFM.

Apart from laboratory tests, in situ wear measurements are carried out on microdevices directly. For example, in situ wear measurements on polysilicon micro turbines were carried out by Gabriel and Bell [67] that were spun at 10 000 rev/s. The reported typical lifetime of these components was nearly one million cycles at 5000 rev/s. Mehregany et al. [68] carried out wear studies on a wobble micromotor by comparing its gear ratio before and after wear test. Indeed, an increase in bearing clearance due to wear will lead to a decrease in gear ratio. Nearly 100 million cycles, equivalent to a rolling distance of 11

km, were achieved with these micromotors. Although in situ measurements give a direct insight on the performance of a device, such tests are expensive and time consuming. Therefore, laboratory-scale tribological tests act as a preliminary materials-selection process and later in situ measurements can be performed to verify their real-time performance.

23.4 Prospective Materials

In recent years, many methods were addressed in order to counteract tribological problems in microdevices. Modifying the surface properties of materials can mitigate adhesion. Common ways to reduce adhesion are: (i) surface roughening or surface texturing, (ii) hydrophobization, e. g., with surface coatings, and (iii) special drying techniques after etch release step, e. g., critical CO_2 drying. Inducing roughness on a surface by chemical means was one of the solutions used for solving adhesion problems. When a surface is roughened, the true contact area between the contacting surfaces lowers and the surfaces have less chance to interact. Romig et al. [27] reported that surface roughening of silicon oxide resulted in a decrease of the work of adhesion from $20\,\mathrm{mJ/m^2}$ down to $0.3\,\mathrm{mJ/m^2}$. Houston et al. [69] roughened the silicon using ammonium fluoride and terminated the surface with hydrogen bonds. This surface modification resulted in a work of adhesion less than $0.3\,\mathrm{mJ/m^2}$. A more recent method of surface roughening was adapted from a lotus leaf [70]. A lotus leaf is waxy and consists of microbumps (Fig. 23.8a [71]) well dispersed on its surface. If a water droplet falls on its surface, the water drop rolls off because the bumps are much smaller than the droplet and so close together that they cannot puncture the droplet. Scientists are modeling the locations of such bumps so that they can be fabricated

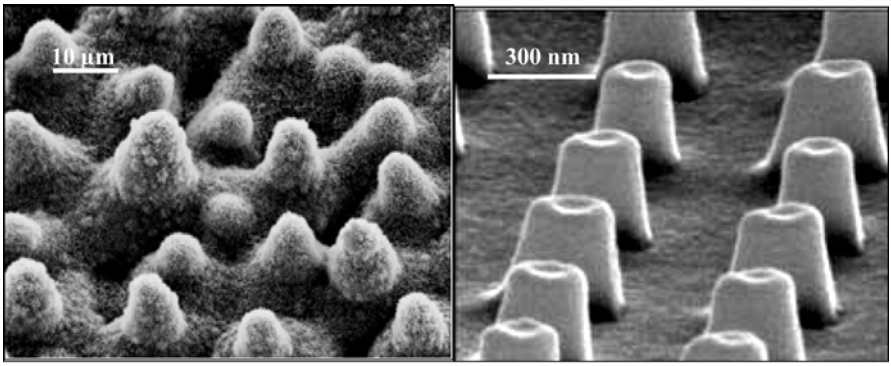


Fig. 23.8. **a** SEM image of a lotus leaf [71] and **b** artificially grown PMMA SAM bumps on silicon wafer [72]

onto the surface of a micromachined device to obtain maximum hydrophobicity. Recent experiments on such lotus-like surfaces have shown promising results. Yoon et al. [72] fabricated nano/micropatterns made of polymethyl methacrylate (PMMA) on a silicon wafer using capillary force lithography (Fig. 23.8b). The patterned samples with certain aspect ratio of the bumps exhibited superior tribological properties at the nano- and microscales when compared to the nonpatterned PMMA film.

Surface roughening is not a complete solution for addressing adhesion issues because most of the chemically roughened surfaces react with water vapor and become smooth over the long term [69]. Apart from this, rough surfaces give rise to high local contact pressures thereby leading to wear in the contact area during operation. Hence, it is extremely important to know the optimum roughness for a given material couple operating under certain contact conditions.

Another way to reduce problems due to adhesion, friction, and wear is by surface modification of materials using special coatings or solid lubricants with low surface energy, low friction, and good wear resistance. Surface modification is an extensively investigated area in MEMS technology [27]. The coatings applied are typically a few molecular layers thick and specially selected such that they do not modify the stress state of the polysilicon substrate. An ideal solid lubricant for MEMS/NEMS must be easy to apply, must chemically bond to micro/nanodevices surfaces, and must be insensitive to the environment.

Much of the earlier attention was diverted towards hard ceramic-based materials like silicon nitride (Si_3N_4), silicon carbide (SiC), titanium carbide (TiC), amorphous carbon (a-C:H), diamond-like carbon (DLC) and diamond coatings [73,74]. Diamond has recently emerged as a promising coating thanks to its highest hardness $\sim$ 100 GPa and high elastic modulus $\sim$ 1100 GPa. DLC coated surfaces sliding against each other have very low coefficients of friction between 0.02 and 0.1 compared to silicon surfaces with coefficients of friction of 0.45 – 0.6. DLC has shown promising results as a bearing material when used as a coating material on single-crystal silicon (SCS) and polysilicon surfaces under large contact stresses. DLC coatings have the disadvantage that their tribological properties are highly dependent on thickness. Thus, DLC coatings can only be applied on components where a large coating thickness can be accommodated. The disadvantage of CVD polycrystalline diamond coating is its high surface roughness (RMS $\sim$ 1 µm) [74]. Most of the ceramic-based coatings are deposited using PVD techniques. Intricate parts of a device remain inaccessible for line-of-sight deposition techniques [74]. Some typical tribological results of ceramic coatings are listed in Table 23.3.

To overcome the problems encountered with hard coatings, research was diverted towards superhydrophobic organic coatings. Typically, these are very thin organic monolayer coatings, either physisorbed or covalently bound to

Table 23.3. Hard coatings used for tribological applications in MEMS [74–79], * deposited by PVD techniques, ** deposited by CVD techniques

Material	Adhesion (nN)	Microscale coefficient of friction using AFM/FFM	Macroscale coefficient of friction using pin-on-disk, fretting equipment	Hardness (GPa)
Silicon Si(001)	50–80	0.04-0.07	0.45–0.6	12
Silicon oxide	35	0.087	0.65–1	–
Silicon nitride*	50	0.06	0.66–0.45	30–50
Silicon carbide*		0.02, 0.06–0.08	0.20	25
Titanium carbide*	–	–	0.16	35
Diamond-like-carbon**		0.02–0.04	0.08–0.1	90 (max)
Diamond**	15	0.01–0.02, 0.05, 0.11	0.20	100
Amorphous carbon**	33	0.02–0.14	0.05–0.20	90 (max)

the surface with water contact angles greater than 90°. These films are self-limiting in thickness, and can penetrate through the liquid or vapor phase to deeply hidden interfaces. There are two types of organic molecules commonly applied on solid substrates, namely Langmuir–Blodgett (LB) films and self-assembled monolayers (SAM). A LB film contains amphilic molecules, i. e. molecules with a hydrophilic head and a hydrophobic tail. A SAM contains organic molecules in a dilute solution that spontaneously adsorb and organize onto a solid substrate to form a self-assembled monolayer. LB films require a flat sample surface, which limits its applicability and SAM, on the other hand, can be easily deposited irrespective of the surface structure. Examples of coatings that were successfully applied in commercial MEMS products include a perfluorodecanoic acid coating on structural aluminum of a digital micromirror device (DMD Texas Ins.) and phenylsiloxane coatings for accelerometers (Analog Devices) that resist a charge build-up and also survive packaging temperatures as high as 500 °C. The tribological behavior of self-assembled monolayers is significantly influenced by their functionality and chain length. Friction tests carried out at nano- and microscales revealed that SAMs with $–CH_3$ functional group exhibit a lower kinetic coefficient of friction and a superior wear resistance than SAMs with –COOH functional groups. Hence, proper knowledge on SAM chemistry is essential to obtain good tribological properties [80].

Presently, organic trichlorosilane ($RSiCl_3$) SAM monolayers like octyltrichlorosilane (OTS) and octadecyltrichlorosilane ($C_{18}H_{37}SiCl_3$, ODTS) are examples of SAM coatings that are highly investigated. The work of adhesion on silicon oxide is 8 mJ/m^2, whereas OTS-coated silicon has a work of adhe-

sion of 0.012 mJ/m^2. Similarly, the coefficient of friction on OTS was 0.073 compared to 1.1 in the case of SiO_2 [40]. The only disadvantage of OTS is its poor thermal stability above 225 °C, whereas in packaging applications the required temperature stability exceeds 400 °C. Therefore, in addition to low surface energy, coatings must be compatible with subsequent device processing, including packaging thermal treatments at 400 to 500 °C. Perfluorinated alkyltrichlorosilanes ($C_8F_{17}C_2H_4SiCl_3$, FDTS), dichlorodimethylsilane ($(CH_3)_2SiCl_2$, DDMS) are examples of monolayers that possess good thermal stability and maintain hydrophobic behavior without dissociating up to 450 °C. FDTS-coated surfaces sliding against each other exhibit a low coefficient of friction at 0.02 compared to OTS-coated surfaces with a coefficient of friction at 0.073. However, FDTS causes a larger particulate contamination and release of harmful chemicals than OTS SAM. On the other hand, DDMS has a good thermal stability but surface properties like high water contact angle (103°), high work of adhesion at 0.045 mJ/m^2, and high static coefficient of friction of 0.28, make it inferior to OTS [81]. Bhushan et al. [41] carried out an extensive investigation of the tribological properties of fluoroalkylsilane and alkylsilane groups using a friction force microscope. They reported that flouroalkylsilanes SAM are superior in terms of wear resistance over alkylsilane SAM. For example, perfluoroalkylsilane (PFTS) has a critical failure load of 56 μN compared to 17 μN for ODMS alkylsilane monolayers. More details on adhesion and wear behavior of different SAMs are given in Table 23.4. SAM coatings suffer from a poor wear resistance since they tend to wear away easily at moderate contact pressures. Each SAM coating has a critical contact load at which it wears off completely [49]. The importance of good wear resistance was illustrated by Liu et al. [49] during LFM/FFM measurements on perfluorodecanoic acid ($CF_3(CF_2)_8COOH$) used in digital micromirrors. Wear on the coating leads to a sudden rise in capillary forces (stiction) and friction, and results in the failure of the micromirror. Polymer coatings like polydimethylsiloxane (PDMS) and polymethylmethacrylate (PMMA) and perfluoropolyether (PFPE) [82] are under investigation as potential candidates in MEMS/NEMS devices (Table 23.4). PFPE is a commonly used lubricant in the magnetic disk drive industry to reduce friction and wear at the head/disk interface [82]. Hybrid coatings like PTFE + Si_3N_4 are being considered because these coatings offer a good combination of low friction and high wear resistance [83]. Furthermore, novel coatings with low friction along with good wear resistance must be introduced. Nanostructured coatings are such materials and can be deposited with desired thickness on surfaces. For instance, "chameleon coating" is a type of coating that contains nanoparticles of hard material like DLC and MoS_x solid lubricant that are embedded in a Au matrix. This coating reacts with the environment and forms lubricious layers in the tribocontact and reduces friction. These coatings are now being used in space applications [84].

Table 23.4. Typical surface properties of on various self assembled monolayer coatings and polymer coatings [40, 41, 81]

Material/coating	Adhesion force (nN)	COF nano/ microscale	COF macroscale	Water contact angle (°)	Surface free energy using CBA (mJ/m^2)	Side-wall coefficient of friction
Si	50–80	0.04–0.07	0.45–0.6	30	–	0.1–0.27
SiO_2	35	0.087	0.65–1	38	> 8	1.1
Octyltrichorosilane (OTS)	–	–	0.14	109	0.012	0.073
Octadimethylsilane (ODMS)	26	0.017	0.14	103	0.045	0.28
Octadecylydimethylsilane (ODDMS)	29	0.018	0.13	103	–	–
Perfluoroalkyltrichlorosilane (FDTS)	–	–	0.12	108	0.005	0.02
PFTS (perfluoroalkylsilane)	19	0.024	0.12	108	–	–
Polyfluoropolyethylene (PFPE) Z-DOL bonded	34	0.04	0.25	97	–	–
Z-15 PFPE lubricant	91	0.09	0.2	52	–	–
Polydimethylsiloxane (PDMS)	37	0.04–0.06	0.2–0.3	105	–	–
Polymethylmethacrylate (PMMA)	25	0.03–0.05	0.3–0.4	75	–	–

23.5 Conclusions

Tribological problems are still hindering the future development of contact MEMS/NEMS. Despite the research effort dedicated to the reliability of NEMS/MEMS, many tribological issues still remain unresolved. The true contact area between the interacting surfaces plays a major role in affecting adhesion and friction. A good understanding of surface roughness and true contact area between surfaces is necessary to optimize surfaces with less interactions. Most of the theoretical models on adhesion and friction are based on the assumption that surfaces remain flat. However, in reality this is not true and new theories taking into surface roughness are necessary.

Analytical solutions and atomistic simulations in tribology are far from real engineering contact situations. There is no analytic tool yet to extrapolate the data obtained at nanoscale (using AFM or SFA) to any contact size of interest. Further development in this domain is necessary to predict tribological phenomena occurring at the nano-/microscales.

An improvement in laboratory testing tools is also needed. Tools that can operate at contact pressures, thermal conditions, and sliding speeds corresponding to real application conditions are necessary for an accurate evaluation of materials. Recently, new test equipment has been developed in that respect.

Limitations from a material point of view must be bypassed. It is necessary to either develop new ways of depositing ceramic coatings efficiently or to improve the chemistry of organic layers so that they offer a better wear resistance. Hybrid coatings as multilayers, nanocomposites, nanostructured, etc., are essential that can offer both low friction and good wear resistance.

A major challenge that still has to be tackled is the in-depth understanding of the evolution of material surface properties when two surfaces interact and rub against each other. Tribo(electro)chemical reactions at contacting asperities may generate surface layers with totally different properties. Physical, chemical, electrochemical, and mechanical properties of material change in a tribocontact depending on the contacting conditions and the time between successive contact events, and evolve with testing time. As long as that aspect is not completely understood, the development of optimized materials and coatings for MEMS/NEMS applications will be a trail-and-error approach rather than a scientifically based one.

References

1. M.A. Huff, A distributed MEMS processing environment, The MEMS Exchange: A Distributed MEMS Processing Environment. Michael Huff, Corporation for National Research Initiatives Commercialization of Microsystems, MST 1998
2. B. Bhushan, CRC Handbook of Micro/nanotribology, 2nd edn, 1999.

3. Courtesy of Sandia National Laboratories.
4. B. Stark, MEMS Reliability assurance guidelines for space applications, Jet Propulsion Laboratory, NASA, JPL Publication 99–1, 1999.
5. A. White, A review of some current research in microelectromechanical systems (MEMS) with defence applications, Weapons systems division aeronautical and maritime research laboratory, DSTO, 2000.
6. Y. Blanter, Nano-electromechanical systems (NEMS), Work sponsored by FOM (o.a. Physics for Technology), NanoNed and by an EU project (CANEL), T.U. Delft, Nederlands.
7. P. Sharke, Eye on the future: Nanotechnology: hybrid NEMS, Mechanical Engineering magazine online, ASME, Feb 2001.
8. M.W. van Spengen, MEMS Reliability stiction, charging and RF MEMS, Ph.D dissertation, IMEC-K.U. Leuven, 2005
9. J.W. Judy, Microelectromechanical systems (MEMS): fabrication, design and applications, Smart Mater. Struct. 10, 2001, pp. 1115–1134
10. J.A. Walraven, Failure mechanisms in MEMS, ITC Intern. Test Conf. 33.1, 2003, 828
11. M.W. van Spengen, MEMS reliability from a failure mechanisms perspective, Microelectron. Reliab., 43, 2003, pp. 1049–1060
12. R.Maboudian, R.T.Howe, Critical review: adhesion in surface micromechanical structures, J. Vac. Sci. Technol. B, 15, 1997, p. 1
13. R.E. Sulouff, "MEMS opportunities in accelerometers and gyros and the microtribological problems limiting commercialization", in: B. Bhushan (Ed.), Tribology Issues and Opportunities in MEMS, Kluwer Academic Publishers, Dordrecht, Netherlands, 1998, pp. 109–120
14. M.R. Douglass, 36th IEEE International Reliability Physics Symposium Proceedings, IEEE, New York, 1998, pp. 9–16.
15. M.W. van Spengen, R. Puers, I. De Wolf. A physical model to predict stiction in MEMS, J. micromech. Microeng., 12, 2002, pp. 702–713.
16. B. Bhushan, Plenary lecture 1st Vienna conference on micro/nano technology, Mar 2005
17. S.L. Miller, G. LaVine, M.S. Rodgers, J.J. Sniegowski, J.P. Waters, P.J. McWhorter, Routes to failure in rotating MEMS devices experiencing sliding friction, Proc. SPIE Micromachined devices and Comp. III., 3224, 1997, pp. 24–30
18. X. Lafontan, F. Pressecq, F. Beaudoin, S. Rigo, M. Dardalhon. The advent of MEMS in space. Microelectron Reliab 43,2003, pp. 1061–1083
19. Courtesy of TIMA micro and nanosystems group, France.
20. N.R. Tas, A.H. Sonnenberg, A.F.M. Sander, M.C. Elwenspoek, MEMS. In: Proceedings, IEEE, Tenth Annual International Workshop on MEMS-97, 26–30, 1997, pp. 215–220
21. M. Urbakh, J. Klafter, D. Gourdon, J. Israelachvili. The nonlinear nature of friction, Nature 430, 2004, pp. 525–528
22. B. Bhushan, Handbook of Nanotechnology, Springer, Berlin/Heidelberg/New York, 2004.
23. B. Bhushan, Applications of micro/nanotribology to magnetic storage devices and MEMS, Trib. Int. 28, 1995, pp. 85–96.
24. D.W. Brenner, Mysteries of Friction and Wear Unfolding: CMS Advances the Field of Tribology, The Amptiac Newsletter, 5, 2001, pp. 1–14

25. C.M. Mate. Force microscopy studies of the molecular origins of friction and lubrication, IBM J. Res. Dev. 39, 1995, pp. 617–627.
26. J.N. Israelachvili, "Techniques for Direct Measurements of Forces Between Surfaces in Liquids at the Atomic Scale," Chemtracts Anal. Phys. Chem. 1, 1, 1989
27. A.D. Romig, M.T. Dugger, P.J. McWhorter. Materials issues in microelectromechanical devices: science, engineering, manufacturability and reliability, Acta Mater. 51, 2003, pp. 5837–5866
28. F.Bowden, D.Tabor, The Friction and Lubrication of solids, Clarendon Press, Oxford 1968.
29. N.Tas, T. Sonnenberg, H. Jansen, R. Legtenberg, M. Elwenspoek. Stiction in surface micromachining, J. Micromech. Microeng. 6, 1996, pp. 385.
30. T. Stifter, O. Marti, B. Bhushan, Phys. Rev. B 62, 2000, pp. 13667–13673.
31. Y. Ando, J. Ino. Friction and pull-off forces on sub-micrometer size asperities, Wear 216, 1998, pp. 115–122
32. K.L. Johnson, K. Kendall, A.D. Roberts. Proc. R. Soc. London A 324, 1971, pp. 301.
33. B. Bhushan, Self-assembled monolayers for controlling hydrophobicity and/or friction and wear, in: B. Bhushan (Ed.), Modern Tribology Handbook, Vol. 2: Materials Coatings and Industrial Applications, CRC Press, Boca Raton, FL, 2001, pp. 909–929
34. B.N.J. Persson, E. Tosatti. The effect of surface roughness on the adhesion of elastic solids, J. Chem. Phys., 115, 2001, pp. 5597–5610
35. S.-H. Yang, S.M. Hsu, Effect of colloidal probe random surface features on adhesion, Proc. World Tribology Congress III, Sep 2005.
36. M.P. de Boer, J.A. Knapp, T.A. Michalske, U. Srinivasan, R. Maboudian. Adhesion hysteresis of silane coated microcantilevers. Acta Mater. 48, 2000, pp. 4531.
37. C.H. Mastrangelo, C.H.Hsu, A simple experimental technique for the measurement of the work of adhesion of microstructures, IEEE Solid-State Sensor and Actuator Workshop, New York, USA, 1992, pp. 208.
38. M.P. de Boer, J.A. Knapp, T.M. Mayer, T.A. Michalske. The role of interfacial properties on MEMS performance and reliability, Proc. SPIE 3825, 1999, pp. 2
39. J.J. Rha, S.C. Kwon, J.R. Cho, S. Yim, N. Saka. Creation of ultra-low friction and wear surfaces micro-devices using carbon films, Wear 259, 2005, pp. 765–770.
40. W.R. Ashurst, C. Yau, C. Carraro, R. Maboudian, M.T. Dugger, Dichlorodimethylsilane as an anti-stiction monolayer for MEMS: A comparison to the Octadecyltrichlosilane self-assembled monolayer, J. MEMS, 10, 2001, pp. 41–49.
41. B. Bhushan, T. Kasai, G. Kulik, L. Barbieri, P. Hoffmann, AFM study of perfluoroalkylsilane and alkylsilane selfassembled monolayers for anti-stiction in MEMS/NEMS, Ultramicroscopy, 105, 2005, pp. 176–188
42. S.L. Miller, J.J. Sniegowski, G. LaVigne, P.J. McWhorter. Friction in surface micromachined microengines, Sandia National Laboratories, (www.sandia.gov/documents).
43. E. Santer, D. Spaltmann. Adhesion of cleaned nanoscopic metal contacts, Tribotest 9–4, 2003, pp. 333–344
44. R.W. Carpick, N. Agrait, D.F. Ogletree, M. Salmeron, Variation of the interfacial shear strength and adhesion of a nanometer-sized contact, Langmuir 12, pp. 3334–3340

45. B.V. Derjaguin, V.M. Muller, Y.P. Toporov. J. Colloids Interface Sci. 1975, pp. 53–314
46. Y. Ando, Y. Ishikawa, T. Kitahara, Friction characteristics and adhesion force under low normal load, Trans. ASME J. Tribol. 117, 1995, pp. 569–574.
47. J. Krim, R.A. Widom, Damping of a Crystal Oscillator by an Adsorbed Monolayer and Its Relation to Interfacial Viscosity, Phys. Rev. B 38, 1988, pp. 12184
48. S. Sundararajan, B. Bhushan. Static friction and surface roughness studies of surface micromachined electrostatic micromotors using an atomic force/ friction force microscope, J. Vac. Sci. Technol. A 19(4), 2001, pp. 1777–1785
49. H. Liu, B. Bhushan, Nanotribological characterization of digital micromirror devices using an atomic force microscope, Ultramicroscopy 100, 2004, pp. 391–412
50. K.S.K. Karuppiah, S. Sundararajan. A comparison of lateral calibration techniques for quantitative friction force microscopy, Proc. World Tribology Congress III, sep 2005.
51. S. Zhang, G. Wagner, S.N. Medyanik, W.-K.Liu, Y.-H. Yu, Y.-W. Chung, Experimental and molecular dynamics simulation studies of friction behavior of hydrogenated carbon films, Surf. and Coat. Technol. 177–178 (2004) 818–823
52. H. Wang, Y.-Z. Hu, T. Zhang, Simulations on atomic-scale friction between self-assembled monolayers: Phononic energy dissipation, Tribology International, 2005
53. M.G. Lim, J.C. Chang, D.P. Schultz, R.T. Howe, R.M. White. Polysilicon microstructures to characterize static friction. In: Proc IEEE Workshop on Micro Electro Mechanical Systems, 82, 1990, pp. 11–14
54. D.C. Senft, M.T. Dugger. Friction and wear in surface micromachined tribological test devices. Proc. SPIE, 1997, pp. 3224–3231
55. M.P. de Boer, D.L. Luck, J.A. Walraven, J.M. Redmond. Characterization of an inchworm actuator fabricated by polysilicon surface micromachining. Proc. SPIE, 4558, 2001, pp. 169.
56. M.Z. Huq, J.-P. Celis, Expressing wear rate in sliding contacts based on dissipated energy, Wear, 252, 2002, pp 375–383
57. D. Drees, S. Achanta, J.-P. Celis, Surface testing into the 21st century, Bridging the gap between nano and macro, 1st Vienna International conference on micro- and nano- technology, Vienna, Austria, 9–11 March 2005.
58. B. Bhushan, Nanotribology and nanomechanics, Wear 259, 2005, pp. 1507–1531
59. R. Kaneko, T. Miyamoto, Y. Andoh, E. Hamada. Microwear, Thin Solid Films, 273, 1996, pp. 105–111
60. S. Miyake, R. Kaneko, Microtribological properties and potential applications of hard lubricating coatings, Thin Solid Films, 212, 1992, pp. 256–261
61. S. Sundararajan, B. Bhushan, Micro/nanotribology of ultra-thin hard amorphous carbon coating using atomic force /friction force microscopy, Wear, 225–229, 1999, pp. 678–689
62. A.R. Machcha. An investigation of nanowear in contact recording, Wear, 197, 1996, pp. 211–220
63. A. Ramalho, J.-P. Celis. High temperature fretting behavior of plasma vapor deposition TiN Coatings, Surf. Coat. Technol, 155, 2002, pp. 169–175
64. Z. Tao, B. Bhushan. Bonding, degradation, and environmental effects on novel perfluoropolyether lubricants, Wear, 259, 2005, pp. 1352–1361

65. U. Beerschwinger, T. Albrecht, D. Mathieson, R.L. Reuben, S.J. Yang. Wear at microscopic scales and light loads for MEMS applications, Wear, 181–183, 1995, pp. 42–35
66. S. Achanta, D. Drees, J.-P. Celis. Friction and nanowear of hard coatings in reciprocating sliding at milli-Newton loads, Wear, 259, 2005, pp. 719–729
67. K.J. Gabriel, F. Behl., *In-situ* friction and wear measurements in polysilicon mechanisms, Sens. Actuators,A21–23, 1990, pp. 184
68. M. Mehregany, S.D. Senturia and J.H. Lang, Measurements of wear in polysilicon micromotors, IEEE Trans. Electron. Devices, 39, 1992, pp. 1136–1143
69. M.R. Houston, R.T. Howe, R. Maboudian Proc. Solid-State Sensors and Actuators-Transducers 95, Stockholm, Sweden. 1995, pp. 210.
70. E.D. Flinn, Lotus leaf yields slick idea for MEMS, Aerospace America/May 2005, pp. 25
71. K. Glasmästar (Eds.), S. Sjödin (Eds.), Final report for the research program, Biocompatible materials, SSF research programme: Biocompatible Materials, Chalmers University of Technology and Göteborg University, 2004, pp. 19–39.
72. E.-S.Yoon, S.R. Arvind, K. Hosung, B. Kim, D.-H. Suh, K.Y. Jeong, H. Eui. Tribological Properties of Nano/Micro-Patterned PMMA surfaces on silicon wafer, Proc. World Tribology congress III, sep 2005
73. D. Liu, G. Benstetter, E. Lodermeier. Surface roughness, mechanical and tribological properties of ultrathin tetrahedral amorphous carbon coatings from atomic force measurements, Thin Solid Films, 436, 2003, pp. 244–249,
74. R. Huber, N. Singer, Out with the old and in with the new, Materials today Jul/Aug 2002, pp. 36–43
75. A.R. Kraussa, Ultrananocrystalline diamond thin films for MEMS and moving mechanical assembly devices, Diam. Relat. Mater. 10, 2001, pp. 1952–1961
76. X. Li, B. Bhushan, Micro/nanomechanical characterization of ceramic films for microdevices, Thin Solid Films, 340, 1999, pp. 210–217
77. G. Radhakrishnan, R.E. Robertson, P.M. Adams, R.C. Cole. Integrated TiC coatings for moving MEMS, Thin Solid Films, 420–421, 2002, pp. 553–564
78. S. Sundararajan, B. Bhushan. Micro/nanotribological studies of polysilicon and SiC films for MEMS applications, Wear, 217,1998, pp. 251–261
79. D. Liu, G. Benstetter, E. Lodermeier. Surface roughness, mechanical and tribological properties of ultrathin tetrahedral amorphous carbon coatings from atomic force measurements, Thin Solid Films, 436, 2003, pp. 244–249.
80. H.-S. Ahna, P.D. Cuonga, S. Park, Y.-W. Kim, J.-C. Lim. Effect of molecular structure of self-assembled monolayers on their tribological behaviors in nano- and microscales, Wear, 255, 2003, pp. 819–825
81. R. Maboudian, W.R. Ashurst, C.Carraro, Self-assembled monolayers as antistiction coatings for MEMS: characteristics and recent developments, Sens. Actuators 82, 2000, pp. 219–223,
82. Nikhil S. Tambe, B. Bhushan, Micro/nanotribological characterization of PDMS and PMMA used for BioMEMS/NEMS applications, Ultramicroscopy, 105, 2005, pp. 238–247.
83. X.-C. Lu, B. Shi, L.K.Y. Li, J. Luo, J. Wang, H. Li. Investigation on microtribological behavior of thin films using friction force microscopy, Surf. Coat. Technol. 128–129, 2000, pp. 341–345
84. A.A. Voevodin and J.S. Zabinski, Supertough wear-resistant coatings with "chameleon" surface adaptation, Thin Solid Films, 370, 2000, pp 223–231

24 Nanotribology in Automotive Industry

Martin Dienwiebel and Matthias Scherge

IAVF Antriebstechnik AG, Im Schlehert 32, D-76187 Karlsruhe, Germany
martin.dienwiebel@iavf.de, matthias.scherge@iavf.de

24.1 Introduction

The automotive industry, although commonly considered to be "old technology" is one of the industrial sectors where nanotechnology is being broadly applied, for example in paintings, windows, tyres and coatings.

In engine development, the importance of small length scales below one micrometer was long underestimated. It is becoming more important as downsizing of the engines continues; The engine that powered the first "automobile", built by Carl Benz in 1885 developed a power of 0.6 kW with a displacement of about 1 liter. Present day passenger car engine have a power density of 60 – 110 kW/l. As a consequence the contact pressures and the thermal load of the material increased significantly and sliding surfaces in modern engines have to withstand much harsher conditions.

That they actually do is less the result of new materials but was mainly made possible by significant improvements in manufacturing skills. One example of this is the finishing quality of sliding surfaces. The rms roughness of bearing surfaces is nowadays often well below 100 nm. Hence traditional methods used in engineering to determine roughness, like tactile profilometry are approaching their limit. Therefore new techniques such as optical profilometry or scanning probe methods are recently being adopted from the nano sciences and semiconductor industry also in engine engineering.

24.1.1 Wear and Length Scales

Engine wear can lead to unwanted noise and vibrations, to an increase in emissions and fuel consumption, to a loss of power or eventually to the end-of-life. A rough back-of-the-envelope calculation illustrates that the wear rates in technical systems are in the nanometer per hour regime: A conrod bearing will produce unwanted noise if the clearance widening is larger than approximately 20 µm. If we aim at a service life of the bearing of 200,000 km and assume an average speed of the car of 80 km/h, this would demand a wear rate of less than 8 nm/h. Table 24.1 lists typical wear rates of sliding friction systems in an internal combustion engine (given as average values over the contact area).

Table 24.1. Typical average wear rates of engine components. From [1]

engine component	wear rate
piston ring	5 to 15 nm/h
small conrod bearing	max. 8 nm/h
large conrod bearing	2 to 10 nm/h
tappet	10 nm/h
cam	5 to 10 nm/h

Wear rates and friction losses of an automotive component are influenced by factors that play on very different length scales from centimeters to a few nanometers. If e. g. due to a design error, the clearance of a crankshaft bearing is too low or to high, processes which lead to a reduction of the friction coefficient and the wear rate during running-in will not take place. Bulk hardness and hardness depth curves as well as e. g. the adhesion properties of coatings are the foundation of the tribo-system. The development of a so-called "third-body" between the tribo-partners is crucial as well for low wear as well as for a high stability of the system against external disturbances. We will discuss this point in detail in Sect. 24.3. Finally lubricant and additive properties determine friction and wear at the molecular level. This chapter considers some experimental and theoretical studies of friction-induced changes of the near-surface volume, that take place when the wear rates are in the range of nanometers per hour.

24.2 Energetic View of Friction and Wear

While atomic scale studies on friction using tools like atomic force microscopy (AFM) or the surface forces apparatus (SFA) usually focus on clean surfaces and model fluids, the technical tribosystem usually consists of "dirty" metal surfaces, formulated oil (often "contaminated" by wear particles and soot), and an ill-defined atmosphere. Because of the complexity that allows many different microscopic processes to occur, it is difficult to generalize mechanisms. Therefore, a very universal approach which is valid at the first hand, is to evaluate the tribological system in terms of the friction power P_F [2,3] generated during sliding:

$$P_\mathrm{F} = \frac{d}{\mathrm{d}t} \int \mathbf{F}_\mathrm{F} \cdot \mathrm{d}\mathbf{s} = \mu \cdot F_\mathrm{N} \cdot \mathbf{v} \,, \tag{24.1}$$

with F_F the friction force, μ the friction coefficient, F_N the normal force and v the sliding speed.

We have to add that a more suitable quantity to describe the stressing of the tribosystem is the friction power density $\rho_\mathrm{F} = P_\mathrm{F}/V_\mathrm{R}$, where V_R is the real volume, in which the frictional energy is dissipated. Unfortunatly, V_R is

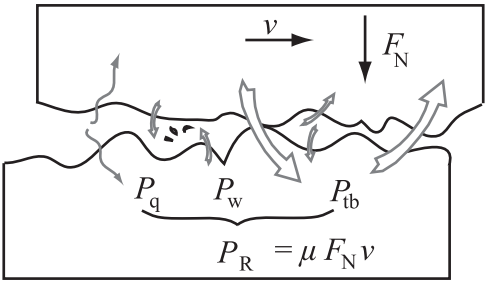


Fig. 24.1. Illustration of the energy dissipation channels in a tribological system. From left to right the arrows denote energy dissipation by heat, wear and change of material. These three contribution form the largest part of total dissipated power of the system

not measurable during the experiment. It is also not possible to determine it ex situ, since the real area of contact is constantly changing during running (see Sect. 24.4.2).

The energy which is fed into the materials is assumed to be dissipated in three main channels [4]. A part P_{q} is consumed for the generation of heat and a part P_{w} for wear particle generation. The remaining friction power P_{tb} will induce changes in the material, leading to the power balance:

$$P_{\mathrm{F}} = P_{\mathrm{q}} + P_{\mathrm{w}} + P_{\mathrm{tb}} \,. \tag{24.2}$$

A similar idea, called the "forced alloy concept" was introduced by G. Martin et al. [5]. He proposed it first to describe phase transformations which occur under irradiation of high energy electrons, during ball milling or during shearing of material. The material is influenced by atomic movements which depend on temperature ("thermal jumps") and atomic movements which depend on external forcing ("ballistic jumps"). In the case of sliding, the magnitude of ballistic jumps will depend on the friction generated in the contact. The competition between thermal and ballistic jumps governs the existence of new phases. One can define an effective temperature $T_{\mathrm{eff}} = T(1+ D_{\mathrm{ball}}/D_{\mathrm{chem}})$, where D_{ball} and D_{chem} are the ballistic and chemical diffusion coefficients. This temperature takes a similar role as the flash temperatures proposed by Blok [6].

24.3 The "Third Body"

The experimental evidence for mechanical intermixing is vast and its existence was already pointed out more than 60 years ago by Bowden and Tabor [7,8] when they introduced their often quoted friction law $F_{\mathrm{F}} = \tau A_{\mathrm{R}}$.

Here, F_{F} is the friction, τ is the shear strength and A_{R} the real area of contact, constituted of of many individual microcontacts. They observed that

the values of τ which was needed to fit their friction data were close to the bulk shear strength of several bearing materials. Consequently, most of the energy dissipation has to be due to plastic deformation.

Later, Kragelski [9] and also Godet [10] introduced the concept of the "third body". A third body can be a particle in between the sliding surfaces but also a continuous layer whose material properties differ significantly from the first and the second body. Since the early 1980's Rigney and coworkers studied the microstructural changes of the third body of several materials [11, 12]. They observed that in copper, nanocrystalline material is formed near the surface. Theses structures were found to be similar to copper samples that had been created by severe plastic deformation (SPD) [13]. The thickness of the third body was found to differ strongly depending on the materials. Gervé [3] observed that implantation of sliding surfaces by carbon or nitrogen ions lead to the development of a sub-surface material composition similar to what is observed on worn engine parts (see Fig. 24.4).

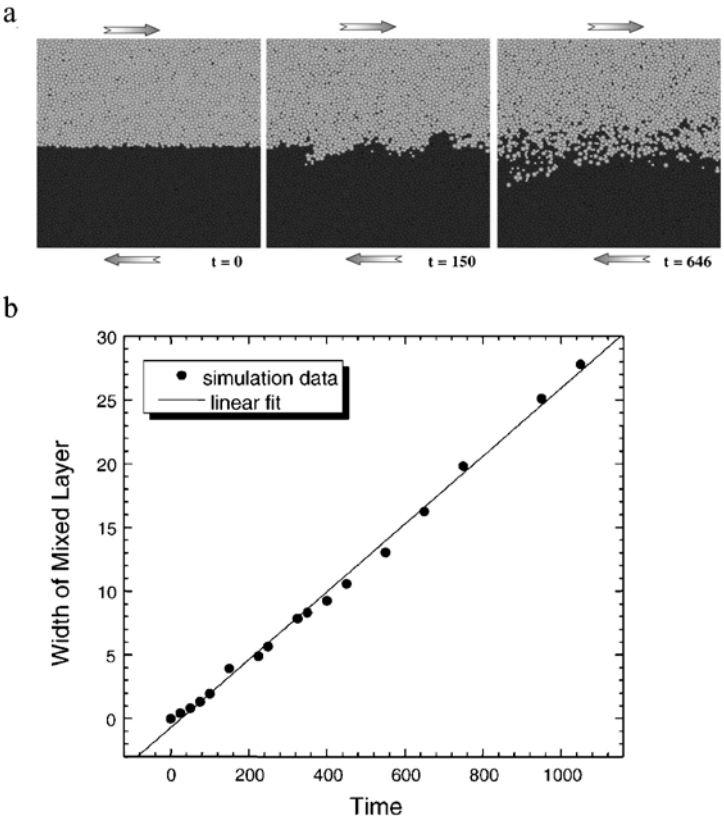


Fig. 24.2. a Snapshots from a MD Simulation by Fu et al. showing the mixing of two sliding bodies. **b** Growth of the mixed layer thickness vs. time. Reprinted from [18] with permission from D. Rigney

The process of mixing was also studied using computer simulations. The simulations allow to observe the creation of the third body but are still far to simple to show all effects that might occur in a lubricated tribocontact. Popov et al. performed mesoscale 2D simulations using the method of movable cellular automata (MCA) [14,15]. The method employs discrete entities with a diameter ranging from 2.5 to 10 nm [15] which have a given density, plastic and elastic properties, yield stress and viscosity of the bulk material and interaction with their neighbors. The MCA simulations showed mechanical mixing and was able to produce depth profiles of the element concentration similar to those found in the cylinder of a combustion engine [16].

Mechanical mixing was also observed in simulations using simple 2D molecular dynamics (MD) calculations by Rigney and Hammerberg [17] and also by Fu et al. [18]. The simulations showed that the friction decreased with running time while the thickness of the mixed layer increased with $t^{-\frac{1}{2}}$. During the mixing the MD simulations show many effects which are known from fluid dynamics, such as convection cells, laminar but also turbulent flow.

Therefore the third body might be envisioned as a viscous fluid which is flowing during the tribological interaction. The analogy with a fluid made Popov assume that wear particles originate from material of the quasi-liquid layer which is extruded at the sides of the contact. Using classical hydrodynamics equations he derived that the wear rate $\dot{W}$ depends quadratically on the thickness of the mixed zone. This would mean that wear should be zero at zero thickness of the third body which is certainly not the case. Therefore this simple assumption might be valid only for a sufficient thickness of the quasi-liquid layer.

24.4 Nanowear

Scherge et al. [19] performed a series of tribometer experiments of a steel (100Cr6) pin resembling a typical piston ring sliding against a gray cast iron disk resembling the cylinder surface of an internal combustion engine. The tribometer was coupled to a high-resolution wear measurement system using the radionuclide technique (RNT). For this technique the pin was labeled with radionuclides and the level of radioactivity in the oil corresponds to the amount of material worn off. The resolution of RNT is usually a few micrograms per liter of oil. Knowing the density of the material and the activity of the sample, the wear can be given as average wear depth over time. A maximum resolution of 0.5 nm/h can be achieved by choosing the appropriate activation parameters. Besides the high resolution, the strength of this method lies in the possibility of measuring the wear continuously during the test. Details on this technique can be found e. g. in [20].

An example of an RNT wear experiment is shown in Fig. 24.3. The two experiments differed only in the normal force that was applied during the first 6 hours. But interestingly the different running-in conditions yielded clearly distinguishable wear rates even after 200 hours of testing. The reason for

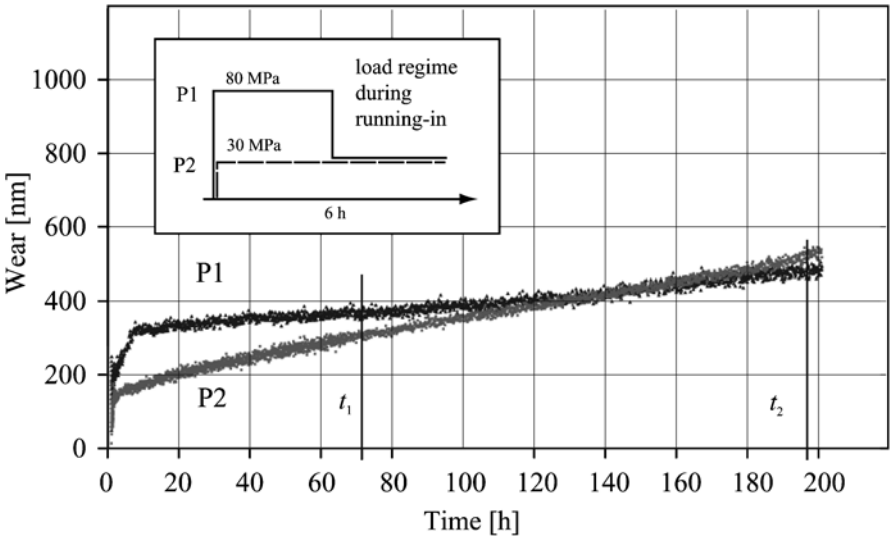


Fig. 24.3. Wear as a function of time for two different running-in procedures. The inset shows the running-in regime. The high load was maintained over a period of 6 hours. Then both systems were run at 30 MPa at a sliding velocity of 5 m/s. From [19]

this is that the difference in the energy dissipation leads to variations of the third body, whose influence is observable even after many hundreds of hours of running time. These variations can be observed in the composition, the topography and the microstructure of the third body as we will show in the remainder of this chapter.

24.4.1 Composition of the Near-Surface Material

Changes in the elemental composition are shown in Fig. 24.4, which were obtained by Auger electron spectroscopy (AES) depth profiling. The upper graph represents the initial state. Due to finishing, the near-surface area has already been changed. The detected oxygen points to oxidation processes. Carbon, zinc and calcium originate from the cooling fluid during machining. At a depth of approx. 20 nm the bulk concentrations are reached. After tribological interaction the distribution of elements changed considerably. The modified zone now exhibits an extension of more than 150 nm. The change of chemical composition of the near-surface volume can be considered as a measure for the degree of mechanical intermixing. In addition, we find a high concentration of calcium, zinc and nitrogen. These elements originate from the additive package of the engine oil. Antiwear additive molecules therefore do not only adhere at metal surfaces to form protective layers, moreover they influence friction and wear by becoming an integral part of the near-surface zone (see also Sect. 24.4.4).

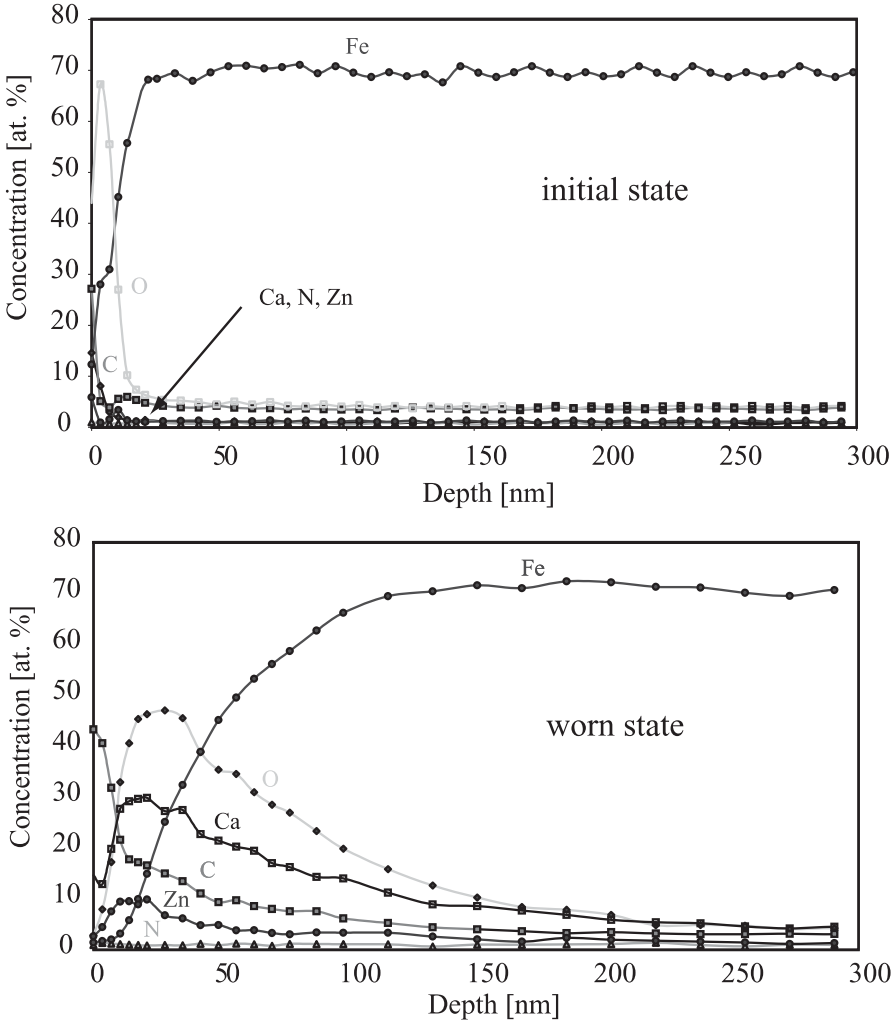


Fig. 24.4. Depth profile by Auger electron spectroscopy of gray cast iron. Upper graph: initial state; lower graph: worn state

24.4.2 Friction- and Wear-Induced Changes of the Surface

As the near-surface volume changes, the surface itself is also subject to permanent change. Very often surfaces in tribological interaction develop a wave-like microstructure as shown in Fig. 24.5. The wavelength of the microstructure was found to correspond to the the applied load meaning that the long-wave portions in the profile of the surface increased with load. One might imagine tribological interaction at contacting surfaces like the impacts of water drops on a sandy beach. A wave-like topography emerges as a result, shown

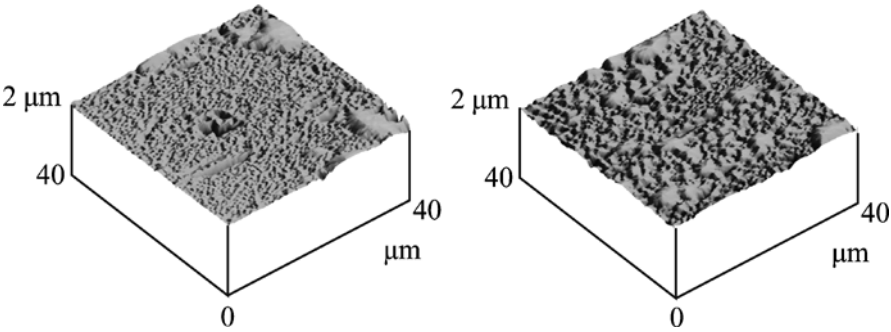


Fig. 24.5. AFM images of two worn gray cast iron surfaces. The normal force was 30 N (*left hand side*)and 90 N (*right hand side*). From [19]

in Fig. 24.5. The size of the impact crater corresponds to the local contact pressure in this very mechanistic sketch. The higher the pressure is, the larger are the waves.

24.4.3 Structural Changes of the Mixed Zone

The focused ion beam (FIB) technique is an ideal tool to study the microstructure of the mixed zone. Cross-section preparation by ion beam milling minimizes the impact of the sample preparation on the grain structure. The imaging capability of the FIB allows the use of either electron or ion induced secondary electrons for image formation. Figure 24.6 shows cross-sections perpendicular to the surface of three brass alloy samples. Brass is used e. g. for piston pin bushings. The images were recorded using ion channeling contrast, which is a function of the grain orientation. Figure 24.6a shows a sample before the test. In order to reduce the effect of finishing on the microstructure, all samples were prepared by electropolishing. Therefore, the microstructure

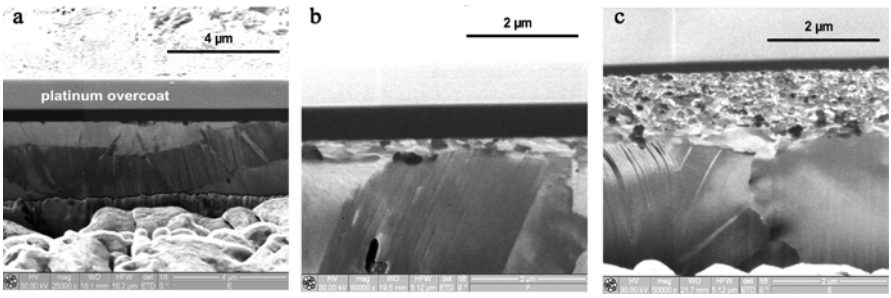


Fig. 24.6. Scanning ion micrographs of brass cross-sections. **a** sample after electropolishing, **b** after tribological stressing with a contact pressure of 4 MPa, **c** after tribological stressing at a contact pressure of 16 MPa

is nearly undisturbed up to the platinum layer, which was deposited prior to the ion milling to protect part of the surface from beam damage.

The samples were tested for 48 h against a stainless steel disk using formulated engine oil as a lubricant. At a contact pressure of 4 MPa, a 170 nm thick zone has developed where the size of the grains are reduced from approximately 1.5 μm to about 50 – 150 nm, with the smallest grains found directly underneath the protective platinum overcoat. Interestingly, the transition to the initial microstructure appears to be very sharp.

At a contact pressure of 16 MPa the thickness of the modified zone is 780 nm. Directly under the surface the grain size is around 30 nm. The grain size increases slightly with depth to around 400 nm. The thickness of the nanocrystalline zone increases nonlinearly with the dissipated friction power which corresponds with the observed changes of the composition of gray cast iron surfaces.

24.4.4 Wear Debris

Wear particles generated during the tribometer experiments were taken from the oil bath. The particles were extracted from the oil by centrifugation, rinsed in n-heptane and deposited on a copper grid, covered by a very thin carbon film (approximately 5 nm thick) for transmission electron microscopy (TEM) observations. TEM micrographs showed particles of different shapes and sizes. A common feature of all particles is their very small thickness, mainly less than 30 nm.

All elements occurring in either the two interacting solids or the oil are present within single wear particles. The element spectrum shows a high content of carbon and oxygen as the result of intense interaction with the lubricant and/or air. P, S and Zn are the fingerprints of additives, especially of antiwear additives, whereas particles with a high concentration of Ca originate from a detergent. The concentration of iron and chromium is extremely low, since the intermixed near-surface zone posesses a low concentration of the original elements (either chromium or iron, see Fig 24.4).

In 1973 Suh [21] proposed a new theory for wear of metals. The theory focuses on plastic deformation and dislocations at the surface, subsurface cracks and void formation. The subsequent joining of cracks by shear deformation as well as the crack propagation lead to particle generation. The theory predicts flakelike wear particles. In continuous wear measurement using radionuclide technique (RNT) wear curves show a strong increase of total wear during running-in. For optimized running-in conditions [19] the increase in total wear becomes gradually smaller and adopts constant increments (stationary conditions). After the running-in, the wear rate is often low but not zero. Assuming that the interaction of both solids with the additivated oil forms a protective film and that the tribological interaction is confined to that film, then the RNT should not be able to detect any wear, i. e., the wear curve should become horizontal. Interestingly, this effect has never been

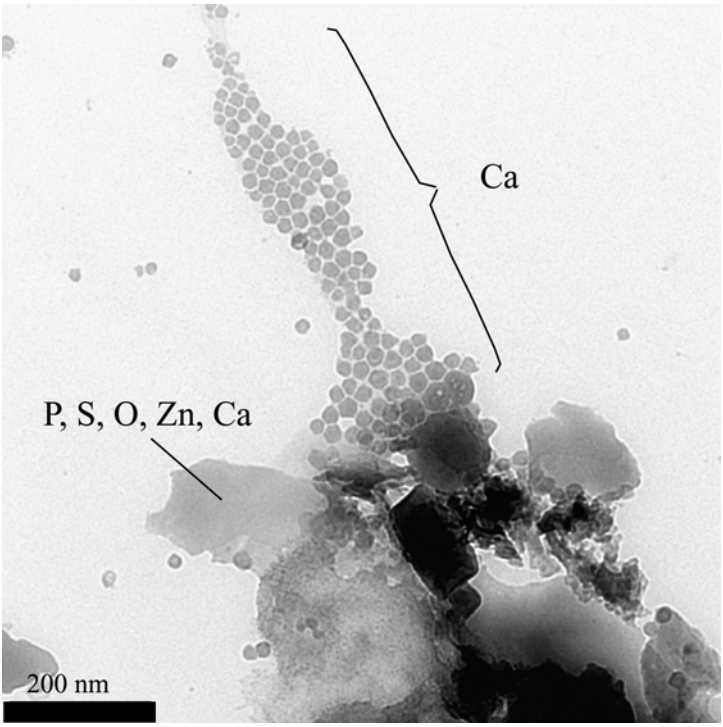


Fig. 24.7. Chemical composition of wear particles at selected locations

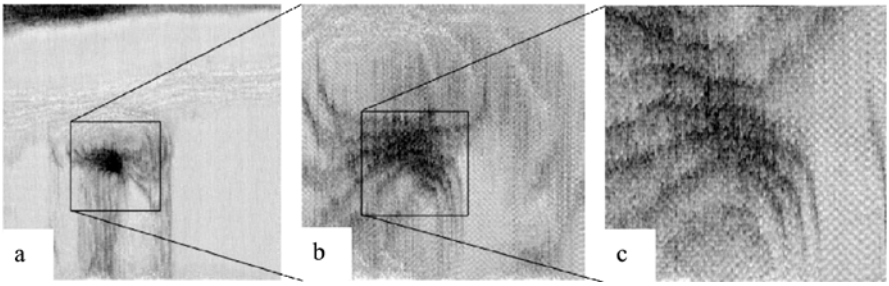


Fig. 24.8. FFM images of a wear groove on KBr after 256 sliding cycles with a normal load of 20.9 nN. Image size **a** 60 nm × 60 nm, **b** 25 nm × 25 nm, **c** 15 nm × 15 nm. From [22]. Reprinted with permission from E. Gnecco

observed so far. Therefore, the delamination theory applied to the formed additive layer alone does not hold. Only when we assume that, due to delamination, also base material (carrying nuclides) is removed, then low but increasing total wear can be achieved. The TEM images have shown that the majority of the wear particles have a thickness in the range of typically less than 30 nm suggesting that the formed film should not be thicker.

24.4.5 Atomic-Scale Wear Studies

Very few experimental studies on wear phenomena have been carried out on the atomic level. Gnecco et al. [22] used an AFM tip to wear down a KBr crystal in UHV. During the experiment ions were removed from and redeposited on the sides of the wear track. On much smaller scale this is also a mixing process as it is observed in macroscopic experiments, here the mixing process occurs only between KBr atoms because of the large difference in hardness between the Si tip and the surface. Repeated scanning over a square area of the surface eventually leads to a wavy surface [23], looking similar to surfaces which were observed by AFM after pin-on-disk experiments. In the macroscopic experiment, processes of removal and redeposition will take place at many asperities in parallel.

24.5 Conclusions

Due to the steady increase of the power density of engines and surface finishing quality, we find that nowadays tribological problems frequently have their origin at the first few hundreds of nanometers of the sliding surfaces. As this trend continues the understanding of wear mechanisms at the nanoscale will become increasingly important in automotive engineering.

A high service life of an internal combustion engine can only be achieved when the wear rates are just a few nanometers per hours. This requires that a third body forms at contacting spots within the first seconds to hours of running. Many experimental and theoretical studies have shown that these plastic processes are alike in a turbulent fluid in which atoms are transported laterally and also perpendicular to the surfaces. This has the consequence that the shape and distribution of asperities is not static but the landscape of asperities is changing constantly during sliding due to plastic flow processes. The composition, the crystal structure and the surface topography are very different from the original material, leading to a reduction of the friction coefficient and the wear rate. The thickness of the third body depends on the friction power dissipated within the contact.

Acknowledgement. We are grateful to the contributions to the results shown in Sect. 24.4 by Dmitry Shakhvorostov, Klaus Pöhlmann, Patrick Thomas, Susan Enders, Oliver Kraft and Jean-Michel Martin. The work was made possible by financial support from IAVF Antriebstechnik AG.

References

1. M. Scherge, J. M. Martin, and K. Pöhlmann, Wear **260**, 458 (2004).
2. G. Fleischer, Schmierungstechnik **16**, 385 (1985).

3. A. Gervé, Surf. Coat. Technol. **60**, 521 (1993).
4. D. Shakhvorostov, K. Pöhlmann, and M. Scherge, Wear **257**, 124 (2004).
5. L. Chaffron, Y. L. Bouar, and G. Martin, C. R. Acad. Sci. Paris **t.2, Serie IV**, 749 (2001).
6. H. Blok, Wear **6**, 483 (1963).
7. F. P. Bowden and D. Tabor, Proc. R. Soc. Lond. A **169**, 391 (1938).
8. F. P. Bowden and D. Tabor, J. Appl. Phys. **14**, 141 (1943).
9. I. V. Kragelski and M. N. Dobycin, *Grundlagen der Berechnung von Reibung und Verschleiß*, 1 ed. (VEB-Verlag Technik, Berlin, 1982).
10. M. Godet, Wear **100**, 437 (1984).
11. P. Heilman et al., Wear **91**, 171 (1983).
12. D. A. Rigney, in *Proc. Materials Week '97* (ASM International, Indianapolis, Indiana, 1997).
13. D. A. Rigney, L. H. Chen, M. G. S. Naylor, and A. R. Rosenfield, Wear **100**, 195 (1984).
14. V. L. Popov, S. G. Psakhie, and A. Gervé, Physical Mesomechanics **4**, 73 (2001).
15. V. L. Popov, S. G. Psakhie, and A. D. amd E. Shilko, Wear **254**, 901 (2003).
16. B. Kehrwald, Ph.D. thesis, University of Karlsruhe, 1998.
17. D. A. Rigney and J. E. Hammerberg, MRS Bulletin **23**, 32 (1998).
18. X. Y. Fu, M. L. Falk, and D. A. Rigney, J. Non-Cryst Solids **317**, 206 (2003).
19. M. Scherge, D. Shakhvorostov, and K. Pöhlmann, Wear **255**, 395 (2003).
20. M. Scherge, K. Pöhlmannand and A.Gervé, Wear **254**, 801 (2003).
21. N. P. Suh, Wear **25**, 111 (1973).
22. E. Gnecco, R. Bennewitz, and E. Meyer, Phys. Rev. Lett. **88**, 215501 (2002).
23. A. Socoliuc, E. Gnecco, R. Bennewitz, and E. Meyer, Phys. Rev. B **68**, 115416 (2003).

25 Nanotribological Studies by Nanoparticle Manipulation

Udo D. Schwarz[1], Claudia Ritter[1], and Markus Heyde[2]

[1] Yale University, Department of Mechanical Engineering, New Haven, USA
[2] Fritz-Haber Institute of the Max Planck Society, Berlin, Germany

25.1 Introduction

In recent years, considerable efforts have been directed towards the clarification of the atomic-scale origins of friction. An in-depth knowledge on the fundamentals of friction on all length scales is crucial for the optimization of materials and processes in a wide variety of technological applications. For example, friction is decisive for the lifetime and the energy consumption of any equipment that features movable parts. Understanding nanoscale friction is believed to be key in order to gain a complete image of the phenomenon "friction", as it would enable a "bottom-up" explanation of macroscale phenomena by atomic-scale effects. The importance of understanding nanoscale frictional behavior is additionally augmented because many old concepts of lubrication break down with the ongoing miniaturization of moving components in technological devices such as computer disk heads in magnetic storage devices (hard disks) or microfabricated motors (so-called micro electro mechanical systems, MEMS).

25.1.1 The Dependence of Friction on Contact Area: Fundamental but Poorly Understood

A central question in this context is how the frictional force F_f experienced at a finite interface featuring nanoscopic dimensions scales with interface size, i. e., how it depends on the actual contact area A_contact. Macroscopically, Amontons' well-known law $F_\mathrm{f} = \mu F_\mathrm{l}$ applies, where F_l represents the externally applied loading force and μ the so-called friction coefficient. Since μ is found to be constant in good approximation for a given material combination, it follows that friction is independent of the *apparent* contact area. If, however, we perform the transition from the apparent macroscopic contact area (left image in Fig. 25.1) to the true dimensions of the nanocontacts that actually support the weight of the slider (middle image in Fig. 25.1), two different scenarios are expected, depending on the exact condition of the contact:

1. If the surfaces are "clean" on an atomic scale, i. e., no adsorbed "third bodies" ("dirt particles") such as small hydrocarbon molecules that ad-

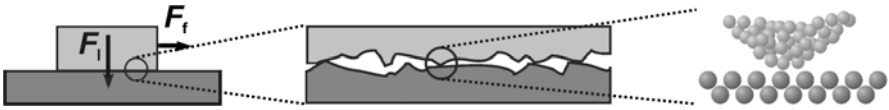


Fig. 25.1. The interface between a sliding macroscopic body and its support (*left*) is rough on a microscale (*middle*) even if it features polished surfaces. Actual contact is only made at relatively few nanoscale contact points, which can be modelled by an FFM tip siding on an atomically flat sample (*right*)

here on any surface under non-ultrahigh vacuum conditions exist between the two sliding interfaces, an effect called *structural lubricity* is expected to occur [1]. This phenomenon, which was originally termed "superlubricity" [2], basically describes the reduction of the shear stresses $\tau = F_{\mathrm{f}}/A_{\mathrm{contact}}$ on atomically flat surfaces with increasing size due to a reduction of the potential barrier between stable states caused by lattice mismatch. The existence of structural lubricity is evidenced by a growing number of theoretical [1–5] and experimental [6–9] studies, but still under debate. Structural lubricity is expected to lead to very low shear stresses between disordered or incommensurate atomically flat surfaces once the contact areas have reached a certain size [1]. This, however, holds only as long as the interbulk stiffness is high enough to ensure that the two bodies move essentially as rigid bodies [1]. Otherwise, energy might be dissipated due to considerable internal elastic deformations of the sliding objects, i. e., elastic jumps between different mechanically stable or metastable configurations. Such processes are often referred to as elastic multistabilities [4, 10].

2. If, however, an adsorbed layer of mobile atoms or molecules (the above-quoted "dirt particles") exists, friction is expected to change drastically [11]. In this case, Amontons' law is reinforced, as an area independent friction coefficient μ is obtained for any surface geometry [3].

While macroscopic investigations could be carried out for centuries, investigations targeting nanoscale frictional behavior have only be enabled in the last 20-30 years by introducing new experimental tools such as the surface force apparatus (SFA) [12, 13], the quartz crystal microbalance (QCM) [14, 15], and the friction force microscope (FFM) [16–18], which essentially represents a scanning force microscope (SFM) modified to allow the detection of lateral forces. The latter one has been an especially popular choice for nanotribological studies as the tip-sample contact in this device is believed to represent actual nanocontacts at arbitrarily rough macroscopic surfaces more adequately as the much larger contact areas of the sliding systems realized in QCMs and SFAs do (see Fig. 25.1 for illustration).

However, experimental attempts to shed light on how friction depends on the actual contact area using FFM have been hindered by the fundamental difficulty to not only accurately determine the true contact area of

the nanoscopic contact, but also reliably and reproducibly vary this contact area by factors large enough to ensure adequate coverage of possible contact area-dependent effects. So far, most respective studies were performed indirectly, studying F_f as a function of F_l by taking advantage of the fact that the contact area of elastically deformable contacts will change with load (see, e. g., Refs. [19–29]). In most reports, $F_\mathrm{f}(F_\mathrm{l})$ has been found to be strongly nonlinear, in contrast to the linear macroscopic behavior.

For a meaningful analysis, it is crucial to find a continuum elasticity model that adequately describes the $F_\mathrm{l}(A_\mathrm{contact})$-dependence of the contact. Various models have been applied in the past; all have essentially been derivatives of the Hertzian contact model (see, e. g., Ref. [30] for an overview on the various approaches). In most cases, it has then been found that a good fit with the experimental data could be achieved if $\tau = const.$, implying that $F_\mathrm{f} \propto A_\mathrm{contact}$ [21–25, 27].

In order to explain this observation, Wenning and Müser [31] used an earlier result [3] suggesting that for dry, amorphous, flat surfaces, $F_\mathrm{f} \propto F_\mathrm{l}$ $A_\mathrm{contact}^{-1/2}$ if the conditions for structural lubricity are matched. Assuming that the structure of the tip apex is amorphous and that the contact shows Hertzian or at least quasi-Hertzian behavior under load with $F_\mathrm{l} \propto A_\mathrm{contact}^{3/2}$, it then follows $F_\mathrm{f} \propto A_\mathrm{contact}$. Thus, within this theory, the $F_\mathrm{f} \propto A_\mathrm{contact}$ dependence is essentially a consequence of the specific mechanical deformation behavior of the contact and not a general property of surfaces.

25.1.2 Manipulation of Nanoparticles by SFM: A New Approach

Summarizing the above discussion, we see that it is important to understand how friction depends on the true contact area at the nanometer scale. Conventionally applied experimental techniques, however, have difficulties addressing this issue adequately. While contacts in SFAs are much too large (some ten thousands of $\mu\mathrm{m}^2$ typ.) to represent actual nanocontacts at sliding interfaces, contacts realized in FFMs might be somewhat too small (some tens of nm^2). Moreover, FFM does not permit a direct and independent determination of the true contact area of the sliding interface, and we always rely on the validity of a specific contact mechanical model and assumptions regarding the exact contact geometry. In contrast, an ideal method would allow the investigation of contacts with contact areas in the range of some hundreds to some hundred thousands of nm^2 whose exact sizes have been accurately and independently determined. A promising approach consists in the change from investigating tip-substrate systems, as it is done in FFM or SFA experiments, to probing the frictional properties of particle-substrate systems.

One way to realize this is by pushing nanoparticles with adequate sizes on atomically flat surfaces using the tip of an SFM while probing the friction and/or energy dissipation during translation. This approach offers a couple of advantages compared to conventional friction force microscopy:

- The actual contact area between a suitable nanoparticle and the substrate surface, which does not change during the experiments as long as the particle is not plastically deformed, can be precisely and directly determined by conventional SFM imaging. In contrast, analysis of FFM experiments always requires assumptions on the exact geometrical shape of the probe tip as well as on the validity of a specific contact mechanical model.
- The study of a much wider range of contact areas is possible. The preparation of nanoparticles with diameters from a few nanometers to several hundreds of nanometers allows varying the actual contact area over several orders of magnitude.
- A multitude of different material combinations are possible, since a considerable flexibility in the possible choices for both the adsorbed particle material as well as the substrate material allow much more combinations than in FFM, where only a limited choice of materials exists as tip materials.

A prerequisite for probing the frictional properties of nanoparticles sliding on a substrate is a controlled manipulation technique. Different approaches to achieve a gentle and controlled motion of nanoparticles have been described in the literature. For instance, iron particles have been translated on a GaAs-GaAlAs surface by switching from dynamic mode SFM (i. e., with oscillating cantilever) for imaging to contact mode SFM for manipulation [32]. While the feedback loop is on in both modes, a distinct reduction of the feedback parameters was necessary for successful pushing of the iron nanoparticles in the contact mode SFM. Ramachandran et al. [33] demonstrated the manipulation of gold nanoparticles on a mica surface by means of a so-called setpoint-change protocol. By reducing the setpoint amplitude in dynamic mode SFM, a switching between imaging and manipulating gold particles with diameters of 5 nm to 15 nm could be achieved.

Most commonly, surface modifications by means of a conventional SFM driven in dynamic mode are realized by switching off the feedback loop. Based on this approach, the successful manipulation of gold particles on mica [34–36], gold particles on SiO_2 [37, 38], gold particles on SiO_2 in liquid environment [39], or GaAs particles on GaAs [40] has been reported. The manipulation process starts just before the tip reaches the particle by switching off the feedback loop and ends by switching it on again after reaching the desired position. Thereby, the tip pushes the particles after the oscillation amplitude decreased as the tip is climbing up the particle. The particle moves at the exact point when the lateral component of the force exerted by the cantilever overcomes the frictional force between particle and substrate surface. In all feedback-off approaches, the lateral force that the cantilever can produce in maximum depends on the previously chosen cantilever amplitude, the particle heights, and a possible sample tilt.

25.1.3 Quantifying Frictional Properties by Controlled Nanoparticle Manipulation

While the above quoted work has succeeded in *manipulating* nanoparticles, only few studies have actually *quantified frictional properties during manipulation* [41–43]:

- In a first landmark experiment, Meyer et al. [41] moved bilayer cadmium arachidate islands on a hydrophobized silicon wafer under ambient conditions by conventional FFM. The Cd arachidate had been deposited using Langmuir-Blodgett techniques. From the movement of an individual island with 70 nm diameter, the shear strength at the interface could be estimated to $\tau = 1 \pm 0.2$ MPa.
- In another landmark paper from the same group, islands of C_{60} molecules with contact areas of the order of several ten thousand nm^2 deposited on NaCl were moved by FFM in ultrahigh vacuum. Most noticeably, a drop of the shear strength from ≈0.1 MPa to about 50% of its initial value was observed during sliding [42].
- The first and so far only study actually presenting somehow statistical quantitative data was performed by Sheehan and Lieber [43], who moved 12 MoO_3 nanocrystals on a MoS_2 substrate with conventional FFM in a nitrogen-filled chamber. Crystals with contact areas ranging from 7000 nm^2 to 26,000 nm^2 showed a linear dependence between contact area and friction, featuring lateral forces between 10 and 30 nN.

One reason for the present virtual absence of statistically solid data covering nanoparticle manipulation for nanotribological studies is a lack of adequate experimental approaches. Here, we present a technique that allows the convenient and reproducible manipulation of nanoparticles over a wide range of contact areas while the energy dissipated during translation is recorded.

25.2 Dynamic Surface Modification

Pushing a nanoparticle by means of a scanning force microscopy tip can in principle lead to translation or rotation in plane as well as out of plane of the substrate surface. In addition, care has to be exercised during manipulation as too much tip-sample interaction might permanently deform the nanoparticle. An illustrative example of how a controlled pushing of nanoparticles can look like is given in Fig. 25.2, demonstrating the manipulation of a single antimony nanoparticle on a graphite surface including translation (a–d and g–h), in-plane rotation (a–c), and cutting (e–g) of the nanoparticle by means of the SFM tip. The method used to achieve these very controlled manipulations of the nanoparticle has been denoted as *dynamic surface modification* (DSM) technique [44] and will be explained below in more detail.

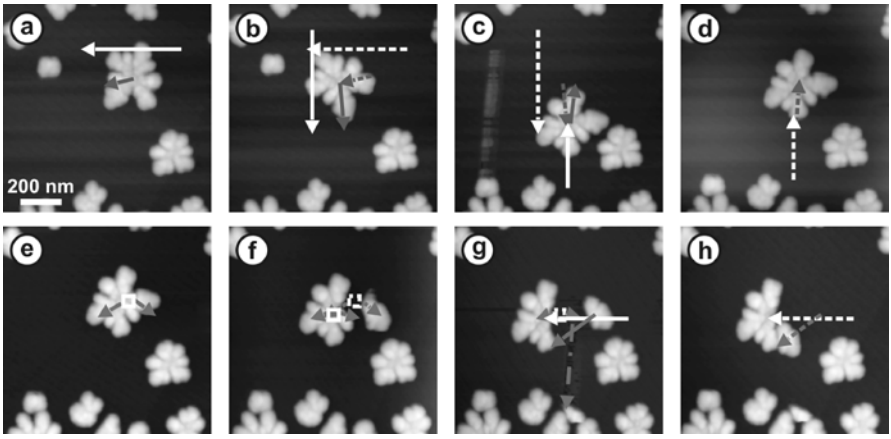


Fig. 25.2. Illustration of the manipulation procedure applying the so-called dynamic surface modification (DSM) technique (for details, see Sect. 25.2.1). Images were acquired under ambient conditions in dynamic (tapping) mode and reflect antimony nanoparticles deposited on graphite. Image size was 1 m × 1 m, the height of the manipulated center particle is 26 nm. Details of the sample preparation are given in Sect. 25.4. **a** Overview of the particle of interest and the surrounding area. A white and a gray arrow indicate the path of the subsequent tip motion and the resulting dislocation of the particle, respectively. **b** Topography after the manipulation. Comparison with **a** shows that the particle experienced a lateral translation of 83 nm and an in-plane rotation of 58°. For the next manipulation step, another contact point between the particle and the tip was selected, visualized again by a white arrow. **c** Result of the second manipulation step, revealing a translation of 211 nm and an in-plane rotation of 77°. To avoid rotation, the point of attack for the third manipulation step was chosen along a line through the center of mass of the particle. **d** Result after the third manipulation step. This time, the motion was almost exclusively translational (175 nm lateral movement). **e**–**g** demonstrate that further increase of the energy input (by about a factor of ten in this case) can be used to purposely cut the nanoparticle in three parts, which are subsequently re-positioned (**g**–**h**)

25.2.1 Basic Concept of Dynamic Surface Modification

Dynamic surface modification (DSM) is a technique that combines the dynamic mode of the SFM, which is under ambient conditions frequently referred to as "tapping mode", with the ability to structure surfaces by increasing the energy input into the cantilever oscillation and, ultimately, into the sample surface. Practically, this is achieved by increasing the amplitude of the dither piezo A_{dither}. Higher energy input into the cantilever oscillation eventually leads to higher energy input into the sample, where it can cause to various types of modifications that range from the controlled movement of adsorbed particles (cf. Fig. 25.2a–d) over the modification of particles or ad-

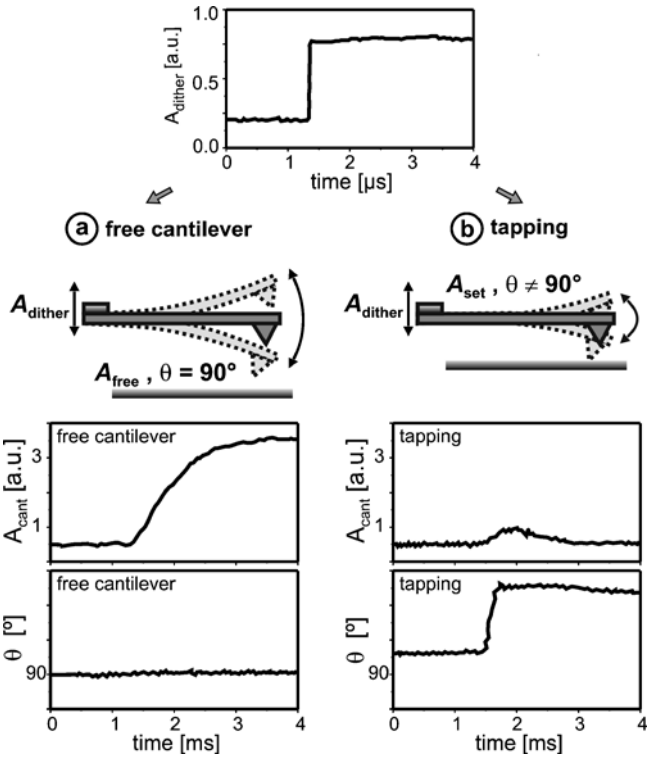


Fig. 25.3. Basic principle of the dynamic surface modification (DSM) technique. If the excitation amplitude A_{dither} of the cantilever is instantaneously increased (topmost graph) while regulating on constant effective oscillation amplitude, two different scenarios are possible depending on whether or not the cantilever makes contact with the surface at some point during the oscillation. **a** No contact ("free cantilever"). The upper graph in this column represents the effective oscillation amplitude signal A_{cant} as delivered by the lock-in amplifier that processes the deflection signal z_{cant} as a function of time. After a transient period, the signal has increased proportional to the increase of the excitation amplitude A_{dither}. The phase difference θ between the driving signal of the dither piezo and z_{cant} (*bottom graph in left column*) remains unaffected. **b** Cantilever in intermediate contact with the surface ("tapping"). Due to the non-zero tip-sample interaction, the active feedback is able to stabilize A_{cant} to the initial value, given by the pre-selected value of the so-called "setpoint amplitude" A_{set}. In contrast, the phase signal θ shows a permanent change, indicating enhanced energy dissipation (cf. Sect. 25.3.2)

sorbed thin films (see, e. g., Fig. 25.2e–g) to the modification of the substrate itself [45], depending on the level of total energy transferred.

The basic principle is elaborated in more detail in Fig. 25.3. For a free cantilever (i. e., with the cantilever not in contact with the sample at any point during an individual oscillation cycle), the sudden increase of A_{dither} results,

with some time delay, in a proportional increase of the effective oscillation amplitude A_{cant} (Fig. 25.3a). Analysis of the cantilever response by lock-in technique reveals that the phase relation between drive signal and cantilever response remains virtually unaffected, as expressed in the so-called *phase signal* θ. If, however, A_{dither} is increased while the cantilever is in contact with the surface and A_{cant} is controlled by an active feedback to maintain a specific value (the *setpoint amplitude* A_{set}), a different behavior is experienced (Fig. 25.3b): After a transient time period, the feedback succeeds to limit the effective oscillation amplitude A_{cant} to the original, pre-selected value A_{set}. The phase signal, however, changes permanently by a significant value.

As discussed later in Sect. 25.3.2, recording θ combined with the knowledge of the pre-selected values for A_{dither} and A_{set} allows the calculation of the power dissipated during manipulation. Thereby, the active feedback loop ensures a steady result (i. e., constant energy input) independent of sample tilt or topographical changes, which enables very precise, controlled, and quantifiable translation experiments independent of the length of the movement or the specific sample corrugations (see, e. g., Fig. 25.5 on page 570 for illustration). This is in contrast to the earlier discussed conventional approaches where the feedback is disabled during manipulation, as these approaches frequently lead to very nonuniform and thus unquantifiable results.

Beside its ability to deliver reproducible quantitative data, a further advantage of the DSM technique is that by varying the excitation amplitude, the energy input can be adjusted over orders of magnitude if necessary. Thus, it is easy to switch with a mouse click between gentle imaging, pushing particles, modifying ("cutting") particles, to extreme energy input where the substrate is plastically deformed. Even if we restrict our discussion exclusively to the lateral movement of structurally unaltered particles, as it is done further below in the friction-related experiments, this flexibility of being able to adjust the input power freely within a wide range is extremely valuable. It allows the non-modifying imaging of even the smallest particles along with high energy input manipulation sufficient to successfully push even very large particles. Please note that the adsorbed particles are displaced as soon as the energy input exceeds the threshold value necessary to overcome the frictional force of the adsorbed particle. Thus, by determining the threshold value of energy dissipation necessary to induce motion for a large number of nanoparticles covering a wide range of sizes, we have a powerful tool for the study of the contact area dependence of frictional forces on the nanometer scale (cf. Sect. 25.3).

An obvious disadvantage of DSM as a nanotribological tool is its inability to directly measure lateral forces, since only the energy dissipation is detected. In contrast, pushing particles in contact mode SFM, as performed in Refs. [41–43], gives direct access to the frictional forces of the adsorbate-substrate interface. Such experiments, however, pose the practical difficulty that it is often not straightforward to cover the full range between the very

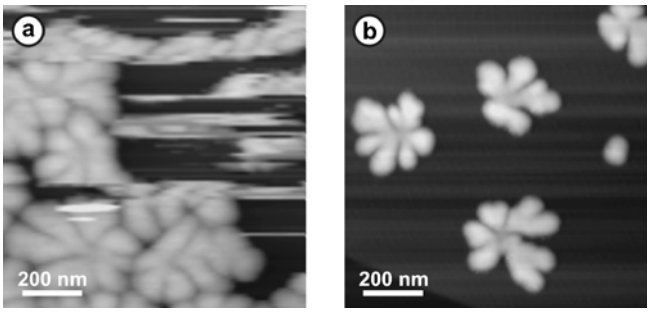


Fig. 25.4. Contact versus dynamic mode imaging. Both **a** and **b** show the same sample consisting of antimony nanoparticles deposited on graphite, but scanned in contact mode (**a**) and dynamic mode (**b**), respectively. While contact mode scanning induced accidental movement of nanoparticles until most particles were essentially "pushed together", stable imaging of differently sized, isolated particles could be easily achieved in dynamic mode

low loading forces required for successful imaging of small particles and the high loads necessary to manipulate large particles (see Fig. 25.4).

25.2.2 Instrumentation for Dynamic Surface Modification Experiments

One of the advantages of the DSM techniques is that it can in principle be employed on any tapping mode-enabled commercial SFM. However, to allow for the convenient, reproducible, and controlled manipulation of islands, essentially three non-standard features have been proven to be advantageous: (1) the use of hardware-linearized piezo scan elements, (2) a separation of the lateral xy- and vertical z-motion, and (3) the manipulation interface of a special, home-written software, which will all be discussed in more detail below. To implement their seamless integration, all experiments described here have been carried out with a home-built beam-deflection SFM system, whose basic components have been described elsewhere [44–47]. Commercially available cantilevers (Nanosensors, Germany) with resonance frequencies $f = 170 - 190\,\mathrm{kHz}$ and spring constants $k = 31 - 71\,\mathrm{N/m}$ have been used to operate the SFM under ambient conditions in the dynamic mode (tapping mode with constant oscillation amplitude).

1. *Hardware-linearized scanner:* Precise and reproducible manipulation requires the ability to position the tip with an absolute accuracy of at least some nanometers anywhere within the scan range. Therefore, hardware-linearized scan elements are mandatory. The ones featured in our instrument were manufactured by Physik Instrumente, Germany, and allow a scan range of $240\,\mu\mathrm{m} \times 240\,\mu\mathrm{m}$. Integrated capacitive displacement sensors suppress nonlinearities and hysteresis, resulting in a repositioning

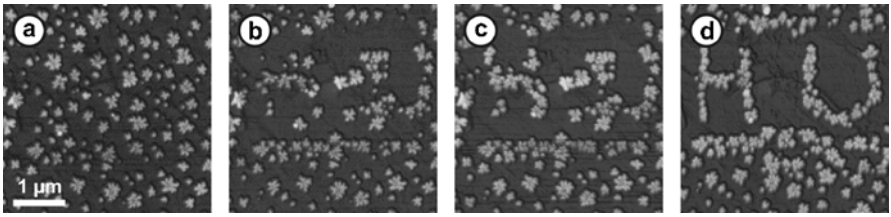


Fig. 25.5. Formation of the letters "H" and "U" using vector-based DSM manipulation, with **a**–**d** reflecting different stages of the formation process. The average particle height is about 30 nm, and the size of all images is $4 \times 4\,\mu\mathrm{m}^2$ each

accuracy of better than two nanometers, i. e., less than 0.001% of the maximum scan range.

2. *Separation of xy- and z-motion:* In common SFM designs, scanner units are usually realized by an integrated piezo element able to achieve displacements in x, y, and z-direction. In contrast, we have implemented a separation of the lateral xy scanning from the positioning in the z-direction by combining the piezoelectric xy scanning table described above with a separate z-piezo. This separation allows a hardware compensation of the sample tilt and thus a more precise operation of the feedback loop due to higher possible gains [44].
3. *Manipulation interface:* The electronic control unit is realized using an ADwin-GOLD system (Keithley Instruments, Cleveland, OH), while the home-written user interface has been programmed in Borland Delphi. Key feature is the manipulation interface of the software, which offers convenient control of all manipulation steps by mouse clicks and enables fast switching between imaging and manipulation and vice versa. Switching between these modes does not entail any scanner movement, and the position of the tip with respect to the sample surface is accurately retained. An application of this vector-based manipulation by mouse control has already been persented earlier (cf. Fig. 25.2).

The four SFM images displayed in Fig. 25.5 illustrate different steps of the assembly process of an artificial pattern (the letters "H" and "U") from virtual disorder. This intentionally composed nanostructure, consisting of 50 individual antimony particles, demonstrates the high accuracy and reliability that the DSM technique can routinely achieve.

25.3 Application of Dynamic Surface Modification for Nanotribological Studies

In the previous section, we introduced the DSM technique as a tool for the manipulation of the shape of surfaces. Here, we will discuss what is needed in

addition to make DSM a valuable tool for nanotribological research. In this context, we have to address three issues: (1) What experimental procedure is suited to determine the threshold energy needed to move an individual particle, (2) how can we determine the dissipated energy from the experimentally measured quantities, and (3) what is the relation between dissipated energy and frictional force?

25.3.1 Experimental Procedure for Threshold Determination

In order to determine the energy threshold at which the particle just starts to move, we have to perform many attempts to push the particle. Each of these attempts is carried out with somewhat higher excitation amplitude A_{dither} than the one before while keeping the setpoint amplitude A_{set} constant, until the particle ultimately displaces. In doing so, we were careful to always attack all particles from the same side. This is to ensure that the relative angle α_{t} between the tip and the plane of the substrate (cf. Fig. 25.9) changes only modestly, if at all, as this angle has an influence on the frictional force exerted on the particle (see Sect. 25.3.3). In fact, we found excellent reproducibility of results acquired even with different tips despite a variation of the particle heights between about 25 – 40 nm as long as always the same type of tip is used (our tips supplied by Nanosensors, Germany, were all from the same batch). In addition, all islands were pushed along vectors that go through the center of mass of the island in order to avoid problems with an eventual rotation of the particle (Fig. 25.2c–d).

To conveniently perform experiments designed to determine the energy threshold in a timely manner, we programmed the following macro. First, an individual island was selected and imaged without manipulation (Fig. 25.6a). Then, a scan line along the x-direction intersecting the center of mass was chosen (white line in Fig. 25.6a), which was profiled over and over again. Between each run, the excitation amplitude A_{dither} was increased by a certain amount, but only within the specific fraction of the scan line. This is illustrated in Fig. 25.6b, which displays the color-coded value of A_{dither} as a function of the x-position on the line (horizontal axis) and the number of attempts to displace the particle (vertical axis). The simultaneously recorded topographical signal is plotted, also as a function of number of attempts, in Fig. 25.6c. Since the particle moves exactly at the point where the energy input due to the increasing excitation amplitude is just sufficient to overcome friction, a convenient determination of the corresponding *excitation amplitude threshold value* $A_{\mathrm{dither}}^{\mathrm{threshold}}$ is possible.

25.3.2 Energy Dissipation during Translational Dynamic Surface Modification

In the above section, we have seen how the excitation amplitude threshold value $A_{\mathrm{dither}}^{\mathrm{threshold}}$ can be found. For a meaningful interpretation of the data,

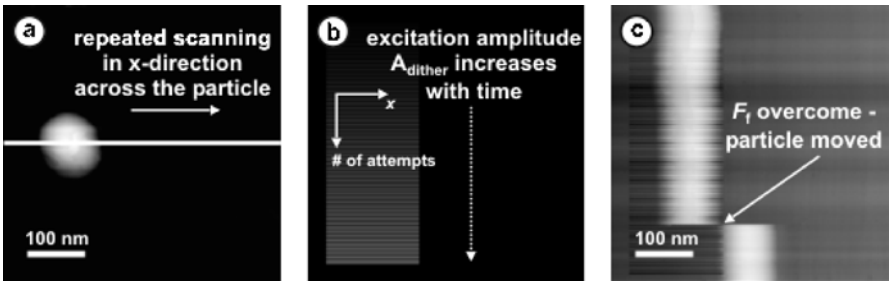


Fig. 25.6. Illustration of the procedure used to determine the excitation amplitude threshold value $A_{\mathrm{dither}}^{\mathrm{threshold}}$ necessary for the lateral translation of a specific individual nanoparticle. **a** Topographical SFM image showing the selected island. **b** Graph representing the applied absolute value of the excitation amplitude A_{dither}, which is displayed color-coded as a function of the x-position along the line indicated in **a** and the number of attempts to move the particle. The computer-controlled macro increases A_{dither} with each attempt, but only within a certain range of x-values visualized by the stripe of increasing brightness. **c** Topographical signal (horizontal axis) recorded along the scan line; the vertical axis corresponds to the same attempt as in **b**. The particle moves during the attempt corresponding to the line indicated by the arrow. Comparison with **b** then allows the determination of the exact value of $A_{\mathrm{dither}}^{\mathrm{threshold}}$

however, this value needs to be linked to the corresponding *minimum power input* P_{min} necessary to successfully move particles. To achieve this goal, we use an elegant method introduced by Anczykowski et al. [48, 49]. Its main advantage is that it does not involve solving the differential equation of motion for the oscillating cantilever, nor depends upon any knowledge of the tip-sample interaction. It is based entirely on the analysis of the energy flow in and out of the dynamic system.

To start our analysis, let us note that in steady-state equilibrium, energy conservation dictates that the average power P_{in} inputted into the cantilever by the dither piezo motion must equal the sum of the average intrinsic power dissipation due to the motion of the cantilever P_0 and the power dissipation due to the tip-sample interaction P_{tip}:

$$P_{\mathrm{in}} = P_0 + P_{\mathrm{tip}} \,. \tag{25.1}$$

Obviously, P_{tip} is the term of interest, as $P_{\mathrm{tip}} = P_{\mathrm{min}}$ if $A_{\mathrm{dither}} = A_{\mathrm{dither}}^{\mathrm{threshold}}$. For its determination, we need to calculate P_{in} and P_0 first. This requires the definition of an appropriate rheological model describing the dynamic system. Instead of considering the complete flexural motion of the cantilever beam, we benefit from the fact that as long as the cantilever motion remains basically harmonic, a simplified model comprising a spring and two dashpots represents a good approximation (see Fig. 25.7). The spring, characterized by the constant k according to Hookes law, reflects the only channel through

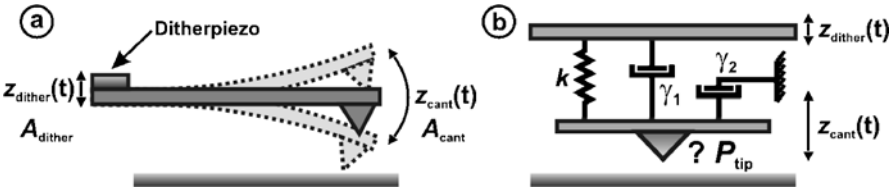


Fig. 25.7. Schematic drawings illustrating the variables and models used in the calculation of the power dissipation. **a** Cantilever oscillating near a surface with $z_{\mathrm{cant}}(t) = A_{\mathrm{cant}} \cos(\omega t + \theta)$ as a result of the motion of its base, which is driven by a dither piezo vibrating with $z_{\mathrm{dither}}(t) = A_{\mathrm{dither}} \cos(\omega t)$. **b** Rheological model describing the dynamic system. Cantilever motion is characterized by the spring constant k and the two damping constants γ_1 and γ_2 [49]

which power P_{in} provided by the motion of the external driver $z_{\mathrm{dither}}(t)$ can be delivered to the tip oscillating as specified by $z_{\mathrm{cant}}(t)$. Therefore, the instantaneous power fed into the dynamic system is expressed by the force exerted by the driver displacement F_{spring} times the velocity of the driver:

$$P_{\mathrm{in}}(t) = F_{\mathrm{spring}} \cdot \dot{z}_{\mathrm{dither}}(t) = k \cdot [z_{\mathrm{cant}}(t) - z_{\mathrm{dither}}(t)] \cdot \dot{z}_{\mathrm{dither}}(t) \ . \quad (25.2)$$

In the present case, we always drive the cantilever sinusoidally, exciting the dither piezo located at the cantilever base with amplitude A_{dither}: $z_{\mathrm{dither}}(t) = A_{\mathrm{dither}} \cos(\omega t)$. Under these circumstances, an in very good approximation sinusoidal steady state response of the cantilever with the amplitude A_{cant} and a phase shift θ relative to the excitation signal will result, i. e., $z_{\mathrm{cant}}(t) = A_{\mathrm{cant}} \cos(\omega t + \theta)$. By integrating Eq. (25.2), the average power input per oscillation cycle over one period $T = 2\pi/\omega$ is found to be

$$P_{\mathrm{in}} = \frac{1}{T} \cdot \int_0^T P_{\mathrm{in}}(t)\,\mathrm{d}t = \frac{1}{2} \cdot k \cdot \omega \cdot A_{\mathrm{dither}} \cdot A_{\mathrm{cant}} \cdot \sin\theta \ . \quad (25.3)$$

This formula contains the well-known result that the maximum power is delivered to an oscillator when the response is 90° out of phase with the drive.

Mainly two effects contribute to the damping term P_0 of the system, marked in Fig. 25.7 by the damping elements γ_1 and γ_2 reflecting intrinsic damping caused by the deflection of the cantilever beam as well as viscous damping due to the motion of the cantilever in the surrounding medium (e. g., air damping if measurements are performed under ambient conditions). Both effects can be combined to an overall effective damping constant $\gamma = \gamma_1 + \gamma_2$, which allows us to write the total average power dissipated by the oscillating cantilever in the following form:

$$P_0 = \frac{1}{2} \cdot \gamma \cdot \omega^2 \cdot A_{\mathrm{cant}}^2 \ . \quad (25.4)$$

Using Eqs. (25.3) and (25.4), Eq. (25.1) can now be solved:

$$P_{\mathrm{tip}} = P_{\mathrm{in}} - P_0 = \frac{1}{2} \cdot k \cdot \omega \cdot A_{\mathrm{dither}} \cdot A_{\mathrm{cant}} \cdot \sin\theta - \frac{1}{2} \cdot \gamma \cdot \omega^2 \cdot A_{\mathrm{cant}}^2 \ . \quad (25.5)$$

Since the damping constant γ can be expressed as $\gamma = k/(\omega_0 \cdot Q_{\mathrm{cant}})$, where Q_{cant} and ω_0 represent the quality factor and the eigenfrequency of the freely oscillating cantilever, respectively, Eq. (25.5) transforms to

$$P_{\mathrm{tip}} = \frac{1}{2} \cdot \frac{k \cdot \omega}{Q_{\mathrm{cant}}} \cdot \left[Q_{\mathrm{cant}} \cdot A_{\mathrm{dither}} \cdot A_{\mathrm{cant}} \cdot \sin\theta - A_{\mathrm{cant}}^2 \cdot \frac{\omega}{\omega_0} \right] \ . \quad (25.6)$$

All quantities required by Eq. (25.6) can be determined experimentally. Note again that no assumptions have been made on how the system is operated, except that the motion of the oscillating cantilever has to remain sinusoidal to a good approximation. FFT spectra of the motion of the cantilever taken at different amplitudes and energy dissipation levels confirm the validity of the argument: Even for manipulation experiments with large interaction, the contribution of higher harmonics was less than 1%, and the oscillation thus remained sinusoidal. In our case, the cantilever is driven at a fixed frequency ω and with a constant drive amplitude A_{dither}, while the oscillation amplitude A_{cant} and phase shift θ may change when the probing tip interacts with the sample surface.

Assuming that the oscillation frequency is chosen to be $\omega_0 = 2\pi f_0$, Eq. (25.6) can be further simplified by considering that the free oscillation amplitude A_{free} is defined by $A_{\mathrm{free}} = Q_{\mathrm{cant}} \cdot A_{\mathrm{dither}}$:

$$P_{\mathrm{tip}} = \frac{1}{2} \cdot \frac{k \cdot \omega_0}{Q_{\mathrm{cant}}} \cdot \left[A_{\mathrm{free}} \cdot A_{\mathrm{cant}} \cdot \sin\theta - A_{\mathrm{cant}}^2 \right] \ . \quad (25.7)$$

If we now keep the oscillation amplitude A_{cant} constant using an external feedback loop so that $A_{\mathrm{cant}} = A_{\mathrm{set}} = const.$, we only need to record the phase shift in order to be able to determine P_{tip}. In Fig. 25.8, the power dissipation has been plotted using Eq. (25.7) from experimentally obtained phase shift data with the set point amplitude A_{set} and the free oscillation amplitude A_{free} as variables in order to visualize the power dissipation for typical values used in the experiments for imaging and manipulation.

25.3.3 Correlation between Energy Dissipation and Friction Forces

It is the goal of this last subsection to establish a relationship between the power dissipation threshold for successful manipulation $P_{\mathrm{tip}}^{\mathrm{threshold}}$ and the static frictional force that a particle experiences before sliding is initiated. This task is anything but trivial, since $P_{\mathrm{tip}}^{\mathrm{threshold}}$ might be correlated other than to the static friction force F_{s} also to the height of the energy barrier between two stable equilibrium positions ΔE or to some kind of viscous (i. e., velocity-dependent) damping. For clarification, a careful theoretical study

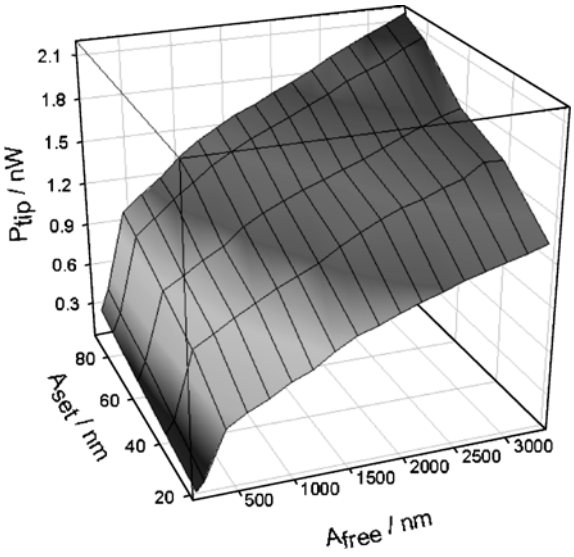


Fig. 25.8. Power dissipation P_{tip} as a function of A_{set} and $A_{\mathrm{free}} = Q_{\mathrm{cant}} \cdot A_{\mathrm{dither}}$, calculated from Eq. (25.7) using experimentally determined values for the phase shift θ. Within the range covered, nanoparticles have been imaged as well as manipulated

of these issues has been carried out, which can be found in Ref. [50]. It is beyond the scope of this contribution to reproduce its details; however, a simple relationship between P_{tip} and F_{s} can be established by intuitive ad-hoc arguments, which is corroborated by the more complete treatment.

We start our analysis based on the one-dimensional Prandtl–Tomlinson model [51–53]. Within this model, a rigid nanoparticle will move from a stable equilibrium position to another if (1) it can lower its total energy, and (2) sufficient energy is provided to overcome the energy barrier between these positions. Denoting the equilibrium distance between the two equilibrium positions with d, an *upper limit* for the minimum energy E_{min} needed to move the particle is $E_{\mathrm{min}} = F_{\mathrm{s}}d$. Note that the distance d is of the order of the lattice constant of the substrate, but that its exact value depends on the relative orientation between the sliding direction and the crystal lattice.

On the other hand, P_{tip}/ω is the energy input into the nanoparticle/substrate system by the tip per oscillation cycle. At the time when the tip impacts the particle, it receives this energy as kinetic energy $(1/2)mv^2$, where m is the mass of the particle and v its velocity, which has components in both lateral and vertical directions relative to the surface. From Fig. 25.9, it becomes apparent that only the component in the lateral direction contributes to E_{min}, but not the one in the vertical z-direction (which results in elastic deformation). Since the lateral component of v can be expressed as $v \sin \alpha_{\mathrm{t}}$, where α_{t} reflects the angle between the sample surface and the dashed line separating tip and particle in Fig. 25.9, $E_{\mathrm{min}} = (1/2)m(v \sin \alpha_{\mathrm{t}})^2$, and thus

$$\frac{P_{\mathrm{tip}}}{\omega} = \frac{E_{\mathrm{min}}}{\sin^2 \alpha_{\mathrm{t}}} \leq \frac{F_{\mathrm{s}} d}{\sin^2 \alpha_{\mathrm{t}}} \,. \tag{25.8}$$

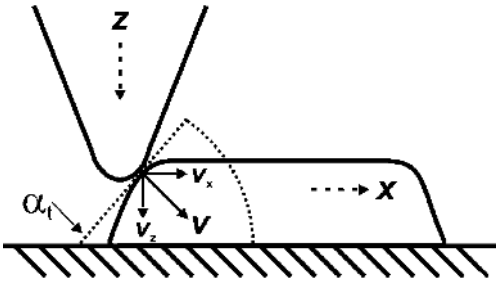


Fig. 25.9. Sketch of the tip-particle coupling. The impact angle α_t between tip and antimony particle determines the normal (z) and lateral (x) components of the acting force

The more detailed treatment carried out in Ref. [50] clarifies that in fact $P_{\mathrm{tip}}/\omega \approx F_{\mathrm{s}} d/\pi \sin^2 \alpha_{\mathrm{t}}$, which is very similar to the equation found above except for an additional factor π in the numerator on the right side of the equation. From this relation, it finally follows that $P_{\mathrm{tip}} \propto F_{\mathrm{s}}$.

25.4 Application Example: Contact Area versus Energy Dissipation Studies of Antimony Islands Sliding on Graphite and MoS_2

In this section, we will illustrate with the example of antimony islands deposited on highly oriented pyrolytic graphite (HOPG) and molybdenum disulfide, respectively, how the DSM method can be applied to the nanotribological studies designed to revealing the contact area dependence of frictional forces. The samples were prepared in an ultrahigh vacuum chamber with a base pressure of less than $6 \cdot 10^{-10}$ mbar. Antimony was deposited by thermal evaporation of the solid material and condensation of the vapor onto the freshly cleaved (0001) surfaces of HOPG and MoS_2 kept at room temperature. Nanometer-sized particles were spontaneously formed by diffusion and aggregation of the deposited material on the surfaces. The HOPG and MoS_2 samples were produced under identical preparation conditions to obtain alike morphologies on both substrate surfaces. Force micrographs showing typical antimony particles produced by this procedure on HOPG have been displayed earlier in Figs. 25.2 and 25.4-25.6.

Results of this nanotribological study can be found in Fig. 25.10. A straight line results in a satisfying fit for both substrate materials, leading to $P_{\mathrm{tip}}^{\mathrm{HOPG}} = (-0.30 \pm 0.11)\ \mathrm{nW} + A_{\mathrm{contact}} \times (2.37 \pm 0.18) \times 10^{-5}\ \mathrm{nW/nm^2}$ and $P_{\mathrm{tip}}^{\mathrm{MoS_2}} = (0.31 \pm 0.17)\ \mathrm{nW} + A_{\mathrm{contact}} \times (2.33 \pm 0.33) \times 10^{-5}\ \mathrm{nW/nm^2}$, respectively. From this result, we find with $P_{\mathrm{tip}} \propto F_{\mathrm{s}}$ (cf. Sect. 25.3.3) that $F_{\mathrm{s}} \propto A_{\mathrm{contact}}$. Since the "loading force" experienced by a particle is almost entirely due to adhesion which scales linearly with A, we finally end up with $F_{\mathrm{l}} \propto F_{\mathrm{s}}$ or, with other words, with a reinforcement of Amontons' law. This can be understood in terms of the arguments discussed earlier in Sect. 25.1.1 as that

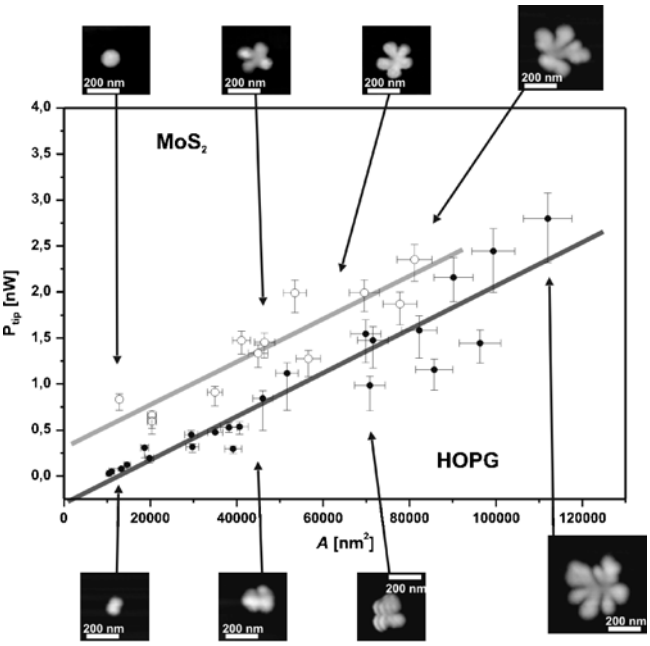


Fig. 25.10. The contact area of translated particles plotted versus the minimum dissipated power $P_{\mathrm{tip}}^{\mathrm{threshold}}$ necessary to successfully move a particle. Full circles represent data for particles grown on HOPG, while open circles reflect measurements carried out on particles deposited on MoS_2. The horizontal error bars indicate the uncertainty in the exact determination of the contact area, while the vertical error bars include both the uncertainty of an individual measurement of the dissipated power as well as the variation of the results for repetitive attempts in moving the identical sample. The inserts show topographical force micrographs of the particles associated with the data points addressed by the arrows and illustrate the structural evolution of the particles from round, amorphous islands to flower-shaped, ramified particles. Please note that the size of the particles translated covers a range from below 10,000 nm^2 to well over 100,000 nm^2, while in Refs. [41–43], moved particles were only of the order of some thousands of nm^2

the sliding interface between particle and substrate is contaminated by mobile adsorbates, which is very plausible considering that our experiments are performed under ambient conditions.

There are, however, more interesting issues to note. For example, we see that the energy dissipated during the manipulation of the smallest particles on graphite is very low, and extrapolation of the linear relationship between contact area and dissipated power leads to a negative offset. In fact, we find that nanoparticles with areas below 10,000 nm^2 are much easier to move than their counterparts with larger contact areas. Preliminary data not shown here reveals also a linear dependence of $P_{\mathrm{tip}}^{\mathrm{threshold}}$ on A_{contact}, but with a slope that is about three times lower than the one for the larger particles.

Speculating about the reasons for the observed behavior, let us note that the onset of considerable energy dissipation during manipulation coincides with a structural transition of the antimony nanoparticles, which has recently been described in a transition electron microscopy (TEM) study [54]. It was found that in the early stage of growth, small antimony particles with spherical shape and amorphous structure were formed (Fig. 25.11a–c). However, when reaching contact areas of 10,000 – 15,000 nm^2, these particles crystallize spontaneously and adopt irregular shapes [55]. Continued deposition leads then to the formation of ramified "flower-like" particles as shown in Fig. 25.11d–f; further examples have already been presented in Figs. 25.2 and 25.5. This structural development is also illustrated in more detail by the insets in Fig. 25.10. TEM studies additionally revealed (1) that the crystalline particles usually consist out of a single domain, but that their crystalline lattice is stressed (indicated by the occurrence of "bending contours" in Fig. 25.11f), and (2) that the surface of amorphous and crystalline parti-

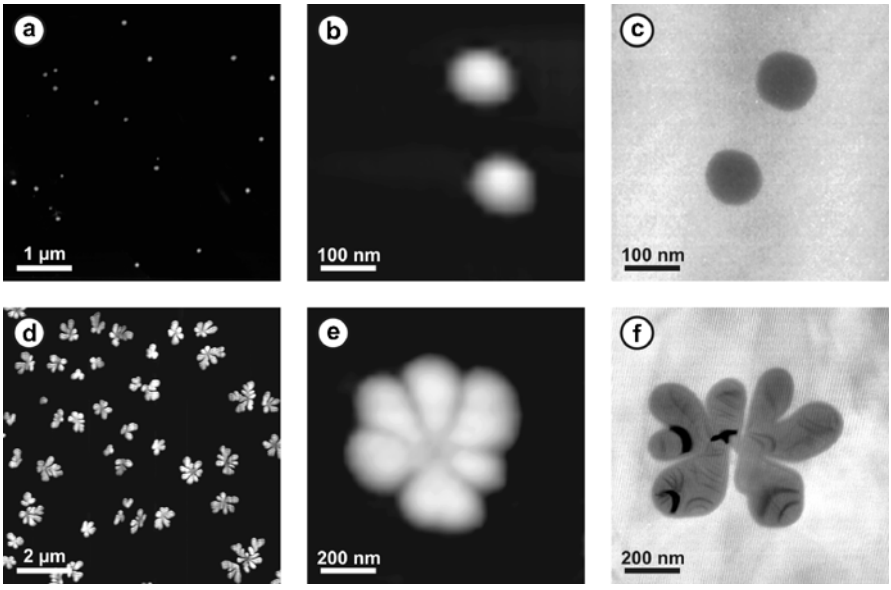


Fig. 25.11. Figure illustrating the typical structure of round, amorphous particles (*top row*), as it manifests for small islands featuring contact areas below about 10,000 – 15,000 nm^2, as well as of larger, crystalline particles, which exhibit flower-like shapes (*bottom row*). **a** and **d** reflect SFM images acquired at low magnification to provide an overall impression of the samples. In contrast, **b** and **d** are taken at ten times higher magnification than their low magnification counterparts to enable a close-up of individual particles. Finally, **c** and **f** represent bright field TEM images, where contrast is mainly based on sample thickness. However, the so-called "bending contours" visible as dark lines in **f** indicate the presence of a stressed crystal lattice for the larger, flower-shaped particles

cles is covered in both cases with amorphous antimony oxide. As the latter finding translates into the fact that the general particle/substrate interface structure should be identical for all islands, we would expect the same frictional behavior of both large and small islands, which is not observed.

The observed behavior could be explained assuming that the larger particles would not longer move as rigid entities, but that internal deformations lead to elastic multistabilities, which add a new route for energy dissipation featuring similar characteristics as the popular Frenkel–Kontorova model of friction [56]. This simple picture would match two of the main characteristics of the data shown in Fig. 25.10, as we would expect not only a linear relationship between $P_{\mathrm{tip}}^{\mathrm{threshold}}$ and A_{contact}, but also the same slope for both HOPG and MoS_2. The first effect is because larger particles will feature more structural instabilities, while the second effect follows because as the energy is almost entirely been released in the antimony nanoparticles, little influence of the substrate on the total energy dissipation is noticeable. However, let us note that particles of sizes similar as used in our experiments still moved essentially as rigid entities in molecular dynamics simulations [57], in contrast to our explanation attempt outlined above.

Another issue that remains to be understood is the larger offset of the linear fit for the particles deposited on the MoS_2 substrate, as opposed to the particles supported by graphite. This could be explained, e. g., by an enhanced chemical interaction between the antimony and the sulphur atoms as opposed to the interactions between the antimony and the carbon atoms at the more inert HOPG surface. Alternatively, it could be related to combined TEM/electron diffraction observations indicating that crystalline antimony nanoparticles grown on MoS_2 are preferentially oriented relative to the substrate lattice rather than distributed randomly. In contrast, such an effect was not observed for particles grown on HOPG.

To conclude our discussion, let us finally note that the values for $P_{\mathrm{tip}}^{\mathrm{threshold}}$ displayed in Fig. 25.10 represent results obtained from particles that have been moved before. About 30% higher values than the ones given were necessary to initiate the dislocation of nanoparticles for the first time. This is most straightforwardly be linked to mobile adsorbates entering the sliding interface, as we postulated them above to explain the observed linear behavior. In addition, it could also mean that the antimony atoms at the interface rearrange in an unknown manner once the substrate-particle bonds have been overcome for the first time, or that the orientation of the particles relative to the substrate changes due to rotation, which reduces the overall interaction strength.

25.5 Conclusion

In summary, we have presented a new approach to study the contact area-dependence of frictional forces on the nanometer scale by the controlled move-

ment of nanoparticles grown on suitable substrates using the tip of a scanning force microscope. During manipulation, the SFM is operated in dynamic mode with active feedback, while the energy needed to overcome friction is provided by successively increasing the cantilever excitation amplitude. This allows to adjust the energy input into the sample by orders of magnitude, enabling a rapid transition from gentle imaging to particle translation or even particle modification ("cutting"). By variation of the growth parameters, particle-substrate contact areas from well below 10,000 nm^2 to significantly above 100,000 nm^2 have been achieved, and their absolute size has been determined independently from the measurement of the energy dissipation during manipulation. The use of a home-build SFM with hardware-linearized scanner and specialized manipulation software made the entire process very controlled and reproducible.

As an application example, the minimum power dissipation necessary to move antimony nanoparticles on HOPG and MoS_2, respectively, was measured as a function of the particle/substrate contact area. For particles with a contact area between 10,000 nm^2 and 110,000 nm^2, a linear dependence has been observed. Within a heuristic model, this result suggests a linear relationship between the frictional force and the particle size, ultimately leading to Amontons' law. In contrast, particles deposited on graphite with contact areas below 10,000 nm^2 were much easier to move compared to their larger counterparts. As this transition coincides with a structural transition within the particles from amorphous to crystalline, we speculate about possible relations between energy dissipation routes and internal particle structure. While more detailed studies need to take place to reach any reliable conclusions, the presented example still demonstrates the usefulness of the new method for nanotribological research.

Acknowledgement. We thank M.H. Müser (University of Western Ontario, London, Canada) for intensive and fruitful discussions and H. Kirmse, R. Schneider, and W. Neumann (Humboldt University Berlin, Institute of Physics) for the TEM analysis. We also would like to acknowledge the support of B. Stegemann and K. Rademann. In addition, U.D. Schwarz acknowledges financial support from from the National Science Foundation (Grant No. DMR-0414944) and the Petroleum Research Fund (Grant No. 42259-AC5), C. Ritter from the Deutsche Forschungsgemeinschaft (Grant No. 494/11-1).

References

1. M. H. Müser, Europhysics Letters **66**, 97 (2004).
2. K. Shinjo and M. Hirano, Surface Science **283**, 473 (1993).
3. M. H. Müser, L. Wenning, and M. O. Robbins, Physical Review Letters **86**, 1295 (2001).
4. M. H. Müser, Tribology Letters **10**, 15 (2001).

5. M. Hirano and K. Shinjo, Physical Review B **41**, 11837 (1990).
6. M. Hirano, K. Shinjo, R. Kaneko, and Y. Murata, Physical Review Letters **78**, 1448 (1997).
7. J. M. Martin, C. Donnet, T. L. Mogne, and T. Epicier, Physical Review B **48**, 10583 (1993).
8. A. Crossley, E. H. Kisi, J. W. B. Summers, and S. Myhra, Journal of Physics D **32**, 632 (1999).
9. M. Dienwiebel, G. S. Verhoefen, N. Pradeep, J. W. M. Frenken, J. A. Heimberg, and H. W. Zandbergen, Physical Review Letters **92**, 126101 (2004).
10. M. H. Müser, Physical Review Letters **89**, 224301 (2002).
11. G. He, M. H. Müser, and M. O. Robbins, Science **284**, 1650 (1999).
12. J. N. Israelachvili, Journal of Colloid and Interface Science **44**, 259 (1973).
13. A. M. Homola, J. N. Israelachvili, P. M. McGuiggan, and M. L. Gee, Wear **136**, 65 (1990).
14. A. Widom and J. Krim, Physical Review B **34**, 1403 (1986).
15. J. Krim and A. Widom, Physical Review B **38**, 12184 (1988).
16. C. M. Mate, G. M. McClelland, R. Erlandsson, and S. Chiang, Physical Review Letters **59**, 1942 (1987).
17. O. Marti, J. Colchero, and J. Mlynek, Nanotechnology **1**, 141 (1990).
18. G. Meyer and N. M. Amer, Applied Physics Letters **56**, 2100 (1990).
19. J. Hu, X. d. Xiao, D. F. Ogletree, and M. Salmeron, Surface Science **327**, 358 (1995).
20. B. Bhushan and A. V. Kulkarni, Thin Solid Films **278**, 49 (1995).
21. R. W. Carpick, N. Agrait, D. F. Ogletree, and M. Salmeron, Langmuir **12**, 3334 (1996).
22. R. W. Carpick, N. Agrait, D. F. Ogletree, and M. Salmeron, Journal of Vacuum Science and Technology B **14**, 1289 (1996).
23. E. Meyer, R. Lüthi, L. Howald, M. Bammerlin, M. Guggisberg, and H.-J. Güntherodt, Journal of Vacuum Science and Technology B **14**, 1285 (1996).
24. M. A. Lantz, S. J. OShea, E. Welland, and K. L. Johnson, Physical Review B **55**, 10776 (1997).
25. U. D. Schwarz, O. Zwörner, P. Köster, and R. Wiesendanger, Physical Review B **56**, 6987 (1997).
26. U. D. Schwarz, O. Zwörner, P. Köster, and R. Wiesendanger, Physical Review B **56**, 6997 (1997).
27. M. Enachescu, R. J. A. van den Oetelaar, R. W. Carpick, D. F. Ogletree, C. F. J. Flipse, and M. Salmeron, Physical Review Letters **81**, 1877 (1998).
28. O. Piétrement, S. Beaudoin, and M. Troyon, Tribology Letters **7**, 213 (1999).
29. O. Piétrement and M. Troyon, Langmuir **17**, 6540 (2001).
30. U. D. Schwarz, Journal of Colloid and Interface Science **261**, 99 (2003).
31. L. Wenning and M. H. Müser, Europhysics Letters **54**, 693 (2001).
32. L. T. Hansen, A. Kühle, A. H. Sorensen, J. Bohr, and P. E. Lindelof, Nanotechnology **9**, 337 (1998).
33. T. R. Ramachandran, C. Baur, A. Bugacov, A. Madhukar, B. Koel, A. A. G. Requicha, and C. Gazen, Nanotechnology **9**, 237 (1998).
34. C. Baur, B. C. Gazen, B. Koel, T. R. Ramachandran, A. A. G. Requicha, and L. Zini, Journal of Vacuum Science and Technology B **15**, 1577 (1997).
35. C. Baur, A. Bugacov, B. E. Koel, A. Madhukar, N. Montoya, T. R. Ramachandran, A. A. G. Requicha, R. Resch, and P. Will, Nanotechnology **9**, 360 (1998).

36. R. Resch, A. Bugacov, C. Baur, B. Koel, A. Madhukar, A. Requicha, and P. Will, Applied Physics A **67**, 265 (1998).
37. R. Resch, C. Baur, A. Bugacov, B. E. Koel, A. Madhukar, A. A. G. Requicha, and P. Will, Langmuir **14**, 6613 (1998).
38. R. Resch, C. Baur, A. Bugacov, B. E. Koel, P. M. Echternach, A. Madhukar, N. Montoya, A. A. G. Requicha, and P. Will, Journal of Physical Chemistry B **103**, 3647 (1999).
39. R. Resch, D. Lewis, S. Meltzer, N. Montoya, B. E. Koel, A. Madhukar, A. A. G. Requicha, and P. Will, Ultramicroscopy **82**, 135 (2000).
40. T. Junno, K. Deppert, L. Montelius, and L. Samuelson, Applied Physics Letters **66**, 3627 (1995).
41. E. Meyer, R. Overney, D. Brodbeck, L. Howald, R. Lüthi, J. Frommer, and H.-J. Güntherodt, Physical Review Letters **69**, 1777 (1992).
42. R. Lüthi, E. Meyer, H. Haefke, L. Howald, W. Gutmannsbauer, and H.-J. Güntherodt, Science **266**, 1979 (1994).
43. P. E. Sheehan and C. M. Lieber, Science **272**, 1158 (1996).
44. C. Ritter, M. Heyde, U. D. Schwarz, and K. Rademann, Langmuir **18**, 7798 (2002).
45. M. Heyde, K. Rademann, B. Cappella, M. Geuss, H. Sturm, T. Spangenberg, and H. Niehus, Review of Scientific Instruments **72**, 136 (2001).
46. M. Heyde, B. Cappella, H. Sturm, C. Ritter, and K. Rademann, Surface Science **476**, 54 (2001).
47. C. Ritter, M. Heyde, B. Stegemann, K. Rademann, and U. D. Schwarz, Physical Review B **71**, 085405 (2005).
48. J. P. Cleveland, B. Anczykowski, A. E. Schmid, and V. B. Elings, Applied Physics Letters **72**, 2613 (1998).
49. B. Anczykowski, B. Gotsmann, H. Fuchs, J. P. Cleveland, and V. B. Elings, Applied Surface Science **140**, 376 (1999).
50. D. A. Aruliah, M. H. Müser, and U. D. Schwarz, Physical Review B **71**, 085406 (2005).
51. L. Prandtl, Zeitschrift für Angewandte Mathematik und Mechanik **8**, 6 (1928).
52. G. A. Tomlinson, Philosophical Magazine Series 7 **7**, 905 (1929).
53. M. H. Müser, M. Urbakh, and M. O. Robbins, Advances in Chemical Physics **126**, 187 (2003).
54. B. Stegemann, C. Ritter, B. Kaiser, and K. Rademann, Journal of Physical Chemistry B **108**, 14292 (2004).
55. B. Kaiser, B. Stegemann, H. Kaukel, and K. Rademann, Surface Science **496**, L18 (2002).
56. Y. I. Frenkel and T. Kontorova, Zh. Eksp. Teor. Fiz. **8**, 1340 (1938).
57. M. H. Müser, private communication.

26 Mechanical Properties of Carbon Nanotubes

Andrzej J. Kulik[1], Andras Kis[2], Branimir Lukic[1], and Kyumin Lee[1], and Laszló Forró[1]

[1] Ecole Polytechnique Fédérale de Lausanne, Institut of Physics of Complex Matter EPFL-IPMC-Station 3, 1015 Lausanne, Switzerland
http://nanotubes.epfl.ch
[2] UC Berkeley, Physics Dept., 125 Birge Hall, Zettl Group

26.1 Technology, Structure and Mechanical Properties

Carbon nanotubes (CNT) can be produced as single-wall (diameter typically 1.4 nm). These can rarely be obtained as isolated single-wall (SW) tubes, rather bundles (ropes) are formed from SW tubes bound by weak Van der Waals forces. Changes of parameters of synthesis lead to the production of multiwall (MW) tubes (diameter in tens of nanometers). CNTs are usually several micrometers long.

In the "early days" the high elastic modulus of the carbon nanotubes was taken as given. The first experiments were performed using tubes obtained by arc discharge or laser ablation methods. These methods produced almost perfect tubes with few structural defects, leading to the elastic modulus approaching theoretical predictions (approximately 1 TPa). Unfortunately, only gram quantities were obtained. Upscaling of production called for the high-yield methods: mainly chemical carbon vapor deposition (CCVD). Multiwall CCVD produced CNTs with deceivingly low Young's modulus – below 100 GPa. Until now, only double-wall CCVD-grown tubes have reached 1 TPa, opening the way of mass applications using their extraordinary mechanical properties: high flexibility, high modulus and high strength combined with low density. Future applications also pushed the development of the measurement techniques of mechanical properties of nanometer-sized objects.

26.2 Dynamic Measurement Methods

Mechanical properties of nanosized objects can be measured by exciting their resonant frequencies. An elegant method was proposed by Treacy and Gibson [20] where thermally excited oscillations were observed using transmission electron microscopy (TEM). Vibration amplitude squared vs. temperature in the range 20 – 800 °C exhibited a linear dependency. Such experiments allowed an estimation of the Young's modulus of the carbon nanotube.

Electrostatic excitation of consecutive modes of vibrations [15] inside TEM allowed not only determination of the elastic properties of the tubes,

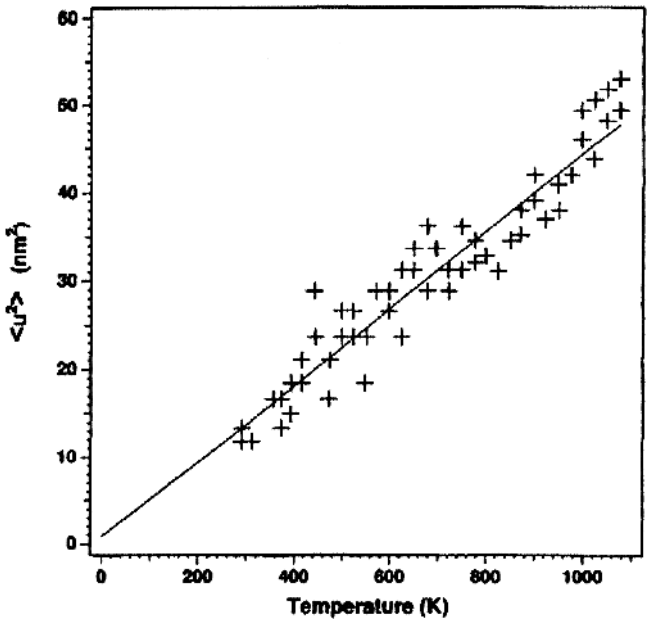


Fig. 26.1. Squared amplitude of thermally induced vibrations of the carbon nanotube as a function of the temperature. Elastic modulus can be derived from this experiment. From [20]

but also served as a nanobalance. The authors could estimate the mass of the attached particle to be in femtogram range.

Since direct observation and analysis of vibrating tubes is difficult to perform, attempts were made to detect resonances using noncontact AFM [22]. These authors directly visualized local stiffness of the suspended coiled tube segment. Fitting the stiffness distribution along the tube axis, the elastic modulus of the tube was obtained. The limited frequency range can be seen as the main disadvantage of this method.

Authors who used dynamic methods to determine mechanical properties, often underlined the precision of their measurements. It is worth remembering that the main source of errors is usually knowledge of the geometry of the specimen, e. g. suspended length, diameter, etc., which can be measured with limited accuracy that in turn seriously limits accuracy of dynamic measurements.

26.3 Quasistatic Measurement Methods

The atomic force microscope (AFM) relies on probing the interaction between a sharp tip located at the end of a force sensor in the shape of a cantilever

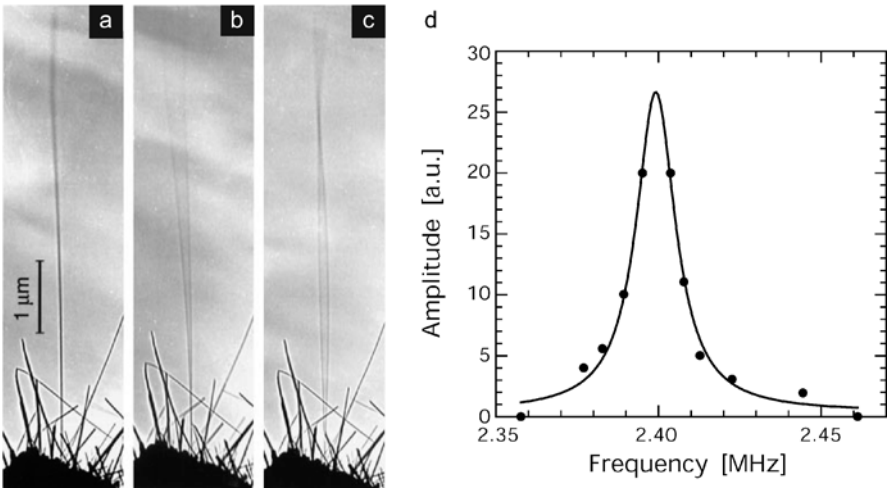


Fig. 26.2. Resonances of the nanotube excited using electrostatic method **a** only thermal vibrations, **b** first flexural mode, **c** second flexural mode, **d** resonant curve. From [15]

and the sample. The tip–sample interaction is detected by monitoring the deflection of the cantilever, most commonly using the laser beam deflection method [14], Fig. 26.3a, that can be used to detect both vertical and lateral deformations of the AFM cantilever (Fig. 26.3b and c). In the constant force mode operation of the AFM, the cantilever's vertical deflection is used as an input to a feedback circuit that displaces the sample in the direction of the z-axis, responding to topographical changes by keeping the cantilever deflection, and consequently, the tip–sample interaction force constant.

Monitoring quasistatic forces and deflections during AFM imaging in contact mode is one of the practical ways of measuring the mechanical properties of nanoscale objects. Several nanomechanical measurement methods incorporating AFMs into the experimental design and involving quasistatic force and deformation measurements have been developed in the last decade and will be presented in this section.

26.3.1 Axial Deformation Using Two AFM Cantilevers

The most direct geometry for measuring the Young's modulus of nanotubes is the one that involves measuring longitudinal deformation of nanotubes under axial strain. Yu et al. have measured the stress–strain response and mechanical strength of individual multiwalled nanotubes [27] and bundles of single-walled nanotubes [28] in such a geometry using a mechanical manipulator operated inside a scanning electron microscope (SEM). Individual nanotubes were attached in situ at opposite ends onto two AFM probes: a compliant

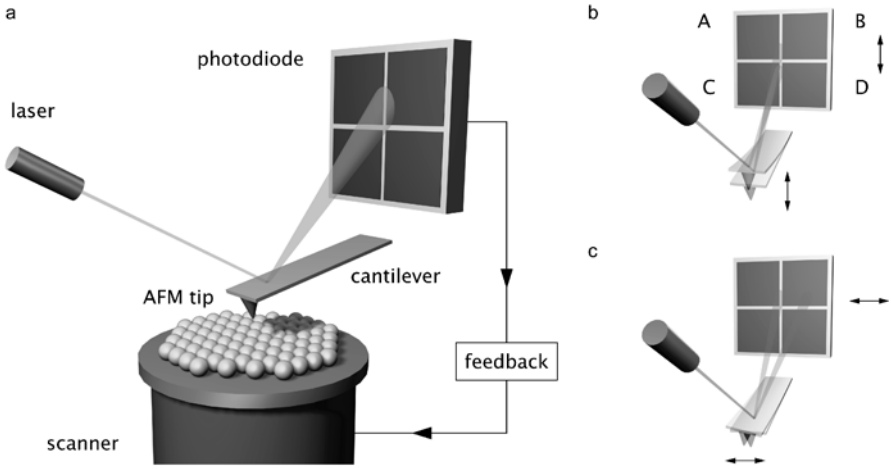


Fig. 26.3. a Schematic drawing of an AFM setup based on the laser beam deflection method. A laser beam is focused on the end of an AFM cantilever and deflected onto a photodiode split into four quadrants, allowing the measurement of vertical (**b**) and lateral (**c**) deflections of the AFM cantilever

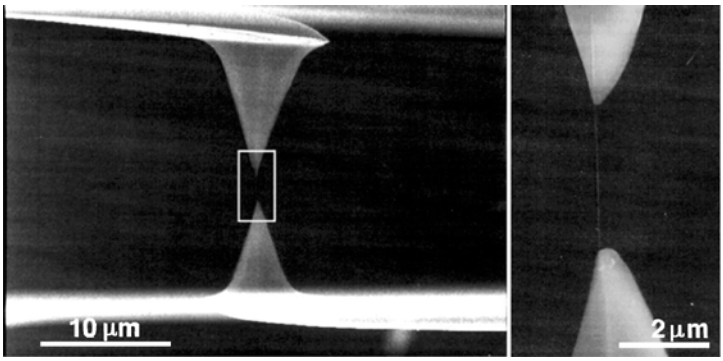


Fig. 26.4. a An SEM image of a multiwalled carbon nanotube linking tips of two opposing AFM cantilevers. **b** A close-up of the region indicated by the rectangle in **a** [27]

one acting as the force sensor with a spring constant $k = 0.1\,\mathrm{N/m}$, and a rigid cantilever ($k = 20\,\mathrm{N/m}$) serving as the second anchorage for the nanotube, Fig. 26.4. Nanotubes were firmly attached using electron-beam-induced deposition of carbonaceous material inside the SEM.

The rigid cantilever was driven using a piezomotor, while the compliant cantilever's deflection and nanotube elongation were simultaneously recorded on video. The tensile force acting on the nanotube is calculated as $F = kd$ where k is the spring constant of the flexible AFM lever and d its displacement

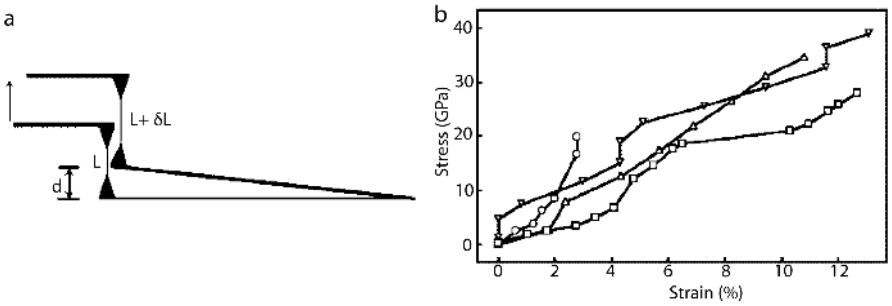


Fig. 26.5. a Principle of the experiment performed by Yu et al. As the top, rigid cantilever is driven upwards, the lower, compliant cantilever bends by d while the nanotube is stretched by δL. **b** Plot of stress vs. strain curves for several individual MWNTs [27]

in the vertical direction. Stress is calculated using the cross-sectional area of the outermost layer of the MWNT. The strain of the nanotube is $\delta L/L$. From the stress–strain curves of nanotubes under tensile loading, Fig. 26.5, Young's moduli in the 270 – 950 GPa range and axial strengths of up to 63 GPa were found. Subsequent examinations of broken tube segments using a transmission electron microscope (TEM) revealed that tubes break with the "sword in sheath" mechanism, with the outer layer carrying most of the mechanical load. Breaking of this outer layer was in most cases accompanied by pullout of inner sections.

26.3.2 Lateral Deformation Using AFM

The first quantitative mechanical measurements on carbon nanotubes using an atomic force microscope (AFM) were carried out by laterally deforming MWNTs deposited on flat surfaces [25]. Wong et al. first randomly dispersed MWNTs on flat surfaces of atomically smooth MoS_2 crystals, chosen because of their low friction coefficient. Friction between tubes and the substrate was further reduced by submerging the sample in water. Tubes were pinned on one end to the substrate under an array of square silicon oxide pads, deposited through a shadow mask, Fig. 26.6a–c. Protruding tubes were geometrically characterized using an AFM. The beam was laterally deformed by the AFM tip, until a certain point where the tip would pass over the tube, allowing it to snap back to its equilibrium position. Multiple lateral-force–distance curves were acquired at different positions along the chosen beam, Fig. 26.6d and e. The tube's maximal deformation was determined by the applied normal load, allowing control over tube breaking. The applied lateral load P in terms of lateral displacement y at the position x along the beam is given by the equation:

$$P(x,y) = 3EI\frac{y}{x^3} + \frac{f}{8}\left(x - 4L - 6\frac{L^2}{x}\right) \tag{26.1}$$

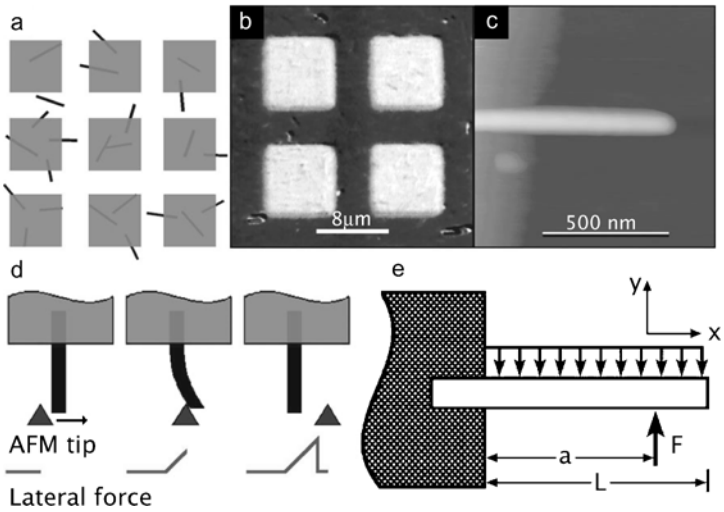


Fig. 26.6. a Nanotubes are dispersed on a substrate and pinned down by SiO_2 pads. **b** Optical micrograph of the sample. **c** AFM image of a SiC nanorod protruding from the pad. **d** The tip moves in the direction of the arrow. The lateral force is indicated at the bottom. During bending the lateral force increases, until the point at which the tip passes over the beam that snaps back to its initial position. **e** The beam of length L is subjected to a point load F at $x = a$ and friction force f [25]

where E is the Young's modulus of the tube, I the second moment of the cross section, equal to $\pi d^4/64$ for a solid cylinder of a diameter d and f the unknown friction force, minimized in this experimental design. Unfortunately, the authors did not calibrate lateral force constants of their AFM cantilevers, so the quantitative value of the lateral force was not known. This uncertainty and the effect of small but nonvanishing friction were eliminated by calculating the nanobeam's lateral force constant, expressed as:

$$\frac{\mathrm{d}P}{\mathrm{d}y} = k = \frac{3\pi d^4}{64x^3}E \tag{26.2}$$

and presented in Fig. 26.6b. Bending curves, Fig. 26.6a, show discontinuities that were attributed to buckling instabilities, first observed by Iijima et al. using a TEM [5]. The mean value of the Young's modulus of MWNTs was $E = 1.3 \pm 0.6\,\mathrm{TPa}$, similar to that of diamond ($E = 1.2\,\mathrm{TPa}$).

Walters et al. [23] also performed mechanical testing of nanotubes in the lateral force mode of the AFM. In their approach, bundles of single-wall nanotubes were pinned beneath metal pads evaporated on top of an oxidized silicon surface. The oxide was first removed by wet chemical etching using HF. The exposed silicon was further etched using KOH in order to produce trenches up to 1.5 μm deep, Fig. 26.8a. The SWNT bundle was deformed in the lateral direction using the AFM, Fig. 26.8b. As the diameter of the nan-

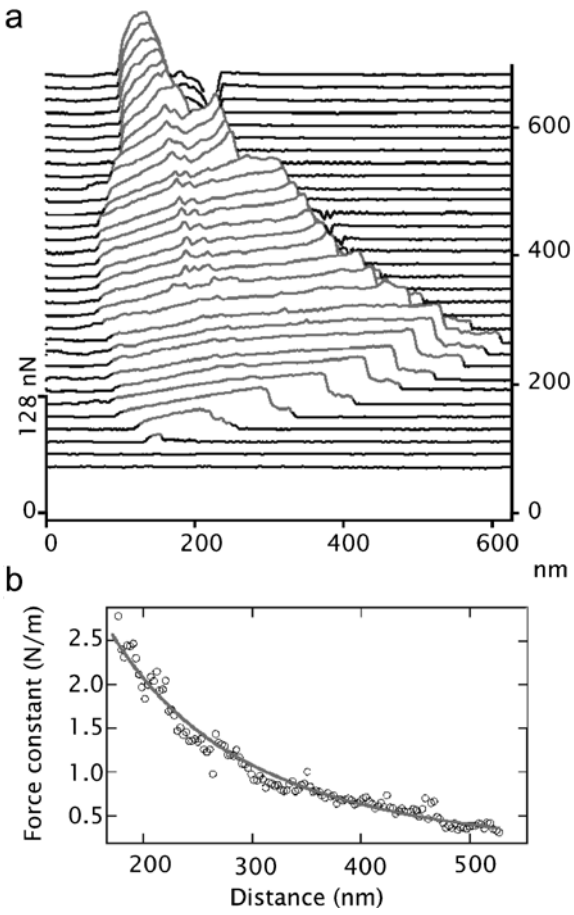


Fig. 26.7. a A series of lateral force–distance curves for different positions along a MWNT. **b** The lateral spring constant as a function of position on the beam. The curve is a fit to Eq. (26.2) [25]

otube bundle was in this case negligible compared to the suspended length, the SWNT rope could be modeled as an elastic string and not an elastic beam. During deformation, all the mechanical strain is manifested as nanotube stretching, until the point at which the nanotubes break. In the case of a tube lying perpendicularly to the trench and the AFM tip deforming it in the middle, the force F exerted on the tube by the AFM tip is given by the expression:

$$F = 2T \sin\theta = 2T\frac{2x}{L} \approx \frac{8kx^3}{L_0} \tag{26.3}$$

where $T$is the string tension, L_0 its equilibrium length, k the spring constant and x the lateral deflection in the middle. The lateral spring constant was not calibrated. The maximal strain at which nanotubes ruptured was $5.8 \pm 0.9\%$. Assuming a value of 1.2 TPa for the Young's modulus, this would correspond to a tensile strength of 45 ± 7 GPa.

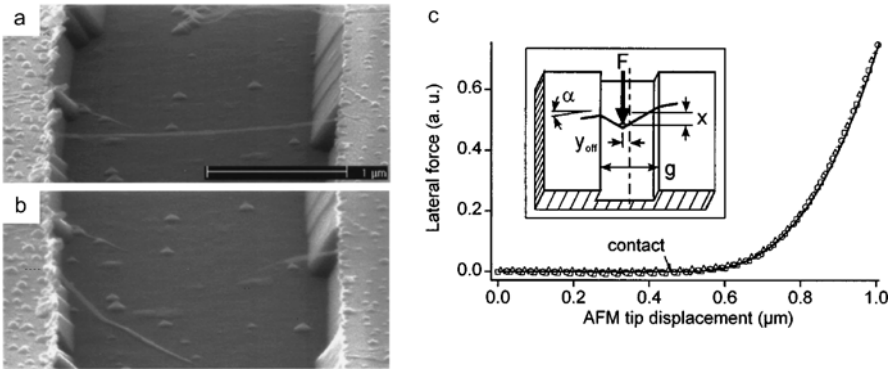


Fig. 26.8. a SEM image of a SWNT rope suspended over a trench in silicon before and **b** after being deformed. **c** Lateral force on a single-walled nanotube rope as a function of AFM tip displacement [23]

Calibrated mechanical measurements in the lateral force mode of the AFM were first performed by Wu et al. on Au nanowires [26]. Their method allows simultaneous measurement of multiple mechanical properties: the Young's modulus, yield strength, plastic deformation and mechanical failure. In this measurement geometry, nanowires are positioned over trenches on a SiO_2 substrate and pinned down on trench edges by electron beam deposition of Pt inside an SEM, Fig. 26.9. Nanowire structures prepared in this manner were subjected to lateral loading by an AFM cantilever that was swept in the horizontal plane with the axis of the cantilever aligned parallel to the nanowire. Figure 26.9 shows a sequence of lateral force–distance curves acquired during mechanical testing. AC-mode AFM images are taken after each loading run. Initially, no permanent deformation can be observed in the AFM images. Increased loading inevitably leads to plastic deformation, manifested as a change of slope in the $F - d$ curves. Further loading ultimately leads to mechanical failure. Linear segments of $F - d$ curves can be analyzed in terms of the elastic beam-bending theory [4], which establishes the relationship between the deformation δ at the midpoint of a doubly clamped beam and the applied force F as:

$$\delta = \frac{FL^3}{192EI} \tag{26.4}$$

where E is the Young's modulus, L the suspended length and I the moment of inertia, given by the nanowire diameter d as $(\pi d)^{4/64}$. The mean value of the Young's modulus of Au nanowires is 70 ± 11 GPa, close to the bulk value (78 GPa) and is essentially diameter independent. Yield strengths are, on the other hand up to 100 times higher than in bulk, approaching the theoretical value of $E/10$ for small, 50-nm diameter nanowires. Such strengthening of gold wires at the nanoscale is most probably due to a reduction in the number of defects accommodated in the nanowire.

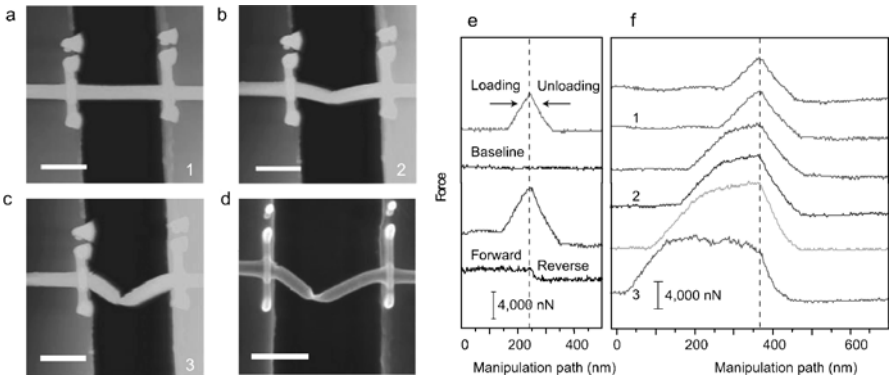


Fig. 26.9. **a** An AFM image of a gold nanowire suspended over a trench in SiO_2. **b**, **c** AFM images of the same nanowire after plastic deformation. **d** SEM image of the wire after the bending test that resulted in the image shown in **c**. **e** Force–distance curves recorded during lateral deformation using the AFM, with and without the contact of the AFM tip with the trench bottom. **f** Force-curves recorded during successive lateral deformations of the wire shown on images **a**–**d** [26]

The experimental geometry also provided the means for calibrating the lateral spring constant of the AFM cantilevers that was performed by pushing tips against the wall of a trench in the SiO_2 covered silicon substrate, following the approach outlined by Schwarz et al. [19].

26.3.3 Vertical Deformation Using AFM

Salvetat et al. performed the first measurement of the elastic properties of carbon nanotubes using a method that involves vertical deflection of nanotubes bridging holes in a porous membrane. This allowed measurements of the Young's modulus of isolated SWNTs and SWNT bundles [16] MWNTs produced using different synthesis methods [18] and the intertube shear modulus of SWNT bundles [17].

Nanotube samples for mechanical measurements are prepared by suspending carbon nanotubes in ethanol and depositing them on the surface of a flat alumina (Al_2O_3) ultrafiltration membrane. Some of the tubes end up suspended over holes in the filter, firmly held in place by van der Waals interaction between the unsuspended parts of the tube and the membrane, Fig. 26.10a. After a suitable nanotube is found, a series of contact-mode AFM images is taken under increasing mechanical load. Line scans across the tube, Fig. 26.10b, reveal the vertical deformation. In the linear elastic regime, the deflection of a thin, long tube is given by the clamped beam formula, Eq. (26.4) [4].

In the case of SWNT bundles, an additional term has to be included in the formula for mechanical deformation (Eq. (26.4)). In the analogy to graphite,

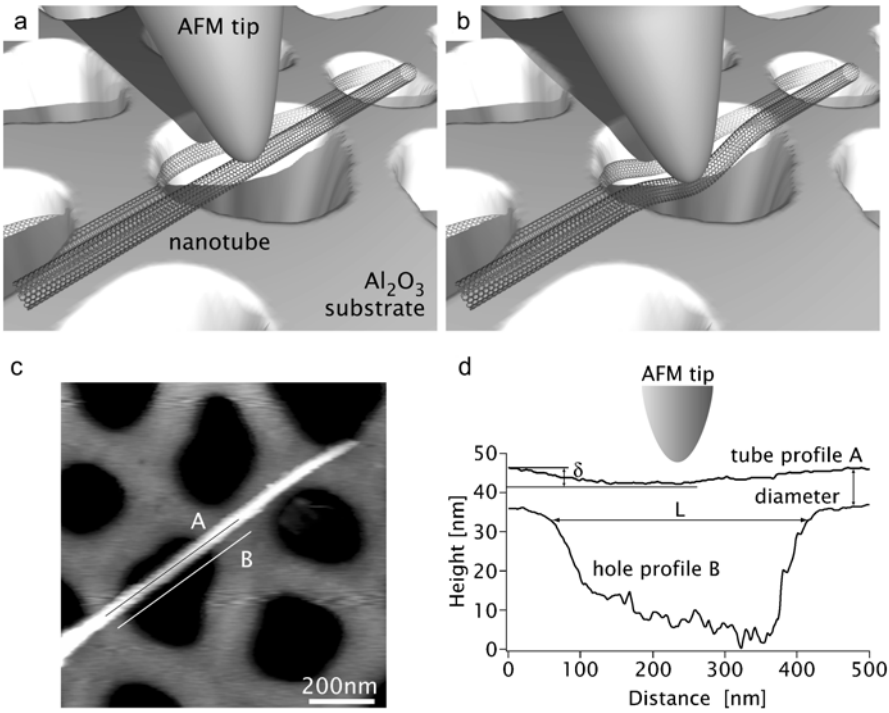


Fig. 26.10. **a** Schematic drawing of a carbon nanotube deposited on a porous Al_2O_3 substrate. **b** During AFM imaging in contact mode, the tube deforms under the applied normal force and this deformation will be recorded as local height in the corresponding AFM image [6]. **c** An AFM image of a 10-nm thick SWNT bundle on a porous substrate. **d** Subsequent analysis of line scans over the tube (*trace A*) and the hole (*trace B*) reveals the nanotube's physical dimensions together with the deformation in the middle [16–18]

where strong sp^2-bonded layers are held together via weak van der Waals interaction, tubes comprising SWNT bundles can easily slide with respect to each other. Bundles, in fact, behave more as a loose assembly of individual tubes than as a compact beam. The deformation can thus be modeled as a sum of deflection due to bending and shear deformation [4]:

$$\delta = \delta_{\text{bending}} + \delta_{\text{shearing}} = \frac{FL^3}{192E_{\text{Young}}I} + f_{\text{s}}\frac{FL}{4GA} = \frac{FL^3}{192E_{\text{bending}}I} \qquad (26.5)$$

where f_{s} is the geometric factor, equal to 10/9 for a cylinder, G the intertube shear modulus and A the area of the beam's cross section. E_{bending} is the effective, bending modulus, equal to the Young's modulus in the case where the influence of intertube shearing can be neglected.

The Young's and the intertube shear modulus can both be determined by measuring E_{bending} as a function of bundle diameter to length ratio. For

thin ropes, the bending modulus thus corresponds with the Young's modulus, while for thick ropes one obtains the value of the intertube shear modulus G, on the order of 1 GPa. Young's modulus is on the order of 1 TPa for both isolated SWNTs and MWNTs, which do not show any dependence of the mechanical response on the tube diameter, see Fig. 26.11. Different synthesis methods can result in widely varying Young's modulus of carbon nanotubes, due to degradation of the graphitic structure with the introduction of defects during production. An average value of $E = 870$ GPa was found for the arc-discharge-grown tubes, while the catalytically grown MWNTs, known to include a high concentration of defects can have a Young's modulus as low as 12 GPa [12, 18], demonstrating the utility of this measurement method in choosing the optimal synthesis conditions for producing carbon nanotubes with favorable mechanical properties [12].

Variations of this technique have been successfully implemented for measurements of other materials with similar geometry like MoS_2 nanotube bundles [9] and components of the cellular cytoskeleton, the microtubules [8].

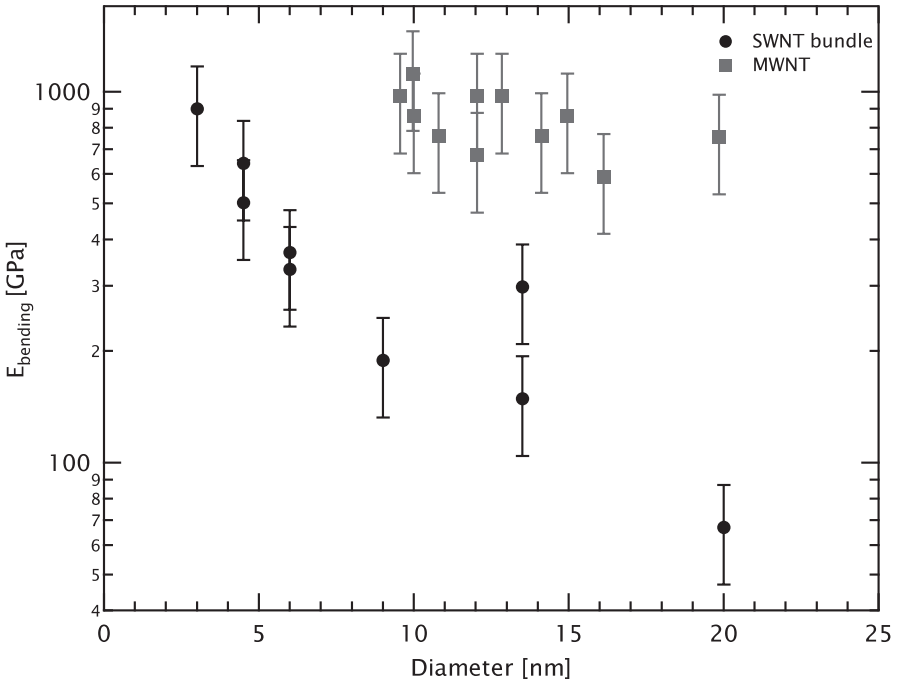


Fig. 26.11. a Values of the bending modulus for 12 SWNT bundles of different diameters. The measured E_{bending} of thin bundles corresponds to E_{Young}, while for thick bundles one obtains the value of the shear modulus G [17]. MWNT data is for arc-discharge-grown tubes [18]

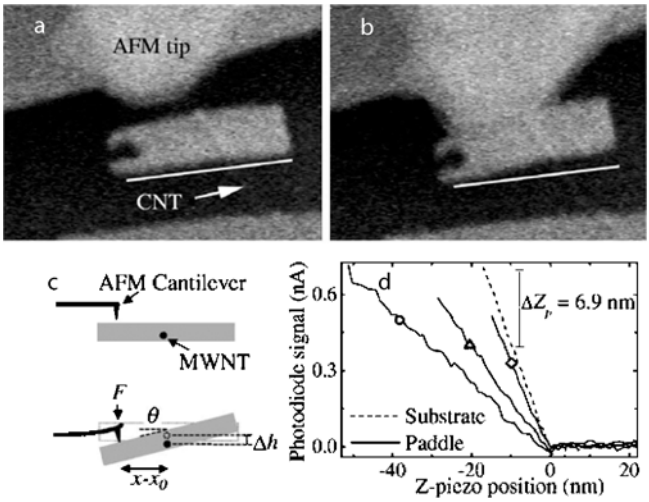


Fig. 26.12. **a** SEM image of an AFM tip close to the paddle of the torsional oscillator **b** The device during measurement. **c** Schematic drawing of the measurement procedure. **d** Force–distance curves on the substrate and at three different positions along the paddle [24]

AFM operated in the vertical deflection mode was also used by Williams et al. [24] to measure the torsional response of MWNTs incorporated as spring elements into "paddle" oscillators. The paddle was fabricated at the center of MWNTs dispersed on silicon substrates with a 500-nm thick oxide layer where they were pinned down by large metal pads. The oxide was etched so that the paddles were completely suspended, but the larger pads remained anchored to the substrate. Measurements were performed using an AFM mounted inside an SEM and the AFM tip was used to apply vertical force on the paddle, causing the MWNT to twist, see Fig. 26.12a and b. Such measurements resulted in force–distance curves like the one on Fig. 26.12c, yielding values of the single-shell shear modulus in the 210 – 830 GPa range.

26.4 Applications

26.4.1 Intertube Binding Using Electron Beam Irradiation

Previous measurements by Salvetat et al. [17] of the elastic properties of carbon nanotube bundles have shown that they behave as a bunch of loosely bound individual tubes. Even though individual nanotubes have mechanical properties comparable to diamond, the weak interaction between the tubes will cause macroscopic structures to fall apart. This structural weakness of nanotube agglomerates is a serious obstacle in their application as reinforcing

elements. It is not only limited to bundles of single-walled nanotubes. Whenever nanotubes aggregate – be it single or multiwalled – the weak intertube interaction will cause them to easily slide with respect to each other. This is clearly illustrated in the example of macroscopic fibers composed of carbon nanotubes [21] connected only via van der Waals interaction or using intercalated polymers. With the highest Young's modulus of 80 GPa for nanotube fibers [3] these remarkable structures still fail to take full advantage of the superior mechanical properties of carbon nanotubes. Introducing stable connections between nanotubes, without significantly degrading their Young's modulus in the process, would bring their application closer to reality. The best mechanical properties would be achieved in the ideal case when the intertube connections would involve only carbon atoms, because of their high strength.

Kis et al. [7] have demonstrated that stable links can be formed between nanotubes in bundles during exposure to electron-beam irradiation. Crosslinking was detected by measuring changes in the mechanical response of SWNT bundles deposited on electron-transparent substrates in the form of silicon nitride membranes. The use of silicon nitride membranes allowed mechanical measurements with AFM, TEM observation and irradiation to be combined for the first time on the same, isolated SWNT bundle, see Fig. 26.13a and b. Mechanical measurements were performed using an AFM by deflecting nanotube bundles deposited over perforations in the membrane.

The same nanotube bundle was located in a transmission electron microscope (TEM), imaged with high resolution and irradiated with a focused electron beam. Measurements of the bending modulus performed after each irradiation step showed a 30-fold increase of the bending modulus on low exposure with gradual degradation of the mechanical properties due to amorphization at longer exposures, see Fig. 26.14. This method of depositing nanotubes on patterned, electron-transparent membranes could enable a host of related experiments in which TEM, SEM and AFM observations on nan-

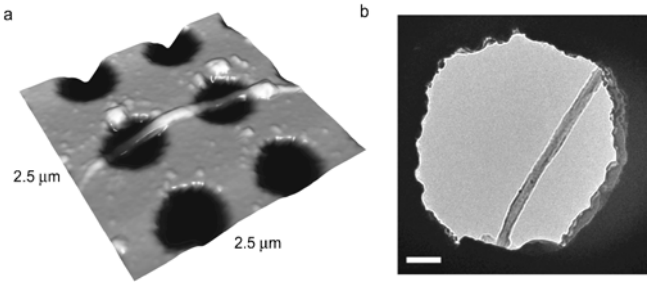


Fig. 26.13. a 3D rendering based on the AFM image of a carbon nanotube rope spanning two perforations on a silicon nitride membrane. Measurements were performed on the right-side hole. **b** TEM image of the same nanotube. The scale bar on the TEM image is 100 nm long

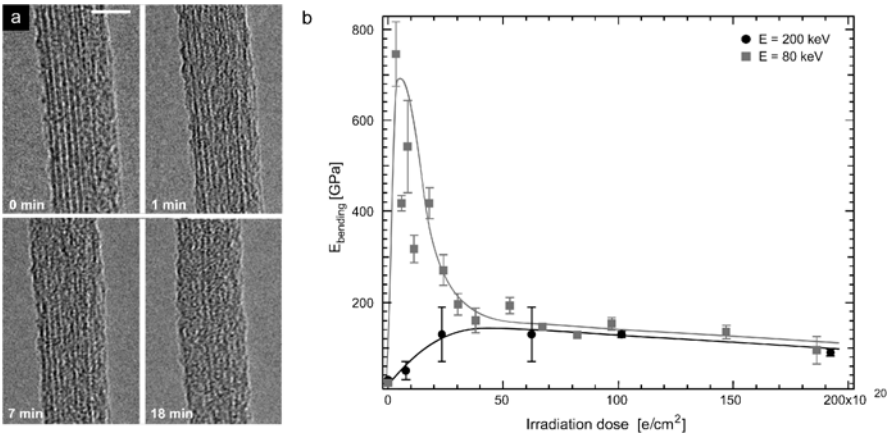


Fig. 26.14. A series of TEM images of a carbon nanotube bundle. The bundle becomes increasingly amorphous as the exposure to electrons is increased. The scale bar is 10 nm long. **b** Behavior of the bending modulus E_{bending} of different carbon nanotube ropes as a function of received dose for two incident electron energies. The bending modulus increases on short exposures due to crosslinking and degrades at higher exposures because of structural damage. The rope irradiated with 80-keV electrons shows a much stronger and sharper increase of the bending modulus [7]

otubes and nanowires could be combined. Fabrication of additional electrical contacts on these electron-transparent membranes would allow in situ measurements of electrical transport and electromechanical properties.

26.4.2 Mechanical Properties of Catalytically Grown Carbon Nanotubes

Carbon nanotubes (CNTs), either single wall or multiwall, should be ideal reinforcing fibers for composite materials, since their Young's modulus can have values of ~ 1 TPa [11]. The method of chemical carbon vapor deposition (CCVD) is the most promising in terms of large-scale production needed for industrial purposes, due to the relative ease of upscaling the preparation and purification methods [2]. It can produce CNTs in quantities of kilograms to tons at low temperatures of around 700 ^{o}C. There, the CNTs grow by precipitation of carbon, dissolved in a metallic catalyst particle that is oversaturated in carbon at one part of the surface [1]. However, such catalytically grown CNTs are generally inferior to CNTs made by other methods, both in terms of structure and the Young's modulus that is below 100 GPa [18].

Lukic et al. [12, 13] have demonstrated that the Young's modulus of catalytically grown CNTs is strongly influenced by the number of walls and diameter. Specifically, the Young's modulus of MWNTs with the diameter between 20 and 50 nm stays below 100 GPa (Fig. 26.15a). When the number

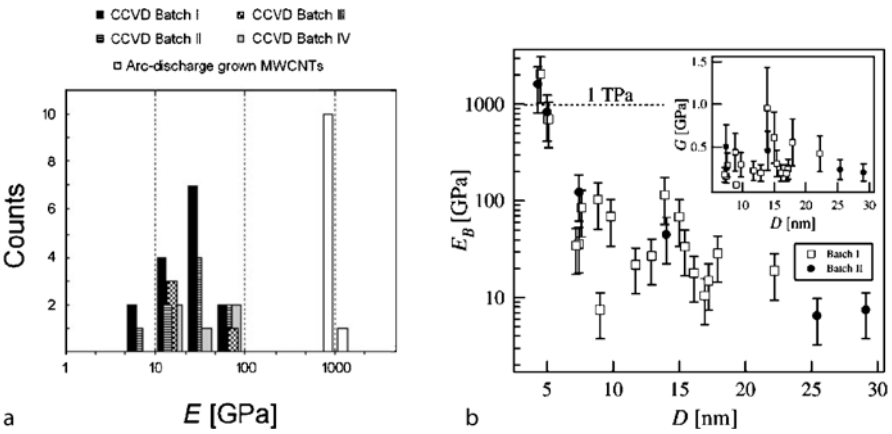


Fig. 26.15. Summary of mechanical measurements of catalytically grown CNTs. **a** Histogram of average values of Young's modulus E for four different growth conditions (batches). There are three bins per decade, where the range of bins is constant on a logarithmic scale. Values for arc-discharge-grown MWNTs are shown separately. **b** Bending modulus E_B versus diameter D for CNT bundles consisting mostly of double-wall CNTs. Theoretical value of 1 TPa is indicated by a dashed line and is reached for small bundles or individual nanotubes. Decrease in E_B with the diameter comes from shearing deformation of the bundles

of walls is decreased to around two and the radius to 5 nm, the Young's modulus reaches the value of 1 TPa (Fig. 26.15b). Changes in growth conditions, like the material of the catalyst or the synthesis temperature, do not play a significant role in determining the Young's modulus.

26.4.3 Diameter Dependence of the Bending Modulus of the MWCNTs

New experimental results by Lee et al. [10] report a strong diameter dependence of the bending modulus of CCVD-grown MWCNTs. In the case of MWCNTs, the bending modulus may be considered as the Young's modulus, as the shear is often very small. The new data (batch 3 on Fig. 26.15a) fill the gap between previous data on MWCNTs [13]. (batch 1 and 2 in Fig. 26.16a, $E_B < 100$ GPa) and data on small-diameter DWCNT ropes [12] (CCVD DWCNTs in Fig. 26.16a, $E_B \approx 1$ TPa).

Figure 26.16b shows a comprehensive plot of the reported data on carbon nanotubes. [15,17,18,20]. The bending modulus of MWCNTs approaches the 1 TPa limit exhibited by arc-discharge CNTs, as the diameter decreases. An empirical fit suggests the 1 TPa value for CCVD-grown MWCNT with diameter below 5 nm. The nanotube diameter is directly related to the catalyst particle size in CCVD, and thus the effect of catalyst particle size on the kinetics of the growth process should be investigated in the future.

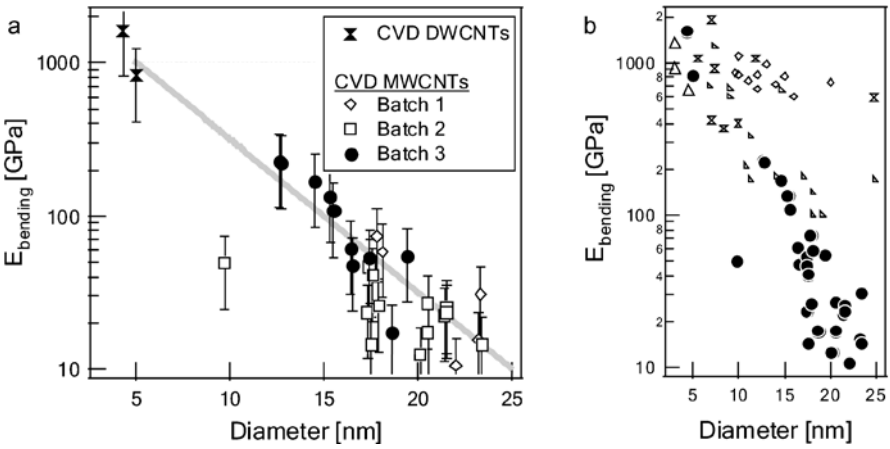


Fig. 26.16. a The bending moduli of different batches of CCVD-grown MWCNTs show a strong diameter dependence. Shear is negligible during the bending of MWCNTs, in which case the bending modulus may be considered as the Young's modulus. The gray line indicates a fit on the MWCNT data, $E_{\text{bending}} = 1\,\text{TPa} \cdot \text{e}^{-(D-5\,\text{nm})/10\,\text{nm}}$. Only the small-diameter (< 5 nm) data have been selected from the DWCNT rope data, as the DWCNT ropes with larger diameter exhibit non-negligible shear. **b** The CCVD-grown MWCNT data (*filled circles*) are shown together with the data for arc-discharge CNTs: sand clock – MWCNT data by [20], diamond – MWCNT data by [18], equilateral triangle – SWCNT rope data by [17], and right triangle – MWCNT data by [15]

References

1. Amelinckx, S., X.B. Zhang, D. Bernaerts, X.F. Zhang, V. Ivanov and J.B. Nagy (1994). "A Formation Mechanism for Catalytically Grown Helix-Shaped Graphite Nanotubes." *Science* **265**(5172): 635–639.
2. Dai, H.J. (2001). Nanotube growth and characterization. *Carbon Nanotubes*. **80:** 29–53.
3. Dalton, A.B., S. Collins, E. Munoz, J.M. Razal, V.H. Ebron, J.P. Ferraris, J.N. Coleman, B.G. Kim and R.H. Baughman (2003). "Super-tough carbon-nanotube fibres." *Nature* **423**: 703.
4. Gere, J.M. and S.P. Timoshenko (1984). Mechanics of materials, PWS-Kent, Boston**:** 690.
5. Iijima, S., C. Brabec, A. Maiti and J. Bernholc (1996). Structural flexibility of carbon nanotubes. *Journal of Chemical Physics*. **104:** 2089–2092.
6. Kis, A. (2003). Mechanical properties of mesoscopic objects. PhD thesis. *Faculte des Sciences des Base*. Lausanne, EPFL.
7. Kis, A., G. Csanyi, J.-P. Salvetat, T.-N. Lee, E. Couteau, A.J. Kulik, W. Benoit, J. Brugger and L. Forró (2004). Reinforcement of single-walled nanotube bundles by intertube bridging. *Nature Materials*. **3:** 153–7.
8. Kis, A., S. Kasas, B. Babic, A.J. Kulik, W. Benoît, G.A.D. Briggs, C. Schönenberger, S. Catsicas and L. Forró (2002). Nanomechanics of microtubules. *Physical Review Letters*. **89:** 248101.

9. Kis, A., D. Mihailovic, M. Remskar, A. Mrzel, A. Jesih, I. Piwonski, A.J. Kulik, W. Benoit and L. Forró (2003). Shear and Young's moduli of MoS_2 nanotube ropes. *Advanced Materials*. **15:** 733–736.
10. Lee, K., Lukic B, Magrez, A. Seo, J. W., Kulik, A.J. and Forro, L. (2005). "Diameter dependence of the elastic modulus of CVD-grown nanotubes." *AIP Conference Proceedings* **786**: 143–145.
11. Lu, J. P. (1997). "Elastic properties of carbon nanotubes and nanoropes." *Physical Review Letters* **79**(7): 1297–1300.
12. Lukic, B., J.W. Seo, R.R. Bacsa, S. Delpeux, F. Béguin, G. Bister, A. Fonseca, J.B. Nagy, A. Kis, S. Jeney, A.J. Kulik and L. Forró (2005a). Catalytically Grown Carbon Nanotubes of Small Diameter Have a High Young's Modulus. *Nano Letters*. **5:** 2074–2077.
13. Lukic, B. S., J.W., Couteau, E., Lee, K., Gradecak, S., Berkecz, R., Hernadi, K., Delpeux, S., Cacciaguerra, T., Beguin, F., Fonseca, A., Nagy, J.B., Csanyi, G., Kis, A., Kulik, A.J. and Forro, L. (2005b). "Elastic modulus of multi-walled carbon nanotubes produced by catalytic chemical vapor deposition." *Applied Physics A-Materials Science Processing* **80**(4): 695–700.
14. Meyer, E., H. Heinzelmann, P. Grutter, T. Jung, H.R. Hidber, H. Rudin and H.J. Guntherodt (1989). Atomic Force Microscopy for the Study of Tribology and Adhesion. *Thin Solid Films*. **181:** 527–544.
15. Poncharal, P., Z.L. Wang, D. Ugarte and W.A. and de Heer (1999). "Electrostatic deflections and electromechanical resonances of carbon nanotubes." *Science* **283**(5407): 1513–1516.
16. Salvetat, J.-P., J.M. Bonard, N.H. Thomson, A.J. Kulik, L. Forró, W. Benoit and L. Zuppiroli (1999a). Mechanical properties of carbon nanotubes. *Applied Physics A*. **69:** 255–260.
17. Salvetat, J.-P., G.A.D. Briggs, J.-M. Bonard, R.R. Bacsa, A.J. Kulik, T. Stöckli, N. Burnham and L. Forró (1999b). Elastic and Shear Moduli of Single-Walled Carbon Nanotube Ropes. *Physical Review Letters*. **82:** 944–947.
18. Salvetat, J.-P., A.J. Kulik, J.-M. Bonard, G.A.D. Briggs, T. Stöckli, K. Méténier, S. Bonnamy, F. Béguin, N.A. Burnham and L. Forró (1999c). Elastic Modulus of Ordered and Disordered Multiwalled Carbon Nanotubes. *Advanced Materials*. **11:** 161–165.
19. Schwarz, U.D., P. Koster and R. Wiesendanger (1996). Quantitative analysis of lateral force microscopy experiments. *Review of Scientific Instruments*. **67:** 2560–2567.
20. Treacy, M.M.J., T.W. Ebbesen and J.M. Gibson (1996). "Exceptionally high Young's modulus observed for individual carbon nanotubes." *Nature* **381**(6584): 678–680.
21. Vigolo, B., A. Penicaud, C. Coulon, C. Sauder, R. Pailler, C. Journet, P. Bernier and P. Poulin (2000). Macroscopic Fibers and Ribbons of Oriented Carbon Nanotubes. *Science*. **290:** 1331–1334.
22. Volodin, A., M. Ahlskog, E. Seynaeve, C. Van Haesendonck, A. Fonseca and J.B. Nagy (2000). "Imaging the elastic properties of coiled carbon nanotubes with atomic force microscopy." *Physical Review Letters* **84**(15): 3342–3345.
23. Walters, D.A., L.M. Ericson, M.J. Casavant, J. Liu, D.T. Colbert, K.A. Smith and R.E. Smalley (1999). "Elastic strain of freely suspended single-wall carbon nanotube ropes." *Applied Physics Letters* **74**: 3803–3805.

24. Williams, P. A., S.J. Papadakis, A.M. Patel, M.R. Falvo, S. Washburn and R. Superfine (2002). "Torsional Response and Stiffening of Individual Multi-walled Carbon Nanotubes." *Physical Review Letters* **89**: 255502.
25. Wong, E.W., P.E. Sheehan and C.M. Lieber (1997). Nanobeam Mechanics: Elasticity, Strength and Toughness of Nanorods and Nanotubes. *Science*. **277:** 1971–1975.
26. Wu, B., A. Heidelberg and J.J. Boland (2005). "Mechanical properties of ultrahigh-strength gold nanowires." *Nature Materials* **4**: 525–529.
27. Yu, M.-F., B.S. Files, S. Arepalli and R.S. Ruoff (2000). Tensile Loading of Ropes of Single Wall Carbon Nanotubes and their Mechanical Properties. *Physical Review Letters*. **84:** 5552–5555.
28. Yu, M.-F., O. Lourie, M.J. Dyer, K. Moloni, T.F. Kelly and R.S. Ruoff (2000). Strength and Breaking Mechanism of Multiwalled Carbon Nanotubes Under Tensile Load. *Science*. **287:** 637–640.

27 Theory of Adsorption and Manipulation of C_{60} on the Si(001) Surface

Natalia Martsinovich, Chris Hobbs, and Lev Kantorovich

Department of Physics, King's College London, Strand, London, UK, WC2R 2LS

27.1 Introduction

Nanotechnology is today widely considered as the key technology of the future. Manipulation of atoms and molecules on crystal surfaces using scanning probe microscopy (SPM) techniques, such as scanning tunnelling microscopy (STM) and atomic force microscopy (AFM), represents an important element of the technique since it provides a way of introducing atoms and/or functionalised molecules to the surface when and where required. Experimentally, manipulation of single atoms and molecules has long been available with STM [1–6]. Recently, first successful atomistic manipulations have also been achieved using Non-Contact (NC) AFM [7, 8].

Several manipulation modes have been identified [2]: attractive, repulsive and sliding. In the attractive regime, the molecule is positioned behind the STM tip and is moved behind it. This regime is only possible if there is a significant attractive interaction between the tip and the molecule. In the repulsive mode, the tip is positioned behind the molecule and the latter is pushed by the tip due to repulsive forces. In the sliding mode the tip positioned somewhere above the molecule drags it along the manipulation direction.

Typical STM tip trajectories during attractive and repulsive manipulation are shown schematically in Fig. 27.1. Actual trajectories recorded in a constant current mode for the manipulation of a C_{60} molecule on the Si(001) surface [9] are shown in Fig. 27.2. Both types of tip trajectories during a constant current manipulation show characteristic sawtooth patterns. In the attractive manipulation, the line scan contains a sharp vertical retraction of the tip from the molecule followed by a gradual approach. The sharp retraction is explained as follows: when the tip is first moved towards the molecule (the lower line-scan in the right panel of Fig. 27.2), the molecule jumps towards the tip due to attractive interaction between them, and the tip is retracted by the feedback system of the STM machine to reduce the current to the set value. When the tip passes over the molecule, the same process happens as shown in the right panel of Fig. 27.1, but this time the molecule hops to the tip in the direction of the tip movement, as the latter moves away from the molecule. In the repulsive manipulation, the left panels of Figs. 27.1 and 27.2,

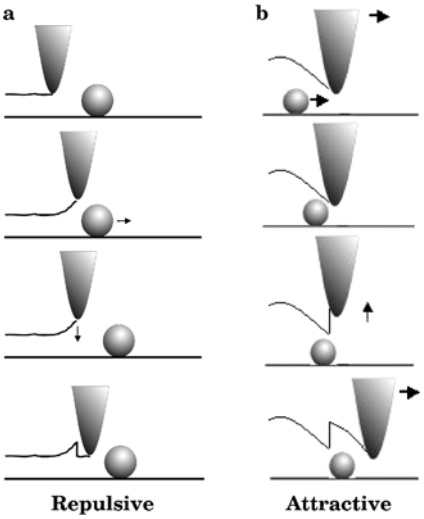


Fig. 27.1. SPM tip trajectories for repulsive (**a**) and attractive (**b**) manipulation regimes

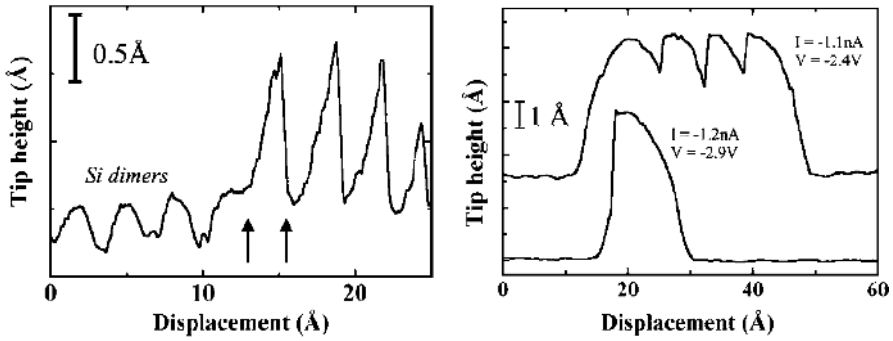


Fig. 27.2. Line scans showing repulsive (*left*) and attractive (*right*) modes of manipulation of a C_{60} molecule on the Si(001) surface (taken with modifications from [9]). The arrows in the left panel mark the tip position where the molecule hops away from the tip. The lower and upper trajectories in the right image correspond to the tip either before or after it passed over the molecule. (Reprinted from *Chem. Phys. Lett.*, Vol. 366, D. L. Keeling et al., Attractive mode manipulation of covalently bound molecules, pp. 300-304, Copyright (2002), with permission from Elsevier)

a gradual retraction of the tip is first observed that is followed by a sharp approach to the surface. In this case when the tip approaches the molecule, the current increases and the tip is gradually retracted from the surface; at the onset of the repulsive interaction at close approach between the molecule and the tip, the molecule jumps away from the tip, and, to increase the tunnelling current to the set value again, the feedback system moves the tip down to the surface.

Attractive or repulsive manipulation can be achieved by changing the tip-surface distance [9]. Attractive manipulation requires the formation of chemical bonds between the tip and the molecule or atom and hence a reactive tip, whereas the repulsive manipulation can even be achieved when the tip is inert with respect to the bonding to the adsorbate.

Most of the STM manipulation experiments have been performed on metal surfaces which interact with adsorbates rather weakly, the molecules easily diffuse across the surface at elevated temperatures and thus manipulation experiments are to be performed at relatively low temperatures [1–5]. The manipulation mechanisms in these cases are well understood. If, however, adsorbates are strongly (i. e. chemically) bound to the surface, then it is more difficult to manipulate them. On the other hand, successful manipulation experiments can be performed at room temperature [6]. In this paper we shall review a combined experimental and theoretical work on one of such systems, namely, the repulsive mode manipulation of C_{60} molecules adsorbed on the Si(001) surface, for which the manipulation mechanism has been proposed only recently [10]. Relevant experimental results are briefly summarised in the next section. In the following sections we shall review the theoretical work done in our group to explain these STM manipulation experiments.

27.2 Repulsive manipulation of C_{60} molecules on the Si(001) surface: a summary of STM results

The Si(001) surface is anisotropic, with surface Si atoms forming rows of dimers. Dimers are buckled at low temperature [11, 12], and in the ground state the buckling directions alternate both along and across the dimer rows (the c(4x2) reconstruction). However, at room temperature the dimers perform a fast flip-flop motion so that on average the dimers can be considered as identical along each row and in the symmetric (not buckled) state.

The C_{60} molecule, when deposited on the surface, was found to prefer adsorption sites above the troughs between dimer rows [13]. In this case the dimer rows serve as rails during the C_{60} manipulation. As one can expect, manipulation along the trough occurs with a higher success rate than across the rows [13]. Note that at higher temperature, after annealing, on-the-row adsorption sites were found as well [13]. In the following, we shall mostly consider the more energetically favourable on-the-trough configurations.

STM manipulation was performed using the constant current mode. Initially, the surface is scanned with a relatively small set current; next, the set current is increased and the tip is moved closer to the molecule following the previously acquired image. Then, the tip is displaced laterally towards the molecule (the repulsive mode), keeping the current constant. During these experiments, the tip height is recorded as a function of its lateral displacement. Finally, the tip is retracted by returning to the initial scanning mode set current, and the scan is repeated to obtain a new image of the surface

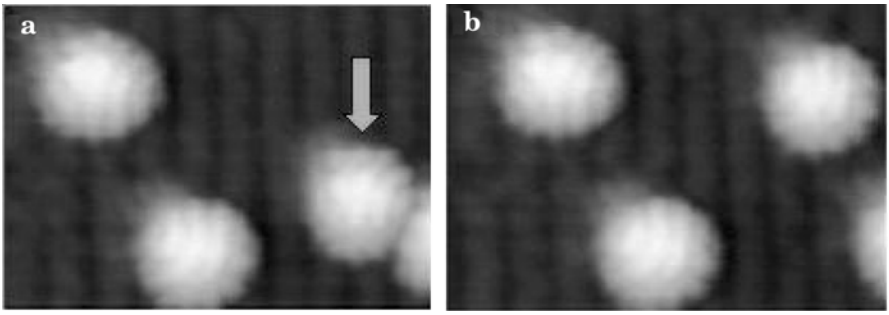


Fig. 27.3. STM images before **a** and after **b** a single manipulation experiment [13]. The C_{60} molecule to be moved is indicated in **a** by the arrow. (Reprinted from *Surf. Science*, Vol. 407, P. Moriarty et al., Translation, rotation and removal of C_{60} on Si(001)-2×1 using anisotropic molecular manipulation, pp. 27–35, Copyright (1998), with permission from Elsevier.)

with the adsorbed molecule. Typical images before and after a single manipulation experiment are shown in Fig. 27.3. An important conclusion that one can make from those high resolution images is that the molecule changes its *orientation* after the manipulation event.

In the following sets of experiments [10], a long-range periodicity was observed in STM tip trajectories during a repulsive along-the-trough manipulation of a C_{60} molecule on the same surface as shown in Fig. 27.4.

Two types of waveforms were observed: with the periodicity of $4a_0$ (tip trajectories (a) and (b)) and $3a_0$ (tip trajectories (c) and a stand-alone waveform in (a)); here $a_0 = 3.84$ Å, the surface lattice constant, is the distance between Si–Si dimers along the row.

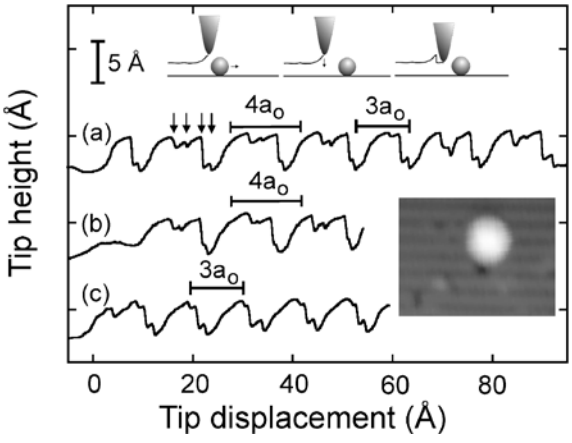


Fig. 27.4. Periodic tip trajectories recorded during three independent repulsive manipulation experiments (a–c) [10]. (Reprinted from D.L. Keeling et al., Phys. Rev. Lett., v. 94, #146104, 2005. Copyright (2005) by the American Physical Society)

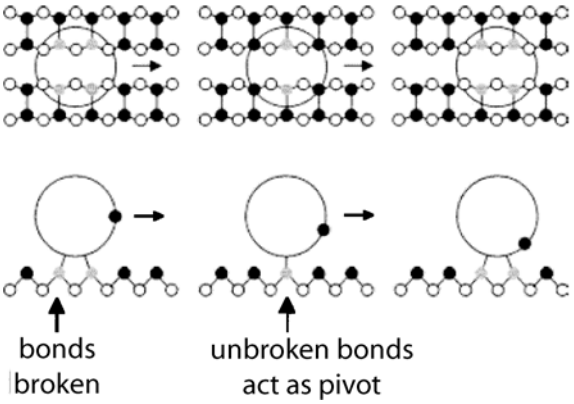


Fig. 27.5. Schematic of the pivoting mechanism of C_{60} molecule rolling along the trough during manipulation [10]. (Reprinted from D.L. Keeling et al., Phys. Rev. Lett., v. 94, #146104, 2005. Copyright (2005) by the American Physical Society)

The observed periodicities imply that the molecule makes four or three hops within a single waveform (as indicated by arrows on one of them) from one stable configuration to another, and then arrives at the configuration equivalent to the initial one. Hence, it makes a 360° rotation during a three or four lattice constants translation. The shape of the waveform substructure suggests repulsive mode of manipulation as in Fig. 27.2 (a). Note that almost identical waveforms can be seen in different independent manipulation experiments shown in Fig. 27.4, curves (a)–(c). This suggests an identical sequence of adsorption sites in all these manipulation experiments. Combining this evidence with the previously mentioned experimental observations that the C_{60} molecule changes its orientation during the manipulation experiment, a *pivoting mechanism* for its manipulation was proposed [10]. According to this mechanism, the translational movement of the molecule along the trough is coupled with its rotation, as shown schematically in Fig. 27.5.

It follows from the previous *ab initio* calculations [14–17] that the molecule forms four covalent bonds with the surface in all its stable adsorption configurations. According to the pivoting model, as the molecule moves across the surface, the back pair of its bonds to the surface are broken, and it pivots over the two remaining front bonds. After pivoting, a new pair of bonds with the surface are formed and the molecule arrives at a new stable configuration with four C-Si bonds. The pivoting mechanism does not require all four molecule-surface bonds to be broken at the same time and thus ensures a relatively low energy barrier. Note that sliding mechanism would require all four bonds to be broken, which would result in a much larger energy barrier.

One interesting observation can also be made in particular about the character of the $3a_0$ waveforms. Firstly, the initial part of the form (see specifically the $3a_0$ waveform in the sequence (a) in Fig. 27.4) looks very similar to the initial part of the $4a_0$ form, which suggests that in both cases the same first two stable configurations are to be expected in the two manipulation sequences. Secondly, the final parts of the two waveforms also look very similar

suggesting that in both cases the molecule arrives at the same stable configuration. Thus, it follows from this comparison that the $3a_0$ waveform differs from the $4a_0$ one in that one intermediate stable configuration is missing, i. e. in the $3a_0$ sequence the molecule performs the complete 360° rotation in three rather than four steps by following a sequence of three stable sites of the $4a_0$ waveform. This means that for some reason one of the configurations of the $4a_0$ waveform does not materialise in the $3a_0$ sequences, otherwise, these sequences are very similar to the $4a_0$ ones and may also be explained by the pivoting mechanism.

Thus, existing experimental data suggest the pivoting mechanism for the manipulation of the C_{60} molecule on the Si(001) surface. However, theoretical calculations are necessary in order to elucidate atomic-scale details of the C_{60} adsorption and diffusion on this surface. We shall first review our results on the C_{60} adsorption [17]. Then, results on the C_{60} manipulation reported in [10, 18] will be considered in some detail.

27.3 Method

In our modelling we use the density functional theory as implemented in the SIESTA code [19], the Perdew-Burke-Ernzerhof (PBE) generalised gradient approximation (GGA) functional [20] for exchange and correlation, norm-conserving pseudopotentials and periodic boundary conditions. A localised double-zeta polarised (DZP) basis set is used to describe the valence electrons, which is composed of a double (split valence) basis for each valence orbital plus polarisation functions for all the atoms. Geometry relaxation was performed until forces on atoms were smaller than $0.04\,\mathrm{eV/\AA}$.

A well-known problem associated with the use of localised basis sets applied to compound systems is the basis set superposition error (BSSE). It arises because each component of the compound system (in our case, the C_{60} molecule and the silicon surface) can use basis orbitals of the other components to expand its wavefunction. Thus, a bigger set of basis orbitals is available to each component in calculations of compound systems compared to the calculations of isolated components, which affects the calculation of the binding energy: the binding energies are overestimated. All energies reported in this paper have the BSSE correction included, calculated according to the Boys-Bernardi counterpoise method [21] as described in [17].

In order to understand the nature of chemical bonding between the C_{60} molecule and the Si(001) surface, electron density difference plots were studied. To do this, electron density difference $\Delta\rho(\mathbf{r})$ was calculated between the densities of the combined C_{60}-silicon system and individual densities of the C_{60} and the silicon slab in the geometry of the combined C_{60}-Si system:

$$\Delta\rho(\mathbf{r}) = \rho_{C_{60}+Si}(\mathbf{r}) - (\rho_{C_{60}}(\mathbf{r}) + \rho_{Si}(\mathbf{r})) \tag{27.1}$$

The electron density difference plots provide invaluable information on the charge transfer between the adsorbate and the surface, as well as on the density redistribution within each subsystem due to the bonding between them.

27.4 Simulation results

27.4.1 Adsorption of C_{60} on Si(001)

As follows from Sect. 27.2, in order to understand the experimentally acquired manipulation sequences, it is important first to find all stable adsorption sites of the C_{60} molecule on the Si(001) surface. In order to address this question, an extensive set of adsorption calculations has been undertaken [16, 17]. The first theoretical studies of this system have been done in [14, 15]. In our studies [16, 17] we have found a significant number of additional adsorption sites as described below in more detail.

The C_{60} molecule can be adsorbed on the surface either above a row or a trough. Four *classes* [15] of adsorption sites can be identified: (i) above the trough centred between four Si–Si dimers (class t4); (ii) above the trough between two dimers (t2); (iii) above the row between two dimers (r2) and (iv) on the row above one dimer (r1). Experimental STM images indicate that C_{60} molecules are mostly adsorbed above the trough rather than above the row [13].

In order to show all configurations, we shall adopt their schematic representation in which only the carbon atoms that are directly bound to the surface are shown, along with the hexagons and/or pentagons of the fullerene cage that are closest to the surface. The dimer rows of the Si surface are indicated by two parallel grey stripes. The way in which the schematic picture of the actual adsorption configuration is derived is shown in Fig. 27.6.

Schematic pictures of all stable adsorption sites found on the through and the row are given in Figs. 27.7 and 27.8, respectively.

The theoretical calculations show that in both cases the most stable classes of configurations $t4a - t4i$ and $r2a - r2h$ have four C-Si bonds: four Si dimers participate in the case of on-the-trough configurations, while in the on-the-row geometries the cage forms four bonds with two dimers. The binding energies of these configurations are given in Table 27.1. They range from –0.87 to –2.63 eV, i. e.from –0.22 to –0.66 eV per Si–C bond relative to the isolated C_{60} molecule and the Si(001) surface in the buckled p(2×1) reconstruction. In all cases the BSSE corrections were found to be significant (1.0 – 1.2 eV) [17].

It should be noted that the formation of Si–C bonds between the molecule and the surface requires some substantial rearrangement of single and double bonds in the lower part of the C_{60} molecule. As an example, we consider the highest binding energy configuration $t4c$ in which the molecule faces the surface with two hexagons. The bond rearrangement within the cage can

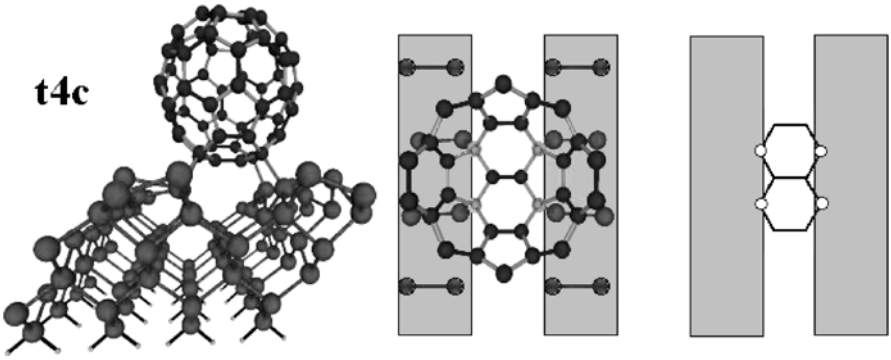


Fig. 27.6. Derivation of the schematic picture of the on-the-trough t4c configuration of the C_{60} molecule on the Si(001) surface from its full 3D configuration (*left panel*), top view (*middle panel*) when only the bottom cap of the cage is shown and the carbon atoms directly bound to the Si dimer atoms shown in a lighter colour, and, finally, its schematic view (*right panel*)

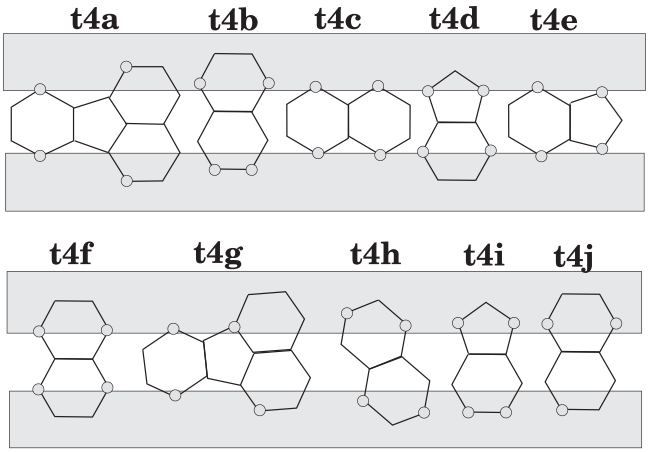


Fig. 27.7. Schematic pictures for sites t4 [17]. The configuration t4j is obtained from t4b by removing a single Si–C bond due to unfavourable orientation of the corresponding surface dimer (see text). (Reprinted from *Surf. Science*, Vol. 591, C. Hobbs et al., An ab initio study of C_{60} adsoprtion on the Si(001) surface, pp. 45-55, Copyright (2005), with permission from Elsevier.)

be analysed either by comparing bond lengths in the C_{60} before and after adsorption [15] or also using the electron density difference [17], as shown in Fig. 27.9. In the Figure, regions where the electron density of the surface-adsorbate system is higher than that of the isolated C_{60} and silicon surface are shown in *red*, and regions with the lower electron density are shown in *black*. One can see that some of the double bonds of an isolated C_{60} transformed into single ones, while some of the single bonds turned into double ones. These

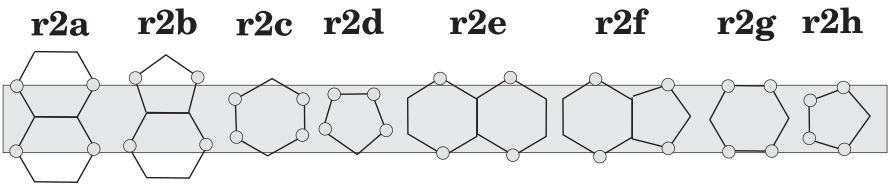


Fig. 27.8. Schematic pictures for sites r2 [17, 18]. (Reprinted from *Surf. Science*, Vol. 591, C. Hobbs et al., An ab initio study of C_{60} adsoprtion on the Si(001) surface, pp. 45–55, Copyright (2005), with permission from Elsevier.)

Table 27.1. Calculated adsorption energies (in eV) for the C_{60} molecule on the Si(001) surface in the trough (*left*) and in the row (*right*) configurations relative to the isolated C_{60} molecule and the Si(001) surface in the buckled p(2×1) reconstruction(taken from [17, 18])

t4a	–2.16	r2a	–2.58
t4b	–2.53	r2b	–2.49
t4c	–2.63	r2c	–2.00
t4d	–1.96	r2d	–0.87
t4e	–1.90	r2e	–1.69
t4f	–2.17	r2f	–2.19
t4g	–2.53	r2g	–1.10
t4h	–1.94	r2h	–0.63
t4i	–1.83		
t4j	–1.98		

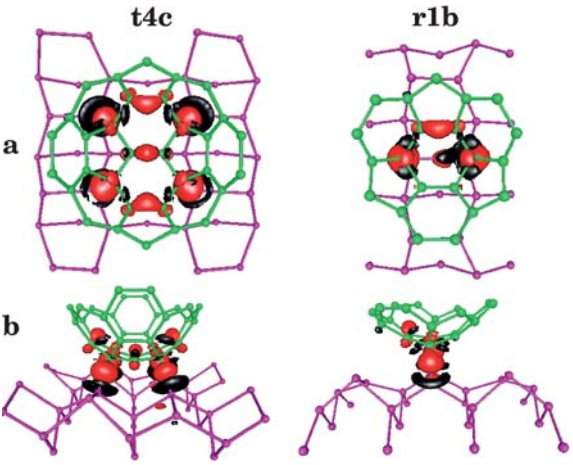


Fig. 27.9. The electron density difference, Eq. (27.1), of the C_{60} molecule adsorbed in the configurations *t4c* and *r1b*: **a** top view; **b** side view. Only the lower part of the cage is shown

bond rearrangements facilitate the formation of the four covalent bonds of the cage with the surface, which are also clearly seen in the Figure.

It was noticed in [15] that more extensive bond rearrangement generally produces higher-energy structures, whereas structures with minimum bond

rearrangement tend to be the lowest-energy ones. For most of the structures this conclusion was also confirmed in our calculations.

Note that adsorption configurations involving three Si–Si dimers above the trough are also possible, that is only three Si–C bonds are formed between the C_{60} and the surface. For example, the configuration $t4j$ in Fig. 27.7 is obtained from $t4b$ by removing one Si–C bond. Such configurations can be formed if one of the Si–Si surface dimers is flipped the opposite way and is therefore separated from the molecule by a sufficient distance for there to be no bonding with that individual dimer. The $t4j$ adsorption configuration with three Si–C bonds is 0.55 eV less stable that its counterpart $t4b$ with four Si–C bonds. This energy difference can be associated with the energy gained by formation of a single Si–C bond between the molecule and the surface. As estimated in [22], the energy barrier to a Si–Si dimer flip is 0.1 – 0.2 eV, which is smaller than the energy gain due to formation of the fourth Si–C bond. Configurations with three Si–C bonds are therefore likely to exist only at low temperatures, when dimers are frozen in their orientations. At higher temperatures dimers are able to flip, and the fourth bond between the surface and the C_{60} will likely be formed. The reverse process of flipping the dimer that is bonded to one of the carbon atoms of C_{60} is unlikely due to the energy cost. Thus, configurations with three Si–C bonds can only be formed during the low-temperature C_{60} adsorption, if surface dimers are initially in an unfavourable orientation for binding to the molecule.

There is also a number of metastable adsorption configurations possible in which the C_{60} is located either above one dimer on the row ($r1$ class of configurations) or between two dimers above the trough ($t2$ configurations) [15, 17]. These configurations have smaller binding energies (between –0.1 and -1.7 eV) and are usually formed with two bonds between the molecule and the surface (see Fig. 27.9), although additional weaker bonds may also be sometimes observed [17]. These configurations are unstable upon small displacement of the molecule along the trough or row and will easily arrive at geometries that form four Si–C bonds with the surface. Although these metastable configurations cannot be observed experimentally, they are of interest as possible intermediate configurations during molecular diffusion and manipulation [10, 18], as will be clear from the following sections.

27.4.2 *Ab initio* modelling of C_{60} diffusion

A part of the C_{60} diffusion path along the trough was modelled in [10,18] using the methods described in Sect. 27.3. Only a $2a_0$-long part of the manipulation pathway was actually modelled due to the significant cost of such calculations. A constrained minimisation technique was used to model the C_{60} diffusion. For this, a single carbon atom of the C_{60} molecule was chosen and displaced in small steps of 0.05 Å along the trough running in the y direction (see the inset in Fig. 27.10). For comparison, in the experiments the STM tip was moved in steps of 0.14 Å [10]. The positions of all other atoms were optimised at

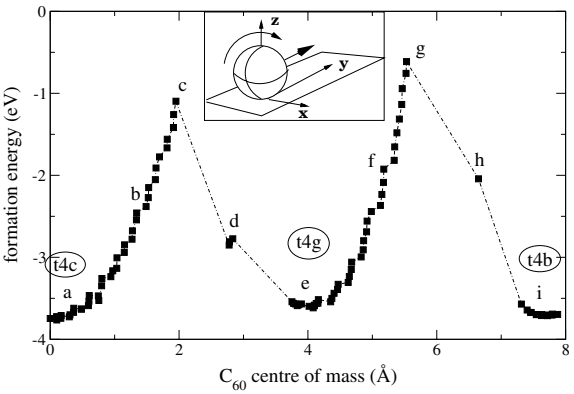


Fig. 27.10. Binding energies of C_{60} along the diffusion path calculated relative to the symmetric p(2×1) reconstructed Si(001) surface (a different reference state of the surface was used in [10]). Stable adsorption configurations *t4c*, *t4g* and *t4b* are marked. Inset: a schematic of the C_{60} rolling

each step, and the position of the displaced atom was optimised in the plane perpendicular to the displacement direction.

Note that this type of calculations describes self-diffusion rather than manipulation of the C_{60}, since an STM tip is not included in the calculations. The issue of C_{60} self-diffusion is interesting in its own right. There have not been systematic studies of the C_{60} diffusion on the silicon surface; however, surface diffusion of C_{60} and higher fullerenes was inferred from fullerene desorption data at temperatures of $300-500^{\circ}$ in an experimental study [23], and jumps of the C_{60} from trough to row positions were observed by STM during annealing at 500° [24]. The experimental results suggest that the barrier to the C_{60} diffusion is very high.

It can be argued, however, that our calculations are relevant to the investigation of the C_{60} manipulation as well. The main difference between manipulation and self-diffusion of the molecule is the presence of an STM tip, which drives the manipulation process. We believe that the tunnelling current in STM between the surface and the tip is not the factor responsible for the manipulation; its role is mainly to control the tip position. Instead, the role of the tip is to push the molecule. Therefore, the computational approach where the movement of the C_{60} is conducted by displacing one of its atoms is an acceptable model to describe externally driven movement of the molecule; note, however, that some specific effects associated with the tip may be missing.

Note that the STM manipulation experiments [10] were conducted at room temperature. It is well known [11, 12] that above 200 K dimers of the Si(001) surface perform a random flip-flop motion, so that on average the surface can be treated as being in the symmetrical p(2×1) phase. Therefore, when calculating the binding energies along the manipulation path, the symmetrical p(2×1) structure of the Si(001) surface was adopted as a reference.

The lowest-energy structure *t4c* was used as the starting geometry for the simulation. Figure 27.10 shows the energy of the molecule, relative to

the isolated C_{60} and the Si(001) surface, along the diffusion pathway with respect to the position of the C_{60} centre of mass. The transition between the configurations $t4c \rightarrow t4g \rightarrow t4b$ (energy minima on the plot) is shown, with the energy barriers between them corresponding to transition states. It can be seen from Fig. 27.10 that the energies at first rise until the barrier of $\sim 2\,\text{eV}$ is reached. After that there is a sharp drop to the next stable configuration. Note that the displacement of the constrained atom was kept constant throughout the simulation. However, the position of the centre of mass of the molecule changed by a large amount, as the molecule fell into the next stable configuration. These few steps were also accompanied by a large energy change. The molecule advanced by $\sim 4\,\text{Å}$, or approximately a_0, as it moved from one stable adsorption configuration to the next one.

The BSSE correction discussed in Sect. 27.3 was added to the raw values of binding energies (Fig. 27.10). The BSSE corrections reduce the calculated binding energies of the system and are equal to $\sim 1.2\,\text{eV}$ along the whole diffusion path.

It is worth noting that the second stable configuration, $t4g$, is asymmetric, despite the initial $t4c$ configuration being symmetric. The two possible alternative configurations, $t4a$ and $t4e$, that can follow $t4c$, are symmetric but are higher in energy than $t4g$.

Three steps can be identified in each elementary translation (by a_0) of the molecule to the next stable configuration. At the first step, between the initial stable configuration and the barrier (points a–c or e–g in Fig. 27.10 for the first or the second elementary translations, respectively), four C-Si bonds are preserved, but both the molecule and the surface undergo strong deformation. At the second step (points c–d or g–h) the two back Si–C bonds are broken, and the molecule has only two bonds to the surface. The configurations d and h belong to the $t2$ group [17]. Thus, the initial $t4c$ configuration is followed by $t2d$ (point d), which is 0.9 eV higher than $t4c$, close to the value of 1.0 eV found in our earlier static calculations [17]. The next $t2$ configuration following $t4g$ (point h), although not identified in [17], has a similar but somewhat higher energy, 1.8 eV higher than $t4c$. Finally, at the last step (points d–e and h–i), a new pair of front Si–C bonds is formed, and the molecule moves into a stable adsorption configuration with four C-Si bonds with the surface dimers.

The energy barriers to C_{60} movement found in these calculations are $\sim 2\,\text{eV}$. Such a high barrier would not allow the diffusion of C_{60} even at room temperature. A displacement of C_{60} has to be driven externally, for example, by means of an STM (or AFM) tip. Note that the calculated barrier is most probably slightly overestimated due to the use of the constrained minimisation technique. This conclusion is also supported by the fact that when alternative carbon atoms were chosen to drive the C_{60} diffusion, the energy barriers were reduced by at least 0.4 eV [18]. Still, the three stages during each elementary translation, as summarised above, should not be affected by

the method used by us to drive diffusion. Interestingly, the highest-energy configurations (points *c*, *g* in Fig 27.10) are not those with two bonds between the molecule and the surface but the configurations where all four Si–C bonds are still retained.

The path between stable configurations with four Si–C bonds via intermediate configurations having two Si–C bonds can be confirmed by electron density difference plots obtained using Eq. (27.1). Such plots can be found in [18]. The electron density plots illustrate the Si–C bond formation between the molecule and the surface and changes in the electron density distribution within the C_{60} cage. The plots clearly show that the highest-energy points are not the structures with two Si–C bonds but the preceding stages just before the back bonds are broken. Structures with only two Si–C bonds have intermediate energies and are immediately followed by new stable structures with four Si–C bonds. An explanation for these energies can also be given based on electron density distribution. At the highest-energy points along the path, the electron density distribution in the lower part of the C_{60} molecule differs strongly from that in the isolated C_{60}. As the Si–C bonds are strongly stretched at these points, the system loses a lot of energy. Intermediate structures with two Si–C bonds (points *d*, *h*) have an electron density distribution rather similar to the isolated molecule. This can balance the effect of breaking two Si–C bonds with the surface and may explain the fact that the highest-energy structures in our calculations are not those with only two Si–C bonds but the immediately preceding ones.

The effect of deformation of the C_{60}-silicon system can be semi-quantitatively estimated by considering separately the deformation of the surface, the C_{60} molecule and the Si–C bonds. Examination of bond lengths and of the geometry of the surface and the C_{60} molecule confirms that the surface atoms are displaced from their stable positions and the molecule is very strongly deformed from its original spherical shape at the highest-energy points.

The energies of the silicon surface and the C_{60} molecule calculated separately in their geometries optimised for the C_{60}-surface combined system along the path are both found [18] to increase by $0.5 - 1.0\,\mathrm{eV}$, as the system moves towards the potential energy barrier, and then drop as the back Si–C bonds are broken. These energies are lowest when only two Si–C bonds remain. Thus, there is a significant contribution to the total energy due to the deformation of the surface and the C_{60} molecule.

The combined effect of the deformation of the silicon surface and the C_{60} molecule and strained Si–C bonds during its diffusion/manipulation strongly reduces the binding energy in the C_{60}-silicon system. This deformation can be relieved by the Si–C bond breaking and subsequent relaxation of the geometry of the C_{60} and the silicon surface, and C_{60} rebonding. This process is accompanied by an increase of the binding energy. The structures with only two Si–C bonds between the molecule and the surface are metastable, and a small displacement of the molecule can cause the pivoting of the molecule

over the remaining two Si–C bonds so that two new Si–C bond are formed, releasing $\sim 1\,\mathrm{eV}$ of energy and resulting in a new stable adsorption configuration.

27.4.3 Diffusion/manipulation pathways

Experiments on the STM manipulation of C_{60} provided a number of manipulation trajectories along the trough with several different periodicities in every case. It is not practical to perform extensive *ab initio* modelling of all possible pathways, since *ab initio* simulations of C_{60} diffusion are very time-consuming. However, it is possible to predict manipulation sequences analytically assuming that: (i) the molecule moves between stable adsorption configurations with four Si–C bonds via metastable transition states having two bonds, (ii) sequential breaking of Si–C bonds and formation of new bonds takes place, and (iii) the molecule pivots between two stable sites over two front Si–C bonds [10].

It can be shown, for example [10, 18], that the pathway described in the previous section is a part of a longer sequence with the $4a_0$ periodicity. An inspection of the bonding in the $t4b$ configuration (the last one obtained in the *ab initio* calculations of diffusion described in the previous section) and all other adsorption configurations above the trough [17] shows that this configuration can only be followed by the $t4g*$ configuration (the mirror image of the $t4g$ configuration preceding the $t4b$ in our simulations). Similarly, of all possible configurations, $t4g*$ can only be followed by $t4c$, i. e. the same configuration as the starting one. Thus, after having gone through four different adsorption configurations, the molecule arrives at the starting configuration. This sequence therefore has a four-fold periodicity. Note that during each translation from one stable configuration to another, the molecule advances by a_0, the surface lattice constant. Therefore, this diffusion sequence has a periodicity of $4a_0$, the same as observed in the experiment [10].

Alternative pathways starting from the configuration $t4c$ can also be conceived: $t4c$ can be followed by configurations $t4a$ or $t4e$, which are geometrically similar to $t4g$, but $0.4-0.6\,\mathrm{eV}$ lower in binding energy. Considering the $t4c \rightarrow t4e$ sequence in more detail, $t4e$ can only be followed by $t4f$, then $t4e*$ (the mirror image of $t4e$) and then the molecule returns to the starting geometry $t4c$ again. These two pathways are illustrated in Fig. 27.11.

Altogether, the analysis of all known adsorption configurations in the trough and the possibilities of transitions between them by means of the pivoting mechanism allows us to identify three sequences with the $4a_0$ periodicity:

$$t4c \rightarrow t4g \rightarrow t4b \rightarrow t4g* \rightarrow t4c \ldots \tag{27.2}$$

$$t4c \rightarrow t4g \rightarrow t4h \rightarrow t4g \rightarrow t4c \ldots \tag{27.3}$$

$$t4c \rightarrow t4e \rightarrow t4f \rightarrow t4e* \rightarrow t4c \ldots \tag{27.4}$$

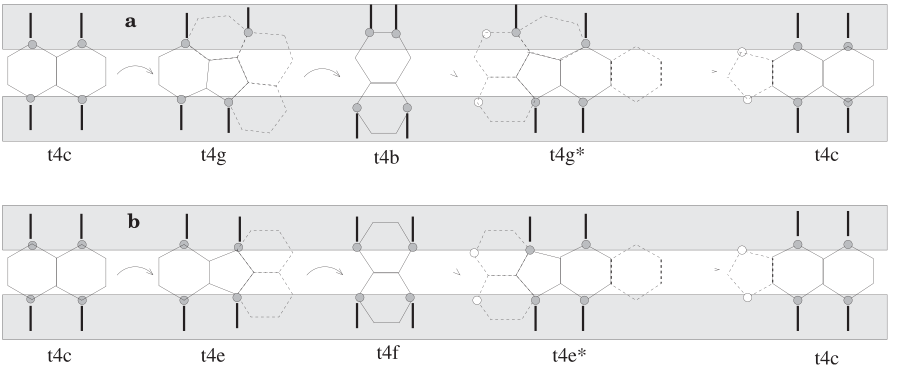


Fig. 27.11. Schematics of two possible mechanisms of C_{60} rolling along the trough on the Si(001) surface: **a** $t4c \rightarrow t4g \rightarrow t4b \rightarrow t4g* \rightarrow t4c$; **b** $t4c \rightarrow t4e \rightarrow t4f \rightarrow t4e* \rightarrow t4c$. Dimer rows are indicated by parallel grey stripes.

and one sequence with the $2a_0$ periodicity: $t4i \rightarrow t4i* \rightarrow t4i \rightarrow \ldots$. The latter sequence is lower in binding energy than the above-mentioned $4a_0$ periodic sequences. A number of sequences involving the configuration $t4a$ are also possible; these do not strictly follow the pivoting mechanism as the configuration following $t4a$ is found unstable [18].

As was discussed in Sect. 27.2, the manipulation trajectories along the trough with $2a_0$, $3a_0$ and $4a_0$ periodicities were observed in the STM experiments [10]. The pivoting mechanism discussed above can only account for the observed $2a_0$ and $4a_0$ periodic sequences, among them the low-energy path ($t4c \rightarrow t4g \rightarrow t4b \rightarrow t4g* \rightarrow t4c \ldots$). However, no $3a_0$ periodic sequence can be identified, based on our assumption of pivoting over two bonds as described above.

As was suggested in Sect. 27.2, there is a part in the $3a_0$ and $4a_0$ manipulation sequences that is common to both of them, and a switch from the $4a_0$ to the $3a_0$ sequence occurs at one of the pivoting stages when the diffusion mechanism is somewhat altered. For instance, the molecule may pivot over only a single bond. We propose that when the molecule is close to pivoting and has only two bonds to the surface, a certain amount of asymmetry in the position of the STM tip can give the molecule a rotational moment about an axis perpendicular to the surface. In this case, the trajectory of the molecule can deviate from the sequences suggested above. For example, a transition from the configuration $t4g$ to $t4g*$ (i. e. bypassing the configuration $t4b$) is possible, which will result in a $3a_0$ periodicity.

To summarise, a number of possible diffusion pathways for the C_{60} translation along the trough were suggested and compared with experimentally recorded STM manipulation tip trajectories. We propose sequences of adsorption configurations that account for the experimentally observed periodic tip trajectories.

27.5 Discussion and Conclusions

Calculations presented above describe adsorption of a C_{60} molecule on the Si(001) surface and its diffusion/manipulation on the same surface. The calculations show that the molecule moves between a series of stable adsorption configurations with four Si–C bonds to the surface via intermediate states with two Si–C bonds. Examination of bonding in the molecule and the surface reveals that the movement of the molecule proceeds via sequential breaking and formation of Si–C bonds to the surface according to a pivoting mechanism: immediately after the transition point, there are two bonds between the molecule and the surface, over which the molecule rolls. Energy barriers of $\sim 2\,\mathrm{eV}$ are involved in such a motion, caused by bond stretching and deformation of the molecule and the surface. By analysing the electron density distribution in the C_{60}-silicon system, we find significant changes in the electron density of the C_{60} molecule during manipulation (self-diffusion).

With the motivation to explain the room temperature manipulation experiments [10], the binding energies were calculated with respect to the symmetric p(2×1) reconstruction of the Si(001) surface. This is a legitimate approximation since it is known that the surface dimers perform a fast flip-flop motion above 200 K [11, 12]. In order to model low temperature manipulation, the c(4x2) surface reconstruction should be taken as the reference. This would give smaller binding energies, since the c(4x2) reconstruction is lower in energy than the symmetric p(2×1) one.

We analysed possible transitions between stable adsorption configurations in order to predict all possible C_{60} manipulation pathways on the surface. We find that a large number of manipulation sequences is possible, although not all of them may be easily achieved due to likely high energy barriers and/or low binding energies of the stable structures involved. The obtained pivoting sequences account for the observed $4a_0$ and $2a_0$ periodicities [10]. However, some additional arguments based on breaking of three, rather than two, C-Si bonds are to be incorporated to explain the observed $3a_0$ sequences: we propose that in those cases the molecule pivots over a single Si–C bond.

A more rigorous treatment of C_{60} manipulation would require an STM (or AFM) tip to be included in the simulations where the tip is moved with respect to the molecule, and each tip displacement step is followed by geometry optimisation. Modelling different manipulation techniques, e. g. pushing and pulling, can be also attempted using different tip models and different positions and orientations of the tip with respect to the molecule.

We expect that the tip positioned off-centre relative to the C_{60} molecule may cause its skewing and consequently a deviation from the pure pivoting pathway discussed in Sect. 27.4.2. In this way, manipulation trajectories with $3a_0$ or even $2a_0$ periodicities may be obtained.

A commonly used tip model [25, 26] is a (111)-oriented atomically sharp tip shown in Fig. 27.12a. Other tip models are also possible, and some of the tips that we are presently using in our manipulation simulations are shown

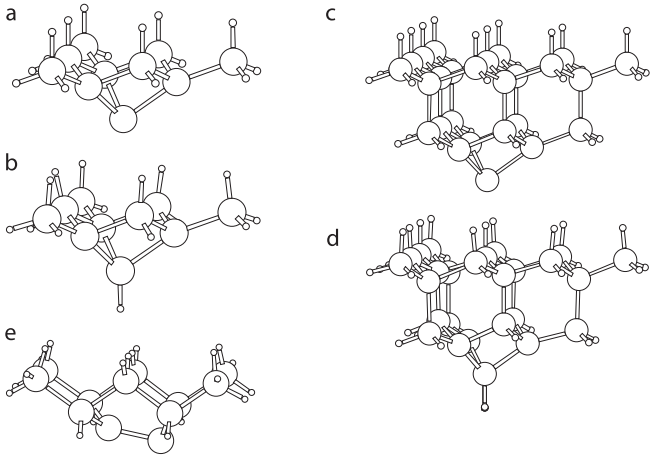


Fig. 27.12. Possible tip models: **a** small dangling bond terminated tip; **b** small hydrogen terminated tip; **c** large dangling bond terminated tip; **d** large hydrogen terminated tip; **e** double dangling bond terminated tip.

in Fig. 27.12b–e, including H-terminated tips b and d and the double-bond terminated tip e. A different tip apex termination will lead to a different chemical behaviour of the tip: a dangling bond-terminated tip should be reactive, whereas a hydrogen-terminated tip with its last remaining silicon dangling bond saturated by a hydrogen atom may be inert. Different tip sizes can also be considered, Fig. 27.12. Our preliminary studies of C_{60}-tip interaction did not show significant differences between small and large tips. The hydrogen-terminated and dangling bond-terminated tips, however, behaved differently. When placed above different sites in the molecule (e. g. above a C atom, above a bond or above a hexagon), the hydrogen-terminated tip did not form chemical bonds with the C_{60}. Binding energies varied between 0 and $-0.5\,\mathrm{eV}$, indicating a weak attractive interaction in some cases. The dangling bond-terminated tip was able to form strong chemical bonds with the C_{60} (with up to 0.9 eV energy gain), if placed with the apex Si atom facing a C atom of the cage. Thus, a dangling bond-terminated tip is a good candidate for a reactive tip that may change the trajectory of the molecule's movement and can be used for both attractive and repulsive manipulation simulations. A hydrogen-terminated tip may be used as an inert tip for pushing the molecule (repulsive regime). Further studies on this system are currently under way, in particular, the effect of the STM tip on the manipulation mechanism is being investigated. Note also that our results on the C_{60} manipulation are also of interest for a possible Atomic Force Microscopy (AFM) assisted manipulation as we believe that the effect of the tunnelling current in the STM experiments [10] on the manipulation pathway is negligible.

The pivoting mechanism of the molecule manipulation/self-diffusion discussed here can be quite general: it may be important for molecules that are

chemisorbed on surfaces, i. e. have strong covalent bonds with them. Note that the situation would be different in the case of physisorption with weak adsorbate-surface interaction: diffusion or manipulation of such adsorbates does not require bond breaking and can proceed by sliding instead of rolling.

Acknowledgement. We would like to acknowledge the computer time on the HPCx supercomputer via the Materials Chemistry Consortium. NM would like to acknowledge the financial support from the NANOMAN (NMP-CT-2003-505660) project, while CH from EPSRC (grant GR/R97023/01). We thank P.H. Beton (Nottingham University) for useful discussions.

References

1. D. Eigher and E. Schweizer, Nature **344**, 524 (1990).
2. L. Bartels, G. Meyer, and K. H. Rieder, Phys. Rev. Lett. **79**, 697 (1997).
3. J. Gimzewski and C. Joachim, Science **283**, 1683 (1999).
4. L. Grill et al., Phys. Rev. B **69**, 035416 (2004).
5. F. Moresco et al., Phys. Rev. Lett. **87**, 088302 (2001).
6. P.H. Beton, A. Dunn, and P. Moriarty, Appl. Phys. Lett. **67**, 1075 (1995).
7. N. Oyabu, Y. Sugimoto, M. Abe, O. Custance, and S. Morita, Nanotechnology **16**, S112 (2005).
8. N. Oyabu, O. Custance, I. Yi, Y. Sugawara, and A. Morita, Physical Review Letters **90**, 176102 (2003).
9. D. L. Keeling, M. J. Humphry, P. Moriarty, and P. H. Beton, Chem. Phys. Lett. **366**, 300 (2002).
10. D. L. Keeling et al., Phys. Rev. Lett. **94**, 146104 (2005).
11. T. Uda et al., Progress in Surface Science, **76**, 147 (2004).
12. J. Yoshinobu, Progress in Surface Science **77**, 37 (2004).
13. P. Moriarty, Y. R. Ma, M. D. Upward, and P. H. Beton, Surface Science **407**, 27 (1998).
14. P. D. Godwin, S. D. Kenny, and R. Smith, Surface Science **490**, 409 (2001).
15. P. D. Godwin, S. D. Kenny, and R. Smith, Surface Science **529**, 237 (2003).
16. C. Hobbs and L. Kantorovich, Nanotechnology **15**, S1 (2004).
17. C. Hobbs, L. Kantorovich, and J. Gale, Surface Science **591**, 45 (2005).
18. N. Martsinovich, C. Hobbs, and L. Kantorovich, Phys. Rev. B **7**, 085304 (2005).
19. J. M. Soler et al., J. Phys.: Condens. Matter **14**, 2745 (2002).
20. J. Perdew, K. Burke, and M. Ernzerhof, Physical Review Letters **77**, 3865 (1996).
21. S. Boys and F. Bernardi, Molecular Physics **19**, 553 (1970).
22. Y. J. Li et al., Phys. Rev. Lett. (2005), submitted.
23. M. Moalem, M. Balooch, A. Hamza, and R. Ruoff, J. Phys. Chem. **99**, 16736 (1995).
24. D. Chen and D. Sarid, Surf. Sci. **329**, 206 (1995).
25. R. Perez, I. Stich, M. C. Payne, and K. Terakura, Physical Review B **58**, 10835 (1998).
26. A. S. Foster, A. Y. Gal, Y. J. Lee, A. L. Shluger, and R. M. Nieminen, Applied Surface Science **210**, 146 (2003).

28 Nanoscale Friction of Self-assembled Monolayers

K. Mougin and H. Haidara

Institut de Chimie des Surfaces et Interfaces (ICSI)-CNRS 15, rue Jean Starcky-B.P. 2488, 68057 Mulhouse Cedex, France

The rapid development of the microelectromechanical industry (MEMs) has stimulated the need and design of molecular assemblies capable to impart new functional surface properties to materials or improve those existing, for the control of their interface behaviors (adsorption, condensation, wetting, lubrication, adhesion, liquid flow). Self-assembled monolayers (SAMs) of either organosilanes or thiols have played and still continue to play a major role in the achievement of this goal, especially in fluidics and as lubrication coatings. For MEMs devices, for instance, the control of adhesion and friction between the moving surfaces is crucial. Indeed the surface-to-volume ratios in these micro- and nanoscale devices are so large that adhesive and friction forces dominate gravitation and inertia. Precise control of the interfacial interactions is thus required to adjust adhesion and friction and prevent wear and stiction. Thus, the basis for molecular design and tailoring of SAMs must include a complete knowledge of interrelationships between the molecular structure and frictional properties of SAMs, as well as an understanding of friction and wear mechanisms of SAMs at a molecular level.

Until recently, tribological studies have not had access to the molecular level detail of interfacial contact. As a consequence, experimental verification of the molecular factors influencing frictional properties has been lacking. Fortunately, development of atomic force microscopy has provided the opportunity to study friction at a nanometer scale, so-called lateral force microscopy (LFM) or friction force microscopy (FFM). LFM provides some images of patterned SAMs but also provides some response for frictional, shear and adhesion properties of monolayers [1].

New interdisciplinary ideas and approaches can be effective and physical-chemical techniques based on self-assembling and self-organization principles are useful for fabrication of new nanomaterials and nanostructures [2]. Indeed, SAMs have great importance in interfacial engineering both for their fundamental and practical interests. For fundamental researches, they provide model systems for studying organized functional molecules and for applications [3], they often serve as the template for the bottom-up fabrication of nanostructures. Indeed, the nature (polarity, hydrophilicity, etc.) as well as the structure of an homogeneous SAM coating on a surface has a strong influence on the frictional properties (tribology, drag force to the liquid flow along a wall). Most of the time, SAMs were self-assembled onto smooth substrates

(e. g. mica, silicon wafer, etc.) to minimize contributions to the measurements from surface topography.

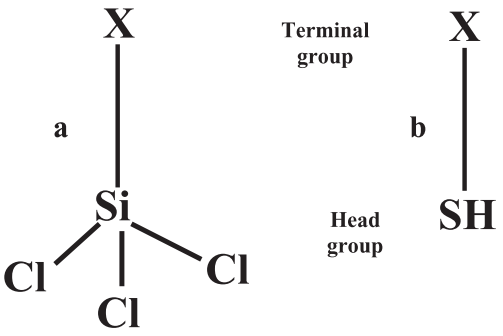
The physical state of thin molecular films can be influenced by a number of forces including intermolecular interactions between molecules making up the film, surface interactions in adsorbed films [4]. The structural characteristics of organic thin films (molecular conformation, packing arrangement and chemical composition) directly influence their lubrication properties.

This chapter is divided into two main parts. The first part will focus on the frictional properties of molecular homogeneous thin films; particularly the influence of chain length, terminal group and packing state of SAMs on the one hand, and that of the environmental and experimental conditions on the other hand, on energy dissipation during friction. Then, we will address the important issue of nanoscale heterogeneities and their influence on the frictional behavior and nanotribological performance of SAMs, using binary molecular films of both regular and random distributions of the heterogeneities. Because real surfaces are often heterogeneous in their chemical composition, these nanoscale heterogeneous SAMs provide good model systems to study and tune the frictional properties of tribological coatings.

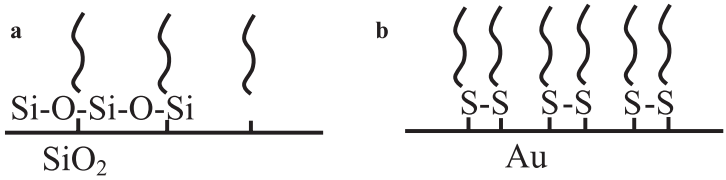
28.1 Homogeneous Organic Molecular Films

The formation of organized monolayer films on a surface by spontaneous adsorption and covalent grafting of molecules from solution or vapor phase is known as self-assembly. Two kinds of organic molecules have been extensively grafted: alkylsilanes onto silicon wafer or glass surfaces and organosulfur onto gold-coated substrates (Scheme 28.1).

On gold substrates, the anchorage of the organosulfur molecule is provided by a covalent bond between the sulfur end group and the gold atom (Scheme 28.2a). On the silicon surface, the attachment is less well defined as previously. From a common point of view, the grafting of the organosilane involves the hydrolysis of the chloro or alkyloxysilyl headgroups with the residual molecular film of water strongly adsorbed onto the silicon substrate.



Scheme 28.1. Schematic representation of an **a** alkylsilane, and **b** organosulfur molecule



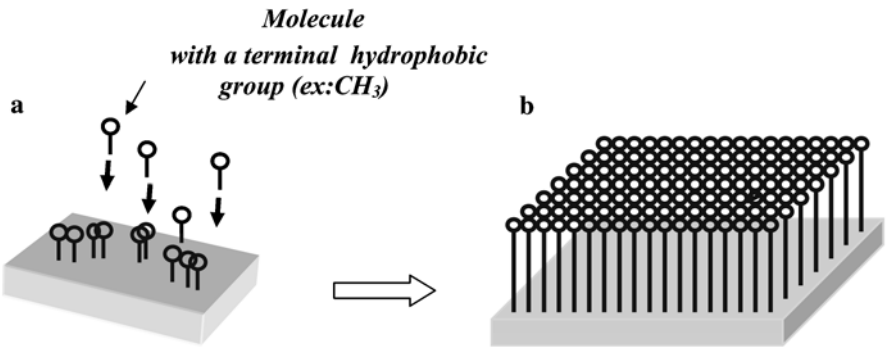
Scheme 28.2. Schematic representation of **a** dimer thiols; **b** siloxane network

The silanols groups ensued from this reaction can either form a covalent bond with the silanol group arising from the silicon surface [5] or create hydrogen interaction with trace water physisorbed to the substrate [6–8]. The stable linkage of the SAMs to the subtrate is provided by a crosslinking process of the silanol groups to form a siloxane network [9] on the extreme layer of the substrate (Scheme 28.2b).

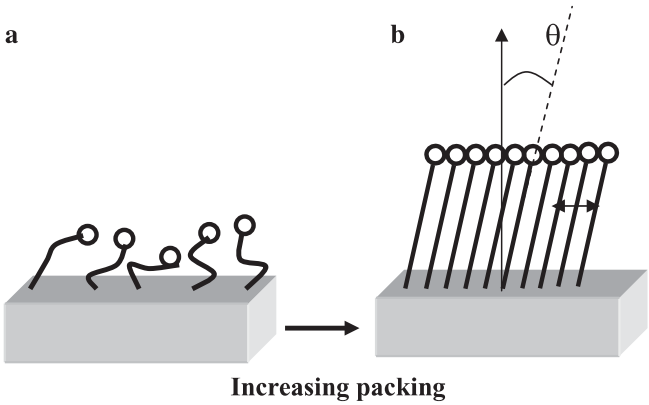
The performance of the self-assembled monolayers is directly dependent on the efficiency of the anchorage to the surface. Granick and coworkers [10, 11] have shown the importance of surface and density of grafting sites to form well-packed monolayers of SAMs.

Two procedures that have been successful for forming SAMs are available. On the one hand, the organic molecules are dissolved in a solvent, and on the other hand, they are dispersed in a paraffin oil and deposited under vaccum ($\sim 5 \cdot 10^{-3}$ Torr, 45 min to 1 h). For both proccesses organic molecules diffuse to the surface and self-asssemble into discrete nanodomains [12] that will grow to complete confluence to form a thin and continuous organic layer as depicted in Scheme 28.3.

The surface coverage and structure of the resulting SAMs depend on several experimental parameters such as the reaction time, the temperature, the hydration state of the substrate [8,13], the nature and polarity of the solvent, or the pressure (vacuum) for vapor-phase deposition. In particular, SAMs pi-



Scheme 28.3. Schematic representation of the formation of a self-assembled monolayer. **a** Nucleation and growth of molecular film in discrete nanodomains **b** compaction and formation of the complete organic thin film



Scheme 28.4. Schematic representation of **a** a liquid-like, and **b** a solid-like structure of a thin molecular film

oneers, Sagiv et al. [14] have shown the correlation that exists between the adequacy of the solvent and the packing density of thin molecular film. These experimental parameters as well as more intrinsic ones have a strong influence on the structure of the monolayer. Indeed, the structural state of the thin film is mainly defined by its uniformity and packing density. Scheme 28.4 presents a well-ordered SAM structure referred to as a solid-like structure and a less-ordered one referred to as a liquid-like structure. As shown by Gerber and coworkers [15], the organic molecules in a well-packed thin film remain in a specific conformation: tilted a few degrees from the surface normal and twisted around their molecular axis to minimize their global free energy at equilibrium.

The possibility of changing the chain length, terminal group and the packing order (disorder) within the molecular film makes self-assembled monolayers attractive model systems to study the nature of frictional interactions at the molecular level.

Along with the intrinsic features of the molecular films, the environment and operating conditions (both experimental and in-service) represent the second class of parameters that can drastically affect the friction, either directly or through the structural changes and response of the molecular film.

The sensitivity of friction to the operating conditions on SAMs arise from the high susceptibility of these molecular films to structural changes that strongly affect the intermolecular surface forces, and hence the threshold lateral force required to slide the contact between two surfaces usually called the fiction force F_f is given by the following equation [16],

$$F_\mathrm{f} \sim \mu(F_\mathrm{ext} + F_\mathrm{adh} + F_\mathrm{c}) \tag{28.1}$$

where F_f represents the friction force, F_ext corresponds to the applied external load, F_adh is the intermolecular adhesion force, F_c is the capillary condensa-

tion force when its exists (Eq. (2)) and, μ is the friction coefficient between surfaces. Obviously, the μ determined under the action of the sole intermolecular adhesive forces may differ, more or less, from that determined under the combined effect of $(F_c + F_{adh})$, or $(F_{ext} + F_{adh})$ above a certain value of F_{ext}. It also should be noted that even in the absence of any external load, the only adhesive forces can be critical as regards nanotribological performance of micromachines (MEMs, NEMs). This results from the space confinement and the high surface-area-to-volume ratio in these systems, to which the adhesion force that tends to stick the different elements is proportional. For vdW interactions between two planar bodies separated by a distance h, for instance, these attracting surface forces scale as $(H/6\pi h^3) \times$ contact area, where H is the Hamaker constant of the interface.

The following section gives a synthetic but critical discussion of the way these chemical, topological and operating parameters affect and control the frictional behaviors of self-assembled molecular films.

28.1.1 Influence of Chain Length and Structure

The films with well-ordered structure exhibited lower friction coefficient than similar films with disordered structure [17, 18]. The contact between tip and sample under a low load is nearly frictionless [19]. Friction is particularly high with short chains of less than eight carbons. Longer chains, stabilized by Van der Waals (vdW) attractions form more compact and rigid layers. This rigidity even results in a much better lubrication [18] effect on frictionless SAMs such as CF_3-, CH_3-terminated films. Indeed, friction depends on the intrinsic properties of the grafted organic molecules: on the one hand, the chemical nature of its end-group that is in direct contact with the counterface and exchange molecular interactions, and on the other hand the density and thickness of the monolayer [20]. However, this lubricating action is lost from a certain external load threshold when wear-induced disorder at the molecular level starts to occur. In particular, Porter et al. have shown the existence of a bimodal dependence of friction coefficient versus n, where n represents the number of methylene units for n-alkanethiolate monolayers. The tribological transition is around $n \sim 12$.

The frictional energy dissipation mechanisms in SAMs has been investigated by different authors and shown to arise from the excitation of the rotational and vibrational modes in the molecules. These excitation modes contribute strongly to the energy transfer to the substrate and thus to the frictional properties. Since short chains are shown to be less well packed than long-chain SAMs, they have structural disorder and defects that promote the excitation modes (rotational, vibrational, etc.) and energy absorption (dissipation), giving rise to a higher friction. The self-organization and close packing of the molecules ensured by Van der Waals interactions between chains plays an essential role in stabilizing thin molecular films and lowering friction. The longer the chains, the greater the cohesive interactions between chains.

As a result, the frictional behavior of well-packed long-chain SAMs might be explained by the strong intermolecular interactions that hold large blocks of molecules together during shear without an effective loss of molecular order, according to the Eyring model [21]. The dissipation mechanism in this case appears to involve the motion of molecular domains. Thus, these blocks that effectively tend to retain the molecular order, globally move with respect to the substrate, resisting the shear induced by the friction and minimizing the relative volume change.

It is then reasonable to suggest that the increase in friction for the short chains ($n < 8$) is due to poor packing of the molecules, which makes possible the excitation of numerous defects and energy-dissipating modes [17]. The excitation of these modes is similar to viscoelastic behavior. The fluid-like chains are more compliant and present a smaller resistance to shear.

As a result, the lack of cohesion in thin films formed with short chains, involves a different frictional process, but sliding corresponds more to a conformational change of the molecules in the film structure. This process is more energy dissipative due to an inelastic and plastic deformation of surface asperities [22] at a macroscopic scale, but at a microscopic level, energy is dissipated through surface phonons. The molecular dynamics study argues that energy dissipation in a long-chain monolayer is associated with vibrational energy through an oscillation around the tilt angle of the chains. The shear-induced disorder of the chains in the shorter monolayers therefore results in a nonpermanent contact because of energy dissipation through bond rotations and vibrations, leading to a higher microscopic friction. So far, these investigations are still confused as additional information is required to provide quantitative tribological results and correlate friction and energy dissipation in organic thin films.

Moreover, the influence of the preparation method and of the underlying substrate's nature on SAMs friction still remain unclear to date. According to Lio et al. [17], the lack of the long-range order in the silanes for chain lengths below $n < 10$, can be explained by crosslinking of the head groups (Scheme 28.1). In thiols, crosslinking is limited to dimer formation, while for silanes, there is an extensive siloxane network along the film/substrate interface [9]. Hence, there must be considerable chain distortion near the interface that is amplified along the network, which may be at the origin of the lack of long-range order in organosilane SAMs. This distortion effect is even more amplified for short chains films, leading to higher disorder and friction coefficient. Finally, Bhushan et al., have proposed to use a C_{18} double-grafted films with a higher area density and hence a stronger chain–chain interaction leading to better nanotribological behavior [23].

Other important questions that need to be answered regarding the frictional behavior of SAMs include: the effect of the lateral order within the film, the presence of heteroatoms (O, N, S, etc.) or branches inside the molecular chain and the existence of double or triple bonds that change the rigidity of the molecules.

28.1.2 Influence of Terminal Group

The influence of the terminal and head groups (Scheme 28.1) on friction and wear properties were also investigated by contact-mode atomic force microscopy [24].

Ahn et al., have shown that at the nanoscale, the frictional behavior of the homogeneous SAMs with different functionality is primarily influenced by the surface energy, which is directly related to the interfacial interactions between the tip and the topmost surface chemistry of the substrate. When using a polar probe as the contacting surface (a clean Si_3N_4 AFM tip, for instance), friction increases from hydrophobic to hydrophilic homogeneous thin films.

Amongst these nanoscale coating films, fluorocarbon-based SAMs have received a great deal of attention, especially for their singular surface energy (wetting) versus friction behavior [25]. In standard organic chemistry textbooks, the sections devoted to fluorocarbon chemistry are typically brief. Nevertheless, this molecule seems to be very interesting as the size of fluorine [26] allows it to replace hydrogen in many organic molecules. In contrast to hydrocarbon or siloxane coatings, fluorocarbon coatings repel both water and oil, and yet, frictional investigations on perfluorocarbon-based SAMs have shown that they exhibit a higher friction coefficient than simple hydrocarbon-based SAMs [27, 28]. This singularity can be explained by the real difference of their both lattices compacity. As mentioned previously, Timmons [29] and Zisman have shown that fluorocarbon-based SAMs are less well packed than hydrocarbon ones. As a result, the space void in a fluorocarbon lattice is larger than in a hydrocarbon one. Indeed, Stoebe et al. [30] have estimated, according to atomic structural data, that the CF_3 group occupied 25% more volume than a regular CH_3 end group. This simple remark shows the real discrepancy of packing density of both lattices. Larger terminal groups in films give rise to increased steric interactions that provide pathways for energy dissipation during sliding. And hence, lateral (and perhaps rotational) motion within the plane of the CF_3 groups is highly cooperative over relatively large distances, involving long-range interactions in the CF_3-terminated films [31]. These long-range interactions might permit the dissipation of energy through phonon modes [32] and thus, induce a higher frictional response for the CF_3-terminated films. These singular nanotribological responses of fluorinated-based SAMs are in total agreement with their wetting properties. Hysteresis of the contact angles of liquids over a fluorinated thin film is actually higher than on a regular hydrocarbon one, given the fact that liquid molecules can penetrate more easily in a CF_3 self-assembled monolayer, less well packed than a CH_3 one [33]. Indeed, wetting theory specifies that hysteresis of contact angle of a liquid onto a SAMs-coated substrate is proportional to the molecular volume of the thin film.

These observations, however, can be rationalized based on earlier models by Yoshizawa et al. [34] assuming (i) that molecular reorientations in the

SAMs are included in the definition of interdigitation and (ii) that SAMs of thiols on gold can be classified as "solid-like" or "amorphous-like", respectively, regarding their tribological properties.

Finally, Overney et al. [35] have observed that fluorocarbon films were more elastic and displayed a higher friction than hydrocarbon films. They have proposed a correlation between friction and elasticity of organic thin films. Garcia-Parajo et al., as well as Bhushan and Liu, have also observed the compression and relaxation of soft and rigid thin films in their loading and unloading tests [36, 37]. At a given normal load, long carbon chain structure such as an alkyl chain can be easily compressed with the tip compared to rigid benzene-ring-structure SAMs. The orientation of the "molecular springs or brush" under normal load reduces the shearing force at the interface, which in turn reduces the friction force. The possibility of orientation is determined by the spring constant of a single molecule (local stiffness), as well as the interaction between the neighboring molecules, which can be reflected by packing density or packing energy. It should be noted that the orientation can lead to conformational defects along the molecular chains, which lead to energy dissipation.

The previous paragraphs have described the impact of intrinsic structural parameters of the homogeneous molecular film onto its frictional response; and yet, other experimental parameters are also able to drastically modify this behavior. These are especially the environmental conditions and sliding velocity, which we will discuss in the following sections.

28.1.3 Effect of Humidity and Temperature

The environment is a crucial parameter of tribological experiments. The frictional results are directly dependent on the humidity and temperature of the surrounding medium. The frictional behaviors of molecular organic thin films relied on the environmental parameters and can completely be reversed if they change drastically. Moreover, humidity and temperature have a direct impact on the structural stability and performance of the SAM coatings. However, only a few studies exist concerning the influence of temperature and water on the monolayers after their formation.

These effects of humidity and temperature on the frictional properties of SAMs were investigated by different authors. The studies of Tian et al. [38] on bare and SAMs-coated mica show that the friction decreases on the mica with increasing relative humidity, while increasing on long alkylsilanes SAMs-coated mica, when operating at room temperature with low external loads. At low humidity (RH $< 5\%$), heating the samples in the temperature range of $20-80\,^{\circ}\mathrm{C}$ was found to induce a negligible effect for both bare mica and SAMs-coated mica. On the other hand, for the same range of temperature but with a high humidity rate (RH $> 50\%$), a significant variation in friction with temperature was observed. Finally, the influence of humidity in these frictional problems is intrinsically related to the capillary condensation of

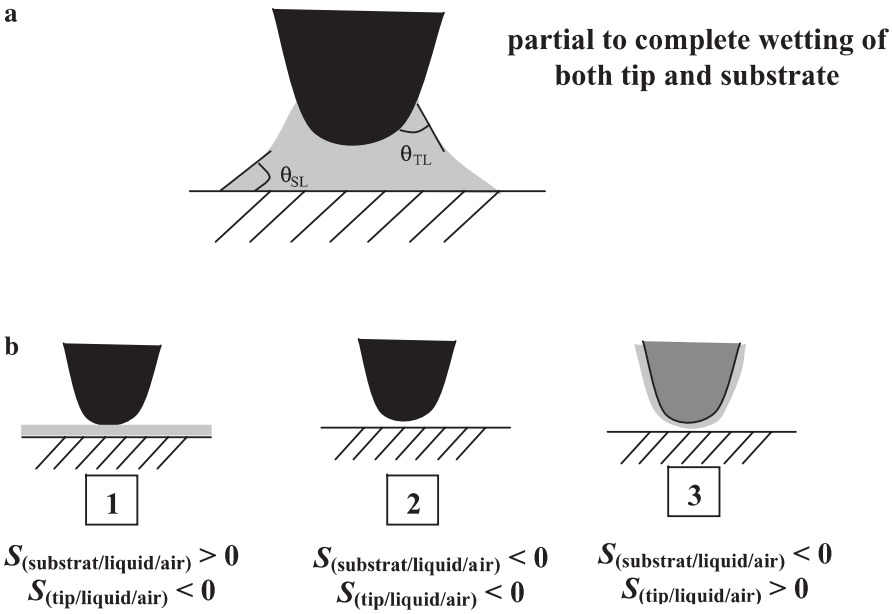
water that bridges the contacting surfaces. Indeed, the capillary bridging force is given to first order by [39]:

$$F_c \sim 2\pi R_T \gamma_L (\cos\theta_{SL} + \cos\theta_{TL}) \tag{28.2}$$

where R_T is the tip radius, θ_{SL} and θ_{TL} are, respectively the static contact angles of the liquid on substrate and tip, and γ_L is the liquid (water) surface tension (Scheme 28.5a).

The resulting water meniscus or layer can either enhance friction through increased adhesion in the contact zone (Eqs. (1) and (2)) or reduce it through the lubricating effect of water [40] (Scheme 28.5b).

The existence and local shape of the liquid condensate around the tip/substrate contact depend on the spreading coefficient S of the system [41], $S_{(solid/liquid/air)} = \gamma_S - \gamma_{SL} - \gamma_{LV}$, where "solid" stands for the tip or substrate, γ_S, γ_{SL} and γ_L representing, respectively, the surface (interface) energy of the bare solid, the solid–liquid and liquid condensate. This physical parameter represents the variation of the surface energy by unit area between



Scheme 28.5. **a** Schematic representation of a thick water bridge connecting the tip and the substrate (wetted hydrophilic SAM and tip, for instance), **b** panel of other possible wetting situations showing 1). a thin continuous water film forms only on the hydrophilic substrate, at low $RH\%$ for instance; 2) there is no water condensation film bridging the hydrophobic contacting surfaces; 3) symmetric case of 1), the continuous condensation and lubricating water film forms only on the hydrophilic tip

the dry and wet solid substrate. It is sensitive to both short- and long-range intermolecular interactions between the solid and the liquid in a given environment. A negative value of S indicates that the surface is not wetted by the liquid, and a positive one, the opposite. For the former case, several wetting situations displayed in Scheme 28.5b, can be investigated as a function of $S_{(\text{tip/liquid})}$ and $S_{(\text{substrate/liquid})}$.

As a result, in a humid environment, the magnitude of friction force is strongly dependent on the capillary force that is related to the wetting intrinsic properties of the interfacial system.

Cohen et al. [42] and Yeh et al. [43] reported some of the thermally induced structural changes of self-assembled monolayers. In these experiments, the monolayers were studied up to 200 °C, and no permanent changes in the structure were observed after the monolayers were returned to room temperature.

In addition, it was shown that even for hydrophobic SAMs, water could penetrate the films, altering their structural order and debonding the molecules from the substrate [37]. It should be noted that this process of "ageing" is strongly activated by the temperature and the mechanical action of the friction. As a result, the tribological properties are affected by these phenomena. While heating SAM coatings at low humidity rate results in no evident changes in their frictional behavior, increasing humidity was shown to generally induce significant structural changes for the same heating conditions. Most of these structural changes of SAMs under the coupled effect of temperature and moisture (T, RH) can be understood and accounted for by the thermal activation of the bonds scission activity of water at the SAM/substrate interface.

28.1.4 Influence of Sliding Velocity

The influence of the sliding velocity on friction, which accounts, at least partly, for the dynamical response of the boundary layer, can be exploited to gain insight into the structure of the investigated monolayers. Because this dynamical response is strongly sensitive to the structure of the topmost surface layer and experimental conditions, the velocity-dependent investigations of friction have often led to contradictory results and discrepancies, even for substrates that were a priori "identical". For instance, molecular films of the same nature, but different structural features (packing, mobility relative to the substrate, etc.) like SAMs and supported Langmuir–Blodgett films will not have the same frictional response to velocity under identical experimental conditions. The same observation may hold for SAMs of identical chemical nature but different packing. For homogeneous SAMs that have compliant long carbon spacer chains, the friction force increases at high (100 µm/s) velocity [13], while for SAMs that have rigid biphenyl chains, the friction force changes in the opposite way. The mechanisms responsible for the variation of the friction forces of SAMs with velocity are believed to be related to the

viscoelastic properties of SAMs. Generally [39, 40, 44], in wearless friction, the nanoscale frictional force (F_{f}) was found to depend logarithmically on the sliding velocity (v) according to

$$F_{\mathrm{f}} \sim F_{v=1} + \beta \ln(v) \tag{28.3}$$

However, the sign of the slope β (increasing or decreasing F_{f} with $\ln v$) is found to strongly depend on both the velocity range, the humidity and the relative hydrophilicity of the contacting surfaces [39, 45]. These different and apparently ambiguous variations of friction with the sliding velocity, humidity and substrate hydrophilicity are quite meaningful and well accounted for by the time dependence (contact time $\tau \sim v^{-1}$) of the magnitude of the capillary adhesive force F_c, which contributes to friction. On hydrophilic substrates and for a given $RH\%$, large contact times (low v) allow larger number density and size of liquid bridges at the tip/substrate contact, increasing the contribution of F_{c} to F_{f}. Since τ decreases in kinetic friction with increasing v, one should expect the slope β in Eq. (3) to be negative on a hydrophilic tip/substrate contact, and positive otherwise [39]: fully hydrophobic and hydrophobic/hydrophilic tip/substrate contacts. Unfortunately, experience shows that the velocity dependence of friction in these nanoscale contacts are still more complex than what one could predict based on the sole contribution and contact-time dependence of the capillary liquid bridge adhesive force. Indeed, the thermal energy produced within the tip/substrate contact can induce molecular excitations and structural transitions in the topmost contacting layers (in SAMs, for instance), the magnitude of which also increases with the sliding velocity. This velocity-dependent conformation and morphological changes can thus affect, through the alteration of the energy dissipation and intermolecular interactions, the overall frictional force in the nanocontact, regardless of the existence of capillary condensation. An illustration of this complex velocity-dependent behavior of nanoscale friction will be provided and discussed in the section dedicated to "heterogeneous molecular SAMs".

28.1.5 Conclusion

The study of the nanoscale friction of monolayer films is an important issue in the world of nanotechnology as it controls the stability under operation, the performance and reliability of the electromechanical microsystems. Control over the intrinsic properties of the coating (chemical and structural ones) and of the environmental operating conditions such as humidity and temperature are the essential leading parameters for the choice and design of SAM layers useful in nanotribilogical applications. However, a unified and comprehensive picture of the energy-dissipation process during friction is still missing, especially at the submicrometer scale. One obvious reason is that one still has a limited access to what really takes place within the contact during sliding,

making quantitative analysis and interpretation rather difficult. The second reason is naturally related to the nonfully understood size effects that show up in nanoscale friction and strongly affect the results.

28.2 Molecular Heterogeneous Thin Films

Nanoscale surface patterns are considered as potential templates and building blocks for nanotechnology [46]. As for nanomaterials in general, these nanoscale surface structures have been of increasing research interest in recent years, due to their unique properties. They are expected to exhibit novel and significantly improved physical, chemical, mechanical and other properties, as well as to offer opportunities for manifestation of new phenomena and processes [47] that, owing to the nanoscale dimensions, are not observed at the macroscopic level. Precise control of nanopatterns is essential to assemble complex two- or three-dimensional structures [48]. As a result, the need for high-precision processes and nonconventional methods of surface patterning has steadily increased. Different techniques such as electrochemical stripping [49], hot embossing lithography [50], nanoimprint lithography [51, 52], edge transfer lithography [53], laser patterning monolayers process [54], dippen nanolithography [55, 56] as well as photolithography techniques [57] are commonly used to fabricate accurate and geometrically controlled micro- and nanostructures.

However, although these techniques make it possible to produce a variety of patterns, these are often limited to well-defined regular geometries (square, circle, strips), with typical sizes still more than 100 nm, except for a few time-consuming and rather expensive techniques. And yet, random patterns of nanoscale heterogeneous surfaces either composed of the discrete distribution of nanoscale domains of one molecular compound in the continuous phase of the second [58] or characterized by the formation of a bicontinuous structure of the two molecular phases [59, 60] represent another route to nanostructure surfaces.

The fundamental issue of this variety of nanopatterning is the way the chemistry, topology and surface fraction can affect or be used to adjust nanoscale frictional properties of these thin molecular films. These nanoheterogeneous surfaces also provide model systems for the understanding on the fundamental level recurrent interface phenomena and processes (wetting, dewetting, nucleation, etc.).

This section will focus on the influence of the distribution (geometrical controlled or its absence) of nanopatterns on the frictional behavior of the substrates. As previously, the frictional responses of these heterogeneous molecular thin films will be studied both versus the intrinsic properties of the thin coating and operating parameters.

28.2.1 Influence of Topology

The tribological behavior of micro- and nanostructured organic molecular thin films was investigated and interpreted as a function of the topology and organization of the surface patterns: from geometrical and well-organized structures to more random ones. So far, little information is available in the literature on this influence of the topological organization. This domain will require further model investigations.

Geometrical Structured Patterns

Nanotribological properties of engineered substrates and especially, geometrical patterned surfaces, were studied versus size and shape of the structures. Micro- and nanopatterned surfaces were prepared (Fig. 28.1) by microcontact printing [61–63], micromolding in capillaries and nanoimprinting pioneered by Whitesides and Chou [64, 65].

Microcontact printing (μCP) involves the use of an elastomeric stamp with micrometer or submicrometer-sized relief features on its surface to print molecular inks onto a substrate that has a higher affinity towards the ink molecules than the stamp. The most popular elastomer used is poly(dimethylsiloxane) (PDMS) and the stamps are prepared typically by replica molding against suitable rigid masters such as silicon surfaces microstructured by etching techniques, photoresist patterns, embossed polymer surfaces, etc.

The molecules used are mainly alkanethiols and alkylsilanes with different terminal functional groups (such as NH_2 and CH_3) and will be printed, respectively, onto coinage metal surfaces (Au, Ag, etc.) and oxidized surfaces (glass, quartz, silicon wafers, etc.). Self-assembled monolayers are thus patterned with line/spacing and dot array patterns with periodicity down to

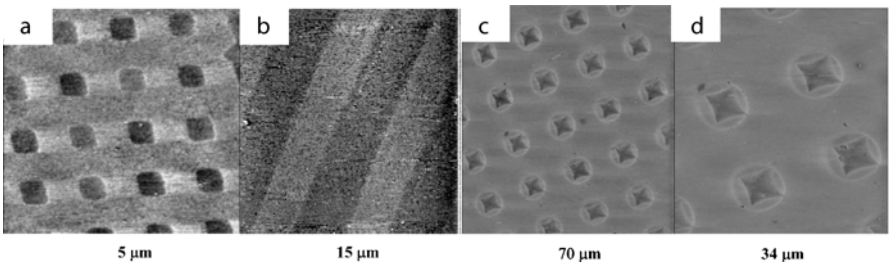


Fig. 28.1. AFM pictures (contact-mode friction) of nano- and micropatterned molecular surfaces. **a** hydrophilic (NH_2-terminated SAM) square nanodomains (*in dark*) surrounded by hydrophobic thin molecular film (CH_3-terminated SAM), **b** alternated microstrips of hydrophilic NH_2 (*in dark*) and hydrophobic CH_3 molecular films, **c** hydrophilic (NH_2) micropatterns (in dark) surrounded by hydrophobic (CH_3) continuum **d** zoom in on **c** patterns

hundreds of nanometers. By using master molds prepared from photolithography and exploiting the same molds as phase-shift masks, we expect to be able to pattern lines and dots as narrow as 200 nm [62]. If necessary, more complicated patterns, such as for instance asymmetric patterns to study ratchet effects, may be produced by designing new photolithographic masks.

Nanoimprinting is based on the use of a rigid mold (usually made out of silicon, metal or glass) bearing micrometer or nanosized relief features on its surface. This method is perfectly appropriate for the fabrication of nanopatterning molds. However, the techniques required in this process are more time consuming and heavier than for the previous one.

The model surfaces presented in Fig. 28.1 are heterogeneous molecular films of alkylsilanes compounds self-assembled onto silicon substrates realized by microcontact printing and nanoimprinting. Two organosilanes molecules, hexadecyltriclorosilanes (referred to as CH_3) and (6-aminohexyl)-aminopropyltrimethoxysilanes (referred to as NH_2) were used to create these patterns. The binary geometrical patterned surfaces were prepared according to a two-step printing process. First, different NH_2 SAMs topologies: stripes, squares and stars structures were printed onto the bare and cleaned silicon wafer with different NH_2 surface fraction: from 75 to 25%; the final and complete heterogeneous surface was obtained by self-assembling a second continuous monolayer of a second molecule (methyl terminated one: CH_3) around the domains in the remaining space. These substrates were then characterized by atomic force microscopy in contact mode (LFM). The frictional experiments consisted in scanning laterally the AFM tip under a constant velocity at (i) a constant load to get the friction contrast images of the patterned surfaces and (ii) at different loads to determine their friction coefficient. The tips used are silicon nitride (Si_3N_4) with a radius of curvature ~ 20 nm.

The average size and surface coverage were determined by analysis of AFM pictures (Fig. 28.1).

These images show the high sensitivity of lateral force microscopy (LFM) to the chemical contrast of the functional groups grafted onto the surface.

The friction experiment consisted in scanning laterally the AFM tip under a constant velocity at (i) a constant load to get the friction pictures of the patterned surfaces and (ii) at different loads to determine their friction coefficient. The tips used are silicon nitride (Si_3N_4) with a radius of curvature of ~ 20 nm. The frictional force between the tip and surface can be estimated from the voltage signal in the scope loop of LFM and friction coefficient can be calculated by multiplying the average signal of trace minus retrace (TMR) during a scan by a calibration factor.

The results of the external load dependent friction response onto square patterns are plotted in Fig. 28.2. For instance, a 50% NH_2 surface fraction is composed of 2 µm edge squares separated from the next pattern by 1 µm distance. The varying parameter in the surface fraction is the periodicity of the length patterning. As expected from Eq. (1), the measured friction force

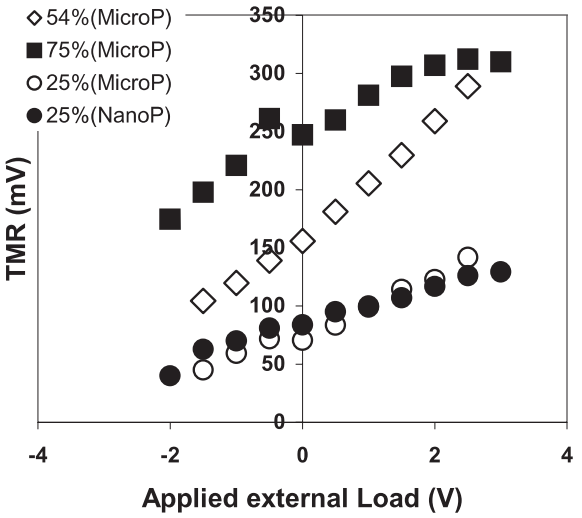


Fig. 28.2. TMR as a function of applied external load for different NH_2 surface coverage on squares micropatterned (MicroP) and nanopatterned (NanoP) binary NH_2/CH_3 surfaces

is linearly proportional to the external applied load. This response is characteristic of a Coulombic behavior. Moreover, the averaged frictional response is found to be more sensitive to the chemistry and relative surface fraction of the molecular domains than to the size and distribution of patterns (micro- or nanoscopic). This is shown by the regular increase of the tip/substrate friction on these (NH_2–CH_3)-patterned surfaces, with the surface fraction of the hydrophilic molecular species (NH_2).The slope of the curves is proportional to the friction coefficient and is represented in Fig. 28.3.

The results show that the friction coefficient decreases linearly as the surface coverage of NH_2 for micropatterned surfaces. The friction forces increase with the magnitude of the adhesive interaction of the SAMs and the tip. As a result, the greater the surface fraction of NH_2 in the thin film, the

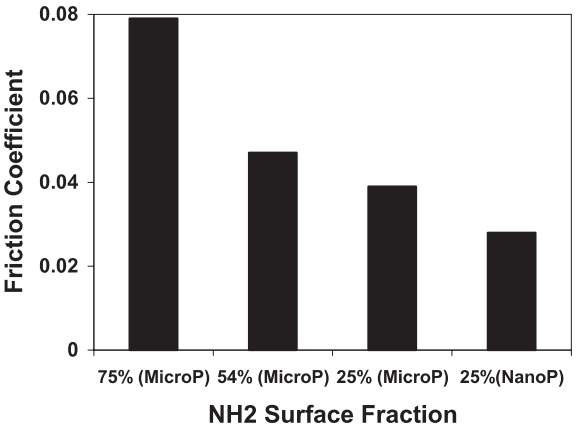


Fig. 28.3. Friction coefficient versus NH_2 surface fraction for micropatterned (MicroP) and nanopatterned (NanoP) binary NH_2/CH_3 molecular surfaces at normal ambient conditions **a** Nucleation and growth of molecular film in discrete nanodomains **b** elaboration of an organic thin film

greater the friction response, in good agreement with the literature [66]. Lopez et al. have also prepared striped micropatterned of methyl and hydroxyl-terminated SAMs and studied correlation between friction and grafted mole fraction [70]. The friction seems to increase steadily with the mole fraction of one molecule. The dependence of friction on the chemical composition of this geometrical patterned coating seems to be correlated by a Cassie-type relation for heterogeneous wetting [67,68]: the frictional response of the patterned surface composed of two uniformly distributed components is proportional to the surface fraction of each component, $\mu_{\text{binary}} \sim (\phi_{\text{NH}_2})\mu_{\text{NH}_2} + (1 - \phi_{\text{NH}_2})\mu_{\text{CH}_3}$. However, at comparable surface fraction, the friction coefficient is lower onto nanopatterned surfaces than on micropatterned ones for the same geometrical design. This result may traduce the coupling effects between nanodomains that lead to a collective frictional response on nanoscale patterned surfaces. Indeed, contrary to microsize patterns where the nanosize tip essentially responds to the discrete feature of the microdomains, the tip/substrate interaction on nanoheterogeneous substrates integrates the force fields emanating from neighboring nanodomains, the extent of which from the tip apex scales with that of long-range forces. This nanosize effect that couples the nearest nanodomains through the lateral overlap of their force fields can involve new properties that are not yet completely understood. Further investigations are required to expand the understanding of these two-dimensional scale effects on interface phenomena.

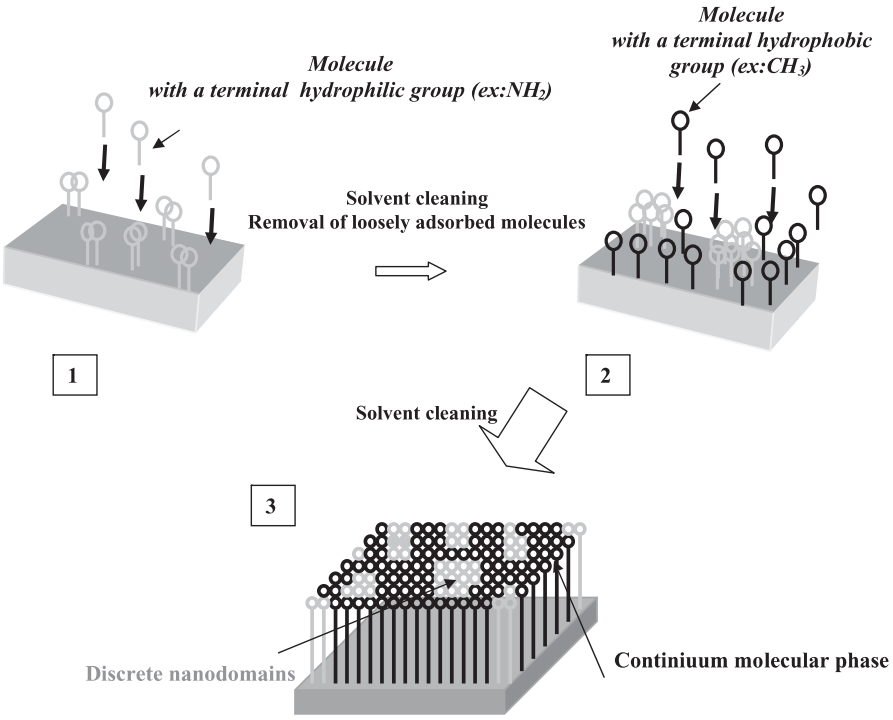
Random Nanopatterns

Random chemical nanoscale patterns and features [58, 69] spanning length scales from a few nanometers to micrometers were fabricated using two different self-assembling methods of alkylsilanes on (Si/SiO_2): a *sequential* and a *coadsorption one,* and nanoscale friction experiments were performed on these nanoheterogeneous molecular films.

1. For the random heterogeneous nanopatterns prepared by *sequential assembling*, the binary surfaces may be composed of a discrete distribution of hydrophilic nanodomains within a continuum of hydrophobic molecules, or vice versa, as illustrated in Scheme 28.6, for CH_3- and NH_2-terminated alkylsilane molecules.

AFM characterizations were performed on these binary heterogeneous molecular films and the results are displayed in Fig. 28.4. The domain size can be easily tuned from tens of nanometers to micrometers and the roughness of the thin coating is of the order of one nanometer as it corresponds to the length difference of both molecules.

As previously, friction experiments were performed on these substrates, and the results are presented in Fig. 28.5. The external applied load (F_{ext}) dependence of friction exhibits the expected Coulombic behavior of linear increase of friction vs. F_{ext}. This result (not plotted here), is similar to the previous one observed on the regular micropatterned molecular thin coatings.



Scheme 28.6. Two-step sequential elaboration of binary molecular thin films. step 1) *solvent phase* adsorption-nucleation and growth of discrete molecular nanodomains (nm to μm), 2) self-assembling of the continuum phase by *vapor- or solvent-phase* adsorption 3) resulting nanoheterogeneous surface

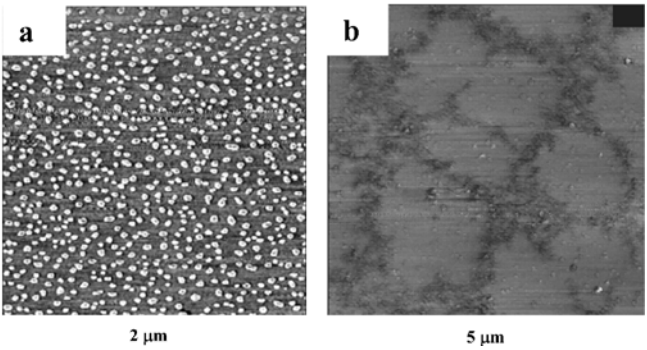


Fig. 28.4. AFM pictures (tapping mode) of random nanoheterogeneous molecular surfaces prepared by sequential adsorption. **a** hydrophobic CH_3 nanodomains (*clear dots*), dispersed in a hydrophilic NH_2 continuum grafted by vapor-phase adsorption, **b** hydrophilic NH_2 microdomains (*in clear*) dispersed in a network of hydrophobic CH_3 continuum, both molecular domains realized by solvent phase coating

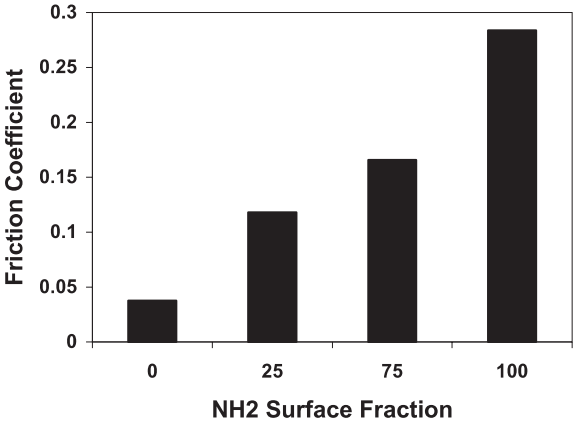


Fig. 28.5. Friction coefficient versus NH_2 surface fraction for nanoscale heterogeneous binary (NH_2/CH_3) molecular surfaces prepared by sequential process

Thus, friction coefficients were determined from the slope of these linear responses and are presented in Fig. 28.5. As one can expect from the intermolecular interactions involved at the (hydrophilic tip/substrate) contacts, the magnitude of this friction gradually and significantly increases with the surface fraction of NH_2, when going from the uniform CH_3 to the uniform NH_2 molecular films. On the homogeneous hydrophobic CH_3-SAM, the interactions with the hydrophilic silicon nitride tip mainly involve the London dispersion forces that have a much lower magnitude as compared to the polar, hydrogen and electrostatic bonds involved in adhesion (and hence friction) of the tip with the homogeneous NH_2-terminated substrates

2. For the heterogeneous surfaces prepared with a coadsorption process, the surfaces are composed of a dissemination of hydrophilic nanodomains in a hydrophobic continuum, and vice versa (Fig. 28.6). The binary heterogeneous surfaces were obtained in this case through a single-step process (sketch of Scheme 28.7). This grafting technique is based on the coadsorption of two molecules (organosilanes or thiols) in the proper solvent, or by vapor-phase deposition.

The AFM images characteristic of such nanoscale heterogeneous SAM film prepared by coadsorption is shown in Fig. 28.6. The domains are smaller and their contours irregular and less well-defined as compared to those obtained by the sequential process.

The nanoscale frictional response of these mixed binary SAMs thus appears to be directly proportional to the surface fraction of their constitutive molecular compounds. This should provide a semiquantitative tool based on the friction-coefficient measurement for the compositional surface analysis [70] of submicrometer-scale heterogeneous substrates

As shown in Fig. 28.8, this assumption is true for mixed SAMs of molecules as diverse as alkyl-terminated CH_3, amine NH_2, hydroxide OH, or acidic group COOH groups, but no longer holds for fluorocarbon-based (or rich) molecules. Indeed, as compared to thin films of hydrocarbon molecules, the

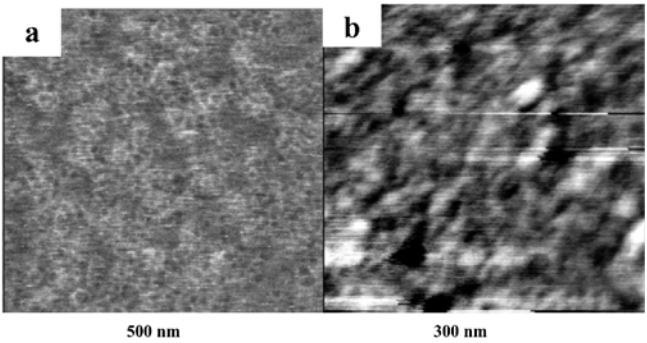
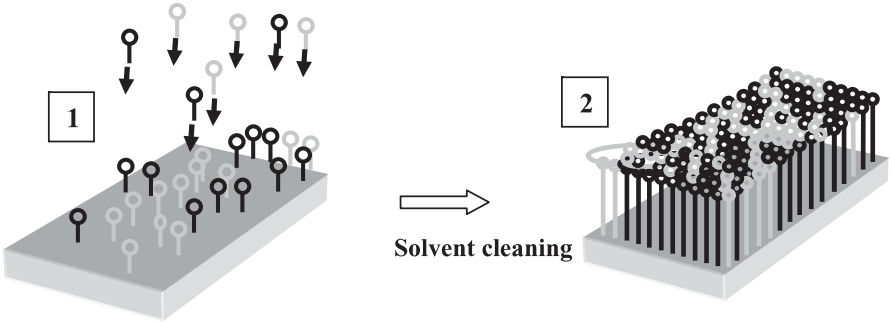


Fig. 28.6. Contact-mode AFM pictures of the random nanoheterogeneous molecular surfaces prepared by coadsorption: **a** hydrophobic CH_3 nanodomains (in clear) dispersed in a hydrophilic NH_2 continuum (both grafted in a single step vapor-phase adsorption from a mixture of the two molecules), **b** nanostructured molecular domains in the binary hydrophilic NH_2 (in clear) and hydrophobic CH_3 heterogeneous SAM (both grafted in a single-step solvent-phase adsorption from a mixture of the two molecules)



Scheme 28.7. Sketch of the coadsorption process allowing preparation of mixed SAMs of different chain lengths or terminal groups (hydrophobic CH_3/hydrophilic NH_2): 1) coadsorption and self-assembling from a solvent or by vapor-phase deposition, and 2) resulting SAMs surface

friction is higher on fluorocarbon molecules [71]. The introduction of the fluorocarbon molecule in a mixed SAM leads to an increase in the frictional response, even for a low nominal bulk concentration or surface fraction. For these fluorocarbon-based binary SAMs the overall frictional response does not show any trivial correlation with the chemical composition of the surface (Figs. 28.7b and 28.8). In this case, the high friction that is observed on the mixed SAMs can reasonably be accounted for by additional energy dissipation modes, which arise due to differences in the "phase state" of the SAMs. For mixed hydrocarbon–fluorocarbon systems, for instance, it has been shown that the fluorocarbon chains impose a nearly perpendicular orientation of the

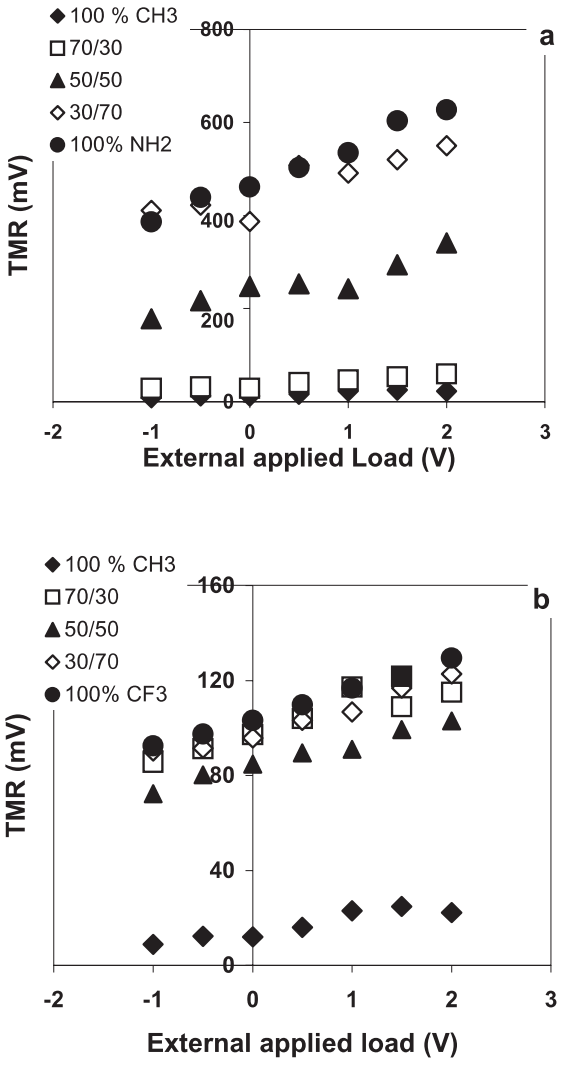


Fig. 28.7. TMR as a function of applied external load for mixed SAMs, for different surface coverage of: **a** NH_2-terminated molecular domains, and **b** CF_3-terminated domains, in the binary NH_2/CH_3- and CF_3/CH_3-terminated SAM, respectively

alkyl chains with respect to the surface normal. In this case, the molecular contrast has created a new structural organization of the molecules within the mixed SAMs inducing (creating) some new tribological properties.

Based on the nanoscale frictional responses on the different patterns, it seems that the chemistry and the surface fraction of the constitutive molecular species are the leading parameters that control the frictional behavior of these heterogeneous coatings. The size, length scales and geometry of the patterns can be used to modulate the fluctuating amplitude of friction from the discrete profile characteristic of each chemical domain on micrometer-size

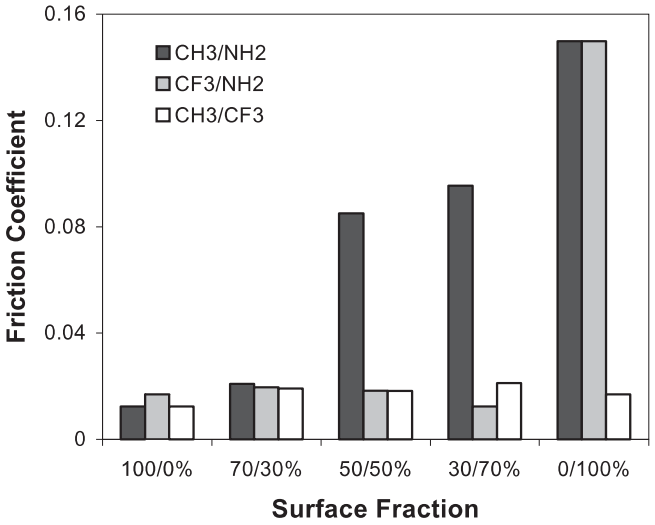


Fig. 28.8. Friction coefficient versus *i* NH_2, *ii* CF_3 and *iii* CH_3 surface fraction, respectively, on heterogeneous NH_2/CH_3, CH_3/NH_2 and CF_3/CH_3 SAM surfaces prepared by coadsorption

patterns, to a smoother one characterized by the coupling effects on nanoscale heterogeneous patterns. Unfortunately, there is still no systematic investigation of these scale effects on chemically heterogeneous SAMs and their impact on friction, especially its environment-dependent behavior (dissipation by domain-edge defects, condensation, etc.). Understanding these scale effects is not only of fundamental importance. This also provides us with a key control parameter for adjusting over different lenghscales (macro, micrometer to nano) the surface chemistry, patterns and properties for a given application. For heterogeneous SAMs that are being increasingly used in almost all the fields of nanotechnology (MEMs, NEMs, molecular electronics, controlled assembling of nanoparticles arrays, etc.), this is a challenge that will focus increasing research interests in the near future.

28.2.2 Influence of Sliding Velocity

Investigating the influence of the tip velocity on the nanoscale friction response of grafted coatings represents the most accessible way to efficiently probe their dynamical behavior and to understand some of the dissipation mechanisms (frequency-dependent ones) involved in the friction process. For random heterogeneous surface patterns [66], either prepared by the sequential or coadsorption methods, two distinct regimes of frictional response to sliding velocity are observed, depending on the nature of the intermolecular forces and structural sate of the SAM coating.

In these measurements, the sliding velocities were varied between 0.2 µm/s and 240 µm/s to cover about 3 orders of magnitude, allowing specific interface phenomena at both low and high molecular stress to be depicted. It should be noted that these measurements were carried out in normal ambient (20 °C, 30%RH) and in wearless conditions for the coating (as verified experimentally).

For the more hydrophobic substrates that mainly interact with the tip through London dispersion forces, a regular and low magnitude increase of the friction with the sliding velocity (v) toward a plateau [72] is observed. On the other hand, the frictional response on partially to highly hydrophilic substrates clearly displays two distinct regimes. In the first regime (low-velocity region, typically $< 20\,\mu$m/s), a steep increase of the friction is observed up to a maximum ($20\,\mu\text{m/s} < v < 50\,\mu$m/s, followed by a smoother decrease toward a stabilization plateau. The magnitude of the transition between these two regimes directly increases with the hydrophilic character (NH_2, SiO_2H), or the surface fraction of the polar molecules for binary heterogeneous surfaces, as shown in Fig. 28.9. This velocity-dependent frictional transition is thus observed for all hydrophilic substrates, independent of their chemical nature (organic, mineral) or composition (homogeneous, heterogeneous). This transition thus seems to essentially rely on the common and unique feature that is the existence of permanent dipoles and hydrogen bonds at the tip/substrate contact. Indeed, in contrast to the London dispersion forces that predominate the hydrophobic CH_3/tip contact, the directional and orientational polar and hydrogen-bond interactions that are involved at the hydrophilic/tip contact are strongly sensitive to thermal effects. Since these thermal effects increase with the sliding velocity, one equally expects the disorientation over the polar and hydrogen bonds in the molecular film to increase, leading from a certain critical velocity to a decreasing adhesive and friction forces. In addition, it was mentioned that a wetting film could form in a hydrophilic contact, even at a low relative humidity (Fig. 28.6). The destruction of such boundary liquid bridge due to thermal effects (dissipation) can also explain the friction inversion observed at high v.

A similar velocity dependence of friction was also reported for other hydrophilic substrates [34, 39, 40]. The slope inversion (friction transition) was explained and attributed to various energy-dissipation processes or capillary effects (already discussed above in Sect. 1.3). Indeed, during sliding, the frictional energy can be partially and irreversibly dissipated into thermal energy, the so-called thermodynamic energy dissipation $E_{\text{dissip}} \sim F_{\text{f}} \cdot v \cdot \mathrm{d}t$, creating an orientational disorder in the end-group dipoles of the molecular film. As is known from the literature, such thermally induced collective disorder involving the entire topmost molecular population (not limited to a single polar group) can drastically reduce the interaction density and energy within the contact zone (dipolar and H-binding ones), especially on SAMs.

Other theoretical interpretations such as conformation and structural changes in the molecular films have been proposed to explain this velocity-

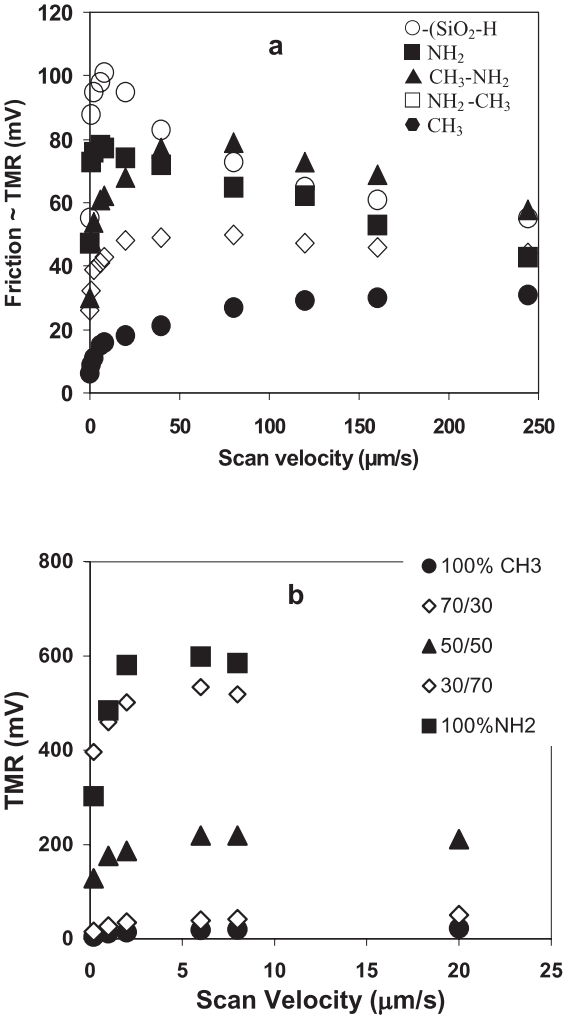


Fig. 28.9. Friction as a function of sliding velocity for random binary nanoheterogeneous NH_2/CH_3 molecular surfaces: **a** prepared by sequential adsorption **b** prepared by coadsorption

dependent frictional transition: the shear-induced transition from solid-like to liquid-like state, the viscous damping giving rise to molecular interdigitations or entanglements of the molecules [73, 74], and the water-condensation effects [39]. However, most of these effects during sliding lead to irreversible interfacial phenomena as dissipation of local heat in organic coatings induce complex film modifications up to severe wear. Nonetheless, this discussion still requires further investigations to acquire a unified picture of this velocity-dependent friction behavior, independently of the nature, chemistry, structure and topology of the contacting substrates (organic, mineral, composites, etc.).

28.3 Wear of SAMs

The frictional properties at the nanometer and micrometer scales have drawn a lot of attention as they provide critical information regarding the mechanisms and origin of wear on these organic molecular films.

During the sliding of a tip onto a SAMs-coated substrate, there are two main factors that create wear: (i) the irreversible disorder induced by the amplification of defects [34] in the thin molecular film and (ii) the rupture of SAMs/substrate bonds in a second step.

The interfacial wear mechanism is mainly governed by a kinetic energy transfer from the tip to the molecular film. This additional energy is dissipated through the film coating by creating modification in the molecules conformations and disorder, leading in extreme cases, to plastic deformations up to breaking of covalent bonds involving real damage in the coatings. In a SAMs coating, the higher the density of the defects, the faster the wear will occur. Indeed, below a certain critical load, in normal operating conditions ($\sim 20\,^{\circ}\mathrm{C}$, $\sim 30\%$ RH) SAMs can undergo orientational modification without damage. From this critical load (above the film cohesion energy), they wear and eventually detach from the substrate, leading to irreversible damage of the molecular film. It is worth noting here that since the intrinsic stability of SAMs (cohesion, structure, molecular mobility) is strongly dependent on the environment conditions (humidity, T, pH), their wear can be drastically accelerated by the coupling of stress to environment conditions. In order to improve wear resistance, Liu et al. have suggested the grafting of rigid spacer chains [24].

28.4 Conclusion

Self-assembled monolayers composed of organosulfur and organosilanes have found widespread applications in surface and interfacial science due to their well-defined and robust structures and the ability to simply modify their chemical composition in order to introduce a variety of modifications to their properties. However, nano- and microtribological investigations on these stable thin organic coatings still remains delicate.

First, the tribological performance of the homogeneous organic grafted thins films mostly depend on their intrinsic properties such as the chain length of the molecules, the area density, the chemistry of the terminal group and their method of preparation. Indeed, the frictional behavior of the SAMs with different functionalities is primarily influenced by their surface energy, which is directly related to the interfacial interactions between the tip and the chemistry of the surface. However, the effects of humidity and temperature have been studied and displayed the impact of these environmental conditions on friction on bare and coated surfaces. In particular, the significant influence

of relative humidity on frictional forces may be explained by the thickness of the adsorbed water layer.

In a second part, random and well-defined nanopatterned heterogeneous molecular thin films have been investigated in terms of nanofriction. Depending on their topological patterns, these surface heterogeneities generate a wide variety of frictional responses. Nanostructuration seems to have a fundamental and crucial effect on the tribological properties of the molecular coating; it induces exclusive and specific tribological properties on the thin coating. Then, the velocity-dependent frictional behavior of patterned SAMs seems to be directly related to the polarization ability of the end functionalities of the grafted molecules. Finally, SAMs represent an easy as well as an active and dry lubricant for nanotechnological devices; they have dramatically reduced friction and adhesion and have found use in various MEMS devices. However, alkyl chains do not sustain high compression and shear stresses, which significantly limits their lifetime of use. These limitations are motivating the development of a new generation of thin coatings based on polymer nanocomposites that display higher nano- and microtribological properties, as proposed by Sidorenko et al. [75].

References

1. H-S. Ahn, P.D. Cuong, S. Park, Y-W. Kim, J-C. Lim, *Wear*, **2003**, 255, 819.
2. M.K. Corbierre, N.S. Cameron, R.B. Lennox, *Langmuir,* **2004**, 20, 2867.
3. J.J. Storhoff, S.S. Marla, P. Bao et al., *Biosensors and Bioelectronics*, **2004**, 19, 875.
4. H.I. Kim, T. Koin, T.R. Lee, S.S. Perry *Langmuir,* **1997**, 13, 7192.
5. M.J. Azzopardi, H. Arribart, *J. Adhesion*, **1994**, 46, 103.
6. R.H. Yoon, D.A. Guzonas, *Colloid Surf. A: Physicochem. Eng. Aspects*, **1994**, 87, 163.
7. C.R. Tripp, M.L. Hair, *Langmuir*, **1992**, 8, 1120.
8. P. Silberzan, L. Leger, D. Ausserré, J.J. Benattar, *Langmuir*, **1991**, 7, 1647.
9. D. Angst, G.W. Simmons, *Langmuir*, **1991**, 7, 2236.
10. C.R. Kessel, S. Granick, *Langmuir*, **1991**, 7, 532.
11. A.G. Carson, S. Granick, *J. Mater. Res.,* **1990**, 5, 1745.
12. D.K. Schwartz, S. Stienberg, J. Israelchvili, J. Azasadzinski, *Phys. Rev. Lett.*, **1992**, 69, 3354.
13. J.V. Davidovits, V. Pho, P. Silberzan, M. Goldman, Surf. Sci., **1996**, 352-354, 369.
14. J. Gun, J. Sagiv, J. Colloid Interface Sci., **1986**, 112, 457.
15. E. Delamerche, B. Michel, C. Gerber et al., Langmuir, **1994**, 10, 2869.
16. K. Kendall, *Nature*, **1986**, 319, 203.
17. A. Lio, D.H. Charych, M. Salmeron, *J. Phys. Chem. B,* **1997**, 101, 3800.
18. X. Xiao, J. Hu, D.H. Charych, M. Salmeron, *Langmuir,* **1996**, 12, 235.
19. M.T. McDermott, J-B.D. Green, M.D. Porter, *Langmuir* **1997**, 13, 2504.
20. D. Dominguez, R.L. Mowery, N.H. Turner, *Tribol. Trans.*,**1994**, 37, 1, 59.
21. H.J. Eyring, *J. Chem. Phys.,* **1935**, 3,107.

22. D. Tabor. In *Fundamentals of Friction: Macroscopic and Microscopic Processes*; I.L. Singer, H.M. Pollock, Eds.; Kluwer Academic: Dordrect, The Netherlands, **1992**; p 3.
23. B. Bhushan, A.V. Kulkarni, V.N. Koinkar et al., *Langmuir,* **1995,** 11, 3189.
24. H. Liu, B. Bhushan, *Ultramicroscopy*, **2002**, 91, 185.
25. A. Kinloch, *J. Adhes. Adhes.* Chapman and Hall: New York, **1987**.
26. C.J. Drummond, G. Georgaklis, C. Chan, *Langmuir,* **1996,** 12, 2617.
27. H.I. Kim, T. Koini, T.R. Lee, S.S. Perry, *Langmuir,* **1997,** 13, 7192.
28. M. Graupe, T. Koini, H.I. Kim, N.G. Arg, Y.F. Miura, M. Takenaga, S.S. Perry, T.R. Lee, *Langmuir,* **1999**, 15, 3179.
29. C.O. Timmons, W.A. Zisman, *J. Colloid. Interface Sci.*, **1966**, 22, 165.
30. T. Stoebe, P. Mach, S. Grantz, C.C. Huang, *Phys. Rev.*, **1996**, 53, 1662.
31. M. Graupe, T. Koini, H.I. Kim, N. Garg, Y.F. Miura, M. Takenaga, S.S. Perry, T.R. Lee, *Colloids Surf. A: Physicochem. Eng. Aspects*, **1999**, 154, 239.
32. B.N.J. Person, *Phys.Rev. B*, **1991**, 44, 3277.
33. M.K. Chaudhury, M.J. Owen, *Langmuir*, **1993**, 9, 29.
34. H. Yoshizawa, Y-L. Chen, J. Israelachvili, *J. Phys. Chem.,* **1993**, 97, 4128.
35. R.M. Overney, E. Meyer, J. Frommer, H-J. Giintherodt, M. Fujihira, H. Takano, Y. Gotoh, *Langmuir,* **1994**, 10, 1281.
36. M. Garcia-Parajo, C. Longo, J. Servat, P. Gorostiza, F. Sanz, *Langmuir,* **1997,** 13, 2333.
37. B. Bhushan, H. Liu, *Phys. Rev. B*, **2001**, 63, 245412.
38. F. Tian, X. Xiao, M.M. Moy, C. Wang, C. Bai, *Langmuir,* **1999,** 15, 244.
39. E. Riedo, F. Lévy, H. Brune, *Phys. Rev. Lett.*, **2002**, 88, 185505.
40. Y. Liu, D.F. Evans, Q. Song, D.W. Grainger, *Langmuir,* **1996,** 12, 1235.
41. R. Yerushalmi-Rozen, J. Klein, *Langmuir*, **1995**, 11, 2806.
42. S.R. Cohen, R. Naaman, J. Sagiv, *J. Phys. Chem.,* **1986**, 90, 3054.
43. M.C. Yeh, E.J. Kramer, R. Sharma, W. Zhao, M.H. Rafailovich, J. Sokolov, J.D. Brock, *Langmuir,* **1996**, 12, 2747.
44. E. Gnecco et al., *Phys. Rev. Lett.*, **2000**, 84, 1172.
45. Y. Liu, T. Wu, D.F. Evans, *Langmuir*, **1994**, 10, 2241.
46. J. Wagner, T. Kirner, G. Mayer, J. Albert, J.M. Köhler, *Chem. Eng. J.*, article in press.
47. G.B. Khomutov, V.V. Kislov, M.N. Antipina, R.V. Gainutdinov et al., *Microelectron. Eng.*, **2003**, 69, 373.
48. W. Li, L. Huo, D. Wang, G. Zeng, S. Xi, B. Zhao, J. Wang, Y. Shen, Z. Lu, *Colloids Surf. A: Physicochem. Eng. Aspects*, **2000**, 175, 217.
49. L.M. Tender, R.L. Worley, H. Fan, G.P. Lopez, *Langmuir,* **1996,** 12, 5515.
50. R. Klauser, M-L Huang, S-C. Wang, C-H Chen, T.J. Chuang, A. Terfort, M. Zharnikov, *Langmuir,* **2004,** 20, 2050.
51. M.M. Alkaisi, W. Jayatissa, M. Konijn, *Curr. Appl. Phys.*, **2004,** 4,111.
52. M. Tormen, L. Businaro, M. Altissimo, F. Romanato, S. Cabrini, F. Perennes, R. Proietti, H-B. Sun, S. Kawata, E. Di Fabrizio, *Microelectron. Eng.,* **2004,** 73–74, 535.
53. O. Cherniavskaya, A. Adzic, C. Knutson, B.J. Gross, L. Zang, R. Liu, D.M. Adams, *Langmuir,* **2002,** 18, 7029.
54. M.R. Shadnam, S.E. Kirkwood, R. Fedosejevs, A. Amirfazli, *Langmuir,* **2004**, 20, 2667.
55. R.D. Piner, J. Zhu, F. Xu, S. Hong, C.A. Mirkin, *Science*, **1999**, 283, 661.

56. S. Hong, J. Zhu, C.A. Mirkin, *Science*, **1999**, 286, 523.
57. O.J. Schueller, D.C. Duffy, J.A. Rogers, S.T. Brittain, G.M. Whitesides, *Sens. Actuators*, **1998**, 78,149.
58. H. Haidara, K. Mougin, J. Schultz, *Langmuir,* **2000,** 16, 7773.
59. J.P. Folkers, P.E. Laibnis, G.M. Whitesides, *Langmuir,* **1992,** 8, 1330.
60. M.J. Wirth, R.W. Fairbank, H.O. Fatunmbi, *Science*, **1997**, 275, 44.
61. S.P. Li, A. Lebib, D. Peyrade, M. Natali, Y. Chen, *Appl. Phys. Lett.,* **2000**, 77, 2743.
62. Y. Chen, A. Lebib, S. Li, A. Pépin, D. Peyrade, M. Natali, E. Cambril, *Eur. Phys. J. AP,* **2000**, 12, 223.
63. M. Natali, A. Lebib, E. Cambril,Y. Chen, I.L. Prejbeanu, K. Ounadjela, *J. Vac. Sci. Technol. B,* **2001,** 19, 2779.
64. Y. Xia, G.M. Whitesides, *Angew. Chem. Int. Ed.* **1990**, 37, 550.
65. S.Y. Chou, P.R. Krauss, P.J. Renstrom, *Appl. Phys. Lett.*, **1995,** 67, 3114.
66. K. Mougin, G. Castelein, H. Haidara, *Tribol. Lett.*, **2004**, 17,1, 11.
67. J. Drelich, J.L. Wilbur, J.D. Miller, G.M. Whitesides, *Langmuir*, **1996**, 12, 1913.
68. F.M. Fowkes, "Contact Angle: Wettability and Adhesion American Chemical Society: Washington D. C, 1964."
69. H. Haidara, K. Mougin, J. Schultz, *Langmuir*, **2000,** 16, 9121.
70. Y. G Zhou, H. Fan, T. Fong, G.P. Lopez, *Langmuir,* **1998,** 14, 660.
71. H. Schonherr, G.J. Vansco, *Mater. Sci. Eng. C*, **1999**, 8–9, 243.
72. S. Gauthier, J.P. Aimé, T. Bouhacina, A.J. Attias, B. Desbat, *Langmuir,* **1996,** 12, 5126.
73. B. Bushan, J.N. Israelachvili, U. Landman, *Nature*, **1995**, 374, 607.
74. M. Schoen, C. Rhykerd, D. Diestler, J. Cushamn, *Science*, **1989**, 245, 1223.
75. A. Sidorenko, H-S. Ahn, D-I. Kimb, H. Yanga, V.V. Tsukruk, *Wear,* **2002**, 252, 946.

29 Sliding Friction of Polymers: The Complex Role of Interface

Sophie Bistac, Marjorie Schmitt, and Achraf Ghorbal

UHA –ICSI 15, Rue Jean Starcky, 68057 Mulhouse, France
sophie.bistac-brogly@uha.fr

29.1 Introduction

Tribology of polymers is complex, due to the fact that they exhibit specific properties compared to other materials. These specificities are due to the molecular structure, and especially chain mobility, at small or larger scales. This movement allows relaxation mechanisms and energy dissipation, notably by internal friction, and the mechanical behaviors of polymers is then viscoelastic, with time and temperature dependences. These dependences (linked by the time–temperature equivalence) are able to directly affect adhesion and friction levels [1, 2]. Adhesion and friction sciences are both characterized by their multidisciplinarity, involving actually several scientific domains, such as polymer chemistry and physics, surface chemistry, rheology and fracture mechanics.

Adhesion reflects the total energy of the substrate/adhesive interfacial bonds and depends only on the nature and density of these interactions. The adherence value is usually greater than the presumed adhesion value, because during separation, a part of the energy is dissipated by internal molecular motions (chains extension, disentanglement, etc.). Other experimental parameters will complicate, unfortunately, the solving of adhesion and friction problems. Changing the type of the adherence or friction test or the samples geometry can, for example, influence the measured values. Polymer friction is also governed by interfacial interactions and dissipation mechanisms located in the interfacial region but also in the bulk, especially in the case of soft materials. The problem with polymer friction is more complex because both materials are in dynamic contact, during which interactions are built and broken up simultaneously.

Relation between adhesion and friction has been investigated for peel tests, during which friction mechanisms can be induced [3, 4]. But the correlation between both properties is usually delicate. However, the challenge is interesting, in order to better predict friction behavior and associated phenomena like wear or lubrication, and also to have the possibility to design smart surfaces able to present specific properties.

Some authors have taken up this difficult challenge [5–10], by defining some pertinent relations between interfacial interactions and friction. This

approach is interesting and original, mainly due to the fact that it is a phenomenological approach, combining mechanical, physical and chemical aspects of adhesion and friction and allows molecular mechanisms at interfaces, closely linked to molecular structures, to be proposed.

In order to investigate the mechanism of boundary friction, adhesion and friction experiments were carried out with a variety of surfactant-monolayer-coated surfaces using the surface forces apparatus. The friction forces were measured as a function of monolayer coverage and phase state, load, temperature, shear rate and relative humidity. The friction coefficient is not directly correlated with adhesion but rather with the adhesion hysteresis, which is a measure of the energy dissipated during an irreversible adhesion (e. g. loading–unloading) cycle. Chain interdigitation appears to be the most important molecular mechanism that gives rise to "boundary" friction and adhesion hysteresis of monolayer-coated surfaces [6].

The role of chain mobility in adhesion and friction has also been studied in the case of elastomers, based on dissipation phenomena during adhesion and friction process in the case of an elastomer in contact with a silicon wafer covered by a grafted layer [11–15]. Some authors have studied the rolling of a cylinder on a flat plate as the propagation of two cracks – one closing at the advancing edge and the other opening at the trailing edge. The difference of adhesion in these two regions, i. e. the adhesion hysteresis, depends on the nonequilibrium interfacial processes in an elastic system. This rolling-contact geometry was used to study the effects of dispersion forces and specific interactions on interfacial adhesion hysteresis. In order to accomplish this objective, hemicylindrical elastomers of polydimethylsiloxane (PDMS) – both unmodified and plasma oxidized – were rolled on thin PDMS films bonded to silicon wafers. Plasma oxidation generates a silica-like surface on PDMS elastomer, which interacts with PDMS molecules via hydrogen-bonding forces. The adhesion hysteresis in the latter case is large and depends significantly on the molecular weight of the grafted polymer, whereas the hysteresis is rather negligible for purely dispersive systems. These results are interpreted in terms of the orientation and relaxation of polymer chains [12]. The effect of interfacial friction on adhesion was also studied, by using a model system, in which a thin strip of silicone elastomer was peeled from a PDMS-treated glass slide. The fluorescent particle-tracking method revealed that the elastomer stretches and slides on glass well before the crack faces open up. Interfacial energy dissipation due to friction, which is the product of the slip displacement and the interfacial shear stress, is found to be a significant contributor to the total fracture energy at various peel configurations. These results suggest that interfacial sliding may provide a mechanism for energy dissipation in the fracture of asymmetric interfaces [11, 15].

For fundamental tribological studies in recent years, self-assembled monolayers (SAMs), formed by chemical adsorption of molecules on solid surface [16,17], were used as boundary lubricants [18–21]. The grafting of mono-

layer on a rigid substrate allows control of the surface chemistry and thus the surface energy without varying the surface roughness. Organized and dense monolayers can be produced by chemical grafting of molecules and by Langmuir–Blodgett deposition. Friction, wear, and indentation of these films on the microscale have been studied using atomic force microscopy friction force microscopy. For comparison, macroscale friction and wear measurements have also been made. Mechanisms of friction, wear, and lubrication of chemical grafted and LB films have been discussed. Moreover, owing to the development of computer-simulation techniques for studying these phenomena at the atomic scale, an understanding is beginning to emerge of the molecular mechanisms of tribology in thin films and at surfaces [20, 21]. During recent decades, much attention has focused on the effect of molecular weight [22, 23], sliding velocity and load [24–26] on friction and wear of materials and especially of polymers. But still little attention has been paid to the role of the interface. In fact, the interface zone is expected to be the field of energy-dissipation processes [27] caused by the contribution of surface energy [28] and opposing surfaces macromolecular interactions, and the mobility of macromolecules in the vicinity of the interface [29]. Moreover, it has been demonstrated that the terminal chain group chemistry of rubbed surfaces and polymers chains mobility can have a significant impact on the friction [30].

Moreover, rheology of polymer surface can be completely different from the bulk [31, 32] and their properties cannot necessarily be known by an extrapolation of polymer bulk. A surface-force apparatus allows investigation of the rheology and tribology of thin liquid films between shearing surfaces, with the measurement of previously inaccessible parameters during frictional sliding, such as the real area of contact, the local asperity load and pressure, and the sheared film thickness. The results show striking noncontinuum, nonbulk-like effects when the thicknesses of sheared films approach molecular dimensions, as occurs under most tribological conditions. A surface-force apparatus allows the viscosity of thin films to be measured at very high shear rates. The results indicate that the nonslip plane, or "shear" plane, is located within less than one molecular diameter of the solid/liquid interface so long as the film thickness is greater than about 10 molecular diameters. Only for thinner films does the viscosity increase markedly from that expected from continuum behavior. This effect correlates with deviations seen in the continuum van der Waals interactions between surfaces at very small separations. The authors conclude that it is not so much the liquid structuring at isolated surfaces that leads to deviations from continuum behavior in both static and dynamic interactions, but rather the close approach of two surfaces that modifies the liquid structure between the surfaces that leads to these effects. A further conclusion is that the structure, interactions, and mobility of liquid molecules adjacent to one surface are not additive properties on approach of a second surface [33, 34].

The previous scientific studies have therefore shown that friction of polymers is able to be greatly influenced by adhesion phenomena. The nature and number of interfacial interactions between the polymer and the substrate can indeed directly affect the friction coefficient. To illustrate the complex role of the interface, polymer friction has been investigated in contact with two types of smooth and rigid substrates being used: hydrophilic silicon wafer (hydroxylated by a piranha treatment) and hydrophobic silicon wafer, obtained by a chemical grafting with a CH_3-terminated silane. Both substrates exhibit identical stiffness and roughness and differ only by their surface chemical composition. Two kinds of polymers are studied: a crosslinked polydimethylsiloxane, PDMS (soft elastomer with a low glass transition temperature) and polystyrene (rigid thermoplastic polymer with a high glass transition temperature). Friction experiments are performed with a translation tribometer for different applied forces and speeds. The evolution of the friction coefficient as a function of speed and normal force is analyzed for both polymers.

Energy-dissipation processes occur at an interface during friction, caused by the contribution of surface energy and also macromolecule mobility. The objectives of the current study are to specify the influence of the interface in friction in order to better understand the complex role of adhesion.

29.2 Materials and Methods

Two polydimethylsiloxanes (PDMS), vinyl terminated, were used (provided by Gelest). The two samples vary by their initial molecular weight Mw: Mw = 6000 g/mol (called PDMS A) and Mw = 17,200 g/mol (called PDMS B). PDMS are crosslinked with tetrakis(dimethylsiloxy)silane, which possess four functional sites, using a platinum catalyst at room temperature. The stoichiometry ratio (crosslinker/PDMS) is 1.1 (10% excess for the crosslinker) to ensure a correct reaction. Increasing the initial molecular weight of PDMS induces major consequences: a lower crosslinking density (higher chains length between chemical crosslinks) and consequently a lower modulus. Glass transition temperature (measured by differential scanning calorimetry at a scanning rate equal to 10 °C/min) and Young's modulus (determined by tensile test at 10 mm/min) of crosslinked PDMS are presented in Table 29.1. PDMS hemispheres (diameter = 16 mm) are used for friction experiments.

Table 29.1. Glass transition temperature T_g and Young's modulus E of PDMS A and B

	T_g (°C)	E (MPa)
PDMS A	–123	1.40
PDMS B	–123	0.42

Table 29.2. Glass transition temperature T_g and Young's modulus E of PS A and B

	T_g (°C)	E (MPa)
PS A	92	2340
PS B	103	2700

Two polystyrenes, varying by their molecular weight, (purchased from Aldrich) are used. The molecular weights Mw are, respectively, equal to 180,000 and 220,000 g/mol for PS A and PS B, with a polydispersity index equal to 2 for both polymers. Glass transition temperature (measured by differential scanning calorimetry at a scanning rate equal to 10 °C/min) and Young's modulus (determined by tensile test at 10 mm/min) of both polystyrenes are presented in Table 29.2. Cylinders (diameter = 2 mm) are obtained by extrusion molding at 190 °C.

Model substrates were composed of a hydrophilic silicon wafer on the one hand, and of a hydrophobic grafted wafer on the second hand. Substrates were (1.5×2) cm^2 slices of polished silicon wafers Si(100) (Silicon Prime wafers). Wafer surfaces were chemically treated so as to obtain hydrophilic and hydrophobic substrates with, respectively, hydroxyl-terminated and methyl-terminated surfaces.

The hydrophilic surface is obtained by a treatment with a piranha solution (H_2SO_4/H_2O_2). This treatment is aimed to produce a high surface silanol density.

The hydrophobic substrate is obtained by grafting a methyl terminated silane (hexadecyltrichlorosilane) on a silicon wafer (previously piranha treated). The terminal group of the grafted layer is then a $-CH_3$ function, which exhibits hydrophobic properties. The formation of homogenous monolayers onto silicon surfaces technique is already well described in the literature. After the cleaning and the activation step of wafers, they were immersed into a solution of 0.1 vol% hexadecyltrichlorosilane (referred to as HTS) in carbon tetrachloride (CCL_4) for 8 h at room temperature. The substrates were then rinsed in CCL_4 solution to remove any excess of physisorbed silanols, rinsed with deionized water and dried under a nitrogen flow.

Friction properties are measured using a translation tribometer (Fig. 29.1). The polymer sample (hemisphere for PDMS and cylinder for polystyrene) is brought into contact with the substrate under a given normal load. The substrate is then moved in translation at a controlled speed (one single passage) and the tangential force, which corresponds to the friction force, is measured. At least 5 friction measurements were performed for each experimental conditions to obtain a mean friction coefficient value. The friction coefficient μ was defined as the quotient of the tangential force F_T divided by the applied normal force F_N, $\mu = F_T/F_N$. All data are collected at ambient conditions.

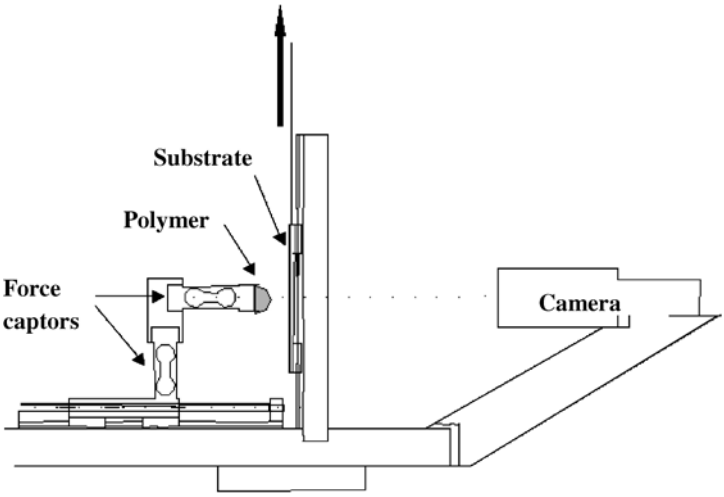


Fig. 29.1. Translation tribometer

29.3 Results and Discussion

Wettability measurements indicate a surface energy close to $20\,\mathrm{mJ/m^2}$ for the hydrophobic wafer (CH_3-grafted surface), with a polar component equal to zero. This result confirms the hydrophobic character of the grafted wafer. A surface energy equal to $79\,\mathrm{mJ/m^2}$ is obtained for the hydrophilic substrate (polar component = $42\,\mathrm{mJ/m^2}$).

PDMS samples have a similar (and low) surface energy (close to $27\,\mathrm{mJ/m^2}$), with a negligible polar (or nondispersive) component. The surface energy of both poly(styrene) is equal to $36\,\mathrm{mJ/m^2}$, with a polar component of $1\,\mathrm{mJ/m^2}$.

Friction tests are performed at different speeds and normal forces. Friction of polystyrenes will be analyzed first. Table 29.3 reports the friction coefficient μ value (tangential force divided by normal force) for PS B in contact with the hydrophilic and hydrophobic wafers. Quite identical values are obtained for PS A.

Table 29.3. Friction coefficient values for PS B in contact with the hydrophilic and hydrophobic wafers, for two normal forces F_N (1 and 2 N) and friction speeds V (1 and 25 mm/min)

	$F_N = 1$ N $V = 1$	$F_N = 1$ N $V = 25$	$F_N = 2$ N $V = 1$	$F_N = 2$ N $V = 25$
Hydrophilic wafer	0.48	0.48	0.50	0.60
Hydrophobic wafer	0.11	0.16	0.13	0.16

Quite identical friction coefficients are obtained for PS A and B whatever the speed and normal force, taking into account the experimental errors. The influence of chain length on friction coefficient is then negligible, in the studied forces, speeds and molecular weights ranges. The absence of a molecular-weight dependence can be explained by both similar mechanical properties and identical surface energy of the studied polystyrenes.

For both systems, the friction coefficient is slightly increased when the speed is increased. The friction coefficient is increased, for example for PS B, from 0.50 to 0.60 for hydrophilic wafer and from 0.13 to 0.16 for hydrophobic wafer when the speed is increased from 1 to 25 mm/min (normal force = 2 N). This effect could be explained by dissipation phenomena that increase at higher speed, due to the viscoelasticity of polystyrene. Tensile tests have shown a slight increase of the polystyrene Young's modulus with speed (PS B modulus is increased from 2300 to 2800 MPa when the tensile speed is increased from 1 to 50 mm/min).

The effect of normal force is quite negligible, in the studied range (from 1 to 5 N).

The main result is the great influence of the nature of the substrate: an important decrease of the friction coefficient is observed for hydrophobic wafer / PS systems. The friction coefficient varies from 0.50 for the hydrophilic wafer to 0.13 for the hydrophobic one (for a normal force = 2 N and a speed = 1 mm/min).

The significant decrease of the friction coefficient (divided by a factor 3) is due to the weak interfacial interactions between the hydrophobic wafer and polystyrene. To resume, friction results obtained for both polystyrenes underline the major role of the chemical composition of the substrate surface, with a great decrease of the friction coefficient for the hydrophobic wafer / PS system.

The influence of the nature of the substrate has also been investigated for PDMS. Both elastomers differ by their degree of crosslinking, which induces a large difference in mechanical properties (modulus). Other structural differences are also induced by varying the degree of crosslinking. The crosslinking reaction is indeed usually incomplete, leading to an imperfect network. Increasing the initial molecular weight of PDMS induces globally two consequences: a lower crosslinking density (higher chain length between chemical nodes) and also a greater quantity of free chains (nonlinked to the network) and pendant chains (linked to the network by only one extremity). Even if they are chemically identical, PDMS differ then by their molecular structure.

Table 29.4 presents friction coefficient values of PDMS A and B, for various normal forces and speeds, and for both substrates: hydrophilic and hydrophobic.

The influences of PDMS molecular weight (or degree of crosslinking), normal force and speed on friction have been described in previous papers [22, 34, 35, 35–39]. Friction coefficients of both PDMS are quite simi-

Table 29.4. Friction coefficient values for PDMS A and B, in contact with hydrophobic and hydrophilic substrates for different friction speeds and normal forces

Speed mm/min		25		50		120	
		Hydro phobic	Hydro philic	Hydro phobic	Hydro philic	Hydro phobic	Hydro philic
1N	A	1.04	1.58	1.29	1.53	1.35	1.36
	B	1.05	1.47	1.10	1.42	1.27	1.23
3N	A	0.81	0.88	0.95	0.86	1.04	0.93
	B	0.76	1.00	0.80	1.02	0.92	0.91

lar whatever the experimental conditions. However, both elastomers exhibit different stiffness (Young's modulus equal to 1.40 MPa and 0.42 MPa, respectively, for PDMS A and B). If we calculate a friction stress (friction force divided by the contact area between elastomer and substrate), significant differences appear between the two polymers, with a higher friction stress for PDMS A, due to its lower stiffness. Explanations based on the respective roles of elastic and adhesive contact have been given to explain the friction-stress differences [22, 33].

Results show, for both PDMS and both substrates, a large effect of normal force with an increase of friction coefficient for low normal forces, as illustrated in Fig. 29.2, which presents the evolution of friction coefficient of both polymers, in contact with hydrophilic substrate, as a function of normal force (a similar evolution is observed for a hydrophobic substrate). The higher friction coefficient observed at low normal force could be explained by the role of adhesion that is magnified at low load (where the bulk contribution is lower). The contribution of interfacial interactions is then high.

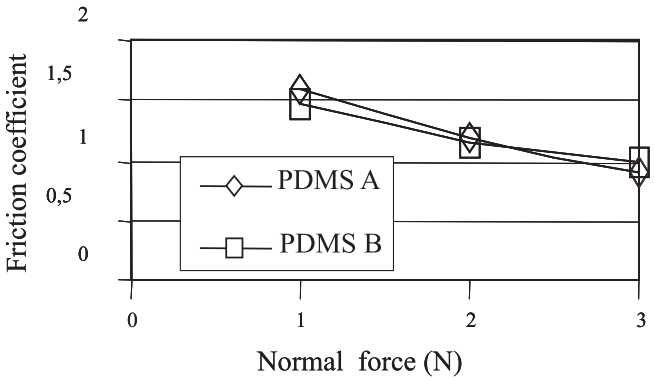


Fig. 29.2. Friction coefficient of PDMS A and B as a function of normal force, for hydrophilic wafer (speed = 25 mm/min)

The discussion will be now focused on the effect of the nature of the substrate and the role of interfacial interactions on friction. The influence of the nature of the substrate appears more complex compared to the previous system (PS).

A significant effect can indeed be observed at lower speed (and low normal force). However, the friction-coefficient differences between hydrophilic and hydrophobic systems are lower compared to whose observed for PS, for which the friction coefficient was divided by a factor equal to 3. For PDMS A, the friction coefficient (at 25 mm/min and 1 N) is equal to 1.04 for a hydrophobic wafer and 1.58 for a hydrophilic wafer (factor equal to 1.5 between both coefficients).

When the friction speed is increased, the difference between friction coefficients of hydrophilic and hydrophobic systems becomes lower, and both coefficients are identical for high speed.

This result means that the speed dependence of friction coefficient is strongly dependent on the nature of the substrate. The friction coefficient is increased as a function of speed for hydrophobic substrate and is decreased for hydrophilic substrate, as illustrated on Fig. 29.3, which presents the evolution of the friction coefficient as a function of speed for both substrates and for PDMS A for a normal force equal to 1 N (the same evolution is observed for PDMS B and for other normal forces).

These results indicate that at higher speed, both friction coefficients (for hydrophobic and hydrophilic systems) are equal, the effect of interface becoming negligible. The rheological behavior of the confined PDMS interfacial layer is able to explain this complex behavior [27]. A competition between interfacial interactions and polymer cohesion can be proposed: at low speeds, interfacial interactions will control the friction (the role of the substrate sur-

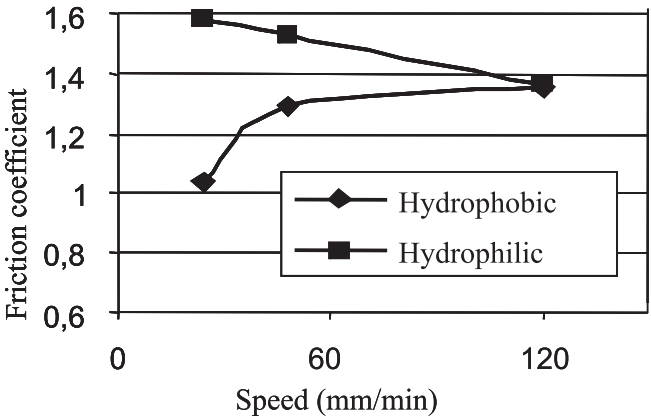


Fig. 29.3. Evolution of the friction coefficient of PDMS A as a function of speed, in contact with hydrophilic and hydrophobic wafer ($F_N = 1$ N)

face chemistry is then significant), and at higher speeds, the influence of the substrate surface becomes negligible and the friction is governed by the intrinsic rheological behavior of the polymer surface. At high speed, chain orientations can occur, induced by the interfacial shear stress. These orientations have been previously evidenced by infrared spectroscopy [36]. Such an orientation will induced a specific rheological behavior of the interfacial polymer layer. At low speed, strong interactions between PDMS and a hydrophilic substrate will activate the dissipation mechanism and bulk deformation, inducing a high friction coefficient. But when the speed is increased, the chains will be oriented at the interface during sliding, avoiding a correct stress transmission to the bulk polymer, and reducing therefore the bulk deformation and consequently the friction coefficient. At high speed, this rheological phenomena will then consume less energy for the hydrophilic system (less dissipation), explaining the decrease of the friction coefficient.

For hydrophobic substrates, interfacial interactions are weak, and when the speed is increased, the same chain orientation will occur at the interface during sliding, but this phenomena will be more dissipative than the effect of interfacial interactions. This competition between interfacial interactions and the interfacial rheological effect is able to explain the increase of the friction coefficient with speed for a hydrophobic substrate.

For the poly(styrene) samples, this complex behavior was not observed: a higher speed induced an increase of the friction coefficient, whatever the nature of the substrate. Polystyrene exhibits a greater cohesion, and a lower friction coefficient (lower interfacial shear level due to less polymer deformation). For polystyrene, the competition between polymer cohesion and interfacial interactions is not equilibrated: friction is still governed by interfacial interactions, whatever the experimental conditions (for the studied normal forces and speeds ranges).

For PDMS, which exhibits a low stiffness, competition between the (low) cohesion and adhesion is more equilibrated: at low speeds, interfacial interactions have a significant effect and partly govern the friction, and at high speeds the influence of the substrate surface becomes negligible and friction is then governed by the polymer intrinsic viscoelastic behavior. The rheological properties of the confined interfacial layer are then able to explain the complex behavior. Analysis of the transfer layer would (observed by atomic force microscopy) also lead to a better understanding of the involved mechanisms [39].

29.4 Conclusion

Experimental results underline the subtle competition between interfacial interactions and polymer rheological properties, especially for PDMS samples. This competition will directly govern energy dissipation (viscoelastic effects) and consequently friction coefficient. The influence of nature of the substrate

on the friction coefficient of both polymers (PS and PDMS) is really different. The effect of surface energy on friction is then not so evident: the friction coefficient of hydrophilic and hydrophobic substrates can be identical or different, depending on the experimental conditions. Friction speed is able to play a major role, through its influence on polymer interfacial rheogoly.

This subtle effect could be a way to monitor friction behavior as a function of speed, with the possibility to have a high friction at low speed and a low friction at high speed (or inversely by changing the substrate chemistry), depending on the application requirements.

Adjustable friction levels could then be proposed, by modifying the competition between interfacial rheology and adhesion.

References

1. Gent, A.N., Schultz, J., J. Adhes., 3, 281 (1972)
2. Bistac, S., J. Colloid Interf. Sci., 219, 210 (1999)
3. Zhang Newby, B.M., Chaudhury, M.K., Langmuir, 13, 1805 (1997)
4. Amouroux, N., Petit, J., Léger, L., Langmuir, 17, 6510 (2001)
5. Israelachvili, J.N., Intermolecular and Surface Forces, Academic Press, London (1991)
6. Yoshizawa, H., Israelachvili, J.N., Thin Solid Films, 246, 71 (1994)
7. Heuberger, M., Drummond, C., Israelachvili, J.N., J. Phys. Chem. B, 102, 5038 (1998)
8. Yamada, S., Israelachvili, J.N., J. Phys. Chem. B, 102, 234 (1998)
9. Chen, Y.L., Helm, C.A., Israelachvili, J.N., J. Phys. Chem., 95, 10736 (1991)
10. Heuberger, M., Luengo, G., Israelachvili, J.N., J. Phys. Chem. B, 103, 10127 (1999)
11. Ghatak, A., Vorvolakos, C., She, H., Malotky, D.L., Chaudhury, M.K., J. Phys. Chem. B, 104, 4018 (2000)
12. She, H., Malotky, D. Chaudhury, M.K., Langmuir, 14, 3090 (1998)
13. Chaudury, M.K., Owen, M.J., J. Phys. Chem., 97 (21), 5722 (1993)
14. Deruelle, M., Léger, L., Tirrell, M., Macromolecules, 28, 7419 (1995)
15. Amouroux, A., Léger, L., Langmuir, 19, 1396 (2003)
16. Ulman, A. An introduction to Ultrathin Organic films, from Langmuir–Blodgett to Self-Assembly, Academic Press: San Diego, CA, pp 245–253 (1991)
17. Sellers, H., Ulman, A., Shnidman, Y., Eilers, J. J. Am. Chem. Soc., 115(21), 9389 (1993)
18. Zhang, Q., Archer, L.A. Langmuir, 21, 5405 (2005)
19. Chandross, M., Grest, G. S., Stevens, M.J. Langmuir, 18(22), 8392 (2002)
20. Bhushan, B., Israelachvili, J.N., Landmann, U. Nature, 374, 607 (1995)
21. Bhushan, B., Kulkarni, A.V., Koinkar, V.N., Boehm, M., Odoni, L., Martelet, C., Belin, M. Langmuir, 11(8), 3189 (1995)
22. Galliano, A., Bistac, S., Schultz, J.J. Colloid Interf. Sci., 265(2), 372 (2003)
23. Lee, S. W., Yoon, J., Kim, H. C., Lee, B., Chang, T., Ree, M. Macromolecules, 36(26), 9905 (2003)
24. Gasco, M. C., Rodriguez, F., Long, T. J. Appl. Polym. Sci., 67(11), 1831 (1998)
25. Zhang, S., Lan, H., Tribol. Int., 35(5), 321 (2002)

26. Rubinstein, S. M., Cohen, G., Fineberg, J., Nature, 430, 1005 (2004)
27. Persson B. N. J. Surf. Sci., 401, 445 (1998)
28. Chen, N., Maeda, N., Tirrell, M., Israelachvili, J. Macromolecules, 38, 3491 (2005)
29. Dao, T. T., Archer, L.A. Langmuir, 17(13), 4042 (2001)
30. Liu, Y., Evans, D. F., Song, Q., Grainger, D.W. Langmuir, 12(5), 1235 (1996)
31. Wallace, W. E., Fischer, D. A., Efimenko, K., Wu, W.-L., Genzer, J. Macromolecules, 34, 5081(2001)
32. Luengo, G., Schmitt, F.-J., Hill, R., Israelachvili, J. Macromolecules, 30(8), 2482 (1997)
33. G. Luengo, J. Israelachvili S. Granick, Wear, 200, 328 (1996)
34. Jacob N. Israelachvili Stephen J. Kott, J. Colloid Interf. Sci., 129, 461 (1989)
35. Galliano, A., Bistac, S., Schultz, J., J. Adhes., 79, 973 (2003)
36. Elzein, T., Galliano, A., Bistac, S., J. Polym. Sci., part B, Polym. Phys., 42, 2348 (2004)
37. Bistac, S., Galliano, A., Tribol. Lett., 18 (1), 21 (2005)
38. Bistac, S., Galliano, S., in Adhesion – Current Research and Application, ed. Wulff Possart, Wiley-VCH (Weinheim, Germany) 59 (2005)
39. Bistac, S., Ghorbal, A., Schmitt, M., Progress in Organic Coatings, 55, 345 (2006)

30 Molecular Origins of Elastomeric Friction

Scott Sills[1,3], Katherine Vorvolakos[2], Manoj K. Chaudhury[2], and René M. Overney[1]

[1] Department of Chemical Engineering, University of Washington, Seattle, WA 98195
[2] Department of Chemical Engineering, Lehigh University, Lehigh, PA
[3] IBM Almaden Research Center, 550 Harry Road, San Jose, CA 95120

Frictional properties of soft elastomers have been in question for over half of a century. Early studies [1–3] on natural rubber originated for the sole purpose of tabulating properties for bulk consumer applications, such as viscoelastic adhesives, [4,5] tires, [6] and windshield wipers, [7,8] to name a few. Empirical tabulation of frictional properties persisted until the early 1950's, when Roth et al. [9] and Thirion [10] began experiments towards a fundamental understanding of rubbery sliding. Quantitative physical analysis began with the observation that the classic Coulombic laws obeyed consistently at rigid body interfaces fail at the interface between a rigid solid and a rubber. Even today, there remains an incomplete understanding of the molecular level parameters that control the frictional behavior of elastomeric surfaces. With this chapter, we explore the historical developments in elastomeric friction and discuss the evolution of an unresolved *triborheological complexity*. We work from an initial macroscopic perspective toward a microscopic one that describes dissipation process in terms of molecular phenomena at frictional contacts. Readers are urged to consider a competition between these molecular processes, where for soft matter, *internal cohesion* is comparable to *interfacial adhesion*. While cohesion may dominate adhesion, or *vice versa*, we develop a picture for elastomeric friction that encompasses both.

30.1 Early Elastomeric Friction Studies

Early studies by Papenhuyzen, [11] as well as Roth et al. [9], showed that the friction force of commercial rubbers on steel increases monotonically with velocity. Beyond a certain velocity, however, sliding becomes unstable and the rubber sample "chatters", or exhibits stick-slip sliding. Thirion [10], on the other hand, observed that the friction increases with normal load, which Schallamach [12,13] attributed to the increase of contact area resulting from the deformation of rubber asperities. Similar suggestions were made by Bowden and Tabor [14]. Assuming the asperities are hemispheres in Hertzian contact with smooth glass, Schallamach predicted that friction should increase with load in a power law manner, with an exponent of two third. Indeed, this prediction was verified over a limited range of loads. However, Schallamach

did not immediately address a crucial implication of his prediction: that the friction force should increase with modulus! If frictional force were to depend on contact area, a softer (lower modulus) material would have a greater contact area at any load, thus exhibiting higher friction, which was contrary to several experimental observations. He moved on to examine the effects of velocity and temperature [19] on rubber friction. As temperature increases, the frictional force decreases. Alternatively, at a given temperature, the friction force increases with sliding velocity. Schallamach showed that the velocity- and temperature-dependent behavior of rubber friction follows Eyring's [15] theory of reaction rates. However, when this activation theory is applied to explain elastomeric friction, interfacial sliding can be described by both tribological (interfacial) and rheological (bulk material) models. While the formation and breakage of adhesive molecular bonds occur at the contact interface, viscoelastic molecular relaxation of the stressed molecules constitutes an internal friction component within the bulk elastomer, i. e. the friction associated with the molecules sliding along themselves. Resolving the respective contribution of each dissipative process to the overall friction remains an ongoing challenge that represents the triborheological complexity.

While Schallamach focused on the molecular processes at the interface, Greenwood and Tabor [16] as well as Bueche and Flom [17] pointed out that the energy of sliding a soft viscoelastic material over a rigid substrate is not spent entirely in breaking molecular contacts at the interface, but at least partially on deforming the soft material. The notion that friction might be a combination of surface and bulk effects prompted Grosch [18] to perform the most systematic study in the field to date. He measured the effects of velocity, temperature, and surface roughness for rubber sliding on optically smooth glass. Grosch observed that the rubber friction increases non-linearly with velocity, much like the shear thinning behavior of high viscous polymers. Above a certain critical velocity, the friction force exhibits a stick-slip behavior with the maximum friction in each pulse decreasing with velocity. Furthermore, at each sliding velocity, friction decreases with increasing temperature. Grosch collapsed his friction-rate isotherms into a single master curve using the well-known Williams, Landel and Ferry [19] (WLF) superposition principle, which is pervasive in bulk rheology. His most intriguing finding was that the velocity corresponding to maximum friction and the frequency corresponding to maximum viscoelastic loss form a ratio that is nearly constant ($\sim 7\,\mathrm{nm}$) for various rubbery materials. He asserted that the interfacial relaxation processes responsible for friction are related to the segmental relaxation of polymer chains. With this, he initiated a debate that continues today: whether the origin of elastomeric friction lies in interfacial adhesion or viscous relaxation of the bulk elastomer.

The critical length of 7 nm represents a molecular length, presumably the characteristic length for molecular jumps in an adhesive picture for sliding friction, or alternatively, the length scale for viscoelastic relaxation in the

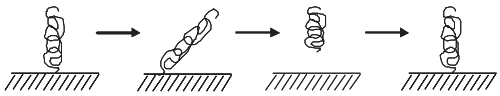


Fig. 30.1. The classic depiction of a polymer chain in contact with a laterally moving countersurface. The chain stretches, detaches, relaxes and re-attaches to the surface to repeat the cycle. Adapted from reference [21]

bulk elastomer. For rough surfaces, the relevant length scale was found to be the characteristic spacing between surface asperities. Grosch's general observations of the dependence of friction on velocity and temperature were also supported by Extrand et al. [8], who examined the more practical geometry of sharp rubber edges against rigid surfaces. Extrand et al. noted that the coefficient of friction depends strongly on the local load and the results are dependent on the surface preparation, i. e. chlorination of natural rubber, which may effect both the viscoelastic moduli and the interfacial adhesion.

Prompted by Grosch's observations, Schallamach [20] refined his model of interfacial friction, since a prediction of a monotonic dependence of friction on velocity was insufficient. He maintained that unlubricated sliding on smooth surfaces is essentially adhesive in nature, mediated by separate bonding and debonding events between the rubber and the rigid surface, depicted in Fig. 30.1.

Schallamach's [20] explanation of Grosch's [18] observations was based on the rate-dependent molecular debonding model of Frenkel [22] and Eyring [15]. In this model, the probability of debonding a polymer chain from a surface is a product of two functions, the first being the frequency factor that increases exponentially with the applied force, and the second being the number of load-bearing chains that decreases with velocity. The solution of the kinetic

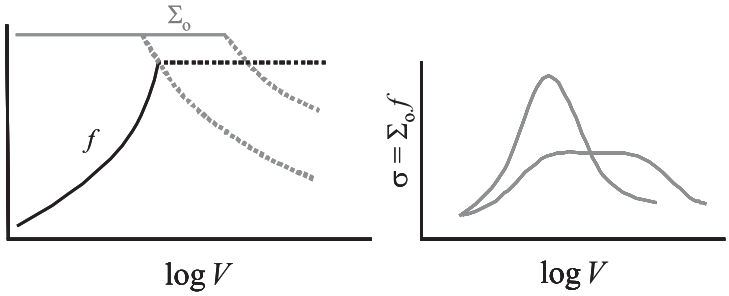


Fig. 30.2. The left figure qualitatively depicts the behavior of the areal density of contact points and the force per adsorption point as a function of velocity. The former decreases, while the latter increases up to a value limited by the interaction strength between the polymer chain and the countersurface. The product of these two quantities yields the shear stress, which increases and subsequently decreases, depicted on the right. Adapted from reference [21]

rate equation resulting from such considerations, leads to an expression for the debonding force that increases with velocity, while the number of the load-bearing polymer chains (Σ) decreases (Fig. 30.2). The net effect is that the total interfacial stress at first increases with velocity, reaches a maximum, and thereafter decreases with velocity.

30.2 Stochastic Adhesive Model

A couple of decades ago, Chernyak and Leonov [23] refined Schallamach's picture of rubber friction by using a steady state stochastic model for debonding kinetics. Within this model, directional stretching of polymer chains occurs as a result of an external force, leading to the detachment of linking chains from the wall. A newly detached chain dissipates the elastic energy accumulated during stretching and re-attaches to the surface. By considering the stochastic nature of detachment forces, Chernyak and Leonov [23] derived the shear stress in dry sliding as given by Eq. 30.1,

$$\sigma_t = \Sigma_\circ \frac{\int_0^\infty \varphi\left(\frac{r(t)}{\delta}\right) p\,(V,t)\,\mathrm{d}t}{V\left[\langle t\rangle_\mathrm{b} + \langle t\rangle_\mathrm{f}\right]} \tag{30.1}$$

where $\Sigma_\circ$ is the areal density of the load bearing chains at zero velocity, $\varphi\left(\frac{r(t)}{\delta}\right)$ is the *elastic* energy stored in the stretched polymer chain, V is the sliding velocity, $\langle t\rangle_\mathrm{b}$ is the mean lifetime of contact, $\langle t\rangle_\mathrm{f}$ is the time the polymer chain spends in free state, $p(V,t)$ is the transition probability of the polymer chain in going from the bonded to the debonded, elastically restored state. The numerator of the Chernyak and Leonov equation (Eq. 30.1) is the work done in stretching the polymer chain to the detachment point, while the denominator represents the mean distance traveled by the chain. Multiplication of this stochastic force with the areal density of the linking chains gives rise to the expression for shear stress. This formulation however, neglects the fact that polymers are not perfectly elastic and can dissipate energy during the stretching process, through the viscous modes associated with internal friction.

Using a steady state stochastic model of bond dissociation, Chernyak and Leonov showed that the mean lifetime of contact $\langle t\rangle_\mathrm{b}$ and the transition probability depend on the sliding velocity as shown, respectively, in Eq. 30.2 and 30.3.

$$\langle t\rangle_\mathrm{b} = \tau_\circ \left\{1 - \exp\left(-\frac{V}{V_\circ}\right)\right\} \tag{30.2}$$

$$p\,(V,t) = \exp\left(-\frac{t}{\tau_\circ}\right)\left\{\delta\,(t - t_\mathrm{b}) - \frac{\theta\,(t_\mathrm{b} - t)}{\tau_\circ}\right\} \tag{30.3}$$

Here, $\delta(z)$ represents Dirac's delta function corresponding to the determinate process of forced break-off, and $\theta(z)$ is the Heaviside step function. With the

above definitions of the bond lifetime and the transition probability, Eq. 30.1 can be integrated for simple Gaussian polymer chains, the elastic energy of which is proportional to the square of the extension. Shear stress can then be expressed as follows:

$$\sigma = \sigma_a \left(m+1\right) \frac{u\left(1+s\right)\left(1-\exp\left(\frac{-1}{u}\right)-\exp\left(\frac{-1}{u}\right)\right)}{m+1-\exp\left(\frac{-1}{u}\right)} \tag{30.4}$$

where m is the fundamental ratio of the lifetimes of the polymer chain in the free and bound states at zero sliding velocity, s is the ratio of the viscous retardation time over the lifetime at rest and u is the dimensionless velocity of sliding defined by Eq. 30.5,

$$u = \frac{V\tau_\circ}{\delta a} \tag{30.5}$$

where $\tau_\circ$ is the lifetime of the bound state at rest, δ is the average distance between the polymer body and the wall, and a is the statistical segment length of the polymer. σ_a is defined by Eq. 30.6.

$$\sigma_a = \frac{kT\Sigma_\circ\delta}{\left(1+m\right)R_{\mathrm{F}}^2} \tag{30.6}$$

R_{F} is the Flory radius of the polymer chain. Eq. 30.4 predicts that the shear stress first increases with velocity in an s-shaped manner. After exhibiting a rather broad maximum, σ usually decreases at very high sliding velocities.

30.3 Viscoelastic Perspective: Contributions from internal friction

Schallamach [20] and Chernyak and Leonov [23] developed their models envisioning purely adhesive sliding. However, Savkoor [24] as well as Ludema and Tabor [25] suggested that even seemingly adhesive sliding could never be purely adhesive. Savkoor [24] proposed a hybrid model, in which the interface consists of discrete patches of asperities of molecular dimensions in adhesive contact with the rubber surface. When a shear force is imposed, the patch stores elastic energy until it overcomes the adhesive energy, causing the propagation of a shear crack. Again, these early models neglected the inherent viscous attributes of elastomers. According to Savkoor [28] as well as Ludema and Tabor [25], sliding may proceed by an activated process, but the extent to which the two surfaces come into contact depends on modulus and sliding velocity. Hidden in more macroscopic terms, these approaches of Savkoor [24] and Ludema and Tabor [25] are similar to the model of Schallamach [20], where the activated nature of the sliding process is still attributed to adhesive debonding.

An activated sliding process does not necessarily identify friction as adhesion dominated. A realistic picture of the contact interface must also consider the viscoelastic nature of the elastomer during the shear process, which from a microscopic perspective involves intermolecular sliding. Intermolecular sliding is an activated relaxation process that can also be described with the Erying model discussed above. Its related local force, i. e., internal friction, is strongly intertwined with the cohesive forces within the bulk elastomer. Considering now both, the intermolecular relaxation in the polymer "bulk" system due to shear, and the adhesive bonding-debonding process, elastomeric friction can be considered a *tribo-rheological* process. If the adhesion force, F_{adh}, exceeds the cohesion force, F_{coh}, the contact undergoes cohesive yield, i. e., "bulk" polymer relaxation properties dictate friction, and *vice versa*. In addition, it has to be pointed out that the two processes possess specific characteristic time scales. Thus, depending on the experimental time window, i. e., the scan velocity, a loss spectrum analysis based on friction-velocity isotherms can be sensitive to either of the two proceses, or both.

When the time scale of the driven surface is much slower than the response time of the material, the contact ages prior to a virtual jump of the driven surface. In this regime, friction may actually increase as velocity decreases because the net adhesion force is proportional to the contact aging time. On the other hand, when the scan velocity is much faster than the interface can respond, the surfaces do not easily slide relative to each other. Instead, the surfaces start peeling locally and detachment waves propagate throughout the entire area of contact, from its advancing to the trailing edge. Schallamach [26] first discovered these waves at high sliding velocities. Roberts and Jackson [27] suggested that when such instabilities occur, the frictional stress between surfaces can be described in terms of the wavelength (Λ) of the Schallamach instability and the adhesion hysteresis (ΔW), i. e. the energetic difference involved in making and breaking interfacial contacts, as $\sigma = \Delta W/\Lambda$. In the region where the experimental time scale is comparable to the characteristic material times, the measured frictional forces are representative of the relevant dissipative mechanisms. However, identifying the dominant friction mechanism requires some characteristic signature for the dissipation process.

In ambient scanning force microscopy (SFM) experiments and computer models, such as molecular dynamic (MD) and Monte Carlo (MC) simulations, molecular scale friction is discussed in terms of mechanical relaxations and internal conformational changes [28–31]. These studies involved highly structured model systems, which were prepared by either self-assembly or Langmuir-Blodgett techniques [32]. Such mono- or multilayered systems provide convenient access for investigations of molecular-scale dissipation mechanisms; e. g. load induced molecular tilts [28], and reversible and irreversible conformation changes [28–30]. Simple frictional models, such as the Tomlinson-Prandtl model [33], could be tested, and the corrugated molecular surface potential compared to the magnitude of discontinuous molecu-

lar stick-slip sliding [29,31]. The jump-distances were found to be stochastic above a critical sliding velocity, which led to a discussion of molecular friction in terms of fluctuations around discrete attractors [31]. This corresponds to recent theoretical treatments, i. e. *creep models* that consider *barrier-hopping fluctuations* of periodic surface potentials with slips occurring at lower energy values than prescribed by the potential barriers [34–36].

Sills and Overney [37] showed that *hindered,* or *frozen*, relaxation states of an amorphous polymer could be activated in the course of a frictional sliding process, and thus, give rise to a barrier-hopping fluctuation not unlike the one observed for highly ordered surfaces. Friction-rate isotherms obtained with a SFM tip on glassy polystyrene could be collapsed to a master curve according to a *ramped* creep model [34,35], Fig. 30.3, which considers a single asperity sliding over a corrugated surface potential that is biased due to the motion of the driven tip, i. e. the probability of a backwards jump is much lower than for a forward jump. The barrier height becomes proportional to a 3/2-power law in the friction force [34, 35], which leads to a distorted power-law friction-velocity relationship of $F = F_c - \Delta F |ln(V^*)|^{2/3}$. Based on Sang et al.'s theory, [35] $V^* \sim V/T$ represents a dimensionless velocity, and $\Delta F \sim T^{2/3}$. F_c is an experimentally determined constant that contains the critical position of the cantilever support. F_c is determined from the intercept of F versus $T^{2/3}$ at a fixed ratio $T/V = 1\,\mathrm{K/(nm/s)}$.

Interestingly, ramped creep scaling is consistent with the solution of the Langevin equation for a perfect cantilever oscillator in the total potential energy E

$$M\ddot{x} + M\beta\dot{x} + \frac{\partial E(x,t)}{\partial x} = \xi(t)\,, \tag{30.7}$$

with thermal noise in the form of the random force, $\xi(t)$, where $\langle \xi(t)\xi(t') \rangle = 2M\beta k_b T \delta(t - t')$, x is the position of the tip on the surface, M is the mass of tip, and β is a linear dampening factor. Equation 30.7 is based

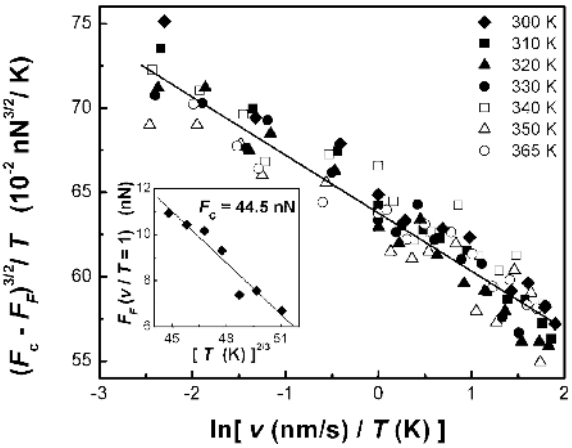


Fig. 30.3. Collapse of poslysyrene friction-rate isotherms to a ramped creep barrier-hopping fluctuation model. (inset) the constant F_c is determined from the intercept of F versus $T^{2/3}$ for a fixed ratio $T/V = 1\,\mathrm{K/(nm/s)}$. Adapted from reference [37]

on a sinusoidal surface potential continuously overcome during the course of frictional sliding, which Sills and Overney experimentally determined as 7 kcal/mol on glassy polystyrene [37]. This activated process corresponds directly with the hindered rotation of the phenyl ring side chains about their bond with the backbone, supporting the contribution of intrinsic molecular relaxation to elastomeric friction. In a similar SFM friction study on poly(methyl methacrylate), Hammerschmidt et al. identified the β relaxation of carboxylate side-chains as the primary dissipation mechanism [38]. In these examples, the effect of adhesion between the silicon tip and polystyrene was negligible. However, it is possible that asperities become coated with elastomer molecules, resulting in a cohesive sliding interface that is chemically indistinct from intrinsic molecular relaxation within the bulk elastomer.

30.4 A Discussion in the Light of Adhesive Stochastic Theory

The above models, all of which offer plausible explanations for interfacial friction, have yet to be rigorously tested experimentally. Decoupling the adhesive and cohesive (viscoelastic) contributions to elastomeric friction requires comprehensive experimental designs to access the energetic signatures of dissipation mechanisms, and the use of model elastomeric networks and counter-surfaces. To maintain a constant adhesion force, the elastomeric contacts would have to be chemically similar, while the internal friction component is varied by confining segmental relaxation through crosslinking [39]. On the other hand, holding the cohesive component constant and varying the adhesion requires counter-surfaces with different surface energies. Both surfaces have to be free of secondary interactions and robust enough to withstand wide ranges of sliding velocity and temperature, without compromising the ideality of the sliding materials.

Model studies of these types have been initiated by several groups. For example, Brown [40] and Casoli et al. [41] examined the pull-out of polymer chains from elastomeric networks and the associated friction. Ghatak et al. [42] studied friction of polydimethylsiloxane (PDMS) on some low energy surfaces as a function of molecular weight of the polymer M and the sliding velocity. They noted that friction decreases with molecular weight. However, it should be pointed out that these studies were limited in the range of molecular weight, and irrespective of the critical molecular weight for entanglements M_e. Vorvolakos and Chaudhury [21] extended these studies using crosslinked hemispheres of PDMS, in sliding contact with two low energy surfaces: a methyl functional self-assembled monolayer (SAM) of hexadecylsiloxane and a thin film of polystyrene (PS), both of which interact with PDMS via dispersion interactions. Using these simple model systems, they carried out the measurements of adhesion and friction to investigate how

Table 30.1. Molecular weight, M, shear modulus, G, and works of adhesion, W, for each PDMS network. The strength of interaction is largely independent of M for $M < M_e \sim 8\,\mathrm{kg/mol}$, but drops significantly for $M > M_e$, which suggests that the degrees of freedom available to molecular interactions are restricted due to entanglements [21]

M (kg/mol)	G ($10^5\,\mathrm{N/m^2}$)	$W_{PDMS-SAM}$ (mJ/m2)	$W_{PDMS-PS}$ (mJ/m2)
1.3	31.4	42	53
1.9	20.3	41	55
2.7	16.0	44	56
4.4	8.8	42	53
8.9	4.3	42	52
18.7	2.6	42	44
52.1	0.6	27	26

the latter depends on surface energy, temperature, velocity, and intercrosslink molecular weight of the elastomer. Contact mechanics measurements in the manner of Johnson, Kendall and Roberts (JKR) yielded the modulus of each network and the works of adhesion when placed in contact with each low-energy surface, as shown in Table 30.1.

Friction against each surface was measured over a velocity range spanning 10^{-7} to $10^{-3}\,\mathrm{m/s}$, comparable to the range studied by Reiter et al. [41]. The friction force F divided by the optically measured contact area A yielded the shear stress σ reported in Fig. 30.4. In all cases, the shear stress increases with velocity up to a certain critical velocity $V_\circ$, then plateaus or drops at higher velocities. For $M < M_e$, the friction peak intensity decreases and broadens with M. The friction above $V_\circ$ reveals an unstable sliding process, Fig. 30.5a, and the contact area shrinks from adhesive (JKR) to Hertzian, Fig. 30.5b.

These observations are qualitatively consistent with the stochastic theory for adhesive dominated friction, discussed above. Consider the inverse relationship between rubber friction and molecular weight in Fig. 30.4, which has similarly been observed in other melts [43, 44] and grafted polymer chains [5, 41]. From a macroscopic perspective, Ludema and Tabor [25] suggested the relationship between the shear stress σ and the areal density ($\Sigma_\circ$) of the contact points as $\sigma = \Sigma_\circ f_\circ$, where $f_\circ$ is the adhesion force of a single polymer chain. This is similar to the prefactor in the Chernyak-Leonov Eq. (30.4) corresponding to the shear stress in the high velocity limit, i. e., where the detachment of the polymer chain from the surface is not stochastic. Within the simple model of Chernyak and Leonov [23], the areal density of the load-bearing chains is $1/Na^2$, where N is the number of statistical segments, each of size a. Thus, the shear stress is proportional to the shear modulus as $\sigma = G\frac{f_\circ a}{kT}$. The areal chain density of a real elastomer; however, scales as $N^{-1/2}$ and the shear stress should follow $\sigma \sim G^{1/2}$. Experimen-

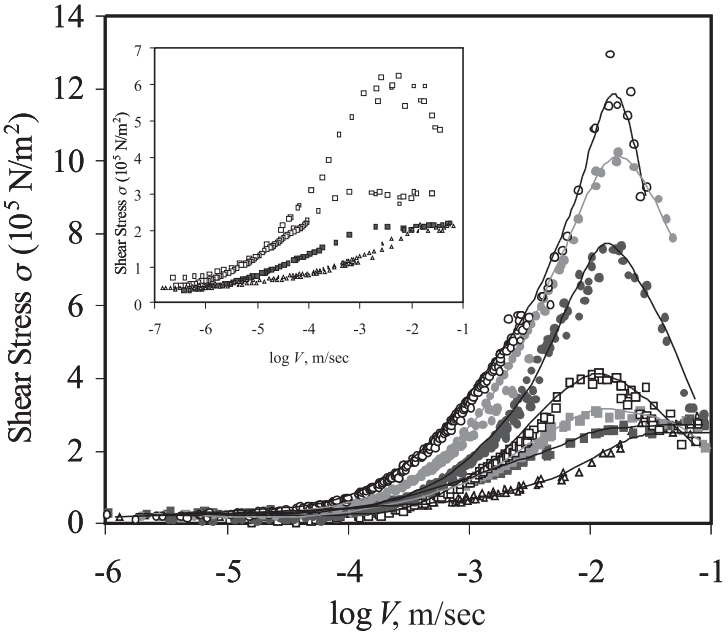


Fig. 30.4. Shear stress as a function of velocity between PDMS and the SAM and (*inset*) PDMS and PS. ○, ●, ●, □, ■, ■ and Δ represent networks with oligomeric precursors of 1.3, 1.8, 2.7, 4.4, 8.9, 18.7, and 52.1 kg/mol, respectively. ($M_e \sim 8$ kg/mol). Adapted from reference [21]

tally, Vorvolakos and Chaudhury observed $\sigma \sim G^{3/4}$. While Grosch [18] did not systematically consider the modulus, he noted that the shear stress is considerably smaller than that expected of two surfaces in true molecular contact and estimated the true contact area at roughly 10% of the apparent area during sliding [18]. More recent SFM studies suggest that a gross mismatch of interfacial contact is not expected based on roughness considerations. Yet, it is plausible that spontaneous roughening of the interface occurs as a result of elastic instability, which ensues from the competition between van der Waals and elastic forces within the first layer of stretched PDMS chains in contact with the surface. If we consider that the dominant wavelength of such roughening scales with the thickness (δ) of the first layer of the polymer chain, then the density of the load bearing sites should scale as $1/\delta^2$ (or $1/Na^2$). In the limit of Rouse dynamics, if one polymer chain remains active in each of the load supporting junctions, one recovers: $\sigma \sim G$. The shear stress should decrease with the molecular weight owing to the fact that the number of load bearing polymer chains decreases with molecular weight. However, above the entanglement molecular weight, entanglement effects dominate the chain dynamics and $\sigma_{\max}$ becomes nearly independent of molecular weight.

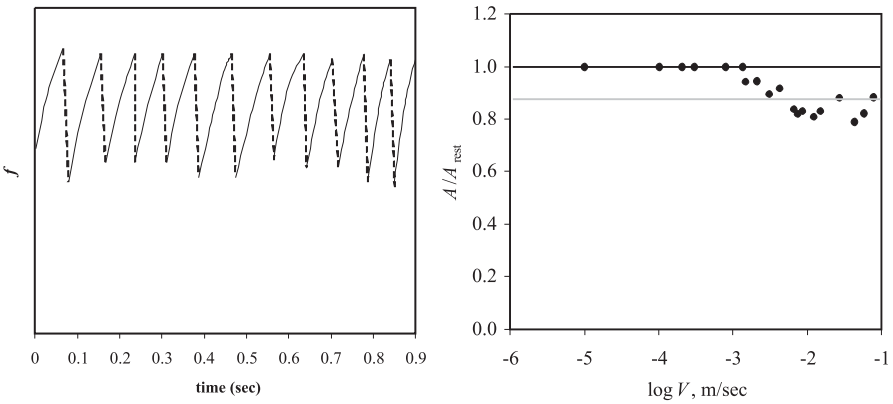


Fig. 30.5. a A typical friction force trace for a PDMS hemisphere sliding across a methyl terminated SAM ($M = 4.4\,\mathrm{kg/mol}$, $V = 2\,\mathrm{cm/s}$). Stick-slip sliding initiates at the critical velocity $V_\circ$ corresponding to the shear stress peak in Fig. 30.4. **b** Change in the sliding contact area A as a function of velocity for the PDMS /SAM interface, normalized by the contact area at rest A_{rest} ($M = 2.7\,\mathrm{kg/mol}$, $R = 2.5\,\mathrm{mm}$, $E = 4.8\,\mathrm{MPa}$, $W = 42\,\mathrm{mJ/mol}$). As the sliding velocity increases, the contact area drops from the JKR prediction (*black line*) to the purely Hertzian prediction (*gray line*). The normal load F_{N} was constant at 48 mN (see Eq. 30.7). Adapted from reference [21]

30.5 A Discussion in the Light of Molecular Relaxation within the Elastomer

It turns out that by utilizing a molecular description, a cohesive process for frictional dissipation can be argued in favor of an adhesive one. Namely, as the molecular weight is increased (still below M_{e}) segmental relaxation times increase, indicating that deeper wells in the potential energy landscape restrict the degrees of freedom for molecular motion. For a given sliding velocity (or contact time), the molecules are less effective in dissipating energy internally through relaxation, and in accommodating the interface adhesively. Hence, friction and shear stress can be expected to decrease. Above M_{e}, the segmental relaxation time is essentially invariant with M, thus F and $\sigma_{\max}$ are independent of M. It becomes apparent that the stochastic theory for adhesive dominated friction, applies almost in direct analogy to cohesive friction, by simply redefining $f_\circ$ as the internal friction force per molecule stretched in the contact (as shown in Fig. 30.1). The same argument applies for the molecular weight, aerial chain density, modulus, and shear stress. Thus, again we are faced with a tribo-rheological complexity, where a distinction between adhesive and cohesive contributions to elastomeric friction require some characteristic signatures of the dissipation mechanism, e. g. the activation energy, relaxation time, or length scale of the process.

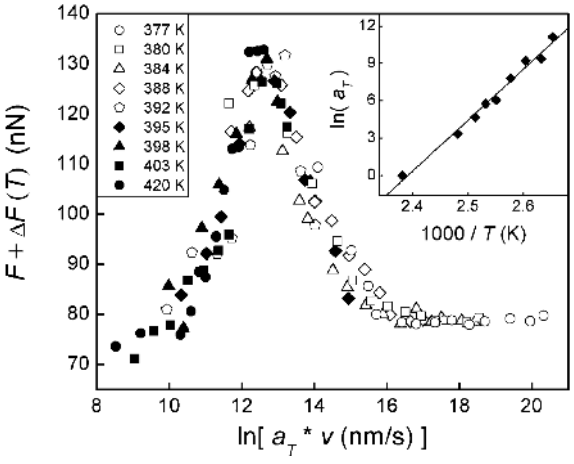


Fig. 30.6. Friction force-velocity, $F(v)$ isotherms for polystyrene ($M = 96.5$ kg/mol) above $T_g = 373$ K, superposed using the method of reduced variables with a reference temperature of 420 K. Inset: From the Arrhenius behavior of a_T, an average activation energy, E_A, of 81 kcal/mol identifies the α-relaxation as responsible for frictional dissipation. Adapted from reference [45]

The characteristic signatures of the dissipation process can be determined by the superposition of friction-rate isotherms and from the critical velocity corresponding to the maximum in the friction force [18, 21, 45]. This is illustrated in Fig. 30.6 with superposed friction-velocity isotherms obtained with SFM measurements on a polystyrene melt [45]. An activation barrier of 81 kcal/mol (3.5 eV) is deduced from the *apparent* Arrhenius behavior of the thermal a_T shift factor in the inset Fig. 30.6. The value coincides with the 80 – 90 kcal/mol activation energy for the α-relaxation process [46], i. e. the segmental relaxation of the PS backbone. In this case, the activation barrier overcome during the course of frictional sliding corresponds directly with the molecular relaxation within the bulk elastomer.

The friction peak is recognized as an analogue to a spectroscopic peak in the frequency space [18, 45]. It reflects the competition between material and experimental time-scales. At low sliding velocities, the contact stress stored in the soft material is capable of relaxing (through internal friction modes) before an asperity can slip to the next contact site. In this region, the friction force increases logarithmically with velocity, which is consistent to an activated molecular relaxation process in the soft material. Above the critical velocity, the probing tip is driven to the next contact site before the material can respond internally through viscoelastic relaxation. Thus with increasing velocity, the tip experiences fewer and fewer dissipative relaxation events per jump, and consequently, the friction force decreases. Hence, molecular scale dynamics can be deduced from local friction measurements.

A nanoscopic description of polymer dynamics involves, in general, only two parameters: an internal, or monomeric, friction coefficient and an appropriate macromolecular length scale [47]. Internal friction dictates the degree of local segmental motion, and thus, is responsible for the bulk viscoelastic and relaxation properties of polymeric materials. Grosch [18] and Ludema and Tabor [48] were the first to combine the velocity at the friction peak with

the frequency for the maximum viscoelastic loss to deduce a characteristic dissipative length scale. However, with their macroscopic techniques, they could only suspect dissipation through segmental relaxation. With the single asperity SFM approach to elastomeric friction, Sills, Gray and Overney [45], identified the specific molecular relaxation process activated in frictional sliding on an entangled polystyrene melt. With the velocity at the friction peak $V_\circ(T)$ in Fig. 30.7a and knowledge of the associated relaxation times $\tau_\alpha(T)$ from dielecrtic spectroscopy, they directly determined the dissipative length scale $X_\mathrm{d}(T)$ of the α-relaxation in a polystyrene melt via:

$$X_\mathrm{d}(T) = V_\circ(T) \cdot \tau_\alpha(T) \,. \tag{30.8}$$

The characteristic length of the dissipation process was found to grow from the segmental scale to 2.1 nm, following a power law behavior that is consistent with predictions for cooperative motion during the α-relaxation, Fig. 30.7b. However, in the vicinity of T_g, dissipation lengths of several tens of nanometers deviate from the above power law behavior, suggesting that long-range processes, i. e. Fischer modes, may couple with the α-relaxation with an apparent non-ergodic time-averaged behavior.

In the inset of Fig. 30.7a, the frictionpeak drops sharply as the melt is cooled toward T_g, indicating that the number of α-relaxation events per jump decreases with the onset of vitrification. Considering, as discussed above, that frictional dissipation on a glassy polystyrene surface is dominated by the lower energy side-chain relaxation, the attenuation of the α-friction-peak intensity in the crossover regime ($T_\mathrm{g} < T < T_\mathrm{c}$) reflects a structurally and dynamically heterogeneous phase, with a spatial distribution of dissipation through independent backbone and side-chain relaxations. One would then expect a simultaneous increase in a β-friction peak at a velocity corresponding to the time scale for side chain rotation. Unfortunately, the frequency (i. e. velocity) bandwidth of the SFM is not sufficient to capture the fast rotational peak of the phenyl rings, which would be expected at a velocity on the order of centimeters per second.

30.6 Dissipation due to Decoupling of Adhesive Bonds versus Segmental Relaxations

The work of Vorvolakos and Chaudhury [21] specifically targeted a decoupling of the adhesive and cohesive components of elastomeric friction. The internal friction associated with the elastomer (PDMS hemispheres) was varied through the crosslink density (i. e. molecular weight between crosslinks M_c), and the interfacial adhesion was catered via the surface energy of the counter-faces: a methyl terminated hexadecylsiloxane self-assembled monolayer (SAM) and a thin film of glassy polystyrene (PS). While both surfaces are essentially non-polar, the critical velocity at the friction peak $V_\circ$ differs by

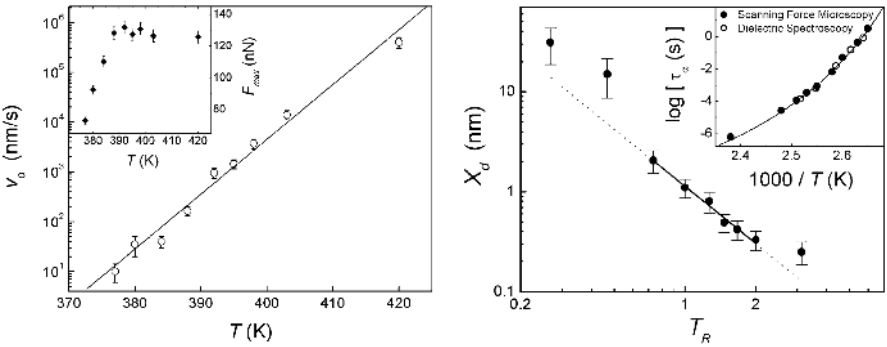


Fig. 30.7. a the velocity v_o corresponding to the friction peak of polystyrene ($M = 96.5$ kg/mol, $M_e \sim 20$ kg/mol) in Fig. 30.6. (*inset*) The corresponding friction peak intensities F_{max}. **b** Dissipation length X_d for the α-relaxation of polystyrene in terms of the reduced temperature, $T_R = (T - T_g)/(T_c - T_g)$ where T_g is the calorimetric glass transition temperature (373 K) and T_c is the crossover temperature of the dynamic glass transition (388 K) identified at the saturation point of F_{max} in the inset of **a**. The power law fit $X_d \sim T_R^{-\phi}$ over the range $0.7 < T_R < 2.0$ (*solid line*) reveals an exponent of $\phi = 1.89 \pm 0.08$. (*inset*) polystyrene α-relaxation times from dielectric spectroscopy [49] (*open circles*) and the corresponding times from $V_\circ$ and X_d and equation 8 (*closed circles*). Adapted from reference [45]

almost an order of magnitude between the SAM and PS surfaces (Fig. 30.4). On the other hand, $V_\circ$ is independent of the intercrosslink molecular weight, above and below the entanglement weight. Initially, a strong adhesive contribution is suspected; however, a shift in $V_\circ$ may be attributed to a change in the timescale of the dissipation process, a change in the lengthscale over which dissipation occurs, or both in the event of a completely different dissipation mechanism.

For adhesive dominated friction, the critical velocity, according to Chernyak and Leonov, appears at $V_\circ = \frac{\delta \cot \chi}{\tau_\circ}$. As $\delta \sim N^{1/2}a$ and $\cot \chi \sim \frac{f_\circ \delta}{kT}$, $V_\circ \sim \frac{f_\circ N a^2}{kT\tau_\circ}$. It was argued by Chernyak and Leonov that the life time $\tau_\circ$ of the bound state of the polymer chain at rest is related to the characteristic time of τ segmental motion as $\tau_\circ = \tau(R_F/a)^3$, which results in the molecular velocity $V_\circ$ decreasing with N following a $1/2$ power law. However, the experimental data of Vorvolakos and Chaudhury [21] indicate that $V_\circ$is nearly independent of molecular weight for PDMS ($M \alpha N$), thus suggesting that $\tau_\circ$ varies linearly with N, not as $N^{3/2}$.

If dissipation through segmental relaxation is considered, the segmental length of PDMS (0.6 nm) combined with the friction peak velocities in Fig. 30.4, gives characteristic times of 10^{-7} and 10^{-8} s, for the SAM and PS surfaces, respectively. These times are considerably larger than the viscous relaxation time (10^{-11} s) of dimethylsiloxane monomer [50]; however, the relaxation dynamics in crosslinked PDMS occur on different time scales. The *fast* mode that follows from the Rouse-like dynamics of chain segments between

entanglements occurs on a time scale of 10^{-9} s at room temperature [51]. On the other hand, the *slow* mode translational dynamics of cross-link junctions through the polymer network occurs at 10^{-7} s, independent of molecular weight [51]. The correspondence of the PDMS friction peak velocity with the slow mode dynamics suggests dissipation through crosslink motion. Interestingly, junction fluctuation theories predict that the mobility of crosslink junctions is of crucial importance in determining the modulus of the network [51].

The dissipation picture is still not complete. The shift in the friction peak velocity due to the surface energy of the opposing SAM or PS surfaces implies a significant adhesive contribution. By the superposition of friction-rate isotherms ($T > T_g = -130\,^{\circ}\mathrm{C}$), Vorvolakos and Chaudhury [21] were able to estimate Arrhenius activation energies of 25 and 27 kJ/mol for dissipation on the SAM and PS surfaces, respectively. While the activation energies did not differ substantially, the pre-exponential factors varied by one order of magnitude, from 10^{-12} to 10^{-11} for the SAM and PS surfaces, respectively. The activation energies are considerably higher than that for the bulk viscosity of PDMS (16 kJ/mol), which suggest that the potential barrier overcome during the course of frictional sliding may be modified by the adhesive interactions. However, fluorescent probe reorientation measurements in PDMS have revealed activation energies of 20 – 28 kJ/mol for both the fast and slow relaxation modes [52]. The barrier enhancement relative to the bulk viscosity was attributed to the unique nature of the silicon-oxygen backbone, which results in a local mobility that is quite distinct from the long-range chain dynamics. Thus, the observed barrier height during frictional sliding is sufficient to activate viscoelastic relaxation in the PDMS elastomer, presumably the slow mode associated with crosslink mobility. For this case, the adhesive contribution appears, not strictly in the barrier height, but in the barrier jump attempt frequency, i. e. in the exponential pre-factor in the Arrhenius equation. This suggests that the intrinsic chemical interactions at the contact interface dictate the lifetime of the bound state at rest, i. e. $\tau_\circ$ in Eq. 30.2, for adhesive debonding.

30.7 Outlook and Closing Remarks

We have discussed the molecular origins of elastomeric friction in the light of two different processes: interfacial adhesion and internal friction. Various studies have highlighted the importance of both processes; however, a unified picture that accurately accounts for both has yet to evolve. On the molecular scale bulk rheological constructs like viscosity and modulus are insufficient to capture the interfacial molecular dynamics. Both processes are rate dependent, and both can be described by Eyring's activation model. While most studies have focused on one or the other aspect, it is unlikely that the two processes are mutually independent. A more realistic picture considers the competition between the two. Resolution of the debate is likely to unfold

by considering the energetic signatures of each process. We should ask what potential barrier is higher, adhesive bonding-debonding, or internal molecular relaxation of the stressed elastomer. If interfacial adhesion dominates internal cohesion, sliding asperities are apt to be coated with a molecular layer of the elastomer, forming a sliding plane within the bulk elastomer and yielding frictional properties related to the bulk material properties. On the other hand, if internal cohesion dominates interfacial adhesion, like in many rigid materials within limits of plastic yield, the frictional outcome is likely to reflect the interfacial interactions.

As we move into the new millennium, precise understanding of the frictional process is necessary for the development of advanced materials. A refined understanding of molecular friction is critical to emerging nanotechnologies, e. g., nano-electromechanical systems (NEMS) [39,53] and biomaterials [54–56]. On this scale, *all* transport processes depend on the degrees of freedom available to molecular motion.

References

1. R. Ariano, *The Coefficients of Friction Between Rubber & Various Materials. Part II – Gripping Friction of Rubber Belting*, India Rubber Journal **79(2)**, 56–8 (1930).
2. J.B. Derieux, *The Coefficients of Friction Between Rubber & Various Materials. Part II – Gripping Friction of Rubber Belting*, J. Elisha Mitchell Scientific Society **50**, 53–5 (1934).
3. T.R. Dawson and B.D. Porritt, *Rubber: Physical and Chemical Properties*, 381–386 (1935).
4. B.Z. Newby, M.K. Chaudhury, and H.R. Brown, *Macroscopic Evidence of the Effect of Interfacial Slippage on Adhesion*, Science **269**, 1407–8 (1995).
5. B. Newby and M.K. Chaudhury, *Friction in Adhesion*, Langmuir **14**, 4865–90 (1998).
6. S.L. Aggarwal, I.G. Hargis, R.A. Livigni, H.J. Fabris, and L.F. Marker, in *Advances in Elastomers and Rubber Elasticity* (Plenum, New York, 1986).
7. A.N. Theodore, M.A. Samus, and P.C. Killgoar, *Environmentally Durable Elastomer Materials for Windshield Wiper Blades*, Ind. Eng. Chem. Res. **31 (12)**, 2759–64 (1992).
8. C.W. Extrand, A.N. Gent, and S.Y. Kaang, *Friction of a rubber wedge sliding on glass*, Rubber Chemistry and Technology **64 (1)**, 108–17 (1991).
9. F.L. Roth, R.L. Driscoll, and W.L. Holt, *Frictional Properties of Rubber*, J. Research of the National Bureau of Standards **28(4)**, 439–62 (1942).
10. P. Thirion, *Les coefficients d' adherence du caoutchouc.*, Revue Generale du Caoutchouc **23(5)**, 101–6 (1946).
11. P.J. Papenhuyzen, *Wrijvingsproeven in verband met het slippen van autobanden*, De Ingenieur **V75**, 53 (1938).
12. A. Schallamach, *The Velocity and Temperature Dependence of Rubber Friction*, Proceedings of the Physical Society B **65**, 657–61 (1952).
13. A. Schallamach, *The load dependence of rubber friction*, Proceedings of the Physical Society B **66**, 386–92 (1952).

14. F.P. Bowden and D. Tabor, *The Friction and Lubrication of Solids* (Oxford University Press, 1950).
15. H. Eyring, *Viscosity, plasticity, and diffusion as examples of absolute reaction rates*, J. Chemical Physics **4**, 283–91 (1936).
16. J.A. Greenwood and D. Tabor, *The Friction of Hard Sliders on Lubricated Rubber: The Importance of Deformation Losses*, Proceedings of the Physical Society **71**, 989–1001 (1958).
17. A.M. Bueche and D.G. Flom, *Surface friction and dynamic mechanical properties of polymers*, Wear **2 (3)**, 168–82 (1959).
18. K.A. Grosch, *The relation between the friction and visco-elastic properties of rubber*, Proceedings of the Royal Society A **274**, 21–39 (1963).
19. M.L. Williams, R.F. Landel, and J.D. Ferry, *The Temperature Dependence of Relaxation Mechanisms in Amorphous Polymers and other Glass-forming Liquids*, J. Am. Chem. Soc. **77**, 3701–7 (1955).
20. A. Schallamach, *A theory of dynamic rubber friction*, Wear **6**, 375–82 (1963).
21. K. Vorvolakos and M.K. Chaudhury, *The effects of molecular weight and temperature on the kinetic friction of silicone rubbers*, Langmuir **19**, 6778–87 (2003).
22. T. Kontorova and Y.I. Frenkel, *On the theory of plastic deformation and twinning*, Soviet Phys. JEPT **13**, 1–10 (1938).
23. Y.B. Chernyak and A.I. Leonov, *On the Theory of Adhesive Friction of Elastomers*, Wear **108**, 105–38 (1986).
24. A.R. Savkoor, *On the friction of rubber*, Wear **8**, 222–37 (1965).
25. K.C. Ludema and D. Tabor, *The friction and visco-elastic properties of polymeric solids*, Wear **9**, 329–48 (1966).
26. A. Schallamach, *How does rubber slide?*, Wear **17**, 301–12 (1971).
27. A.D. Roberts and S.A. Jackson, *Sliding friction of rubber*, Nature **257**, 118–20 (1975).
28. J.I. Siepmann and I.R. McDonald, *Monte Carlo Simulation of the Mechanical Rexation of a Self-Assembled Monolayer*, Phys. Rev. Lett. **70**, 453–6 (1993).
29. T. Bonner and A. Baratoff, *Molecular dynamics study of scanning force microscopy on self-assembled monolayers*, Surf. Sci. **377–379**, 1082–6 (1997).
30. P.T. Mikulski and J.A. Harrison, *Packing-density effects on the friction of n-alkane monolayers*, J. Am. Chem. Soc. **123**, 6873–81 (2001).
31. R.M. Overney, H. Takano, M. Fujihira, W. Paulus, and H. Ringsdorf, *Anisotropy in friction and molecular stick-slip motion*, Phys. Rev. Lett. **72**, 3546–49 (1994).
32. A. Ulman, *An Introduction to Ultrathin Organic Films* (Academic Press, New York, 1991).
33. G.A. Tomlinson, Philos. Mag. Ser. **7**, 905 (1929).
34. O.K. Dudko, A.E. Filippov, J. Klafter, and M. Urbakh, *Dynamic force spectroscopy: a Fokker-Planck approach*, Chem. Phys. Lett. **352**, 499–504 (2002).
35. Y. Sang, M. Dube, and M. Grant, *Thermal Effects on Atomic Friction*, Phys. Rev. Lett. **87**, 174301/1–4 (2001).
36. F. Heslot, T. Baumberger, B. Perrin, B. Caroli, and C. Caroli, *Creep, stick-slip, and dry-friction dynamics: experiments and a heuristic model*, Phys. Rev. E **49**, 4973–88 (1994).
37. S. Sills and R.M. Overney, *Creeping friction dynamics and molecular dissipation mechanisms in glassy polymers*, Phys. Rev. Lett. **91**, 095501(1–4) (2003).

38. J.A. Hammerschmidt, W.L. Gladfelter, and G. Haugstad, *Probing polymer viscoelastic relaxations with temperature-controlled friction force microscopy*, Macromolecules **32**, 3360–7 (1999).
39. B. Gotsmann, S. Sills, U. Duerig, J. Frommer, and C. Hawker, *Controlling nanowear in a polymer by confining segmental relaxation*, NanoLett. (2006).
40. H.R. Brown, *Chain Pullout and Mobility Effects in Friction and Lubrication*, Science **263**, 1411–3 (1994).
41. A. Casoli, M. Brendle, J. Schultz, A. Philippe, and G. Reiter, *Friction Induced by Grafted Polymeric Chains*, Langmuir **17 (2)**, 388–98 (2001).
42. A. Ghatak, K. Vorvolakos, H. She, D. Malotky, and M.K. Chaudhury, *Interfacial Rate Processes in Adhesion and Friction*, J. Physical Chemistry B **104 (17)**, 4018–30 (2000).
43. Y. Inn and S.-Q. Wang, *Hydrodynamic slip: Polymer adsorption and desorption at melt/solid interfaces*, Physical Review Letters **76 (3)**, 467–70 (1996).
44. S. Hirz, A. Subbotin, C. Frank, and G. Hadziioannou, *Static and kinetic friction of strongly confined polymer films under shear*, Macromolecules **29 (11)**, 3970–4 (1996).
45. S. Sills, T. Gray., and R.M. Overney, *Molecular dissipation phenomena of nanoscopic friction in the heterogeneous relaxation regime of a glass former*, J. Chem. Phys. **123**, 134902 (2005).
46. G.D. Patterson, C.P. Lindsey, and J.R. Stevens, *Depolarized Rayleigh spectroscopy of polystyrene near the glass-rubber transition*, J. Chem. Phys. **70**, 643–5 (1979).
47. V.N. Pokrovskii, *Mesoscopic theory of polymer dynamics* (Kluwer Academic Publishers, Dordrecht, 2000).
48. K.C. Ludema and D. Tabor, *The friction and visco-elastic properties of polymer solids*, Wear **9**, 329–48 (1966).
49. C.M. Roland and R. Casalini, *Temperature dependence of local segmental motion in polystyrene and its variation with molecular weight*, J. Chem. Phys. **119**, 1838–42 (2003).
50. M. Appel and G. Fleischer, *Investigation of the chain length dependence of self-diffusion of poly(dimethylsiloxane) and poly(ethylene oxide) in the melt with pulsed field gradient NMR*, Macromolecules **26**, 5520–5 (1993).
51. P.B. Leezenburg, M.D. Frayer, and C.W. Frank, *Photophysical studies of probes bound to crosslinked junctions in poly(dimethyl silozane) elastomers and nanocomposites*, Pure & Appl. Chem. **68**, 1381–8 (1996).
52. A.D. Stein, D.A. Hoffmann, A.H. Marcus, P.B. Leezenburg, and C.W. Frank, *Dynamics in Poly(dimethylsiloxane) Melts: Flouresence Depolarization Measurements of Probe Chromophore Orientational Relaxation*, J. Phys. Chem **96**, 5255–63 (1992).
53. S.E. Sills, R.M. Overney, B. Gotsmann, and J. Frommer, *Strain shielding and confined plasticity in thin polymer films: Impacts on thermomechanical data storage*, Tribo. Lett. **19**, 9–15 (2005).
54. H.B. Dong, T. Blawert, C. Mordike, B. L., *Plasma immersion ion implantation of UHMWPE*, J. Mater. Sci. Lett. **19 (13)**, 1147–9 (2000).
55. J. Wang, E. Stroup, X. Wang, and A. J. D., *Study of PEO on LTI Carbon Surfaces by Ellipsometry and Tribometry*, Proc. SPIE-Int. Soc. Opt. Eng. Int. Conf. Thin Film Phys. Appl.**Pt. 2**, 835–41 (1991).
56. T. Murayama and C.R. McMillin, *Dynamic mechanical properties of elastomers for use in circulatory assist devices*, J. Appl. Polym. Sci. **28 (6)**, 1871–7 (1983).

31 Nanotribological Perspectives in Tissue Engineering

Mario D'Acunto[1], Gianluca Ciardelli[2], Alfonsina Rechichi[3], and Franco Maria Montevecchi[2], and Paolo Giusti[1,3]

[1] Department of Chemical Engineering and Materials Science, University of Pisa, via Diotisalvi 2, 56100 Pisa, Italy
[2] Department of Mechanical Engineering, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129, Torino, Italy
[3] CNR, Institute of Composite and Biocompatible Materials, via Diotisalvi 2, Pisa, Italy m.dacunto@ing.unipi.it

31.1 Introduction

Successful strategies for complete regeneration of adult human tissues are the main objective of the field of regenerative medicine. This field has been labeled over the past 20 years as tissue engineering [1–5]. A damaged tissue is regenerated by the migration of cells to the injured area, induced by the release of chemotactic signals from the damaged tissue. Tissue regeneration and healing is eventually accomplished by a combination of cell proliferation and cell-matrix synthesis to form the regenerated tissue [5, 6]. Full achievement of this target may not be possible for many decades, although a number of related goals could be reached earlier. One goal, currently progressively achieved, has been the development of biomaterials able to offer ideal interfaces with cells and tissues. Again, the knowledge of how cells dynamically interact with substrates and the subsequent proliferations and differentiations are other fundamental objectives. The ultimate goal of regenerative medicine will require extracellular matrix engineering, a task that necessitates more learning about self-assembly, supramolecular chemistry and physics, biomimetic material properties and, as far as contacts between living systems are concerned, particular features of tribology on the typical molecular or supramolecular scale of protein and molecular motors should be addressed.

As the basic unit of life, cells are complex biological systems. Cells must express genetic information to perform their specialized functions: synthesize, modify, sort, store and transport biomolecules, convert different forms of energy, transduce signals, maintain internal structures and respond to external environments. All of these processes involve mechanical, chemical and physical processes. Mechanical forces play a fundamental role in cell migration, where contractile forces are generated within the cell and pull the cell body forward. In protein secretion for example, protein molecules are packaged in vesicles and transported to the cell membrane by means of a molecular motor running along filaments and cells. On the other side, mechanical forces and

deformations induce biological response in cells, and many normal and diseased conditions of cells are dependent upon or regulated by their mechanical environment [7]. The effects of applied forces depend on the type of cells and how the forces are applied to, transmitted into, and distributed within cells.

A cell is formed by a cytoskeleton wrapped by the plasma membrane and trapping inside a nucleus surrounded by the soup-like cytoplasm, Fig. 31.1a. The cytoskeleton is a system of protein filaments – microtubuli, actin filaments and intermediate filaments – that give to the cell shape and the capability for directed movements. Cells are attached to the extracellular matrix (ECM), a complex network of polysaccharides and proteins secreted by the cells – that serves as a structural element in tissues. Proteins in the ECM, including collagen, elastin, fibronectin, vitronectin and laminin, also

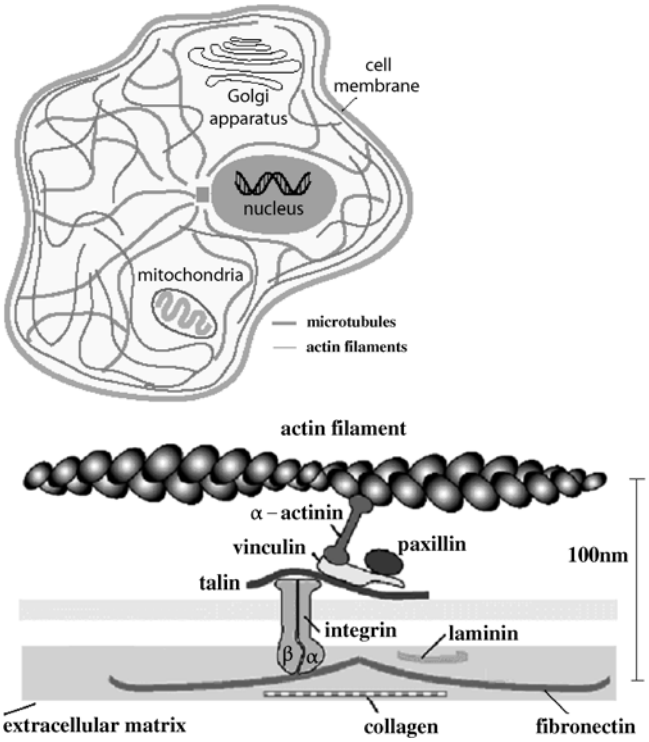


Fig. 31.1. a An animal cell is a structure with an average size of 10 – 40 μm, composed of cytoplasm including nucleus, mitochondria, Golgi apparatus, and other organelles, sustained by the cytoskeleton, and wrapped up by the cytoplasmatic membrane. The cell cytoskeleton consists of microtubules (25 nm diameter), actin filaments (8 nm diameter), intermediate filaments (13 nm diameter) and other binding proteins. **b** The cells adhere to the extracellular matrix (ECM) via focal adhesion (FA) complexes consisting of integrin and other binding molecules (adapted from [7], Copyright (2002), with permission from Elsevier)

play a regulatory role of cellular function through binding to various receptor proteins, found on the cell surface. Some of these receptors are members of the integrin family of transmembrane proteins. They are composed of two units, α and β, and are expressed on the membrane of a wide variety of cells. When attached to the cell cytoskeleton, integrins are critical to the mechanical stability of cell adhesion to ECM and to other cells. Integrins also serve as biochemical signaling molecules in normal and diseased states of cells, and are involved in regulating cytoskeletal organization.

Tribological properties such as adhesion and friction play a fundamental role in cell motion, proliferation and omeostasis, both in the interaction with the external environment or in cell–cell interactions. For example, epithelial cells that form tissues that cover the internal and external surfaces of organs, such as skin cells, lining of the lungs, intestines, must adhere to substrates under a wide variety of conditions. Their adhesion properties can be regulated by the cell – or the system of cells – which simultaneously senses the chemical and mechanical properties of their environment. While these tribological properties of biological cells must ultimately be described by the known laws regulating tribological systems, there are important differences between conventional inorganic ("dead") materials and wet living matter. A complex combination of shear forces and adhesive features is the key for understanding the ability of cell to proliferate, the phenotype characterization, and the ultimate ability to regenerate a tissue. Recent experiments show that, contrary to artificial vesicles that exert only normal forces when adhering to a substrate, adhering cells show both normal and lateral forces [8–10]. The normal forces arise from the action of either specific adhesion molecules or van der Waals interactions or macromolecular adsorption, while the lateral force arise from elastic deformations of the adhesion region by cytoskeletal forces. These lateral forces regulate the size and shape of the finite-sized, discrete adhesion regions (focal adhesion, FA) and allow a cell to probe and to adjust the strength of the adhesion to its physical environment. For mechanically active cells like fibroblasts, there could be hundreds of FAs, the forces associated with the sites distributed along the cell rim keep the cell under tension. The forces that arise from the tension in the actin cytoskeleton tend to polarize the tense actin filaments. It is possible to sum over all the focal FA and model such adhering cells as a pair of nearly equal and oppositely directed contraction forces (elastic force dipoles) with typical forces of 100 nN over a scale of tens of micrometers.

This chapter is organized as follows: after the present introduction (Sect. 31.1), in Sect. 31.2, some features of cell motility, including molecular motors will be discussed. In Sect. 31.3, the adhesion forces at the cell-extracellular matrix are analyzed and some related experiments for the characterization of biomimetic scaffold surface involving cell proliferation are reviewed. Sect. 31.4 focuses the attention on experimental strategies for the measurements of cell–ECM adhesion force.

31.2 Fundamental Features of Cell Motility

The phenomena involved in the movement of cells and of their internal components are the subject of a great deal of investigations, due to the intrinsic interest of the processes and because of the medical importance not limited only in TE [11]. Most cancers, for example, are not life-threatening until they metastasize and spread throughout the body. Metastasis occurs when previously sessile cells in a tumor acquire the ability to move and invade nearby tissues and circulate in the bloodstream or lymphatic system. A treatment that could impede the ability of tumor cells to acquire motility would largely prevent metastasis. The field of nanotribology offers a growing interdisciplinary area in which a detailed knowledge of the mechanisms of cell motility might prove useful. Cells have spent several billion years developing highly efficient machinery to generate forces in the piconewton-to-nanonewton range that operate over distances of nanometers to micrometers and function in aqueous environments. The better we understand the mechanics of cell motility, the more we will be able to adapt the cell movement machinery for medical treatments and, in general, for bioartificial systems.

A wide variety of cell movements has been characterized by biologists and biophysicists in relation to different degrees of molecular and mechanical features [13]. Movements of whole cells can be roughly divided into two functional categories: swimming, when the movement is through liquid water, and crawling, when the movement is across a rigid surface. Since viscous forces are many orders of magnitude greater than inertial forces at the speeds, viscosities and length scales experienced by swimming cells, for bacterial cells, the rotation of a helical or corkscrew-shaped flagellum in bacterial cells has been particularly well characterized in the biophysics field [14]. The flagellum is a long filament constructed by the noncovalent polymerization of hundreds of identical protein subunits, called flagellin. The speeds of flagellar swim-

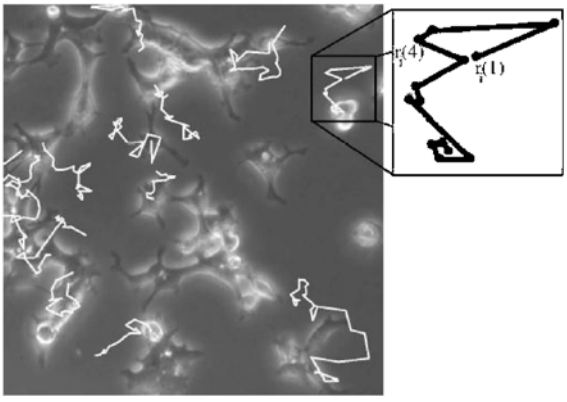


Fig. 31.2. Sketch of cell trajectories superposed on the initial frame. The inset shows 13 subsequent positions along the trajectory of the cell examined. The mean square displacement $\langle x^2(t)\rangle$ varies as Dt^{α}, where D is an effective diffusivity, α is an exponent that indicates normal diffusion, like a random walk ($\alpha = 1$), or anomalous diffusion ($\alpha \neq 1$). (Adapted from [12])

ming range from about 10 or 100 micrometers for second. Some unicellular eukaryotes swim by gradually changing the contour of their surface. This movement, called metaboly [15], shows that the reshifting of cellular content to generate a local increase in drag that propagates from the front of the cell to the back, is sufficient to pull the cell center of mass forward through the viscous aqueous environment.

The best characterized movement of cells across a rigid surface is the crawling motility, or amoeboid motility. It is a general process where a cell attached to a rigid substrate extends forward a projection at its leading edge that then attaches to the substrate. Long thin projections are called filopodia; flat veil-shaped projections, lamellipodia; and thick knobby projections, pseudopodia. All three types of projections are filled with assemblies of cytoskeletal actin filaments. Some difference can be observed in the cytoskeleton strategy for movement: lamellipodia are often associated to continuous advancing due to a rolling mechanism, while filopodia and pseudopodia are characterized by a protrusion, sticking and pulling strategy. After protrusion and attachment, the crawling cell then contracts to move the cell body forward, and movement continues as a threadmilling cycle of front protusion and rear retraction. The speed of such amoeboid movement can range from less than one micrometer per hour to more than one micrometer per second, depending on the cell type and its degree of stimulation. Several kinds of movements in which cells slide across a rigid substrate are known as gliding movements. Gliding appears in some bacteria to be driven by a low Reynolds number analog of jet propulsion, in which a sticky and cohesive slime is extruded backwards to push the cell forward [16].

31.2.1 The Role of Molecular Motors and Friction Force

Molecular motors denote a biological mechanism that converts chemical energy into mechanical energy, used by a cell to generate directed motion [11,17–19]. Cytoskeletal motors bind to the filaments of the cytoskeleton and then walk along these filaments in a directed fashion. This class of motors is essential for intracellular transport, cell division and cell locomotion. Cells generally store chemical energy in two forms: high-energy chemical bonds, such as the phosphoanhydride bonds in ATP (adenosine triphosphate); and asymmetric ion gradients across membranes, such as the electrical potential seen in nerve cells. These sources of chemical energy drive all cell processes, from metabolism through to DNA replication. The subset of cell proteins and macromolecular complexes that convert chemical energy into mechanical forces are generally called molecular motors. Their wide variety reflects the diversity of cell movements necessary to life. Known biological molecular motors may be divided into three principal groups: (i) rotatory motors; (ii) linear stepper motors; (iii) assembly and disassembly motors [14, 20].

All the various cell movements are performed by ensembles of molecular motors that fall into these categories. One of the best characterized motor

in the bacterial species is the tiny rotary motor that enables bacteria to swim [23]. This motors uses an ion flux along an electrochemical gradient to drive the rotation of the long, thin helical flagellum at a frequency of about 100 Hz. All known biological rotary motors use energy stored in an ion gradient to produce torque [24]. Most use the gradient of hydrogen ions that is found across the membranes of living cells. A best-known example is the ATP synthase. The ATP synthase is a special motor enzyme, which can either pump poptons across an insulating membrane against the electrochemical gradient using ATP hydrolysis or ATP from ADP and phosphate using the energy derived from a transmembrane proton gradient [19], for more details see glossary.

Linear stepper motors are much more common in eukaryotic forms of motility. These motors move along preassembled linear tracks by coupling binding to the track, ATP hydrolysis, and a large-scale protein conformational change (see glossary for the connection between ATP hydrolysis and molecular motors). The first linear stepper motor to be characterized was myosin, the motor that drives filament sliding in skeletal muscle contraction [21]. The track for myosin is the actin filament, a helical polymer formed by noncovalent self-association of identical globular subunits. Currently, there are at least 18 different classes of myosins known, and each class may comprise dozens of different members even in a single organism. Figure 31.3 shows a schematic representation of Myosin V double arm acting between an actin filament and a cargo and the correspondent simulated step-like motion. Various forms of myosin in humans are responsible for biological movements as diverse as muscle contraction, cell division, pigment granule transport in the skin, and sound adaptation in the hair cells of the inner ear. The exact mechanical features of each type of myosin motor appear to be carefully tuned to their biological functions. Similar functional tuning of motor mechanical properties is found in another abundant and diverse family of linear stepper motors, the kinesins, which use microtubules rather than actin filaments as a track [24]. Kinesins are involved in many types of intracellular transport, including transport of organelles along nerve axons and chromosome segregation. Another family of linear stepper motors that walk on microtubules, the dyneins, is less well characterized.

Microtubules and actin filaments can assemble and disassemble rapidly to change the shape of the cell and to produce force on their own [25]. In these forms of biological force generation, the chemical energy comes from nonequilibrium protein polymerization, although ultimately the cellular pools of polymerizing actin and tubulin subunits are maintained in a steady state far from chemical equilibrium due to a coupling between protein polymerization and ATP hydrolysis. The force generated by actin polymerization is responsible for the movement of certain kinds of bacteria pathogens and the major driving force for cell protrusion at the leading edge in amoeboid motility.

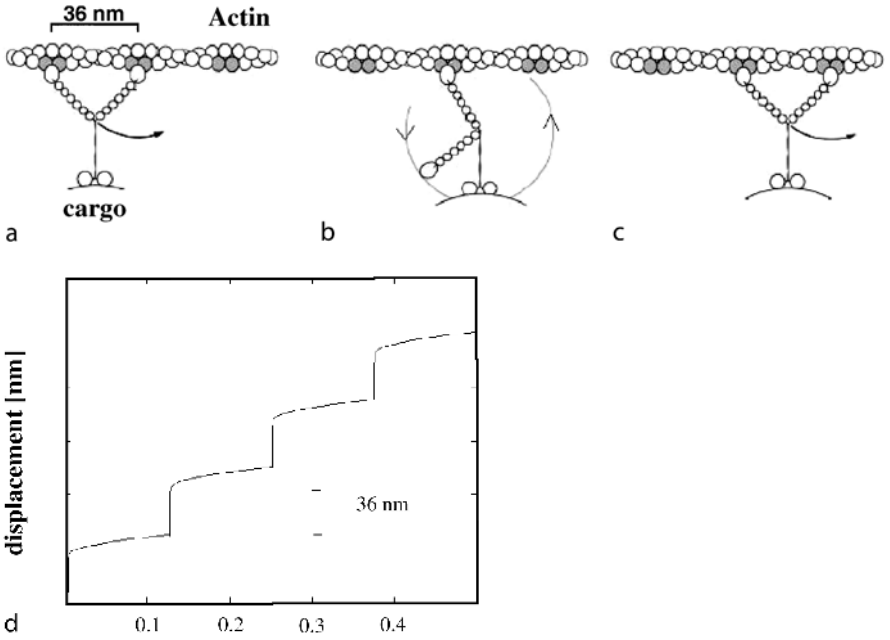


Fig. 31.3. Sequence for swinging lever-arm model proposed for the walking of a myosin-V dimer along actin filaments [21]. Changes in the actin-binding domain during the ATPase cycle makes possible the rotation of the lever arm resulting in movement. The binding sites on actin filament (8 nm diameter) are marked in gray. In **d** is represented the step-like motion from a numerical simulation [22]

Whereas rotary motors and linear stepper motors have been characterized in great physical detail and force measurements have been performed on single molecules of each class, assembly and disassembly motors are poorly understood. Only a few measurements of the amount of force generated by single microtubules have yet been made, and there is to date no direct measurement of the force generated by polymerization of a single actin filament [13]. Recently, Filippov et al. proposed a model for the motility of a bead driven by the polymerization of actin filaments [26]. It is remarkable that such a model exhibits the rich spectrum of behaviors similar to those experimentally observed, and that experimental observations can be reproduced taking into account only one type of filament interacting with the bead: the detachment filaments that push the bead.

Molecular motors have a size of 10 – 100 nm. On the contrary, the cargo particles that are moved by cytoskeletal motors have a typical size of up to a few micrometers. This implies that all of these movements are necessarily overdamped and dominated by friction, as a consequence a discrete amount of kinetic energy is changed in heat. Generally, aspects of friction-dominated motion have been discussed within Brownian motion and Brownian ratchets

in nonequilibrium chemical processes. It is assumed that, in analogy to Brownian motion where a particle is subject to collisions with water molecules and reflects the thermal motion of such molecules, the cytoskeletal motor is the particle, and that this particle is attracted towards a filament, Fig. 31.3. The simplest description for Brownian motion in one dimension is provided by the Langevin equation. The dynamical laws governing the time evolution of the system must predict the state of the system at a time $t + \Delta t$, given the corresponding state at time t. The particle moves in a potential field $U[\sigma(x)]$, where the potential at any arbitrary location x is determined by the local density $\sigma(x)$ of chemical elements used by the elements for communicating within themselves and experiences viscous friction and noise (random forces) caused by the surrounding medium

$$\ddot{x} = -\gamma\dot{x} - \nabla U\left[\sigma(x)\right] + \eta(t) \tag{31.1}$$

where $\eta(t)$ is a Gaussian white noise with the statistical properties $\langle\eta(t)\rangle = 0$ and $\langle\eta(t)\eta(t')\rangle = 2\gamma k_{\mathrm{B}}T\delta(t - t')$, where T is the temperature, and k_{B} is the Boltzmann constant, the strength $2\gamma k_{\mathrm{B}}T$ of the noise determines also the degree of determinacy by which the particle would follow the gradient of the local potential. In a ratchet, the potential field is a sawtooth potential, which depends on the one-dimensional coordinate and is periodic but asymmetric in x. The Langevin equation (31.1) is difficult to study and it is more convenient to describe the system in terms of the time-dependent probability distribution for the position x and velocity $(\mathrm{d}/\mathrm{d}t)x$ of the Brownian particle. In the strong friction limit, one may essentially ignore the inertial term and it is sufficient to study the time-dependent probability distribution $P(x, t)$ for the particle position x alone. This probability distribution satisfies the conservation law $[(\partial/\partial t)P(x, t)+(\partial/\partial x)J(x, t)] = 0$, where the current J has the Smoluchowski or Fokker–Planck form $J(x, t) = -D[(\partial/\partial t)U(\sigma(x), t) + (\partial/\partial x)]P(x, t)$. The Fokker–Planck equation shows that the sawtooth potential modulates the position probability of the Brownian particles on small length scales, which are comparable to the period of the sawtooth potential, but does not change their motion on larger scales that still is purely diffusive without any preferred direction. As a consequence, if the whole system is in equilibrium, one cannot extract useful work from such a device in accordance with the second law of thermodynamics. Temperature differences or gradients are not usable in the context of biological systems that are essentially isothermal, than the biological systems are driven by nonequilibrium chemical processes. These processes arise from concentration gradients across membranes, in most cases they are provided by exergonic chemical reactions such as ATP hydrolysis. Thus, if we add ATP, the ratchet moves in a directed fashion. This process will continue until the ATP hydrolysis has produced a sufficient amount of ADP and P and the forward and backward reactions balance each other. Thus, the external field to which the motor/filaments system is coupled does not arise from a spatial gradient but from an imbalance in the concentration of ATP, ADP and P. A fundamental role in the whole processes is played by

the rate of ATP hydrolysis; it is rather slow in the absence of any catalyst or enzyme for a reaction involving a large energy barrier. The only catalytic action, which reduces this energy barrier is provided by the motor molecules themselves. The chemical free energy stored in the ATP molecules is released when such molecules are adsorbed by the molecular motors. In this way, the ATP molecules provide a spatially distributed source of energy, which is only tapped at those locations where it can be directly transformed into mechanical work.

31.2.2 Dynamic Friction in Large-scale Cell Motility: the Case of *Dictyostelium* Slug

The cellular slime mold Dictyostelium discoideum provides an experimentally accessible and simple model system to investigate a combination of biological and tribological processes including chemotaxis, friction and adhesion, cell signaling and cell movements within cellular aggregates [27–29]. Dictyostelium slug movement results from the coordinated movement of its 10^5 constituent cells. It can be considered as a tissue that undergoes permanent reconstruction, as all cells move all the time. There are only two major cell types, prespore and prestalk cells and strongly limited cell division occurs at the slug stage. These properties make it one of the simplest systems to study.

Wijer and coworkers [27] considered the description for cell movement similar to the one referring to densely packed masses of cells. They represented the cellular aggregate as a sphere of influence around a center of mass with the constraint that the distance between the centers of mass can vary over a limited range during the movement. In this way, shape changes during movement are taken into account. Further, it is supposed that the cell movement is affected by four main forces: chemotaxis, adhesion, pressure, friction. Chemotaxis, see glossary, is the most important one as it is responsible for both directed cell motion and coordinated behavior of cells. Two opposing forces, pressure and adhesion, are involved in keeping the volume of the slug constant. Friction consists of two terms, a term dependent on the movement of surrounding cells and a term that takes into account that the cells are slowed down due to friction with the ECM. Cells influence each other when the distance between them is less $15\,\mu\mathrm{m}$, at a distance of $10\,\mu\mathrm{m}$ the balance of pressure (repulsion between cells) and adhesion (attraction between cells) comes into play. The layer of slug less than $10\,\mu\mathrm{m}$ feels the substrate boundary that is at $z = 0$. It has been demonstrated that the cell moves chemotactically when it detects a temporal increase in cAMP concentration and moves in the direction of the concentration gradient. The force due to chemotaxis is $F_{\mathrm{chemotaxis}} = mk_{\mathrm{chemotaxis}}\Omega(\mathrm{cAMP})$, where m is the mass of Dictyostelium, $k_{\mathrm{chemotaxis}} = 0.015\,\mu\mathrm{m/s}^2$ if $\partial(\mathrm{cAMP})/\partial t > 0.004\,\mathrm{s}^{-1}$ and $k_{\mathrm{chemotaxis}} = 0$ otherwise and $\Omega(\mathrm{cAMP})$ is a function that quantifies the concentration of cAMP in the surrounding of the cell aggregate [27]. Pressure and adhesion are two opposing forces responsible for the stability of the

aggregate of cells and play a fundamental role for mechanical response in relative cell–cell movements, while for cell motion along the ECM–substrate adhesion force overcomes the pressure force. Cells experience a viscous friction, which depends on the relative speed of neighboring cells and a friction depending on interactions with the ECM. Wijer and coworkers, as in more recent models, considered the two main friction sources (the slug–ECM friction and the cell–cell friction) as proportional to the velocity. Effectively, the role of friction in an aggregate of cells such as Dictyostelium is complicated by cell–cell interactions. Since a cell inside an aggregate gets traction from its neighboring cells it seems likely that friction depends on the motion of neighboring cells, if the neighboring cells move in the same direction friction is less than when they move in opposite directions. Nevertheless, cell motion in a whole aggregate is highly coordinatedd. As a consequence, the average velocity is increased as cell–ECM friction in decreased. Moreover, the velocity is dependent by the cAMP wave concentration, and a residual velocity permains between two cAMP distribution peaks. The increasing of velocity involves higher values for adhesion and pressure forces. Recently, Rieu et al. [28,29] showed that tip friction and trail friction (see Fig. 31.4) scale with the velocity as $v^{-1.31}$ and $v^{-0.96}$, respectively, and the overall friction scales as $v^{-1.26}$. In dry solid frictional sliding between a rough surface and a smooth surface, one often observes a stick–slip regime in the range of low velocities. In this regime, the friction coefficient is described by a logarithmic decreasing function of velocity. Moreover, they observed that at relatively high velocity ($v > 1\,\mu$m/s), the friction coefficient increases again logarithmically. The fact that they measured a friction decreasing with slug velocity in the range $0.1 - 0.5\,\mu$m/s, suggests that cell/ECM and/or sheath/substrate friction is compatible with a stick–slip mechanism. Moreover, since the friction and traction forces decrease with slug velocity, it is reasonable to suppose that the dynamics of the cell/ECM adhesion sites may regulate the slug velocity. Too many sites create large traction, but also large friction. In addition, this behavior must be combined with the probability needed for an adhesion site to be formed, a probability that is inversely proportional to the velocity, so Dictyostelium single cells show slow motion, as do fibroblast cells. In fact, friction is ultimately due to the break over and recovery cycle of transmembrane molecular links within cells or within cells and substrate, and its dependence on velocity reflects the dependence on velocity on actin network capability for remodeling.

A theoretical approach for Dictyostelium spreading process has been recently proposed by Chamaraux et al. [30]. They have supposed that the kinetics of spreading is controlled by the membrane–cytoskeleton attachment. The spreading process is described quantitatively by the contact area for a hemispherical cap cell as a function of the time and following the relation $A(t)/A(t = \infty) \cong \tanh(\alpha t)$, where $A(t)$ is the area covered by the cell, $A(t = \infty)$ is the asymptotic value around which the contact area oscillates

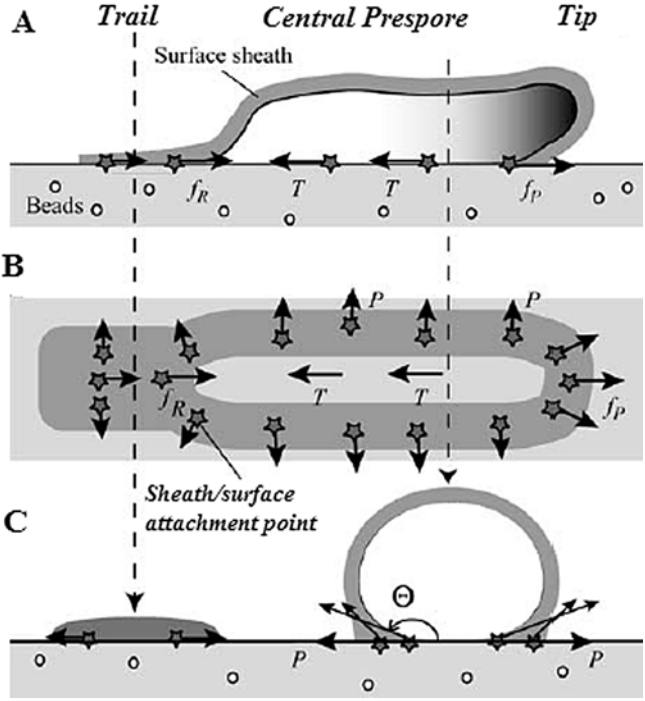


Fig. 31.4. Model of a migrating slug; forces are those exerted by the slug on the external flexible substrate. The surface sheath covering the slug is pronounced in the tip; it immediately anchors to the substrates at various attachment points. In the tip, this anchorage combined with the slug migration creates a resistive force f_{P} pushing the elastomer forward, **A** and **B**. In the central prespore region, a traction force T due to cell traction along the slug migration direction is localized in the inner close contact area. The perpendicular force P visualized by the flexible substrata method in the inplane component of the stretching force tangential to the slug profile exerted by the sheath at the various lateral attachment points, **C**. In the tail, the collapse of the sheath creates a resistive force f_{R}, **A** and **B**. Frozen sheath only partially realizes the perpendicular force in the trail of the slug, **B** and **C**. (Reprinted from [29], Copyright (2005), with permission from Biophysical Journal)

when the cell is motile, and α is a characteristic time for the spreading process of the Dictyostelium aggregate. The driving force for spreading is believed to be the polymerization of the actin network at the extreme margin of the contact area. The model pointed out by Chamaraux et al. couples the kinetic coefficients of the actin cytoskeleton polymerization to the mechanical shearing stress sensed by the adhesive bridges at the extreme margin of the contact area. Moreover, from such a model it turns out that the stress is proportional to the membrane tension, increasing with the contact area. Since mechanosensitivity depends on the tangential stress, Chamaraux et al.

assume that the shearing stress controls the (de)polymerization rate of the actin network. Such shearing stress decreases with the distance to the extreme border as $\tau(x) = \tau_0 \exp(-\lambda x)$, with $\lambda = [\mu/(Ehd)]^{1/2}$, where $\mu \approx 100\,\mathrm{Pa}$ is the shear modulus of the adhesive belt, $E \approx kPa$ is the Young modulus, h is the characteristics distance between the surface and the actin network, d is the thickness of the actin cytoskeleton. Generally, $h \approx 20\,\mathrm{nm}$, $d \approx 500\,\mathrm{nm}$, one finds $\lambda^{-1} \approx 0.4\,\mu\mathrm{m}$. The relation describing the shearing stress $\tau(x)$ at the extreme border of the adhesive belt of Dictyostelium is useful in understanding how the compliance of the substrate influences cell spreading and motility. The shear modulus μ of an elastic substrate can be modeled with two springs in parallel. As a consequence, the shearing stress is less concentrated for a soft substrate than for a hard one. Since the concentration of this stress is at the origin of the actin polymerization, thus substrate stiffness favors cell spreading. Moreover, the model provides a quantitative expression for the characteristic time α that is dependent on the relative difference of speed filamentous actin (de)polymerization rates and stress calculated at the margins of cell spreading.

31.3 Cell–ECM Adhesion Forces

Cells adhere to a surface initially by attaching to a preadsorbed protein network called the extracellular matrix (ECM) or to neighboring cells. The cells spread out and their shapes are influenced by the surface topography and contribute to their phenotypic behavior. The nature of the ECM influences major cellular perspectives of growth, differentiation and apoptosis and its composition will ultimately determine which cellular functions will be selected. In Fig. 31.5 it is schematically shown how cells feel and react to the ECM by means of integrin-dependent focal adhesion (FA) sites. The ingredients of a FA site are schematically sketched in Fig. 31.1b. FAs were first observed to form between cells and solids by Ambrose (1961) [31] and later Curtis (1964) [32] found that their distance of closest approach was approximately 10 nm.

The FA complex is composed of a high density of proteins that attach the extracellular portion of the cell to the intracellular cytoskeleton portion. Transmembrane proteins, such as integrins, attach to the ECM and connect indirectly to the actin filaments through protein assemblies of talin-paxillin-vinculin, see Figs. 31.1 and 31.5. Recently, it has been shown experimentally and by a theoretical approach that mechanical forces can influence the association of integrin with the cytoskeleton to form the FA complexes [34–37]. In such complexes, integrins are likely a major force transmitter, since they provide the mechanical linkage between the cell cytoskeleton and ECM. Furthermore, cells may be sensible to mechanical forces or deformations through both integrin-mediated cell–ECM interactions and the subsequent force balance within the cytoskeleton. This balance may play a crucial role in regulat-

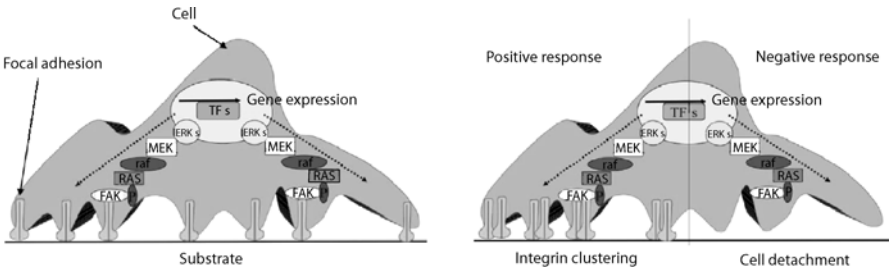


Fig. 31.5. Schematic representation of how integrin-mediated activation of signal transduction pathway may regulate the cell/substrate interaction. Once the cell comes into contact with the substrate it forms focal adhesions. The integrins are thought to relay signals to the nucleus though the MAPKase (mitogen-activated protein kinase) pathway that is a cascade of proteins abbreviated FAK = focal adhesion kinase, P = phosphorylation, TFs = transcription factors, RAS and raf monomeric GTPases, ERKs = extracellular signal related kinases, MEK = MAP kinase. In the right image, it is shown that if the signals relayed by the focal adhesions to the nucleus are positive integrin clustering occurs, increasing the area of cell adhesion to the substrate, if the signals are negative then matrix metalloproteinases are related causing integrin substrate detachment decreasing the area of cell attachment. (Adapted from [33], with permission)

ing the shape, spreading, crawling, and polarity of cells. A still open question is how the mechanical force balance is recognized by cells and transduced into biological responses; consequently, the exact molecular mechanisms responsible for mechanochemical transduction in living cells remain unknown. Different mechanisms have been proposed. One proposed mechanism regards the ion transport in the cell membrane. Ion transport can be changed by mechanical forces, especially tension in the cell membrane, thus changing the biochemical processes in cells [38]. Another proposed mechanism is based on the experimental evidence that cell cytoskeleton components, such as actin filaments and microtubulues, deform under mechanical forces inducing conformational changes of other proteins attached to them and altering their functions. Furthermore, the role of centrosome, of microtubules and of related motor protein has been discussed in relation to mechanotransduction and movement [39]. Moreover, ECM biopolymers such as fibronectin may deform under force, changing their interactions with cell-surface receptors including integrins. Fibronectin is able to contract to a fraction of its original length, and this property serves as a mechanosensitive control of ligand recognition. Existing experimental results suggest that the binding specificity and affinity between a receptor and a ligand can be changed by mechanical forces.

Living cells exert directional, lateral forces on adhesive sites. Since the adhering cells are rather flat, the forces exerted on the substrate can be considered to be tangential to the plane of the substrate surface. These forces

originate from the interaction of the contractile cytoskeleton (actin filaments and their associated myosin motors) and the FA adhesion sites. Such adhesion sites respond dynamically to the local stresses: increased contractility leads locally to larger adhesions, on contrary, FAs are disrupted when myosin in inhibited [40]. Typical forces at mature FAs of human fibloblasts were found to vary between 10 and 30 nN [10]. Since each cell can have several hundred FA, the overall force exerted by a single cell goes up to the μN range. A linear relationship between magnitude of force and FA area was found [41]. The direction of force usually agreed with FA elongation, for larger areas, force grows in proportion with area, with a stress constant of $5.5\,\mathrm{nN}/\mu\mathrm{m}^2$. The relation between force and size of the FA indicates that they act as mechanosensors: forces are used to actively probe the mechanical properties of the environment.

Several models have been introduced to describe within physical condensed-matter schemes how the anchorage-dependent cells constantly assemble and disassemble FA sites, thereby probing the mechanical properties of their environment. For example, the concept of force dipoles has been used to model cells in an elastic environment [42, 43], see Fig. 31.6.

The force dipoles model applied to the cellular case can be summarized as follows: anchorage-dependent cells probe the mechanical properties of the soft environment though their contractile machinery. Actin stress fibers (lines in Fig. 31.6a) are contracted by myosin II molecular motors and are connected to the environment through FA (dots in Fig. 31.6a). Independent of the cell shape, different stress fibers probe different directions of space and compete with each other for stabilization of the corresponding FA. As a consequence, the probing process can be modeled as an anisotropic force contraction dipole.

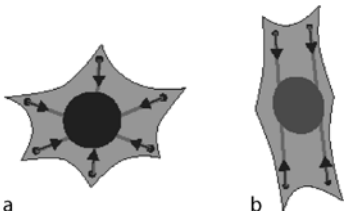


Fig. 31.6. Schematic representation of cellular force dipoles. **a** Cells feel the mechanical properties of the soft environment through their contractile sensors. Actin stress fibers (*lines*) are contracted by myosin II molecular motors and are connected to the environment though FA (*dots*). Different stress fibers probe different directions of space and compete with each other for stabilization of the corresponding FA. Such a probing process can be modeled as an anisotropic force contraction dipole. **b** The cell morphology becomes elongated in response to anisotropic external stimuli, during locomotion or spontaneously during times of strong mechanical activity. As a consequence, most stress fibers run in parallel and the whole cell acts as an anisotropic force contraction dipole. (Reprinted from [10], Copyright (2005) with permission from Elsevier Science)

Cell morphology becomes elongated in response to anisotropic external stimuli during locomotion or spontaneously during times of strong mechanical activity, Fig. 31.6b. Then most stress fibers run in parallel and the whole cell acts as an anisotropic force contraction dipole. One main advantage of such a model is that since cells are modeled as anisotropic force dipoles, calculations are in general similar to calculations for isotropic force dipoles. For example, the calculation of the energy interaction for a cell characterized by a distance d to the surface, an angle θ with respect to the surface normal and a force dipole moment P, can be calculated exactly, giving the result [10]:

$$W = \frac{P^2}{Ed^3}(a + b\cos^2\theta + c\cos^4\theta) \tag{31.2}$$

where E is the Young's modulus and where the coefficients a, b, c are functions of the Poisson ratio v. Minimizing the interaction energy, W, predicts that cells orient parallel and perpendicular to soft or stiff domains of the substrates, respectively, as observed in many experiments. This result is fundamental for a rational design of biomimetic scaffolds in TE, as will be discussed in the next section.

Recently, a simple two-spring model has been proposed by Schwarz et al. [44] making predictions regarding the way cells perceive extracellular rigidity, Fig. 31.7. A FA complex is a structure based on an adhesion cluster that can be schematically represented as a certain number of bonds. The rupture and rebinding mechanisms of such an adhesion cluster play a fundamental role in mature FA formation.

In the two-spring model, the ECM and the force-bearing intracellular structures are represented by harmonic springs with spring constants K_c and K_i, respectively. Being in series the overall stiffness is mainly determined by the softer spring, which in a physiological situation should be the ECM. Tension in the actin stress fibers is generated by myosin II molecular motors. As the motors pull, the springs get strained; for the static situation, the stored energy is $W = F^2/2K$, where K is the overall system stiffness. For the dynamic situation, the power $\mathrm{d}W/\mathrm{d}t$ is given by $\mathrm{d}W/\mathrm{d}t = (F/K)\,\mathrm{d}F/\mathrm{d}t = Fv(F)$, where $v(F)$ is the force–velocity for the molecular motors, this implies that the major contribution to the power is generated by the molecular motors. For a typical force–velocity relation as $v(F) = v_0(1 - F/F_\mathrm{s})$, so that the expression for the dissipation power can be readily integrated giving $F = F_\mathrm{s}(1 - \mathrm{e}^{-t/t_\mathrm{K}})$ with $t_\mathrm{K} = F_\mathrm{s}/v_0K$. This relation implies that the pulling force saturates at $F = F_\mathrm{s}$, but the stiffer the environment (the larger K) the faster a given threshold in force can be reached. If the cell pulls on a material with a bulk modulus of kPa, then the corresponding spring constant on the molecular level can be expected to be of the order of $K = pN/\mu m$ and the typical time scale t_K is in seconds. If the bulk modulus is of the order of MPa, then $K = pN/nm$ falls in the typical range for protein stiffness and the time scale t_K is in milliseconds. In the two-spring model the internal structure of the FA is represented by one biomolecular

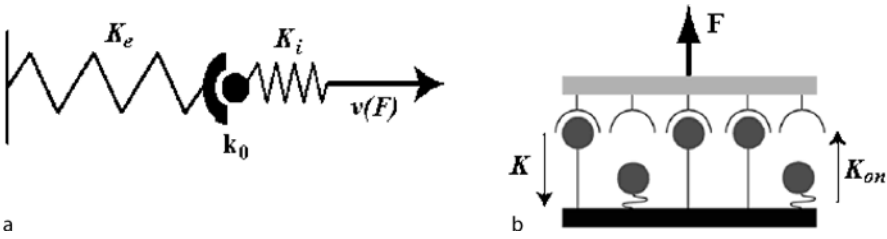


Fig. 31.7. a Two-spring model sketch for mature FA site, K_e represents extracellular elasticity and spring constant K_i represents the mechanical properties of the intracellular structure. Force generation by the actin cytoskeleton is represented by the linearized force–velocity relation $v(F)$ for a single myosin II molecular motor. The internal state of the focal adhesions is represented by a biomolecular bond that opens in a stochastic manner with a dissociation rate k_0. **b** Schematic representation of an adhesion cluster under force before mature FA formation. Closed bonds rupture with a force-dependent rate and open bonds close with a force-independent rebinding rate k_{on}. (Reprinted from [45], Copyright (2005), with permission from Elsevier Science)

bond with an unstressed dissociation rate k_0, Fig. 31.7a. As in the general picture of receptor–ligand bonds, the rupture under mechanical force of such bonds determines many properties of FA sites. Rupture dynamics of an adhesion cluster under a pulling force is schematically outlined in Fig. 31.7b. Such a model is a stochastic version of an earlier model by Bell [45]. Briefly, the model assumes that N receptor–ligand bonds have been clustered on opposing surfaces, of which the upper one acts as a rigid transducer that transmits the constant force F homogeneously onto the array of bonds. At each time, i bonds are closed and $N - i$ bonds are open, see Fig. 31.7b. Closed bonds are assumed to rupture with a force-dependent rupture rate $k = k_a \, e^{F/iF_b}$, where k_a is the unstressed rupture rate (typically around 1/s) and F_b the internal force scale (typically a few pN) of the adhesion bonds. The exponential dependence between force and rupture rate results from a Kramers-type description of bond rupture as escape over a transition state barrier. The factor i results because force is assumed to be shared equally between closed bonds, which holds true when the transducer is connected to a soft spring, whereas in the opposite limit of a stiff spring, all bonds feel the same force. Open bonds are assumed to rebind with a force-independent rebinding rate k_{on}. The probability that i bonds are closed at time t is described by a master equation with appropriate rates for the reverse and forward sates between possible states i. The force F destabilizes the cluster, rebinding stabilizes it. In the case of a two-spring model, the single bond under a stall constant force F_s has the average lifetime $t = (1/k_0) \, e^{-F_s/F_b}$. For large K, the bond experiences constant loading with stall force F_s. In the case of small K, loading is approximately linear with loading rate F_s/t_K. Between these two limiting case, the average force that is built up until bond rupture has given by the

relation [44]

$$\langle F\rangle = \int_0^\infty p(t)F(t)\,\mathrm{d}t = \frac{F_\mathrm{S}}{1 + k_0 F_\mathrm{S}/v_0 K} \tag{31.3}$$

where $p(t) = \mathrm{e}^{-k0t}k_0\,\mathrm{d}t$ is the assumed probability that the bond beaks at time t in a interval $t + \mathrm{d}t$. It is interesting to note that in the relation (3) the level of force reached is essentially determined by the quantity k_0F_s/v_0K. Since unstressed dissociation k_0, still force F_s and maximal motor velocity v_0 are molecular constants, the only relevant quantity in this context is the external stiffness K. By using the preview data, the average force $\langle F\rangle$ is larger by a factor 2 in a stiff environment. If a cell is pulling at several focal adhesions with a similar investment of resources, then those contacts will reach the level of force putatively required for activation of the relevant signaling pathways that experience the largest local stiffness in their environment. As a consequence, growth of contacts in an elastically anisotropic environment might then lead to cell polarization and locomotion in the direction of maximal effective stiffness in the environment, which has been observed in many experimental situations.

31.3.1 Biomimetic Scaffolds, Roughness and Contact Guidance for Cell Adhesion and Motility

All TE constructs are composed of two major components: a scaffolding material that provides the mechanical and structural properties required, and site-specific cells. One fundamental target of TE is the production of biomimetic scaffolds [46–48]. Polymeric fibers and cellular solid scaffolds should be engineered to include the opportunity of biomolecular signals [49]. Since information that is introduced on the scaffold material surface is processed as biomechanical and biochemical signals through receptors, which are nanometer-sized entities on cell surfaces, it is important that this information be presented on the same length scale as occurs in nature. Generally, an assembly of functionalized particles could serve as a versatile tool for imparting texture and chemical functionality on a variety of surfaces. Such modified surfaces can be tuned to posses tethered or covalently adsorbed biomolecules such as peptides, proteins and biopolymers and can serve as a platform for engineering biomimetic interfaces to modulate cellular behavior toward implants and scaffolds in TE [50]. In the early 20th century Weiss [51] observed that cells preferentially orient along ECM fibers, an organization principle he termed *contact guidance*. Further, Weiss observed that two tissue explants reorganize the collagen gel between them into aligned parallel fiber bundles and that cells leaving the explants migrate and orient along the aligned fibers. Contact guidance therefore could serve both as a cue for organization on cellular scales and as a large-scale organization tool in tissue development by guiding motile cell along ECM-bundles. More recently, the community of

scientists and bioengineers involved in tissue reconstruction and remodeling agree that cells use spatial variation of adhesiveness and frictional force to favor orientation of cells along thick fiber bundles [52]. Adherent cells can respond to mechanical properties of their environment, and the use of a sophisticated elastic scaffold has provided strong evidence that cells respond to purely elastic features in their environment [36, 53].

One emerging aspect in biomimetic scaffolds realization is the role of scaffold roughness for contact guidance on nanopatterned surfaces [54, 55]. Surface roughness has an enormous influence, such as in contact mechanisms, friction, sealing and adhesion, in a large variety of situations [56], also in cellular dynamics surface roughness plays a fundamental role. For single cells, the cellular response to alterations in surface topography, down to the nanometer scale, has been documented especially for grooved topography [57]. Most cells follow the discontinuities of grooves and ridges, and attain an elongated shape due to surface-induced rearrangements of the cytoskeleton. The development of techniques to produce surface structures in the nanometer range has revealed that cells also respond to such nanostructures. Macrophage-like cells can react to steps in the nanometer range. Endothelial and fibroblast cells are sensitive to patterns with features down to 10 nm, and the importance of symmetry and discontinuity of the nanometers patterns was pointed out.

A range of techniques can be used to create well-defined topographical and chemical cues for cell patterning. Many of these approaches rely on photolithography and reactive ion etching of the substrate and can be also fol-

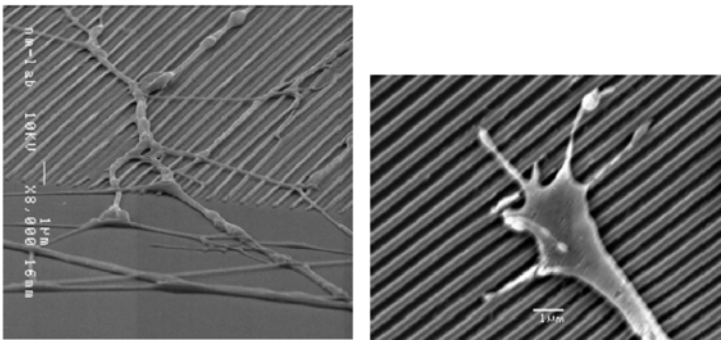


Fig. 31.8. *On the left*, scanning electron microscopy picture showing axons that after a few μm align to the ridges. The ridges are 100 nm high and 100 nm wide and they are of the same dimensions as the axons. It is interesting to note the contact guidance role of the ridges, all axons are growing on top of the ridges and not in the grooves between the ridges, unit 1 μm. *On the right*, scanning electron microscopy showing filopodia extending both perpendicular and parallel to the underlying ridges pattern, scale bars 1 μm in both images. (Reprinted from [57], Copyright (2006), with permission from Elsevier)

lowed by UV and glow discharge treatments [58]. Microcontact printing is also a popular technique, and other methods including inkjet printing and diamond cutting have been successfully exploited [59]. These techniques are generally suitable only for micropatterning. To go down in size, photolithography results are limited by light diffraction effects. In fact, its resolution is on the order of the wavelength of the light used for exposure (typically $> 200\,\text{nm}$). Electron-beam lithography can be used to produce nanoscale patterns, but it is expensive and time-consuming. A simple method has been reported to fabricate nanoislands of $13-95\,\text{nm}$ in height on a large scale based on phase separation of polystyrene and poly(4-bromostyrene) spin-coated on silicon wafers [60]. However, the reliability to produce nanofeatures of controlled size and geometry based on such phase separation phenomenon is relatively poor. Recently, a rapid AFM-based method to pattern polymeric scaffold surfaces forming regular and tunable ripples with a periodicity of grooves varying in the $20-200\,\text{nm}$ range [61] has been found, and these nanopatterned surfaces are going to be tried for cell-proliferation tests.

Molecularly tailored surfaces utilizing peptide segments provide a route to mimic cell or ECM environments without the use of an entire protein. Molecular biologists have identified many minimal sequences of amino acids of proteins that are responsible for adhesion to a particular receptor. These minimal peptide sequences, are much more stable than entire proteins and the surface concentration of the specific ligand may be much higher. Here, we limit the discussion to describe typical JKR test performed on such bioactive surfaces [62]. The Langmuir–Blodgett technique is used to create amphilic bilayers of controlled surface and peptide density, in Fig. 31.9 is presented a sketch of the contacting surfaces. The contact mechanical method used is that referred to as the JKR method, developed by Johnson, Kendall, and Roberts [63]. The JKR theory represents a balance of elastic energy associ-

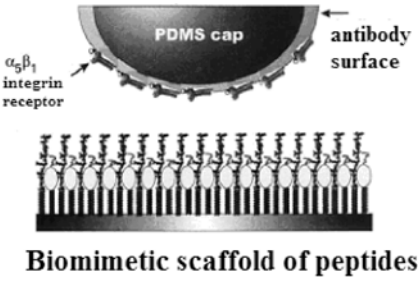


Fig. 31.9. Schematic diagram of JKR apparatus and biomimetic surface in contact with integrin-immobilized lens. The antibody layer serves to block the integrin receptor. The polydimethylsiloxane (PDMS) lens has a radius of curvature of 1 – 1.5 mm, and applied loads are less then 400 mg, maximum time of contact is 10 min, the contact area fall in the range $0 \div 2 \cdot 10^{-3}(\text{mm}^3)$. (Reprinted from [62], Copyright (2001), with permission from Elsevier)

ated with compressing the elastic sphere, potential energy associated with displacing a normal load, and surface energy associated with forming contact area. The contact area will increase under an applied load (P), such that at mechanical equilibrium and energy balance the contact radius a is a function of the adhesion energy (G), the elastic modulus of the materials (K), and the radius of curvature of the sphere R is given by the expression

$$a^3 = \frac{R}{K}\left[P + 3\pi GR + \sqrt{6\pi GR + (3\pi GR)^2}\right] \tag{31.4}$$

where $1/K = 3/4\left[\left(1 - v_1^2\right)/E_1 + \left(1 - v_2^2\right)/E_2\right]$ and the radius of curvature of the two spheres is calculated as $1/R = 1/R_1 + 1/R_2$ (generally, for a sphere-contacting flat surface, $R_2 \to \infty$, $R_1 = R_{\text{sphere}}$). The adhesion energy (G) is a function, among other things, of the surface energy and the interfacial energy between the two surfaces. At equilibrium, G is equal to the thermodynamic work of adhesion that is equal to $\gamma_1 + \gamma_2 - \gamma_{12}$. Using the JKR method, the contact area between an elastic sphere and a rigid flat surface can be measured optically and plotted as a function of the measured normal load. By modifying the surfaces of materials and biomolecules that comply, the JKR method can be used to measure the specific adhesion force between integrin protein and mimetic surfaces. In the method of Tirrell et al. [62], one can measure the contact area during loading and unloading, any hysteretic differences can be used to distinguish between specific and nonspecific adhesion. The adhesion in the loading process is due mainly to nonspecific van der Waals interactions. Once the solids are in contact, short-range specific interactions may occur (including hydrogen bonds, specific lock-and-key-type bonds associated with ligand–receptor interactions, and molecular rearrangements). Once these bonds have formed, additional energy is required to separate the surfaces, which results in a hysteresis in the unloading curve. The degree of hysteresis provides information about the strength of adhesion due to interactions such as molecular rearrangements, hydrogen bonds and other specific interactions formed at the interface. The method used provides a useful tool to select peptides on the mimetic surface increasing adhesion to integrin and revealed that peptide orientation, spatial arrangement and binding site accessibility area are crucial for effective ligation to integrin receptors.

Different time phases can be distinguished in the culture of cells on surface scaffolds [55]. The first one is the initial cell-adhesion phase involving nonspecific electrostatic forces and passive formation of ligand–receptor bonds, followed during the hours following the cell-spreading phase involving receptor recruitment, clustering to anchoring sites and interactions with cytoskeletal elements. The second phase concerns the proliferation and differentiation phases involving the ECM formation. During the first phase, when the ECM is not already formed, the surface properties of the scaffold play a major role in the subsequent proliferation phase. In particular, the surface polymer presenting a consistent texture of gradient elasticity can induce contact guidance

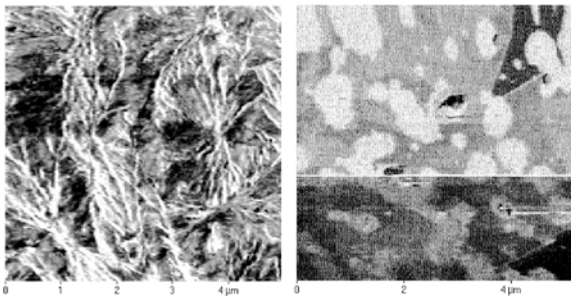


Fig. 31.10. AFM modulation force images of triblock polyurethane scaffold surfaces present Young's moduli maps (*brighter domains* correspond to stiffer moduli, and *darker areas* to soft moduli). The characteristic interconnection of harder and softer domains consequent to phase separation involves a high gradient of elasticity that on the first cell seeding should provide a contact guidance for the cells. (Reprinted from [66], Copyright (2006), with permission from Nova Science Publisher)

for cells. Tri-block polyurethane family surfaces have been extensively used in TE recently [64]. They present a wide variety of surface interconnection between harder and softer domains and it has revealed recently that a higher interconnection of phase separation in the material improves cell proliferation [65], also if the connection between the increased contact guidance due to phase separation and the increased cell proliferation in the first phase of seeding presents still many open question and unknown experimental and theoretical mechanisms.

31.4 Experimental Strategies for Cell–ECM Adhesion-force Measurements

Cell adhesion is one of the initial events essential to subsequent proliferation and differentiation of cells before tissue formation. Consequently, many *in-vitro* evaluations of cell adhesion on substrates have been performed in order to discern the main surface properties influencing the cell response to implant surfaces. To determine cell adhesion many techniques have been evolved, such as functionalized latex beads moved with optical tweezers [67], interferometric techniques [68], centrifugation experiments [69,70]. Viscoelastic properties of cells were measured with AFM in either force-modulation mode or more recently by force-volume technique [71,72]. Adhesion between single cells was measured in the past using mechanical methods, such as micropipette manipulations [73,74] or hydrodynamic stress [75,76]. Many new techniques for measurements of adhesion force between cell and substrate or cell-to-cell or between ligand and receptor are based on the combination of traditional optical microscopy with AFM technique. The AFM offers partic-

ular advantages in biology: measurements can be carried out in both aqueous and physiological environments, and the dynamics of biological processes can be studied *in vivo*. Many cellular functions require the accurate knowledge of ligand binding to receptors, single ligand–receptor pair measurements include biotin–avidin [77], antibody–antigen [78], cellular proteins, either isolated or in cell membranes [79]. The general strategy is to bind ligands to AFM tips and receptors to probe surfaces (or vice versa), respectively. In a force–distance cycle, for example [80], the tip is first approached to the surface whereupon receptor–ligand complexes are formed, due to the specific ligand–receptor recognition. During subsequent tip–surface retraction a temporarily increasing force is applied to the ligand–receptor connection until the interaction bond breaks at a critical force (unbinding force). AFM-based investigations on cellular mechanics were performed on elastic and mechanical properties of cells including platelets, osteoblasts, glial cells, macrophages, endothelial cells, epithelial cells, fibroblasts, bladder cells, [71,72,81–86], generally, Young's moduli of living cells vary in the range $1-100\,\mathrm{kPa}$ depending by the cell type and softest parts, 1 kPA, are located around the nucleus.

Sato et al. [87] performed indentation test on endothelial cells exposed to shear stresses of 2 Pa. They obtained a relationship between external applied force F and indentation depth δ, as $F = a\delta^2 + b\delta$, where a and b are two constants representative of the mechanical response. Viscoelasticity of leading-edge fibroblast lamellopodium was measured by Mahaffy et al. [88] using an AFM-based microrheology device. They quantified viscoelastic constants as elastic storage modulus, viscous loss modulus, and the Poisson ratio. Lamellopodium are thin regions ($< 1000\,\mathrm{nm}$) of cells strongly adherent to a substrate with an elastic strength of $\sim 1.6 \pm 0.2\,\mathrm{kPa}$ and with an experimentally determined Poisson ratio of ~ 0.4 to 0.5.

Many experiments have been carried out to reveal the effect of lateral forces on adhesion sites, one important technique is the so-called elastic substrate method [89]. This technique provides that a thin elastic film over a fluid is produced, under cell traction the film shows a wrinkled pattern, which is characteristic of the pattern of forces exerted. From the first semiquantitative tests, quantitative analysis of elastic substrate data was pioneered by Dembo and Wang [90]. Using linear elasticity theory for thin elastic films and numerical algorithms for solving inverse problems, the surface force exerted by keratocytes was reconstructed. Recently, a novel elastic substrate technique to measure cellular forces at the level of single FA was proposed [9]. Wang et al. [91] carried out experimental testing on various adherent cell types cultured on deformable substrates and revealed specific patterns of cell reorientation in response to cyclic stretching of the substrate. They showed that under uniaxial deformations cells were found to elongate perpendicular to the stretch direction, Fig. 31.11, whereas in cases where the substrate was laterally unrestrained (biaxial deformations) cells were found to elongate at an angle to the stretch direction. The alignment directions in both cases

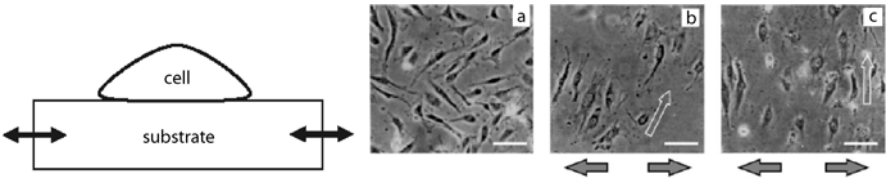


Fig. 31.11. On the left, typical set up for cell orientation under cyclic substrate deformation tests, the arrows indicate the direction of cyclic substrate deformation. On the right, representative phase-contrast microphotographs of endothelial cells: unstretched (**a**), after 3 h of simple 10% elongation (**b**), and after 3 h of pure uniaxial stretching along the x-axis with a maximum strain 10% and constant frequency of 0.5 Hz (**c**). Unstretched cells did not appear to orient in any specific direction, but with simple elongation and pure uniaxial stretching, the cells oriented about 70 ° and 90 °, respectively (*arrows in the photographs*). (Reprinted from [91], Copyright (2001), with permission from Elsevier Science)

correspond to directions of minimum substrate strain. McGarry et al., [92] performed a finite element study on a similar system in order to investigate the role of cell viscoelasticity in cell debonding and realignment under the conditions of cyclic substrates stretching. The characteristic length scale used in the simulation is based on the length of the receptor–ligand bonds at the cell–substrate interface, receptor–ligand bond strength used varying in the range 5 – 40 pN and the bond density approximately 50 sites/μm^2. Such a 2D computational model revealed that discrete cell–substrate contacts at focal adhesion sites result in a completion of debonding in fewer cycles and that permanent debonding at the cell–substrate interface occurs due to the accumulation of strain concentrations in the cell.

The role of shear strength on cell–material adhesion has been investigated also by Yamamoto et al. [93] and Athanasioiu and coworkers [94, 95]. Such a method was based on the design of a cytodetachment instrument able to quantify the force required to displace and to detach cells attached to a substratum, Fig. 31.12. The cytodetacher allows the force required to detach cells from a substratum to be directly measured and the ability of different substrata to support cell adhesion to be indirectly determined.

A cell adheres to a material in a medium inside a dish on an XY stage. The stage can be moved at a speed of 20 μm/s in the direction of a tip attached to a cantilever. The distance between the pointed head of the tip and the material's surface is controlled to be nearly 0.2 μm/s prior to each measurement. Then, the tip touches the cell and a lateral load is applied to the cell. The cantilever is deflected corresponding to the deformation of the cell from the material. Finally, the cell is completely detached from the material. The deflection of the cantilever is measured and the shear force applied to the cell is calculated as the product of the force constant of the cantilever and the deflection of the same. The shear force applied to the cell is recorded as a function of the displacement of the XY stage, which is named

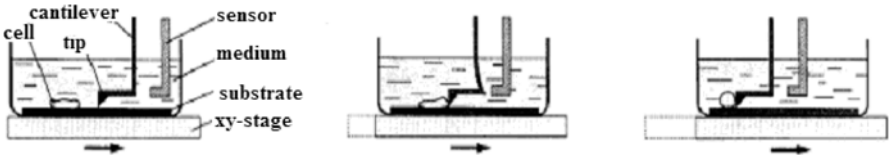


Fig. 31.12. Schematic representation of probe-cell detachment instrument as proposed by [93]. **a** A cell adhered to a material on the XY stage is moved to the tip attached to a cantilever. **b** The cell touches the tip and a lateral load is applied to the cell. The cantilever is deflected corresponding to the shear strength of the cell–material adhesion. **c** The cell is detached from the material. The deflection of the cantilever is measured and recorded. The magnitude of the shear force applied to the cell is given as the product of the force constant of the cantilever and the deflection of the cantilever

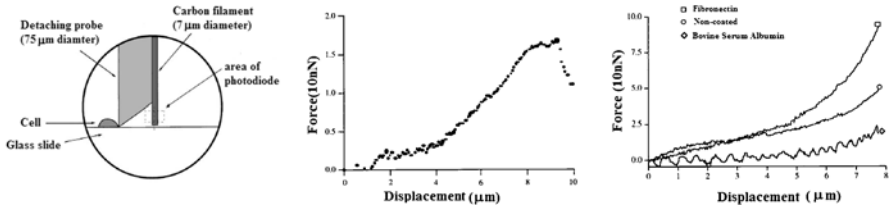


Fig. 31.13. On the left is presented a zoom of the cytodetachment probe by Athanasiou et al. [94]. Deflection of the detaching probe is used to calculate the resistance force offered by the cell. Contact of the cell by the detaching probe causes the probe to bend in response to the resistance offered by the cell. The difference between the displacement of the driving arm and the carbon filament marker beam is used to calculate the resistance force given by the cell to the detaching probe. In the middle, typical force versus displacement graph of a single chondrocyte under shear force applied by the cytodetacher. Mechanical adhesiveness is defined as the maximum force encountered by the cytodetaching probe. On the right, it is shown that the adhesiveness increases when cells are seeded onto substrata that enhance cellular attachment, such as fibronectin. (Reprinted from [94], Copyright (1999), with permission from Elsevier Science)

as a force–displacement curve. The averages of the cell detachment shear strengths measured on murine fibroblasts after 24 h of incubation of glass by Yamamoto et al. fall in the range 500 – 750 kPa, and shear force in the range 350 – 580 nN. Athanasiou and coworkers used the cytodetachment to measure single-cell adhesiveness and the effect of protein on cell adhesiveness [94, 95]. They found that chondrocytes present an adhesion force around 20 nN and that the fibronectin strongly increases the adhesion of the cell to the substrate, see Fig. 31.13.

An example of single-protein role measurement in a complex cell adhesion process has given by Wu et al. [96], who have recently measured the effect of a protein such as focal adhesion kinase (FAK, is a regulator of integrin-

mediated cellular functions (see Fig. 31.5)) on the normal adhesion force making use of an AFM-based cytodetachment technique in connection with optical trapping and an optical tweezer. They investigated the effects of FAK on adhesion force during several stages such as initial binding phase (5 s), beginning of cell spreading (30 min), spread out (12 h) and migration phase (> 12 h). They found that a high concentration of FAK in a cell culture increased the adhesion forces at all the different stages. In fact, the detachment force required for single cells was 343.2 ± 43.4 nN for large presence of FAK and 228.8 ± 36.6 nN for minor presence of FAK at the stage of spreading after 30 min, and 961.0 ± 64.1 and 800.0 ± 75.5, respectively, for major or minor presence of FAK at the stage of spread out after 12 h.

Another diffuse technique is the combination of a JKR apparatus with an AFM design. Generally, the standard AFM tip when used on soft materials such as cells and/or cellular constituents can produce damaging stress. To prevent such unwanted effects, a microsphere bead should be attached to the AFM cantilever. The advantage is also to study adhesion force with the microbead coated with proteins or components of ECM, like collagen, fibronectin, laminin, or entire cells, see Fig. 31.14. This technique is equivalent to the JKR apparatus described in the preview section, with the advantage that such AFM-based instruments work on reduced scales because the microsphere radius is generally varying in the $1-15\,\mu$m range. Making use of such a JKR method, Canetta et al. [97] determined both the local elastic modulus of endothelial cells ($0.2-0.8$ kPa) and adhesion energy $0.3-3.0 \cdot 10^{-7}$ J with a bead attached to the cantilever with a radius of $15\,\mu$m on endothelial cells.

An AFM-based adhesion-force spectrometer combining AFM design and light microscope to investigate cell-to-cell interactions *in vivo* down to the molecular level has been proposed by Benoit and coworkers [98–100]. Single cells or a monolayer of epithelial cells were immobilized on a cantilever. The cells can be monitored using a light microscope during the entire experiment. Adhesion-force measurements were performed via *deadhesion force versus*

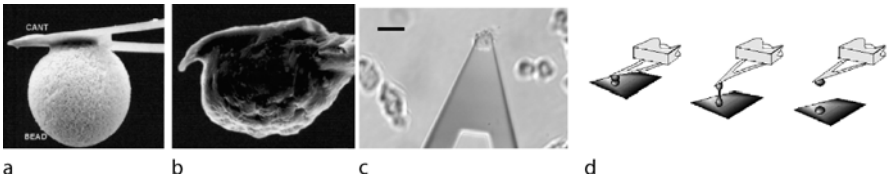


Fig. 31.14. Microbead-mounted cantilevers imaged by scanning electron microscopy. The microbeads glued to the cantilevers were coated both with serum bovine albumin (**a**), or with human trophoblast-type JAR cells (**b**), scale bars $10\,\mu$m. Light-microscope image of a cantilever-mounted cell before being brought into contact with another cell, scale bar $20\,\mu$m (**c**). In **d**, the typical approach–retraction cycle for the deadhesion-force spectrometer is shown. (Reprinted from [98] and [99], with permission from HUMREP and Nature, respectively)

piezo position traces, which are analogous to the force–distance curves with a cell functionalized cantilever. The force resolution was less than 20 pN [99].

The adhesion-force measured with the AFM-based spectrometer is characterized via deadhesion force versus piezo position traces [100], Fig. 31.15. The z piezo velocity was typically set between 1 and 7 µm/s. Low velocities can interfere with drift effects basically caused by cell movement, while at higher velocities hydrodynamics influence the measurement. If a sphere is lowered onto a soft cell surface, the area of interaction increases with the indentation, which leads to an enhancement of the adhesion signal. The adhesion strength is dependent both on the indentation force and on the duration of the contact. This could be due to the fact that the cell shape adapts to the sphere's surface making it possible for more and more molecules to interact with the underlying surface during the contact. Benoit et al. also observed that changing the velocity of retraction leads to a fairly linear relation between separation speed and adhesion in the range 2 – 27 µm/s. The deadhesion force curves show more or less pronounced single steps in the order of 100 pN in the region of descending adhesion indicating ruptures on the molecular level.

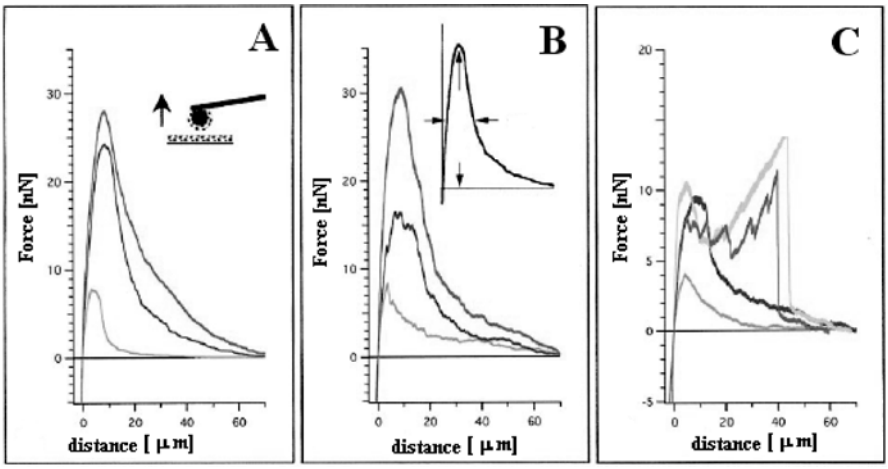


Fig. 31.15. Typical adhesive force curves for human epithelial cells resulting when a bovine serum albumin (BSA)-coated microbead (**A** and **B**) or JAR-coated microbead (**C**) was retraced after periods of 1 – 40 min of contact. The x-axis shows the cantilever displacement, the y-axis shows the force acting on the microbead. In figures **A** and **B**, the rupture events correspond to, 1, 10, 40 min of interaction time, from the lower to the higher, respectively. In figure C, different rupture events corresponding to 20 and 40 min interaction time were observed. Such curves were characterized by increased adhesion force at distance of 20 – 30 µm from the zero force. (Reprinted from [98], copyright (1998), with permission from HUMREP)

31.5 Conclusions

Recent advancements in many biomedical applications require a multidisciplinary approach. The connection between tribology and biochemical and structural aspects on molecular and supramolecular scales could address new significant knowledge in medicine. Among such biomedical applications, tissue engineering (TE) plays a fundamental role to improve or replace biological functions of human tissues. A damaged tissue can be regenerated by the migration of cells to the injured area, as a consequence many questions in TE involve cell dynamics and their proliferation on substrates composed of extracellular matrix (ECM). Animal cells have an average size of $10-30\,\mu\text{m}$, but the adhesion with substrates is limited to sites with dimensions in the nanometer range. As a consequence, a basic understanding on the nanoscale level should be generated to have a satisfactory knowledge of such processes. All TE constructs are composed of two major interacting surfaces: a scaffolding material simulating mechanical and structural properties and specific cell adhering and spreading on it. TE is clearly evident. Mechanical forces play a fundamental role in cell migration, where contractile forces are generated within the cell and pull the cell body forward. For living systems such as complex system cells, the tribological features are more complicated than in the inorganic world because living systems tend to adapt continuously the forces required for their movements by converting some other forms of energy into mechanical energy. For example, molecular motors that denote biological mechanisms converting chemical energy into mechanical energy are responsible for cell locomotion and intracellular transport. Such systems have a size of $10-100\,\text{nm}$ and the cargo particles moved by cytoskeletal motors have a typical size of up to a few micrometers. This implies that all of these movements are necessarily overdamped and dominated by friction. Dry friction behavior, recently evidenced by particular cell aggregates such as Dictyostelium, has been addressed to clarify the role of friction and shearing stress in large-scale cell dynamics. In turn, cell–ECM adhesion-force measurement strategies were discussed.

Glossary

ATP Following Fig. 31.16, in system A, held at constant temperature and pressure, ATP molecules hydrolyze to form the products adenosine diphosphate, ADP and inorganic phosphate P_i. ATP hydrolysis can occur in solution (straight arrow in A) or through an enzyme-catalyzed pathway (curved arrow though the motor enzyme system, B). Independent of the pathway, when one mole of ATP molecules is hydrolyzed, the free energy of A changes by a quantity ΔG_{ATP}. In B, molecular motors (double ovals) move the cargo (helix) against a force, F, as they catalyze the ATP hydrolysis reaction, generating external work, w_{ext} and heat, q.

ATP synthase it is perhaps the world's smallest rotary engine. The general structure of ATP synthase is shown schematically in Fig. 31.17.

cAMP 3',5'-Cyclic AMP (abbreviated cAMP) is used by cells as a transient signal. The cAMP was discovered in the 1950s, it is a small cyclized monophosphate and it is produced from ATP by the enzyme adenylate cyclase. Also many different hormones working in many different cells utilize cAMP. For a brief review on cAMP visit: http://Figurecox.miami.edu/∼lfarmer/BIL265/CAMP.HTM

Chemotaxis The directional translocation of cells in a concentration gradient of some chemoattractant or chemorepellent substances.

Contact guidance The directional translocation of cells in response to some anisotropic elastic property of the substratum.

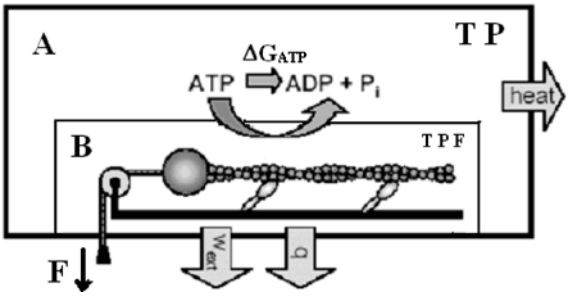


Fig. 31.16. Schematic sketch for adenosine triphosphate, ATP, hydrolysis and energy transfer

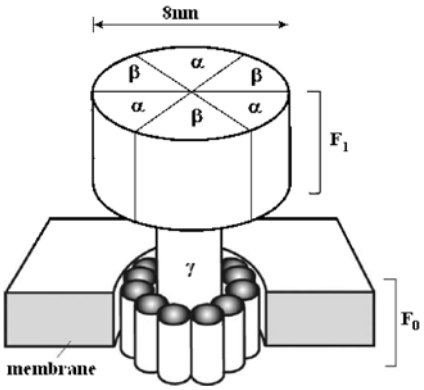


Fig. 31.17. Schematic sketch of the structure of ATP synthase. It is composed by a transmembrane portion F_0, and a soluble component, F_1, which contains catalytic sites located at the $\alpha\beta$ interfaces. Hydrolysis of ATP induces hinge motion of the β domain, which in turn drives the rotation of the γ-subunit in the F_1. In analyzing energy transduction in ATP synthase, it is critical to understand how the protein enzyme converts the free energy of nucleotide binding into elastic energy. (Reprinted from [7], Copyright 2002, with permission from Elsevier Science)

Extracellular Matrix (ECM) The class of components (collagen, fibronectin, etc.) presents in the extracellular space of tissues that serves to mediate cell adhesion and organization.

Haptotaxis The tendency of cells to translocate unidirectionally up a steep gradient of increasing adhesiveness of the substratum.

Integrin Cellular transmembrane protein that acts as a receptor for adhesive extracellular matrix proteins such as fibronectin. The tripeptide () RDG is the sequence recognized by many integrins.

Motility Directed motion driven by energy-consuming motors (opposite to mobility, that is passive motion driven by thermal or other external mechanisms).

Molecular motor biological mechanism that converts chemical energy into mechanical energy, used by a cell to generate directed motion. A single motor can be as simple as a single polypeptide chain or as complicated as a giant macromolecular complex with hundreds of protein and proteoglycan constituents. Chemical energy used by molecular motors is generally either in the form of a high-energy chemical bond (as in ATP) or an ion gradient across a membrane.

Peptide organic compound composed of amino acids, (amino acids are a class of 20 simple organic compounds containing carbon, hydrogen, oxygen, nitrogen, and in certain cases sulfur. These compounds are the building blocks of proteins) linked together chemically by peptide bonds. The peptide bond always involves a single covalent link between the α-carboxyl (oxygen-bearing carbon) of one amino acid and the amino nitrogen of a second amino acid. In the formation of a peptide bond from two amino acids, a molecule of water is eliminated. Compounds with molecular weights of more than 50 – 100 amino acids are usually termed proteins.

Taxis A directed response to some vector-like property of the environment. Translocation is usually biased unidirectionally, either along the field vector or opposite to it.

Acknowledgement. The authors gratefully acknowledge K. Anselme, Université du Littoral Côte d'Opale, Berck sur mer, C. Ascoli and P. Baschieri, IPCF-CNR Pisa; S. Danti, University of Pisa; P. Pingue, Scuola Normale Superiore, Pisa; J.P. Rieu, University Claude Bernard Lyon and CNRS; and U. Schwartz, University of Heidelberg; for useful discussions and suggestions.

References

1. Langer, R., Vacanti, J.P., 1993, Science, 260, 920–926.
2. Peter S.L. Miller M.J., Yasko A.W., Yaszemski M.J., Mikos A.G., J. Biomed. Mater. Res. B, **43**, 422 1998.
3. Griffith M., Osborne R., Munger R., Xiong X., Doillon C.J., Laycock N.L.C., Hakim M., Song Y., Watsky M.A., Science, **286**, 2169 1999.

4. Burg, K.J.L., Porter S., and Kellam J.F., Biomaterials, **21**,2347 2000.
5. Palsson, B., Hubbell, J.A., Plonsey, R., Bronzino, J.D., Tissue Engineering, CRC Press, London, UK, 2003.
6. Anderson J.M., Cardiovasc. Pathol. **2**, S33, 1993.
7. Bao, G., J Mech. Phys. Solids, **50**, 2237 2002.
8. Geiger B., Bershadsky A.D., Cell, **110**, 139 2002.
9. Balaban N.Q., Schwarx U.S., Riveline D., Goichberg P., Tzur G., Sabanay I., Mahalu D., Safran S.A., Bershadsky A.D., Addadi L., Geiger B., Nature Cell Biol. **3** 466, 2001.
10. Safran S.A., Gov N., Nicolas A., Schwarz U.S., Tlusry T., Physica A, **352**, 171, 2005
11. Bray D., *Cell Movements: From Molecules to Motility* 2nd edn, Garland Science ed., New York, 2001.
12. Diambra L., Cintra, L.C., Schubert, D., and da Cinta, L.F., http://xxx.lanl.gov/q-bio.CB/0503013, 2005.
13. Fletcher D.A., and Theriot J.A., Phys. Biol. **1**, T1 2004.
14. Jones C.J. and Aizawa S., Microb. Physiol. **32,** 109 1991.
15. Purcell E.M., Am. J. Phys. **45**, 3 1977.
16. Wolgemuth, Hoiczyk E., Kaiser D., and Oster G., Curr. Biol. **12**, 369 2002
17. Howard J., *Mechanics of Motor Proteins and the Cytoskeleton*, 1st edn, Sinauer Associates, New York 2001.
18. Schliwa M., and Woehlke G., Nature **422**, 759 2003.
19. Ait-Haddou R., and Herzog W., J. Electromyogr. Kinesiology, **1**2, 435 2002.
20. Lipowsky R. and Klumpp S., Physica A, **352**, 53 2005.
21. Schott D.H., Collins R.N., and Bretscher A., J. Cell Biol., **156**, 35 2002.
22. D'Acunto M., unpublished.
23. Berg H.C., Ann. Rev. Biochem. **72**, 19 2003.
24. Kojima S. Yamamoto K., Kamagishi I. and Homma M., J. Bacteriol. **181**, 1927 1999.
25. Theriot J.A., Traffic, **1**, 19 2000.
26. Filippov A.E., Klafter J., Urback M., J. Physics: Condens. Matter **17**, S3929 2005.
27. Bretschneider T., Vasiev B., and Weijer C., J. Theor. Biol. **199**, 125 1999.
28. Rieu J.P., Barentin C., Sawai S., Maeda Y., and Sawada Y., J. Biol. Phys., **30**, 345 2004.
29. Rieu J.P., Barentin C., Maeda Y., Sawada Y., Biophys. J., **89**, 3563 2005.
30. Chamaraux F., Fache S., Bruckert and Fourcade B., Phys. Rev. Lett., **94**, 158012 2005.
31. Ambrose E.J., Exp. Cell Res. **8**, 93 1961.
32. Curtis A.S.G., J. Cell Biol. **20**, 199, 1964.
33. Owen, G.R., Meredith, D.O., Gwynn, I., Richards, R.G., Europ. Cells Mater, **9**, 85, 2005.
34. Chicurel M.E., Chen C.S., Ingber D.E., Curr. Opin. Cell Biol. **10**, 232, 1998.
35. Chicurel M.E., Singer R.H., Meyer C.J., Ingber D.E., Nature **392**, 730, 1998.
36. Lo, C.M., Wang, H.B., Dembo, M., Wang, Y.L., Biophys. J, **79**, 144 2000.
37. Schwarz, U.S., Balaban, N.Q., Riveline, D., Addadi, L., Bershadsky, A., Safran, S.A., Geiger, B., Mater Sci Eng C, **23**, 387, 2003.

38. Hu, H., Sachs, F., J. Mol. Cell Cardiol., **29**, 1511, 1997.
39. Koonce M.P., Kölher J., Neujahr P., Schwartz J.M., Tikhonenko I., Gerish G., EMBO J., **18**, 6786 1999.
40. Riveline D., Zamir E., Balaban N.Q., Schwarz U.S., Ishizaki T. Narumiya S., Kam Z., Geiger B., Bershadsky A.D., J. Cell Biol. **153**, 1175 2001
41. Merkel R., Nassoy P., Leung A., Ritchie K., Evans E., Nature **397**, 50 1999.
42. Bischofs I.B., Schwarz U.S., Proc. Natl. Acad. Sci. **100**, 9274 2003.
43. Bischofs I.B., Safran S.A., Schwarz, U.S., Phys. Rev E, 69, 012911 2004.
44. Schwarz U.S., Erdmann T., Bischofs I.B., Biosystems, in press.
45. Bell G.I. Science, **200**, 618 1978.
46. Bio-Implant Interface: Improving Biomaterials and Tissue Reactions, CRC Press, 2003.
47. Peter S.J., Miller M.J., Yasko A.W., Yaszemski M.J., Mikos A.G., J. Biomed. Mater. Res., **43**, 422, 1998.
48. Hutmacher D.W., Biomaterials, **21**, 2529, 2000.
49. Curtis A., and Riehle M., Phys. Med. Biol., **46**, R47 2001.
50. Dillow A.K., Lowmann A.M., Biomimetic Materials and Design, M. Dekker Publishing, New York 2002.
51. Weiss P., J. Exp. Zool. **100**, 353 1945.
52. Schwarz U.S. and Bishofs I.B., Med. Eng. Phys., **27**, 763, 2005.
53. Wong JY., Velasco A., Rajagopalan P., Pham Q., Langmuir **19**,1908 2003.
54. Meyer U., Büchter A., Wiesmann H.P., Joos U., Jones D.B., Eur. Cells Mater. **9**, 39, 2005.
55. K. Anselme and M. Bigerelle, Biomaterials, in press.
56. Persson B.N.J., Albohr O., Tartaglino U., Volokitin A.I., Tosatti E., J. Phys., **17**, R1, 2005.
57. Johansson F., Carlberg P., Danielsen N., Montelius L., Kanje M., Biomaterials, **27**, 1290, 2006.
58. Winkelmann M., Gold J., Hauert R., Kasemo B., Spencer N.D., Brunette D.M., Textor M., Biomaterials, **24**, 1133, 2003.
59. Curtis A. and Wilkinson C., Trends Biotechnol., **19**, 97, 2001.
60. Dalby M.J., Yarwood S.J., Riehle M.O., Johnstone H.J.H., Affrossman S., Curtis A.S.G., Exp. Cell Res. **276**, 1, 2002.
61. D'Acunto M., Napolitano S., Pingue P., Beltram F., Giusti P., Rolla P., to be published.
62. Dillow A., Ochsenhirt S.E., McCarthy J.B., Fields G.B., Tirrell M., Biomaterials **22**, 1493 2001.
63. Johnson K.L., Kendall K., and Roberts A.D., Proc. R. Soc. London A, **324**, 301 1971.
64. Ciardelli G., Rechichi A., Cerrai P., Tricoli M., Barbani N., Giusti P., Molecular Symp., **218**, 261, 2004.
65. D'Acunto M., Ciardelli G., Narducci P., Rechichi A., Giusti P., Mater. Lett., **59**, 1627 2005.
66. D'Acunto M., in Nanophysics, Nanocluster and Nanodevices, Nova Science Publisher, in press
67. Cloquet D., Felsenfeld D.P., and Sheetz M.P., Cell, **88**, 39 1997.
68. Briunsma G., Behrisch A., and Sackmann E., Phys. Rev E, **61**, 4253 2000.
69. John N., Linke M., Denker H.W., In Vitro Cell. Dev. Biol., **29A**, 461 1993.
70. Suter C.M., Errante L.E., Belotserkovsky V., and Foscher P., J. Cell Biol. **141**, 227 1998.

71. Radmacher M., Fritz M., Kacher C.M., Cleveland J.P., Hansma P.K., Biophys. J., **70**, 556 1996
72. Domke J., Dannöhl S., Parak W.J., Müller O., Aicher W.K., and Radmacher M., Colloids Surf. B, Biointerfaces, **19**, 367 2000.
73. Evans E. A., Biophys. J., **48**, 185 1985.
74. Evans E., in Handbook of Biological Physics, Lipowsky R. and Sackmann E. eds., Elsevier Science, 723, 1995.
75. Curtis A.S.G., Symp. Zool. Soc. London, **25**, 335 1970.
76. Chen S., and Springer T.A., J. Cell Biol. **144**, 185 1999.
77. Florin E.L., Moy V.T., Gaub H.E., Science, **264**, 415 1994.
78. Willemsen OH, Snel MM, van der Werf KO, de Grooth BG, Greve J, Hinterdorfer P, Gruber HJ, Schindler H, van Kooyk Y, Figdor G., Biophys. J., 75, 2220 1998.
79. Atomic Force Microscopy in Cell Biology, Jena B.P. and Horber J.K.H. eds., Academic Press Amsterdam, 2002.
80. Butt H.J., Cappella B., and Kappl M., Surf. Sci. Rep., **59**, 1, 2005.
81. Henderson E., Haydon P.G., Sakaguchi D.S., Science, **257**, 1944 1992.
82. Rotsch C., Braet F., Wisse E., and Radmacher M., Cell Biol. Int. 21, 685, 1997.
83. Braet F., Rotsch C., Wisse E., and Radmacher M., Appl. Phys. A, 66 S575, 1997.
84. A-Hassan E., Heinz W.F., Antonik M. D., D'Costa N.P., Nagaswaran S., Schoenenberger C.A., and Hoh J.H., Biophys. J., **74**, 1564, 1998.
85. Rotsch C., and Radmacher M., Biophys. J., **78**, 520, 2000.
86. Lekka M., Laidler P., Gil D., Lekki J., Stachura Z., and Hrynmiewicz A.Z., Eur. Biophys. J., **28**, 312, 1999.
87. Sato N.K. Kataoka, N., Sasaki M., and Hake K., J. Biomech. **33**, 127 2000.
88. Mahaffy R.E., Park S., Gerde E., Käs J., Shih C.K., Biophys. J., **86**, 1777 2004.
89. Beningo K.A. and Wang Y.L., Trends Cell Biol. **12**, 79 2002.
90. Dembo M. and Wang Y.L. Biophys. J. **76**, 2307 1999.
91. Wang J.H.C., Goldschmit-Clermont P., Wille J., Yin F. C.P., J. Biomech. **34**, 1563 2001.
92. McGerry J.P., Murphy B.P., McHugh P.E., J. Mech. Phys. Solids, **53**, 2597 2005.
93. Yamamoto A., Mishima S., Maruyama N., Sumita M., Biomaterials, **19**, 871 1998.
94. Athanasiou K.A., Thoma B.S., Lanctot D.R., Shin D., Agrawal C.M., LeBaron R.G., Biomaterials, 20, 2405 1999.
95. Hoben H., Huang W., Thoma B.S., LeBaron R.G., and Athanasiou K.A., Ann. Biomedical Eng. 30, 703 2002.
96. Wu C.C., Su H.W., Lee C.C., Tang M.J., Su F.C., Biochem. Biophys. Res. Comm., **329**, 256, 2005.
97. Canetta E., Leyrat A., and Verdier C, Math. Comp. Model., **37**, 1121 2003.
98. Thie M., Röspel R., Dettmann W., Benoit M., Ludwig M., Gaub H.E., Denker H.W., Human Repr. 13, 3211 1998.
99. Benoit M., Gabriel D., Gerisch G., and Gaub H.E., Nature Cell Biol., **2**, 313 2000.
100. Benoit M. in Atomic Force Microscopy in Cell Biology, Jena B.P. and Horber J.K.H. ed., Academic Press, Amsterdam, 2002

Index