INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®

DISSERTATION

USE OF SCANNING PROBE MICROSCOPIES TO STUDY DOPANTS AT SEMICONDUCTOR SURFACES

Submitted by
Mark W. Nelson
Department of Chemistry

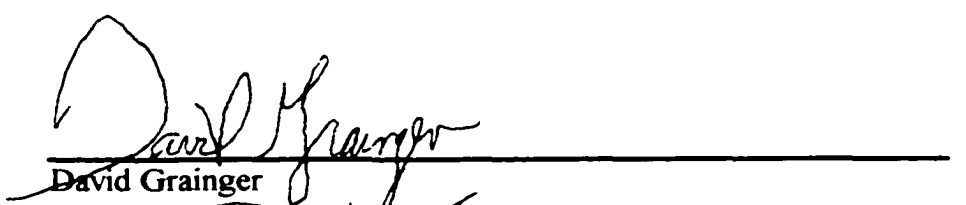
In partial fulfillment of the requirements
for the degree of Doctor of Philosophy
Colorado State University
Fort Collins, Colorado
Spring 2000

COLORADO STATE UNIVERSITY

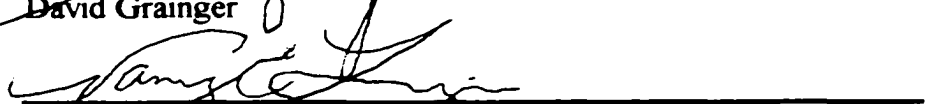
November 30, 1999

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY MARK W. NELSON ENTITLED: USE OF SCANNING PROBE MICROSCOPIES TO STUDY DOPANTS AT SEMICONDUCTOR SURFACES, BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work



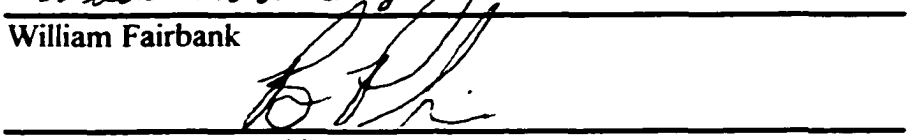
David Grainger



Nancy Levinger

Debbie Crans

William Fairbank



Adviser – Bruce Parkinson

Department Head

ABSTRACT OF DISSERTATION

USE OF SCANNING PROBE MICROSCOPIES TO STUDY DOPANTS AT SEMICONDUCTOR SURFACES

Dopants, in semiconductors, are detected as either protrusions or depressions in scanning tunneling microscopy (STM) images. Measured dopant heights for layered semiconductors are considerably larger than for conventional semiconductors. This is interpreted as the influence of dopant induced electrostatic forces between the tip and the sample leading to a structural deformation of the surface around the dopant atoms. To investigate the influence of electrostatic forces, we performed STM measurements on p-type MoS_2 at different bias voltages. The bias dependence of the STM images indicates the presence of electrostatic forces. Additional measurements with current imaging tunneling spectroscopy (CITS) show that changes in the density of states at dopant sites play only a minor role and cannot account for the large protrusions observed. Atomic force microscopy (AFM), with an applied D.C. voltage between the cantilever and sample, also confirms the role of electrostatic forces.

Recently, we developed a new Tappingmode® AFM (TMAFM) based dopant profiling method based on an electrostatic mechanism similar to the STM imaging of dopants in layered semiconductors. TMAFM with an applied bias was used to spatially resolve areas of different doping type and density on silicon patterned via ion implantation. The application of a D.C. bias between the cantilever and sample during the measurement results in a Coulomb interaction between the tip and sample, whose magnitude depends on the spatial variation in the doping density. This effect was utilized to detect areas differing in doping by monitoring the phase angle between the drive frequency and cantilever response while scanning over areas of differing doping density. Measurements at various bias voltages are presented to demonstrate that the phase contrast observed between differently doped areas is directly connected to the bias induced surface potential (band bending) present on these areas. A quantitative investigation of the contrast mechanism was performed by measuring deflection (force), amplitude, and phase versus distance curves for a typical cantilever under an applied bias, on a gold film.

Dopant profiles using BAAFM have been demonstrated on commercial integrated circuits in both a top-down and cross-sectional manner. Our experiments demonstrate that this method allows for distinguishing between p- and n-doped areas as well as distinguishing between areas of doping density ranging from at least 10^{16} to 10^{20} dopants/cm^3. From measurements on several commercial devices, it was determined that BAAFM is capable of resolving regions differing by 1/10 an order of magnitude at a doping density of 10^{17} cm^{-3}. Sample preparation is discussed and a comparison scanning capacitance microscopy (SCM) and BAAFM is made.

Mark W. Nelson
Department of Chemistry
Colorado State University
Fort Collins, CO 80523
Spring 2000

Acknowledgments

I am grateful to the Integrated Circuit Business Division of Agilent Technologies for partial financial support of this research. C.W. Almgren at Advanced Energy and J.J. Kopanski at NIST are gratefully acknowledged for providing the ion implanted silicon standard samples used in this work. I would also like to thank Digital Instruments for upgrading our scanning probe instrumentation. T.A. Hill and D.L. Hetherington at Sandia National Laboratory are acknowledged for providing insight into this study and for providing the SRAM samples that were investigated. I would like to thank P.G. Schroeder for his assisting with the work presented in Chapter 4. D.W. Niles at Agilent Technologies is gratefully acknowledged for his theoretical assistance pertaining to the distance dependence of the measured phase signal, and for performing photoelectron spectroscopy measurements on various cantilevers. The Parkinson Research Group at Colorado State University and Professor R. Schlaf at SUNY-Binghamton University (Department of Physics) are gratefully acknowledged for their support and for discussions pertaining to this work. Partial support for this work from the U.S. Department of Energy Division of Chemical Sciences under grant number DE-FG-96ER14625 is gratefully acknowledged.

Table of Contents

Index of Figures

Index of Tables

List of Acronyms

AFM: atomic force microscope
BAAFM: bias applied atomic force microscopy
CB: conduction band
CITS: current imaging tunneling spectroscopy
CMOS: complimentary-metal-oxide-semiconductor
CV: capacitance-voltage
D.C.: direct current
DOS: density of states
FET: field effect transistor
FIO: field-induced oxidation
HDS: hydrodesulfurization
HDY: dehydrogenation
HOPG: highly oriented pyrolytic graphite
IC: integrated circuit
LDOS: local density of states
MIS: metal-insulator-semiconductor
MOSFET: metal oxide semiconductor field effect transistor
NC-AFM: non-contact atomic force microscopy
NIST: National Institute of Standards and Technology
NTRS: National Technology Roadmap for Semiconductors
RF: radio frequency
RH: relative humidity
RIE: reactive ion etching
SCM: scanning capacitance microscopy
SEM: scanning electron microscopy
SIA: Silicon Industry Association
SIMS: secondary ion mass spectrometry
SKFM: scanning Kelvin force microscopy
SPM: scanning probe microscopy
SRAM: static random access memory
SRP: spreading resistance probe
STM: scanning tunneling microscopy
TCAD: technology computer aided design
TEM: transmission electron microscopy
TMAFM: Tappingmode® atomic force microscopy
TMD: transition metal dichalcogenide
UHF: ultra high frequency
UHV: ultra high vacuum
UPS: ultraviolet photoelectron spectroscopy
VB: valence band
VDW: van der Waals
VT-STM: variable temperature scanning tunneling microscopy
XPS: x-ray photoelectron spectroscopy

Objectives:

Electric fields are present between the tip and a surface being probed by scanning tunneling microscopy (STM).[1] These electric fields should result in an attractive electrostatic force between the probe and sample. Examples of force interactions in STM imaging include etching of layered materials at high tunneling biases,[2,3] controlled nanostructuring via STM,[4-7] and pulling up sample edges in layered materials.[8] Recently, we have observed that STM imaged dopants in layered semiconductors, such as MoS_2, appear as protrusions or depressions with respective heights or depths approaching 1 nm.[9,10] We believe that the observed protrusions and depressions are a result of bias induced band bending at the semiconductor surface, which results in a decrease in the screening of fixed dopant charges from the mobile charges in the in the layered material. These dopant charges produce image charges in the tip resulting in an attractive Coulomb force between the dopant and the tip. If stronger than the van der Waals forces holding the layers of the layered material together, the Coulomb force can lift the layer containing the dopant atom toward the surface. An alternative to the force based mechanism is that the dopant results in a change in the local density of states (LDOS). An increase in the LDOS could result in an increase in the tunneling current making the dopant appear to have a increased height relative to the surroundings.[11]

If Coulomb forces, present in STM, can be used to image individual dopants, is it possible to use a similar electrostatic force mechanism along and the atomic force microscopy (AFM) to image the doping in three-dimensional semiconductors? We would like to develop a new technique for probing doping density variations on silicon surfaces based on Tappingmode® AFM.

References:

[1] R. Weisendanger, *Scanning Probe Microscopy and Spectroscopy: Methods and Applications* (Cambridge University Press, New York, 1994).

[2] E. Delawsky and B. A. Parkinson, J. Am. Chem. Soc. **114**, 1661 (1992).

[3] B. A. Parkinson, J. Am. Chem. Soc. **112**, 7498 (1990).

[4] T. Schimmel, R. Kemnitzer, J. Küppers, H. Fuchs, and M. Lux-Steiner, Thin Solid Films **254**, 147 (1995).

[5] A. Sato and Y. Tsukamoto, Adv. Mater. **6**, 79 (1994).

[6] I.-W. Lyo and P. Avouris, Science **253**, 173 (1991).

[7] D. M. Eigler and E. K. Schweizer, Nature **344**, 524 (1990).

[8] R. Schlaf, P. G. Schroeder, M. W. Nelson, R. Stübner, S. Tiefenbacher, H. Jungblut, and B. A. Parkinson, Thin Solid Films **331**, 203 (1998).

[9] R. Schlaf, D. Louder, M. W. Nelson, and B. A. Parkinson, J. Vac. Sci. Technol. **A15**, 1466 (1997).

[10] T. W. Matthes, C. Sommerhalter, A. Rettenberger, P. Bruker, J. Boneberg, M. C. Lux-Steiner, and P. Leiderer, Appl. Phys. A **66**, S1007 (1998).

[11] D. A. Bonnel, in *Scanning Tunneling Microscopy and Spectroscopy: Theory, Techniques, and Applications* (VCH-Verlagsgesellschaft, Weinheim, 1993).

Chapter 1: Introduction to Semiconductors in Integrated Circuits

1.1 Drive for investigating doping

Doping involves the controlled introduction of certain impurity atoms into a semiconductor lattice to alter the material's electronic properties. Typically dopants are introduced by epitaxial growth or via an ion implantation process (followed by annealing).[1] These processes allow for a great deal of precision in the lateral dopant placement and in the doping density attained. Areas of different doping densities are placed side by side or in subsequent layers during microelectronic device fabrication processes. These adjacent areas of varied doping produce energy barriers to current flow in devices. Altering the height of these doping-induced barriers, via electrical biasing, results in the switching, signal amplification, and charge storage properties that are required to produce microelectronic devices. Since electrical properties of doped regions have a huge impact on the operation of semiconductor devices such as microprocessors, memory circuits, and optoelectronic devices, there is often a need to profile the number of charge carriers or the local doping density in semiconductor devices.

Development of analytical tools for high spatial resolution two-dimensional dopant profiling of semiconductor devices is becoming increasingly important as device sizes continue to shrink. Techniques such as secondary ion mass spectrometry (SIMS)[2,3] spreading resistance profiling (SRP)[4], and capacitance-voltage (CV)[4] measurements provide dopant or carrier concentration information with high precision and accuracy. These techniques, however, lack the spatial resolution required to analyze modern devices. Several high spatial resolution two-dimensional dopant profiling techniques based on the atomic force microscope (AFM) have recently been introduced to the semiconductor industry.[5-8] Each of these new techniques has its advantages as well as limitations, however, they are quickly becoming invaluable tools for the development, manufacturing, and testing of semiconductor devices.

Measuring doping profiles is essential for all aspects of semiconductor manufacturing. Process development teams use doping information to improve and calibrate models used for process design. Doping information is used by failure analysis teams for determining if device failures are related to doping-related problems such as a blocked implantation, multiple dosing, or diffusion/annealing temperature related problems. In the remainder of this chapter, basic concepts such as band formation, doping, the Fermi level dependence on doping, and basic device operation will be developed. These concepts are essential for understanding microelectronic device operation and will also aid in understanding dopant detection mechanisms used in AFM based dopant profiling techniques. Deeper understanding of AFM[9] and semiconductor physics[4,10] related concepts can be attained from several topical textbooks.

1.2 Bands and Doping

A central concept underlying the description of the electronic structure of solids is that valence electrons, donated by atoms in the solid, can spread throughout the entire structure. This concept is usually discussed in a more formal sense by treating the problem as an extension of molecular orbital theory in which the solid is viewed as an infinite molecular lattice.[11] Overlap of atomic orbitals leads to the formation of molecular orbitals where the number of states provided by the atomic orbitals are conserved in the molecule. The overlap of a large number of atomic orbitals leads to molecular orbitals that are closely spaced in energy and thus form a continuous band covering a range of energies.

Band formation can be understood by considering a line of atoms containing an s-orbital which overlaps the s-orbital of its immediate neighbor (Figure 1.1(a)).[10] When the line consists of only two atoms the atomic orbitals are split into a bonding and antibonding molecular orbital. Introduction of a third atom in the line can result in three orbitals: bonding, antibonding and nonbonding. As more atoms are introduced into the line of atoms, each one contributes an atomic orbital and one more molecular orbital is formed. When there are N atoms in the line, the number of molecular orbitals equals N. As N approaches infinity, which is the case for a solid, the number of molecular orbitals within a finite energy range approaches infinity and the energy separation between orbitals approaches zero. This results in formation of a band (Figure 1.1(a)).

The energetic width of the band depends on the strength of the interaction between neighboring atoms. Stronger interactions between neighbors results in a greater energy width of the band. Bands can be constructed from atomic orbitals other than s-

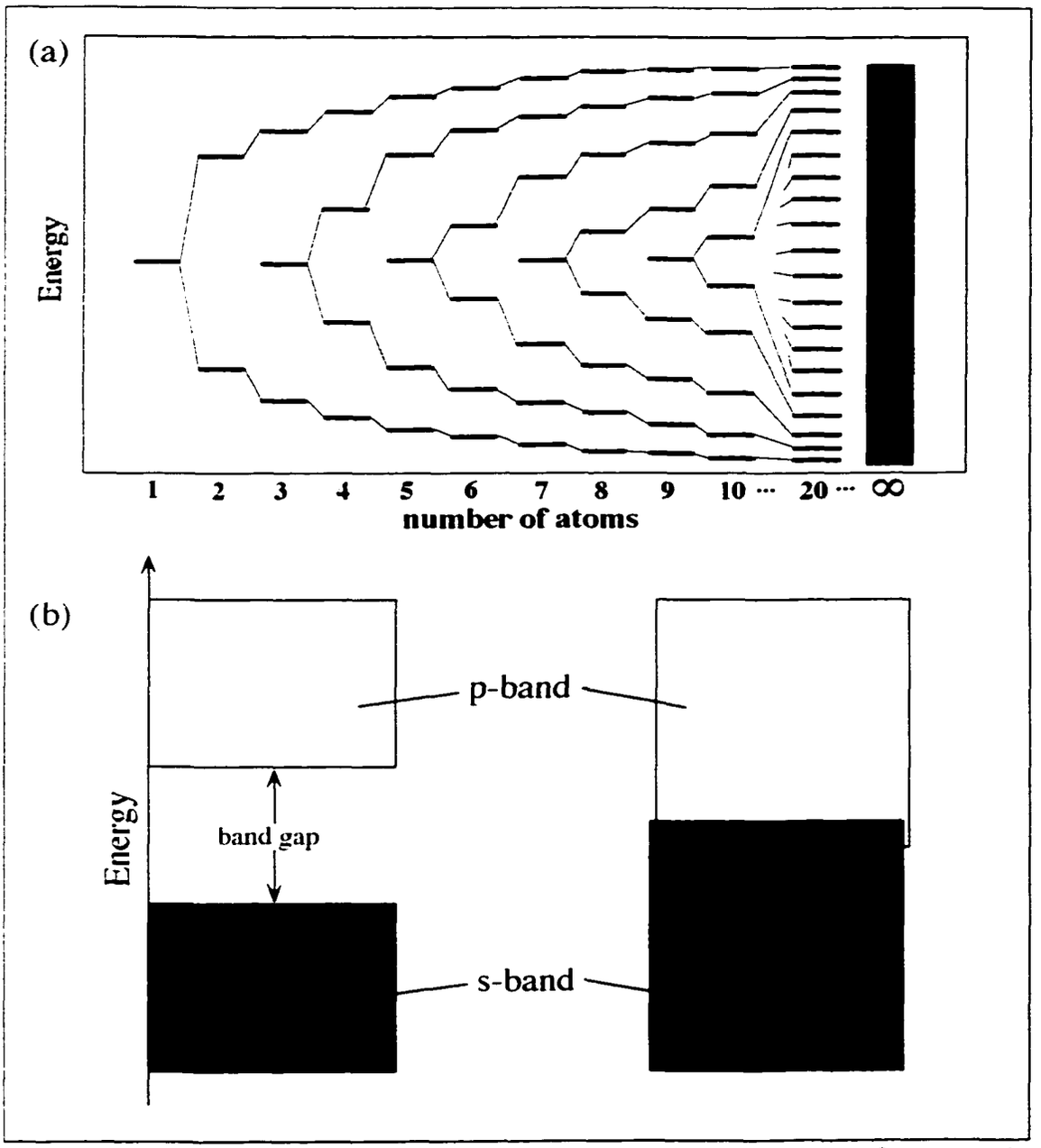


Figure 1.1: (a) Scheme indicating the formation of bands (right) from atomic orbitals (left). As the number of atoms in the solid increase the atomic orbitals from each of the individual atoms adds to the solids molecular orbital. (b) The left diagram shows how the formation of a band gap would be observed or in an insulator and the right image shows the overlap of bands as would occur in a metal.[4,11]

orbitals resulting in p-bands and d-bands. There are often large gaps in the ground and excited states in a material. When incorporated into a solid this often results in an energy gap between bands called a band gap (Figure 1.1(b)).[10] If however the bands span a large energy range and the atomic orbitals forming the bands are close in energy the bands will overlap (Figure 1.1(b)).

The number of states per energy unit is defined as the density of states (DOS). Since there are more energy levels at some energies than others, the DOS is typically not uniform across a band.[4] Within the band gap the DOS should theoretically be zero, meaning that there are no energy levels within the gap. Semiconductors are solids in which all of the states below the band gap are full and those above the band gap are empty at absolute zero in temperature. Since all of the states in the valence band (below the band gap) are full there are no empty states for electrons to migrate through the solid. At temperatures above absolute zero the probability of promoting electrons from the valence band into the conduction band is determined by Fermi-Dirac statistics.[4] As the temperature increases it becomes more probable that more electrons will be thermally promoted into the conduction band. Promotion of electrons into the conduction band results in unpaired electrons in both the valence and conduction bands providing empty states in both bands into which the electrons can move, thus allowing for electrical conduction.

Semiconductors containing zero states in the band gap only exist theoretically and are called intrinsic semiconductors.[11] Intrinsic semiconductors result from semiconducting solids, which contain absolutely no impurity atoms or defect sites (Figure1.2(a)). Impurity atoms and defects result in the introduction of impurity states

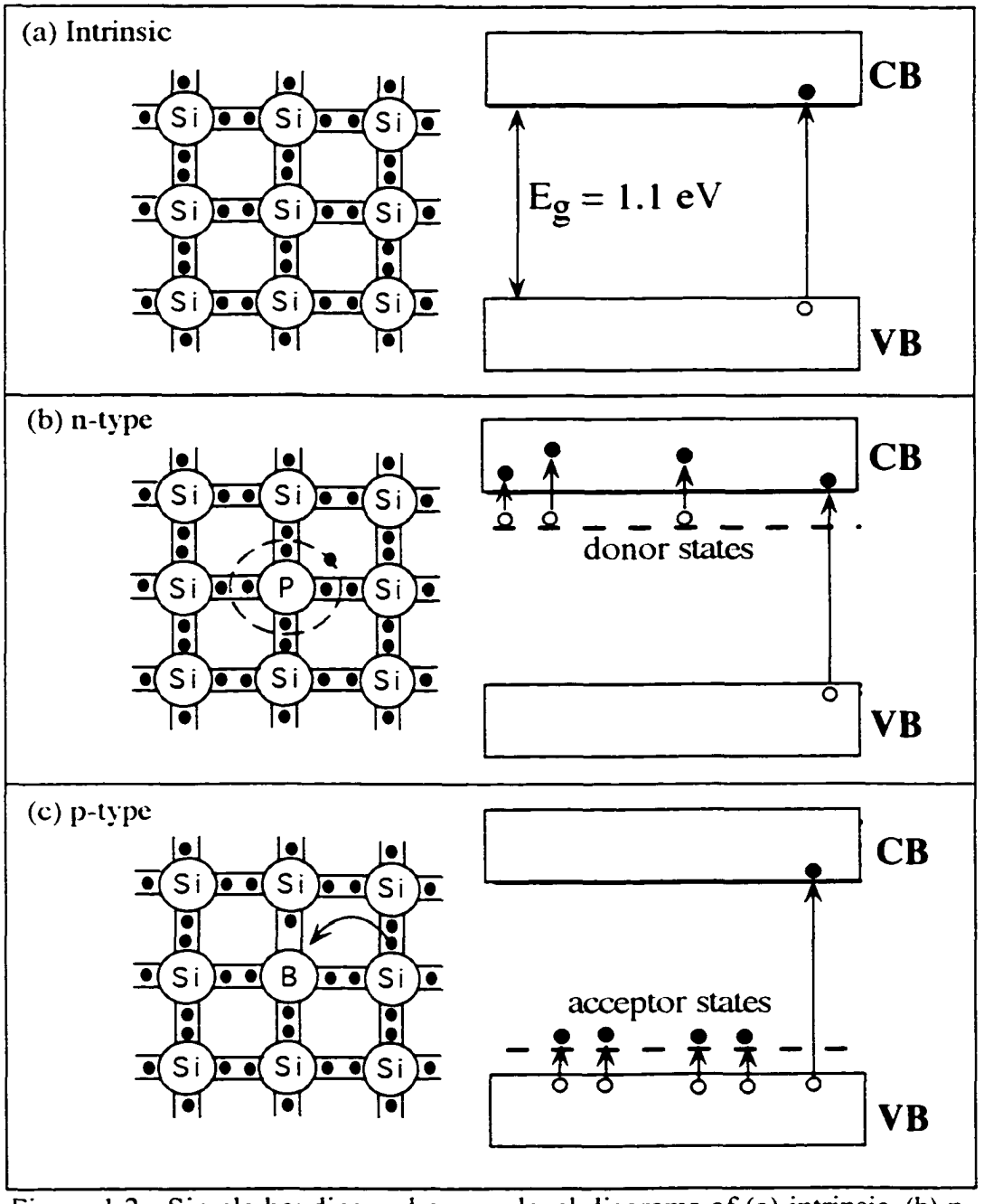


Figure 1.2: Simple bonding and energy level diagrams of (a) intrinsic, (b) n-type, and (c) p-type silicon. The conduction band (CB) and valence band (VB) edges are separated by a 1.1 eV band gap (E_g).

into the band gap of the semiconductor.[10] Doping involves the intentional introduction of certain impurity atoms to the crystal lattice for the purpose of increasing the number of charge carriers in the semiconductor. Doping arsenic or phosphorus atoms into the silicon lattice results in the introduction of impurity states which lie 0.054 eV and 0.045 eV below the conduction band edge, respectively (Figure 1.2(b)).[10] These states are called donor states because they introduce one additional electron to the lattice for each additional dopant atom. Since these levels lie energetically close to the conduction band edge, the electrons associated with the dopants are easily thermally promoted into the conduction band. At room temperature a large fraction of the electrons from the donor states populate the conduction band and can conduct charge through the semiconductor. This means that the carrier concentration is roughly equal to the doping density. Doping with As or P is referred to as n-type doping, the "n" indicates that the majority charge carriers are negative electrons.

An alternative to doping with n-type atoms is to introduce group III atoms to the crystal lattice, such as boron, which have fewer valence electrons than silicon (Figure 1.2(c)). This type of doping is referred to as p-type since the holes represent positively charged carriers. Addition of p-type dopants results in the introduction of impurity states that lie 0.045 eV above the silicon valence band edge for boron impurities and 0.067 eV for aluminum.[10] These impurity states are empty at absolute zero. However, at room temperature and above a large portion of these stated are populated by electrons from the valence band. As the dopant levels fill, "holes" are created in the valence band. These holes or empty states are sites that neighboring electrons can fill thus allowing for conduction. As was the case with the n-type dopants the number of charge carriers is

virtually equal to the doping concentration. With n-type materials the majority carriers are electrons while holes are the majority carrier in p-type materials.

1.3 Fermi Level, Band Bending, and Junctions

An important concept for describing semiconductor systems is that of the Fermi level (E_F), defined as the energy where the probability of a state being occupied by an electron is ½.[11] The population probability (P) of the energy levels is given by the Fermi-Dirac distribution (Equation 1.1), a version of the Boltzman distribution, taking into account the Pauli exclusion principle:

$$P = \frac{1}{1 + e^{(E-E_F)/kT}} \tag{1.1}$$

The quantity E_F is the Fermi level, which in doped semiconductors depends on temperature.[4]

For intrinsic semiconductors, the Fermi level lies half way between the conduction and valence bands within the band gap region. Figure 1.3 shows how the Fermi-Dirac distribution changes with the introduction of dopant atoms. For an intrinsic semiconductor neither electrons or unfilled levels exist in the vicinity of the Fermi level. In doped semiconductors, the location of the Fermi level depends on the doping density. As dopants are introduced to the semiconductor the probability of finding an electron at a given energy level changes. In n-type materials energy levels are introduced near the conduction band that contain some electrons. Since these energy levels contain electrons and have also donated some electrons to the conduction band the probability of finding an electron at energies close to the conduction band increases, thus shifting the Fermi-Dirac distribution up in energy (Figure 1.3(b)). P-type materials introduce empty impurity

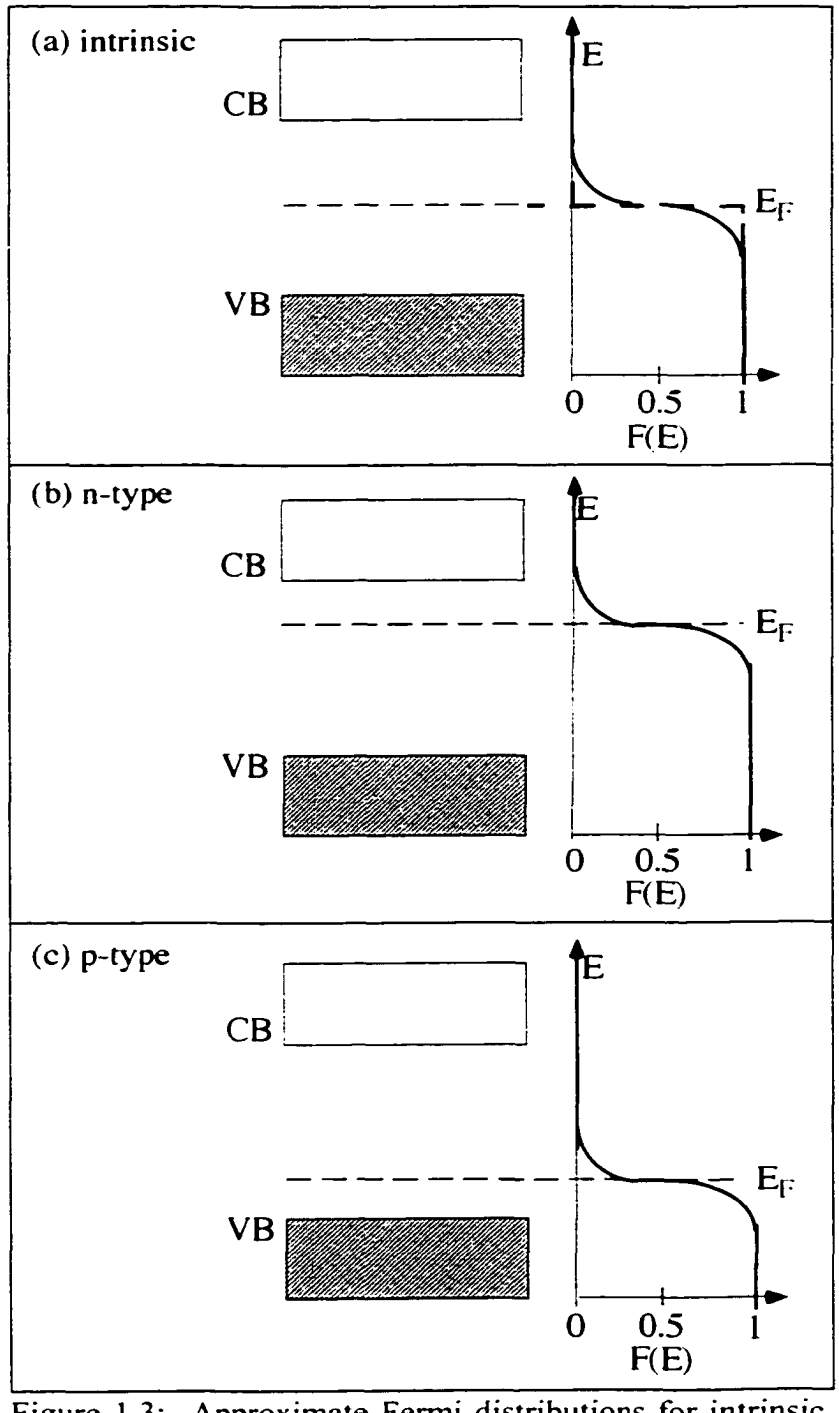


Figure 1.3: Approximate Fermi distributions for intrinsic, n-type, and p-type semiconductors. Fermi level positions are indicated by the dotted line and occur where the Fermi distribution passes through ½.[11] Note that the Fermi energy changes as a function of doping type and density.

states near the valence band. These states are easily populated at room temperature by valence band electrons, thus increasing the probability of finding an electron near the valence band energy. The Fermi level thus moves toward the valence band (Figure 1.3(c)).

Figure 1.4 shows the calculated Fermi level position for silicon at various temperatures and doping levels. The Fermi level of n-type materials lies above that of the intrinsic Fermi level (E_i), which is directly in the center of the band gap. P-type semiconductors have Fermi levels lying lower in energy than the intrinsic Fermi level (E_i). At low doping densities the Fermi level lies near the intrinsic Fermi level (E_I). As the doping density is increased E_F moves closer to the valence band for p-type semiconductors and closer to the conduction band for n-type semiconductors.

Fermi energies are important for predicting the behavior of semiconductor systems. When two materials are brought, together their respective Fermi energies determine what will occur electronically and energetically. Upon electrical contact of two materials Fermi level equilibration will occur, making the electron potential equal across the two materials.[10] Biasing one material relative to the other causes the Fermi levels of the two materials to shift in energy relative to one another. We will look at a biased metal-semiconductor junction as an example of how altering the Fermi level can change a semiconductor's electrical behavior.

Junctions occur at the interface between two electronically different materials. When considering semiconductor junctions, there are several basic conditions that can exist within the semiconductor near the junction. These conditions are depletion, accumulation, flat band and inversion. All of these conditions occur after contact and

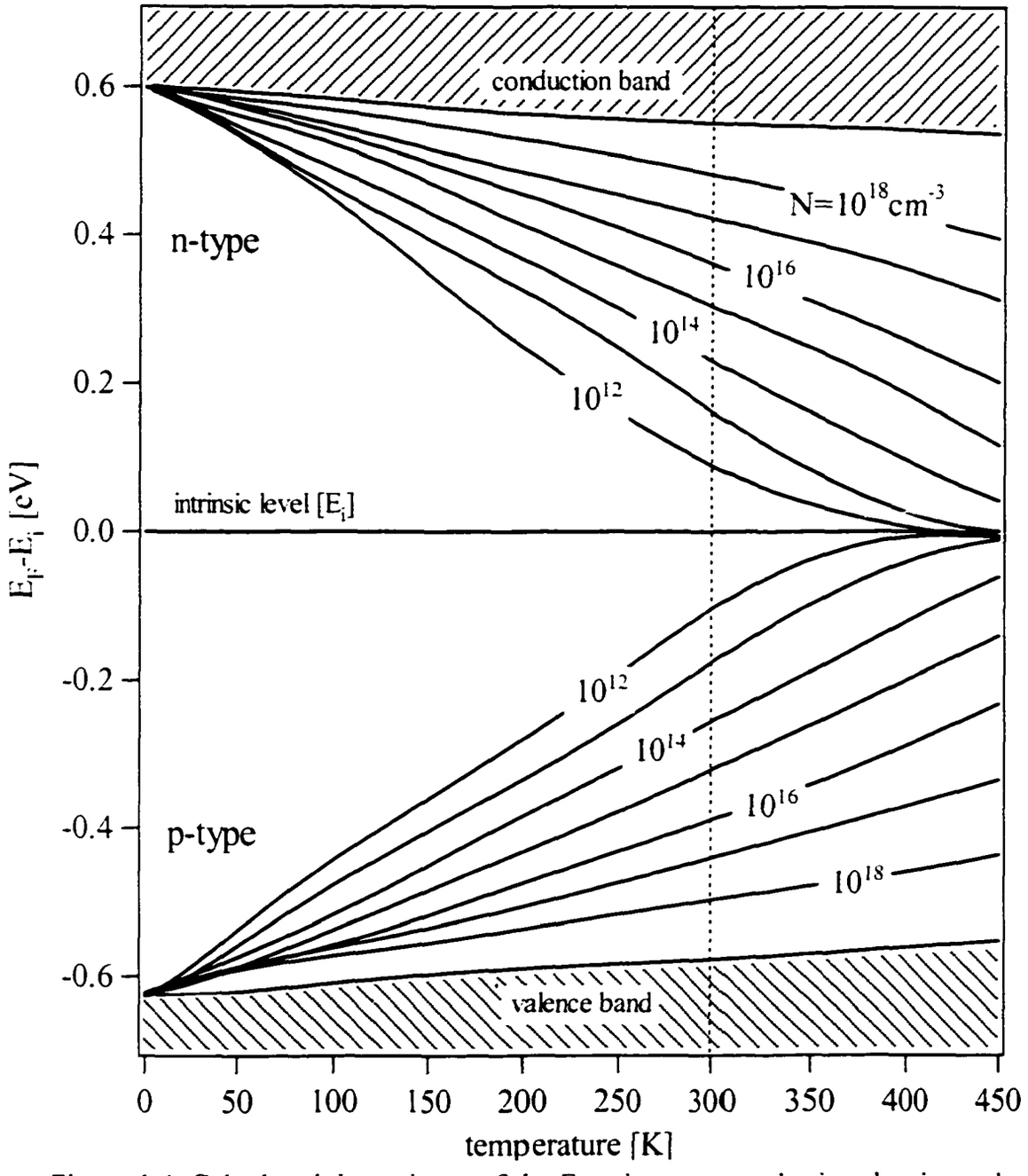


Figure 1.4: Calculated dependence of the Fermi energy on doping density and temperature.[4] Note that the Fermi energy for n-type materials lies above the intrinsic Fermi energy (E_i) while the p-type Fermi energy lies below E_i.

electronic equilibration of two materials (one being a semiconductor) with different work functions or as a result of electrically biasing these materials. Figure 1.5 shows several potential energy diagrams of a biased metal-semiconductor (n-type) interface depicting these conditions. In this case the two materials are separated by a small insulator gap. However, it is assumed that charge carriers can readily cross the gap allowing for electrical equilibration (for all cases in Figure 1.5 except for (a)). When isolated the metal and semiconductor's Fermi levels are not equilibrated (Figure 1.5(a)). While isolated, there is an offset in the Fermi energies, of the materials, which is equal to the work function difference (ϕ_S-ϕ_M) between the semiconductor (ϕ_S) and the metal (ϕ_M). When the two materials are brought into electrical contact, Fermi level equilibration occurs. Depending on the work function difference of the materials, electrical contact can result in major or minor changes in the semiconductor at the surface or junction interface. In the case of Figure 1.5(b) equilibration results in a slight amount of band bending which places the semiconductor into a lightly depleted state. Band bending occurs when an electric field is set up in the semiconductor as a result of a potential gradient between the surface and the bulk of the semiconductor.[10] If the bands are bent downward, from the surface into the bulk, the electric field will force electrons away from the surface.[10] Sweeping majority carriers away from the surface, as is the case with the n-type material shown in Figure 1.5(b), results in depletion of surface charge. Carriers are essentially swept from the interface and ionized dopant atoms are left behind, fixed in their lattice sites.

Applying a positive bias to the semiconductor produces an upward shift in the Fermi level of the semiconductor with respect to the metal.[10] This results in bending of

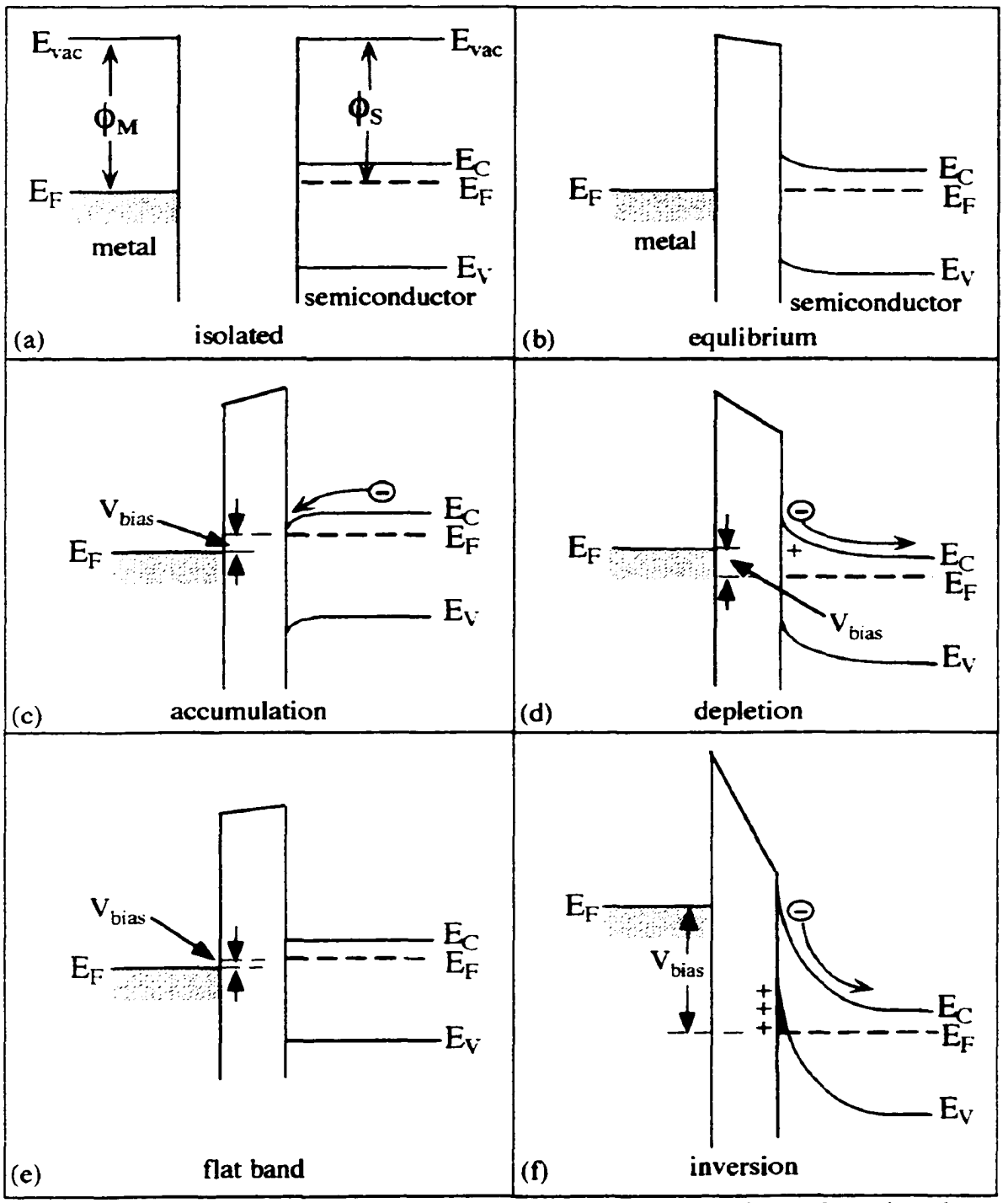


Figure 1.5: Band diagrams showing (a) an electrically isolated metal-semiconductor (n-type) junction and basic of conditions which occur at the surface of a biased semiconductor. These conditions include: (b) equilibrium, (c) accumulation, (d) depletion, (e) flat band, and (f) inversion.

the bands downward at the semiconductor surface, producing accumulation conditions, as is shown in Figure 1.5(c). Accumulation occurs in n-type materials when the conduction band energy of the interface or surface of the semiconductor is lower in energy than that of the bulk. Under these conditions an electric field is set up in the semiconductor which pulls carriers (electrons in an n-type material) from the bulk to the surface.[4] Charge carriers from the bulk accumulate at the surface, hence, the term accumulation. Accumulation conditions in p-type materials result from accumulation of holes at the surface as a result of band bending in a direction in which the surface conduction band is at a higher energy than in the bulk semiconductor.

Depletion in an n-type material occurs when a bias is applied which results in a shift of the Fermi level of the semiconductor down with respect to the metal (Figure 1.5(d)).[4] This results in the bending of the bands downward which results in the development of an electric field at the semiconductor surface. The field forces the electrons from the surface into the bulk of the semiconductor, leaving behind depleted dopant ions in the surface region. Band bending is greater in the case of the biased junction shown in Figure 1.5(d) than that of the equilibrium case shown in Figure 1.5(b). This results in deeper depletion of the surface charge for the case of the biased junction. The name "depletion" is arrived at since the surface is depleted of the mobile charge carrier under these conditions. Increasing the degree of band bending, via biasing the junction, results in a larger potential drop across the interface region. This in turn results in a larger electric field which can deplete charge from or accumulate charge at the surface more readily. Having no band bending in the interface is referred to as flat band

(Figure 1.5(e)). This condition marks the division between accumulation and depletion. Biasing to either side of flat band results in the onset of accumulation or depletion.

Inversion occurs in n-type materials when the Fermi level is forced through the valence band edge.[4] Under increased negative bias the surface region of the semiconductor seems to change in character from n-type to p-type. This results when the semiconductor is under inversion conditions and the minority carrier concentration at the surface exceeds the bulk majority carrier concentration. The region that is shaded red in Figure 1.5(f) represents an area in which the valence band is bent through the Fermi level. Since the valence band is greater in energy than the Fermi level virtually all of the majority carriers (electrons) are forced out of the red region. Holes flow into the area, making the surface region appear p-type instead of n-type. In a p-type semiconductor, inversion results in the surface appearing as n-type.

pn-junctions:

The pn-junction, located at the juncture between a p-type and an n-type material, is the basic building block for many semiconductor devices. The most important characteristic of pn-junctions is that they are rectifying (i.e. they permit electric current to pass in only one direction). When a negative bias is applied to the n-region of the junction, a current begins to flow at a very small applied bias. In contrast, when a positive bias is applied to the n-region, virtually no current flow occurs. Eventually, at large enough positive biases, current does begin to flow. This condition is referred to as junction breakdown and often leads to the destruction of the device.

Many of the important characteristics of a pn-junction are associated with a space charge region formed between the n-type and p-type regions. At the transition between

n- and p-type material a potential gradient is formed which separates the regions with high electron and hole concentrations.[10] The n-type and p-type materials before electrical contact are illustrated in Figure 1.6(a). As was described previously, an n-type semiconductor is characterized as containing a large concentration of electrons, while a p-type semiconductor contains a large concentration of holes.[11] If the two materials are brought together into intimate contact, a flux of electrons and holes will flow in directions in such a way as to even out the large concentration gradients existing between the two regions. When a net flow of electrons from the n-region to the p-region and holes from the p-region to n-region takes place, a "space charge" is formed due to the ionized donors and acceptors which are left behind since they are bound in the semiconductor lattice.[11] Thus, after a certain number of electrons and holes have flowed from one region to the other, an electric field will be set up which prevents further flow of carriers. At this point equilibrium conditions have been established (Figure 1.6(b)) and the space charge region is virtually void of mobile carriers.

At equilibrium, the Fermi level of the semiconductor is constant across the pn-junction resulting in a potential energy offset of the p-type and n-type material's bands. The total potential variation in the junction at equilibrium is equal to the sum of the absolute value of the Fermi level positions of the two isolated materials. This potential, referred to as the junction's built-in-voltage, produces a barrier to electron flow in the device ($E_{barrier}$). Under equilibrium conditions, electron-hole pairs are constantly thermally generated everywhere in the semiconductor. In the absence of an applied voltage virtually all of these electron-hole pairs recombine resulting in zero net current.

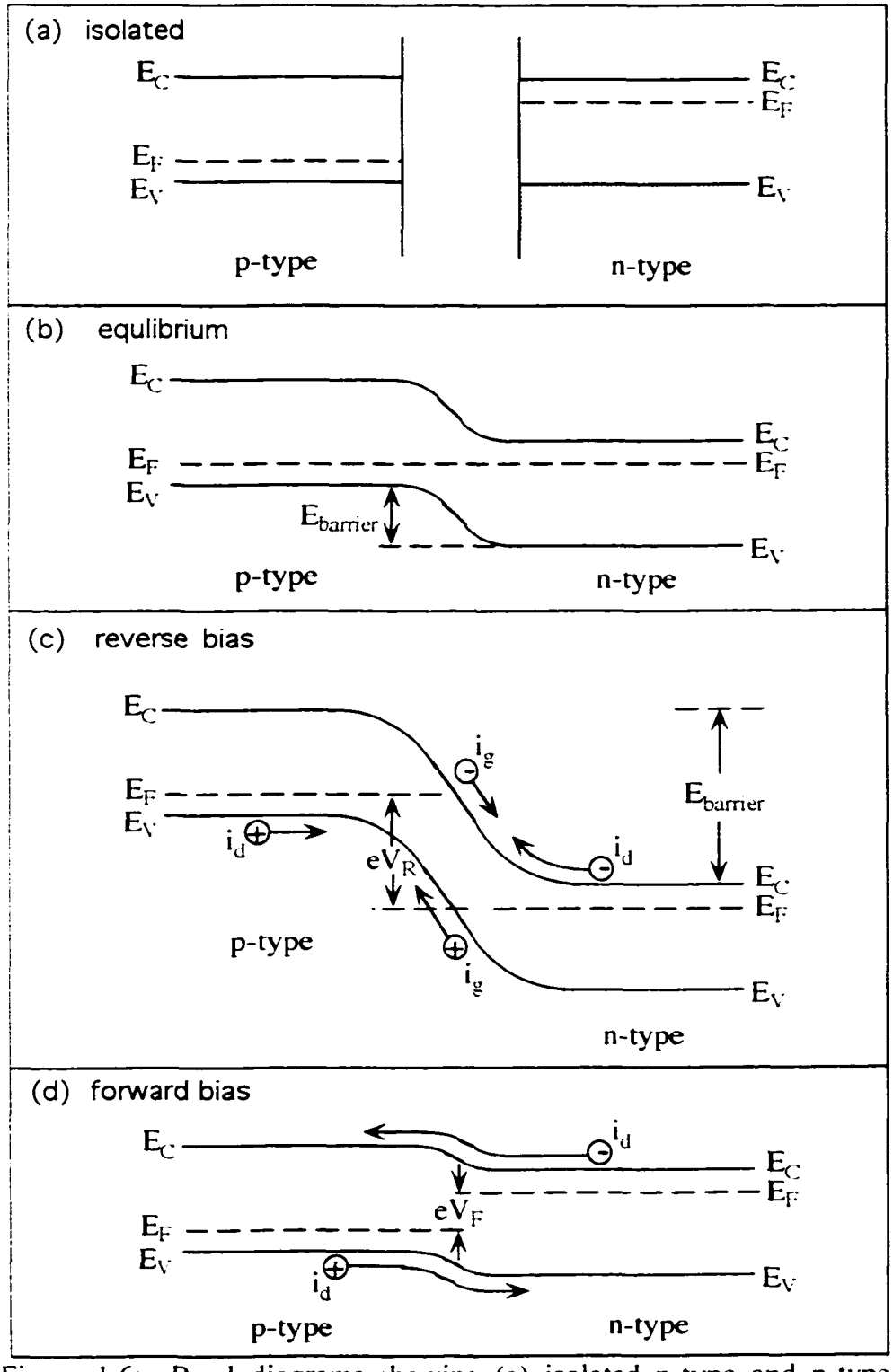


Figure 1.6: Band diagrams showing (a) isolated p-type and n-type semiconducting materials, (b) electrical equilibration of the two materials, (c) reverse biasing (d) and forward biasing of the a pn-junction.

If we apply a positive bias (V_R) to the n-region with respect to the p-region the total the band bending and resulting potential energy barrier will increase by eV_R, where e is the charge on an electron and V_R is the reverse bias potential.[10] This biasing condition is referred to as reverse bias (Figure 1.6(c)). Under reverse bias the electron-hole pair, once generated, is separated and their probability of recombination is diminished. When an electron hole pair is generated within the reverse bias depletion region the resulting current is referred to as the generation current (i_g).[10] The generation current results from carriers being swept energetically downhill, out of the depletion region. When carriers are generated in the neutral parts of the semiconductor they must diffuse to the reverse bias junction. This results in an additional current referred to as the diffusion current (i_d) which flows in the opposite direction of the generation current.[10] The total current in the reverse bias case results from the summation of the generation and diffusion currents. Since diffusion current must surmount the potential barrier the generation current dominates and a very small net back current appears in the circuit.

As a result of the increased band bending, due to reverse biasing, the thickness of the space charge region on both sides of the junction will increase. Since the depletion region contains very few charge carriers it is a region of high resistivity. A substantially increased depletion width means a substantially increased resistance, which is consistent with the small value of the reverse bias current. This back current remains small for increasing reverse biases until the breakdown bias is reached and a very large current onset occurs resulting from thermal instability, tunneling, and avalanche multiplication processes.

In Figure 1.6(d) a forward bias has been applied, resulting in an upward shift in the energy of the bands. This shift results in decrease in the energy barrier which, in turn, makes it easier for diffusion current to flow across the junction. The lowered barrier height results in an increased diffusion current, and a large net forward current results. As the potential barrier is decreased so does the space charge width. This results in a lowering of the junction resistance, which is consistent with the large forward current generated. In summary forward biasing of a pn-junction results in a very small reverse current, which remains virtually negligible at all reverse biases. Forward biasing results in a rapid increase in forward current with increasing forward bias. Thus, applying a bias of one polarity produces a substantial current while biasing with the opposite polarity results in a negligible current, a phenomena called rectification.

1.4 Simple Devices

Transistors are the basis for the switching logic in computers. In these devices a small bias signal applied to one of the device terminals can result in large changes in current passing through the other two terminals. The concepts of doping, Fermi level, band bending, and junctions, which were discussed above, are essential to understanding the operation of silicon based IC devices. These silicon IC devices are essential parts of practically all of today's electronic devices ranging from computers to medical equipment to household appliances.

All silicon based microelectronics devices are fabricated using photo-lithographic technology in which a light sensitive "photoresist" polymer is spin coated onto the surface of the wafer. Exposing the polymer to light through a patterned mask results in a change in the chemistry of the exposed polymer.[12] Development of the wafer produces

patterned regions on the semiconductor wafer's surface in which some areas still have photoresist on them and others expose the bare wafer surface. The wafers can now be run through one of a number of processing steps such as: oxidation, ion implantation, and material deposition.[12] During these steps only the bare wafer regions are affected by the processing step. Etching off the mask and repeating the process allows for patterning of different materials on top of each other. This eventually results in production of working microelectronic devices.

Field effect transistors (FET) are composed of a source, a gate, and a drain which are created by doping the silicon in each region of the device differently. The source essentially supplies charge carriers to the device and the drain is a collecting place or drain for these charge carriers. The gate, which sits between the source and drain, is responsible for controlling the amount of charge that flows from the source to drain. By modulating the voltage on the gate it is possible to control the current of carriers which flow through the device. It is this gating characteristic that makes signal amplification, charge storage in memory cells, and the binary logic of computers possible. The components of junction transistors (bipolar) which are analogous to the source, gate and drain in a FET are the emitter, base, and collector respectively. Operating principles of two of the most common types of IC devices (bipolar transistor and MOSFET) are discussed below.

Bipolar Devices:

One of the most important solid-state devices today is the junction transistor or bipolar transistor. Invention of the bipolar transistor brought about unprecedented growth of research in the area of solid state physics and engineering. The bipolar junction

transistor consists of two pn-junctions in series. A schematic of a bipolar device is shown in Figure 1.7. The two configurations for bipolar devices are p+, n, p and n+, p, n abbreviations for doped semiconductor regions that comprise the emitter, base, and collector, respectively. In the case of the pnp device, shown in Figure 1.7, the current is injected into a highly doped p-type emitter. Under proper biasing conditions the carriers, in this case holes, are swept into the n-type base and onto the lightly doped p-type collector. If the above mentioned bias condition is not met, the emitter-collector current will not flow.

Potential energy band diagrams pertaining to a pnp transistor are shown in Figure 1.7 (b &c). Figure 1.7 (b) shows how the bands behave under equilibrium conditions where the emitter base and collector contacts are tied together. Under these conditions the electrostatic potential varies from each region (emitter, base, collector) in such a way as to balance out the flux of electrons and holes due to their concentration gradients. This results in equilibration of the Fermi level across the device. Under these conditions holes injected at the emitter reach an energy barrier at the emitter base junction. The potential energy barrier for holes is opposite in energy to that for electrons.

When external voltages are applied to the contacts on the transistor the junctions may become forward or reverse biased. The most important biasing condition exists when the junction between the emitter and base is forward biased and the junction between the base and collector is reverse biased. A band diagram for this biasing condition is shown in Figure 1.7(c). Forward biasing of the emitter-base junction results in a decrease in the junction barrier height to holes. This results in the injection of a large number of holes into the base. If both junctions in the device are close to one another

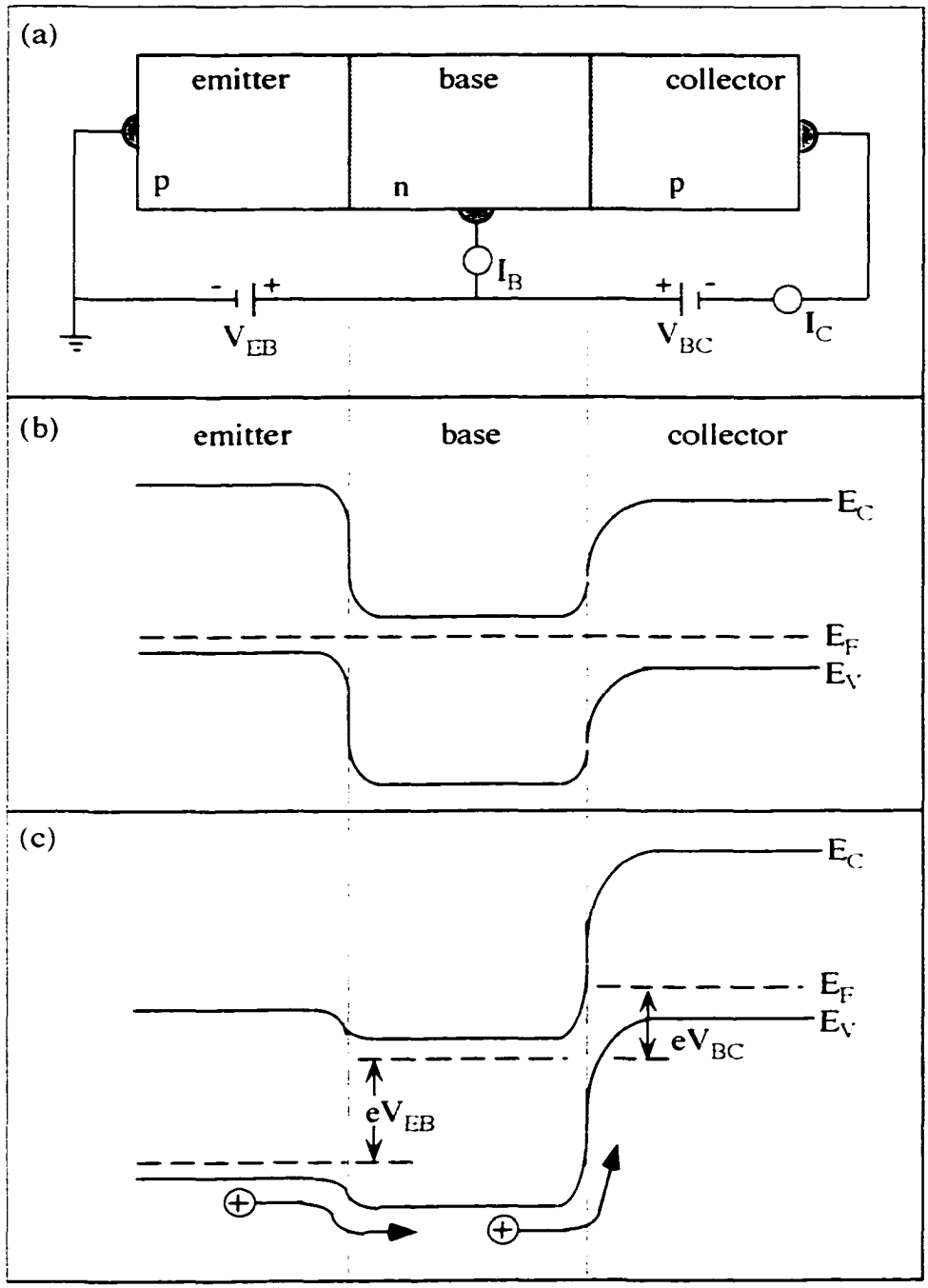


Figure 1.7: (a) Diagram indicating the basic design and operation of a pnp bipolar transistor. (b) Potential energy diagram of the device under equilibrium conditions. (c) Potential energy diagram of the device operating under bias. Holes are injected into and collected from the base.

most of the injected holes will reach the collector-base junction. The reverse bias on this junction produces an electric field in the junction region which results in a lowering of the barrier for holes to pass from the base to the collector. As holes reach the base collector junction they will be swept across under the electric field at the base-collector junction and will be collected in the p-type collector. Modulation of the bias conditions on the junctions results in modulation of the current making its way to the collector.

MOSFET Devices:

Another key component of today's microelectronics is the FET. Many types of FETs are used in modern devices, however, the metal-oxide-semiconductor field effect transistor (MOSFET) is among the most widely used of all field effect transistors. A schematic of the MOSFET device is shown in Figure 1.8(a). The operation of FETs is based on a bias-induced change in resistance of the semiconductor material under the gate. In an operating MOSFET device a conduction path between the source and drain called the "channel" is formed under the oxide by applying a bias to the metal gate (Figure 1.8(b)). Modulation of the gate voltage results in either pinching off or allowing current to flow through the device.[10] Pinching off current flow for a certain time is analogous to a 0 in the binary code. Allowing electrons to pass through the channel during the same duration gives a binary signal of 1. The clock speed of the device is determines the rate at which electrons can traverse the channel thus allowing the gate to open and close to create discrete on and off signals.

The process of channel formation can be understood by looking at band diagrams of the device shown in Figure 1.8 (c,d, &e)). When no bias is applied to the gate, the source-to-drain electrodes correspond to the two pn-junctions connected back to back.

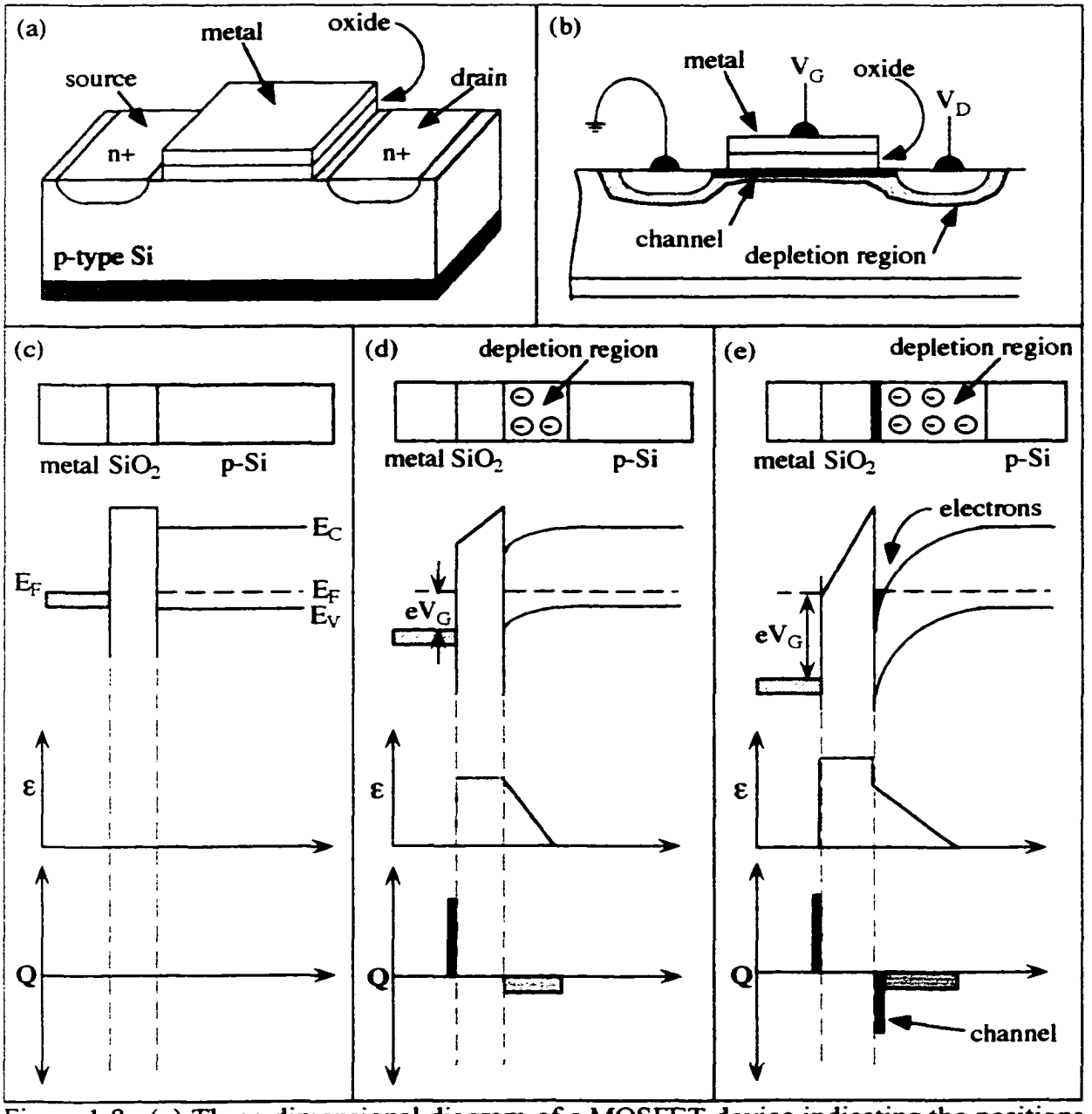


Figure 1.8: (a) Three dimensional diagram of a MOSFET device indicating the positions of essential parts such as the metal gate, gate oxide, source, and drain. (b) Two dimensional diagram showing a MOSFET's channel and depletion region under operating conditions. Band diagrams, electric field distribution, and block charge diagrams for a MOSFET under (c) flat band conditions, (d) a small positive bias, (e) and after the gate voltage has exceeded the threshold for channel formation.

Under these conditions the only current that can flow from source to drain is the reverse leakage current. Under a small positive bias, the p-type region under the gate oxide becomes depleted of mobile charge carriers (holes).[10] Holes are swept away from the interface under the force of the electric field (Figure 1.8(d)). This forms the beginning of the depletion region seen in Figure 1.8(b). An increase in the gate voltage results in more voltage appearing across the semiconductor and the bands bend until the free electrons accumulate at the semiconductor surface. The gate voltage at which this occurs is referred to as the "turn on" or "threshold" voltage. Under the inversion conditions produced at the threshold voltage, a negative sheet charge is formed at the semiconductor oxide interface. This inversion layer connects the source to the drain and forms a channel to current flow between the two regions.

1.5 Chapter Conclusion:

As can be seen by the above discussion, doping is essential to the operation of all semiconductor devices. Determining doping levels and junction positions is therefore an essential part of semiconductor device process design, manufacturing and failure analysis. Small errors in one of the doping related processing steps can often lead to device failure, resulting in the need to scrap large quantities of devices. Since most semiconductor fabrication facilities around the world are currently operating at maximum capacity and the demand for devices is great, yield loss can throw production schedules behind and drive up individual device costs.

A major goal of integrated circuit design is to pack devices more closely together and speed up their operation frequency without sacrificing cost or reliability. Device

operation speed depends on the transit time, the time required for an electron to move from source to drain in a FET.[10] For a gate length, L, the transit time, T_r, is

$$T_r = \text{length/velocity} = L/v. \quad (1.2)$$

Shorter gate lengths result in faster transit times and faster devices. As devices become smaller and gate lengths decrease, it is important to increase doping concentration in order to maintain a constant electric field strength across the junction. Current IC devices have gate lengths on the order of 250 nm.[13] If doping densities or junction positions are slightly out their specified range, device failure is eminent. This can easily happen during fabrication when a mask is misaligned, annealing temperatures are off, or an implant is blocked. Dopant profiling devices of this size requires a technique with spatial resolution on the order of 10-20 nm. As device sizes continue to shrink this resolution requirement will become even more strict. Currently, only techniques based on scanning probe microscopies show promise in achieving this spatial resolution.

Much of the development of today's integrated circuit (IC) devices relies on advanced process and electrical device simulation programs.[13,14] Technology Computer Aided Design (TCAD) offers the potential for making and testing new technologies in the "virtual fab." TCAD is a generic name for software with a set of virtual tools used for the development of new integrated circuit devices and processes. These tools have the potential for determining how well a certain design or fabrication step will function without ever producing a single semiconductor part. The potential advantages of TCAD are huge, however, direct two-dimensional measurements on devices are needed to accurately calibrate and verify models built into the simulation program. Direct two-

dimensional measurements can also provide rapid feedback to improve control of manufacturing processes and decrease the product development time.

Decreasing device size and the need for high yields has produced a demand for new techniques for spatially profiling dopants in patterned silicon circuits with progressively smaller lateral resolution. The National Roadmap for Semiconductor Technology (NRST) has recently specified several requirements for dopant profiling: spatial resolution of 10 nm, uncertainty of ±10%, sensitivity to dopant concentrations from 10^{16} cm^{-3} to 10^{20} cm^{-3}, as well as reproducibility and repeatability.[13] Scanning probe microscopy based methods are currently the only methods showing promise for high spatial resolution dopant profiling. This dissertation focuses on a newly developed method for dopant profiling which is based on imaging doping density dependent local surface charge density variations using Tappingmode® atomic force microscopy.

References:

[1] H. R. Camenzind, *Electronic Integrated System Design* (Van Nostrand Reinhold Company, New York, 1972).

[2] J. Snauwaert, L. Hellemans, I. Czeck, T. Clarysse, W. Vandervorst, and P. Pawlik, J. Vac. Sci. Technol. B **12**, 304 (1994).

[3] X. Liu, S. Goodwin-Johansson, J. D. Jacobson, M. A. Ray, and G. E. McGuire, J. Vac. Sci. Technol. B **12**, 116 (1994).

[4] S. M. Sze, *Semiconductor Devices: Physics and Technology* (John Wiley & Sons, New York, 1985).

[5] A. Born and R. Wiesendanger, Appl. Phys. A **66**, S421 (1998).

[6] A. C. Diebold, M. R. Kimp, J. J. Kopanski, and D. G. Seiler, J. Vac. Sci. Technol. B **14**, 196 (1996).

[7] T. Hochwitz, A. K. Henning, C. Levey, C. Daghlian, J. Slinkman, J. Never, P. Kaszuba, R. Gluck, R. Wells, J. Pekarik, and R. Finch, J. Vac. Sci. Technol. B **14**, 440 (1995).

[8] C. Shafai, D. J. Thomson, and M. Simard-Normandin, in *Two-dimensional Delineation Of Semiconductor Doping by Scanning Resistance Microscopy* (U.S Department of Commerce, Gaithersburg, MD, 1994).

[9] R. Weisendanger, *Scanning Probe Microscopy and Spectroscopy: Methods and Applications* (Cambridge University Press, New York, 1994).

[10] J. W. Mayer and S. S. Lau, *Electronic Materials Science: For Integrated Circuits in Si and GaAs* (Macmillan Publishing Company, New York, 1990).

[11] C. Kittel, *Introduction to Solid State Physics*, (John Wiley & Sons, New York, 1976).

[12] A. S. Grove, *Physics and Technology of Semiconductor Devices* (John Wiley and Sons, Inc., New York, 1967).

[13] *S. I. A. The National Technology Roadmap for Semiconductors, 4300 Stevens Creek Boulevard, San Jose, CA.*, (1997).

[14] J. Kim, J. S. McMurray, C. C. Williams, and J. Slinkman, J. Appl. Phys. **84**, 1305 (1998).

Chapter 2 - Review of 2D-Dopant Profiling Techniques

2.1 Industry Requirements

Since 1992, the Semiconductor Industry Association (SIA) has published "The National Technology Roadmap For Semiconductor Technology Needs" (NTRS) which focuses on the technological needs for mainstream and leading edge silicon based microelectronics.[1] The roadmap points out opportunities for inventing solutions to the microelectronics industry's long range technological problems. In 1997 the NTRS laid out several needs for development of two and three-dimensional dopant profiling of devices. One of the major requirements set forth by the NTRS is high spatial resolution mapping. In addition to high spatial resolution, doping profiles of both n and p-type must be detectable over a sensitivity range of 10^{16} to 10^{20} cm^{-3} with 10% dopant concentration precision for analysis of current 250 nm technology.[1] Profiling techniques should also be reproducible, and preferably non-destructive. Currently, all of the existing dopant profiling techniques fail in one ore more of the above mentioned criteria. Great efforts are being made, however, to reach the above stated goals. In the remainder of this

chapter most of the currently available dopant profiling methods available to the micro electronics industry will be discussed.

2.2 Secondary Ion Mass Spectrometry

Secondary ion mass spectrometry (SIMS) is currently the industry standard for 1D dopant profiling. SIMS is used to quantify standards used for calibration of most of the scanning probe based dopant profiling methods, which will be discussed later in this chapter. SIMS uses an ion gun which accelerates energetic ions, usually Ar^+, at the sample surface. Atoms and molecules in both neutral and charged states are ejected from the substrate after interacting with the energetic ion beam. These ionized atoms and molecules (secondary ions) are collected and sorted, via mass to charge ratio, in a high sensitivity mass spectrometer. SIMS has high detection sensitivity for a majority of the elements ($<10^{-4}$ of a monolayer), capabilities of depth concentration profiling of trace elements with depth resolutions $\leq$ 50Å, and lateral resolution of approximately 500 nm.[2]

SIMS is one of the few doping characterization techniques, which allows direct measurement of dopant chemical type and concentration. Dopant concentrations can be detected either chemically, as is done in SIMS measurements, or by the change of some electrical property they produce in the semiconductor. Electrically and chemically determined doping profiles often differ in detected dopant concentration and junction position. Variations in dopant concentrations detected usually result from dose loss, which is the concentration of an electrically active dopant species versus the total chemical concentration of the dopant species. Dose loss results from a dopant being electrically neutralized by compensating defects or by a bonding deficiency between the dopant and semiconductor material. Junction position variations between chemically and

electrically detected junctions result from changes in electrical properties in the space charge region. Both chemical and electrical doping information are useful, and comparing the two results provides a method to determine dose loss in the ion implantation process. Chemical dopant distribution information is often used to calibrate the diffusion models in process simulators which depends on location of chemical species. SIMS is ideal for detecting this type of chemical information, however, falls short of the current industry needs in that it has poor spatial resolution and is destructive. In addition, SIMS requires ultra high vacuum conditions which can be expensive and difficult to maintain, and requires expert operators.

2.3 Introduction to AFM/STM

Virtually all of the high spatial resolution analytical techniques for profiling two-dimensional doping structure are based on scanning probe microscopy (SPM). SPM refers to all techniques implementing a mechanism for scanning a sharp probe tip over a sample surface at very small distances in order to obtain a two-dimensional image of some physical property of the sample, such as topography, with nanometer scale resolution. Scanning is performed by means of piezoelectric elements, with the distance between the probe tip and sample maintained by feedback control. Probes can be set up to monitor many physical properties such as quantum mechanical tunneling into or out of electronic states in the sample[3], interatomic forces[4], electrostatic forces[5,6], frictional forces[7], magnetic forces[8,9], work function variations[10,11], and local sample resistance.[12-14] The first techniques to use the SPM concept were scanning tunneling microscopy (STM)[3] and atomic force microscopy (AFM)[4]. These two basic imaging methods are the

basis for all of the other SPM techniques, therefore, a background in STM and AFM is necessary for the introduction of SPM techniques for dopant profiling.

To produce an STM image a bias is applied between a sharp metal tip and the sample to be investigated (metal or doped semiconductor).[3] When the tip is brought to within a couple of Ångströms of the sample surface a tunneling current starts to flow due to quantum mechanical tunneling.[15] Monitoring the magnitude of the tunneling current allows for measuring localized physical properties at the sample surface (i.e. local density of states (LDOS) variations) as well as sensing small changes in tip sample separation. Tunneling current is exponentially dependant on the tip-sample separation, therefore slight changes in this distance can easily be monitored as tunneling current variations. Scanning the tip over the sample surface while keeping the tunneling current constant by means of a feedback loop, results in the tip following the sample contour, to a first approximation. This mode of operation is called constant current mode since the tip keeps roughly the same separation distance from the sample when the tunneling current is kept constant. By monitoring the vertical position of the tip as a function of lateral position it is possible to produce a three-dimensional image of the sample surface with lateral resolution on an atomic scale. Images may also be obtained using constant height mode in which the tip is rastered over the sample surface at a constant height and changes in the tunneling current are monitored.

AFM was developed shortly after STM with the intention of making imaging of insulators as well as semiconductors and metals possible. Force interactions between a sharp probe tip (mounted at the end of a cantilever) and sample are used to produce topographs of the scanned area. In most AFM's a diode laser, which is reflected off the

back of the cantilever onto a split photodiode detector, is used to precisely determine the deflection of the cantilever as it interacts with the sample surface. Figure 2.1 shows a schematic of an AFM diode laser detection system, as well as the cantilever, piezoelectric scan tube, and control electronics. Deflection of the cantilever by surface force interactions causes the reflected laser beam to change its position on the split diode detector, thus changing the illuminated area on each half of the detector. The difference in signal between the top and bottom diodes indicates exactly how far the end of the cantilever has been deflected.

In contact mode, the probe is dragged across the sample surface and the deflected laser signal is monitored. A constant force image, which monitors surface-tip forces, is produced by using the deflected laser signal as a feedback signal. The feedback loop adjusts the piezo z-position (perpendicular to the surface) in order to keep the force interacting on the cantilever constant while rastering the tip over the sample. It is also possible to obtain an image by monitoring the cantilever deflection as the tip is scanned over the surface at a constant average height. Contact mode imaging can often result in destructive wear of the soft sample surfaces or the tip itself due to frictional forces, which result from the dragging motion of the cantilever across the surface.

Since contact mode AFM can result in sample damage to soft surfaces a much more gentle method of AFM operation has been introduced. Tappingmode® AFM (TMAFM) also referred to as non-contact AFM (NC-AFM) uses a cantilever which is oscillating near its natural resonance frequency. As the probe is brought near the surface, the amplitude of the cantilever's oscillation is dampened by interactions with the surface. TMAFM height images are produced by adjusting the z-piezo displacement in order to

keep a constant cantilever amplitude. The cantilever taps on the sample surface as it is scanned to produce an image. Destructive frictional forces are reduced since the cantilever hops along the surface and is not dragged.

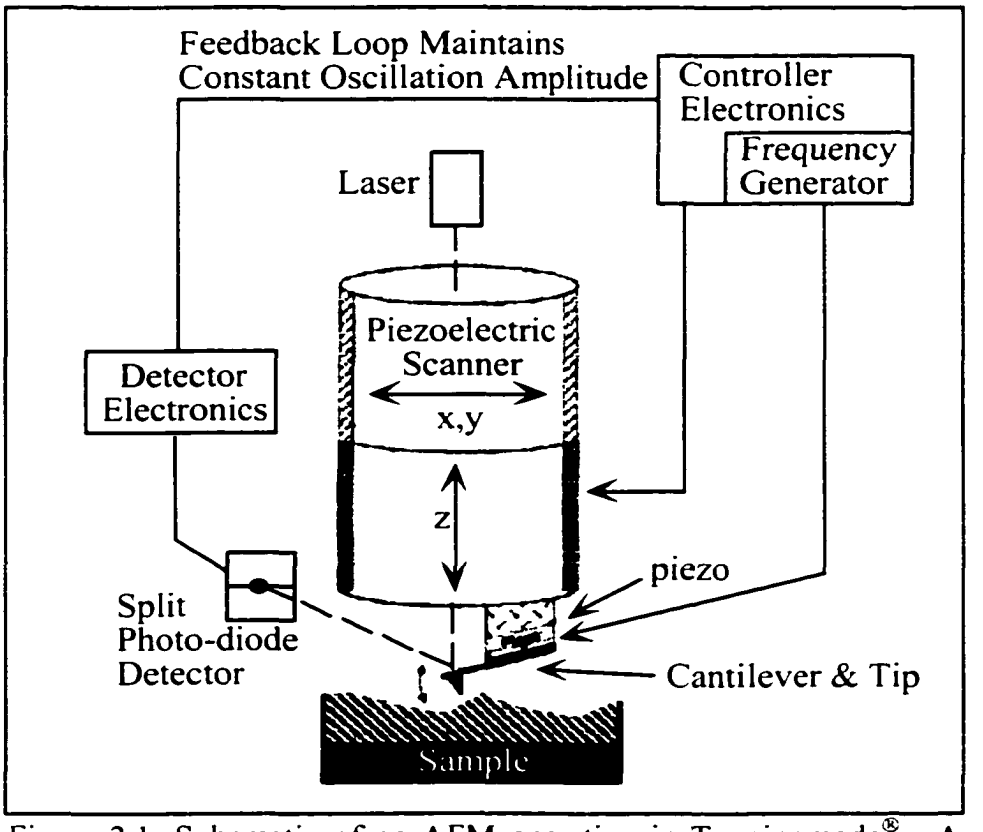


Figure 2.1: Schematic of an AFM operating in Tappingmode®. A diode laser is reflected off the back side of a cantilever-tip assembly. The position of the reflected beam on a split photo-diode detector allows determination of changes in the cantilever deflection, or amplitude. Feedback on the cantilever's amplitude signal during lateral scanning of the cantilever is used to produce topographic images.

It is convenient to use the AFM to position a probe at a surface and then monitor signals which are not related to the samples topography such as; capacitance, magnetic force, cantilever related phase and frequency, or contact potential. Monitoring alternative signals allows mapping of properties other than topography on a nano-scale. Innovations

in this area have become a key to the popularity of scanning probe microcopies. A few of the dopant profiling related innovative SPM techniques are discussed below.

2.4 Differential Etching

Methods of selective chemical etching of doped regions of a device combined with a microscopy method to image the etched surface have been used since the early days of device fabrication as a method to determine dopant profiles. The principle reaction for the doping selective etching of silicon by nitric and hydrofluoric acid is given by the following chemical reaction:[16]

$$Si + 2HNO_3 + 6HF \rightarrow H_2SiF_6 + 2HNO_3 + 2H_2O. \qquad (2.1)$$

In this reaction nitric acid produces silicon oxide which is removed by hydrofluoric acid. The etching process is faster in areas of the sample containing higher doping densities.[17] Calibration of the silicon etch rate as a function of doping density should theoretically allow for a two-dimensional etch profile to be converted into a quantitative two dimensional doping profile.

Scanning electron microscopy (SEM)[18], transmission electron microscopy (TEM)[19-21], scanning tunneling microscopy (STM)[22], and atomic force microscopy (AFM)[23,24] have all been used for qualitative and quantitative dopant profiling by measuring the depth of etch pits. Transmission electron microscopy profiling involves difficult and time consuming sample preparation, however, 5 nm spatial resolution of etch depth information may be obtained indirectly by studying the thickness fringes of the TEM image.[25] Sample preparation for scanning probe based methods is not as difficult as with TEM, however, drawbacks to SPM profiling by etching still exist.

All of the above mentioned imaging methods require calibration of the etch rate as a function of doping density. Calibration, however, is not simple since little is known about the mechanism of silicon etching in $HF:HNO_3$ mixtures. It appears that etch rate is non-linear in time and also varies with dopant type, temperature, etching solution concentrations, light, agitation, surface contamination, and many other factors.[17] Under these circumstances it seems that calibration should be nearly impossible. In addition to calibration problems it is often difficult to determine a reference doping level to compare with other etch depths. Selective etching appears to provide a good high spatial resolution method for chemically based qualitative dopant profiling.

2.5 Scanning Capacitance Microscopy

Scanning capacitance microscopy (SCM) uses a contact mode AFM to position a conducting tip at a semiconductor surface. The local metal-insulator-semiconductor (MIS) capacitance is measured, through a dielectric layer grown on the surface, with the aid of a sensitive capacitance sensor, which is electrically connected to the tip. The MIS capacitance changes as a function of the samples doping density. As the tip is scanned, capacitance and topographic information are gathered simultaneously to produce both topographic and doping density related images.

An ultra high frequency (UHF) capacitance sensor, operating at frequencies in the megahertz regime, is the basis of SCM signal detection.[26] The resonator is connected to the metal coated cantilever/tip assembly via a transmission line. When the resonating probe tip is put into contact with the oxide coated semiconductor the sensor, transmission line, cantilever/tip assembly, and near surface carriers in the semiconductor become part of the resonator system. Variations in the tip-sample capacitance will result in changes in

the load at the end of the transmission line, resulting in a change in the resonance frequency of the system. Small changes in the semiconductor capacitance result in large changes in the resonance frequency of the system, thus allowing detection of capacitance changes in the attofarad (10^{-18} farads) range.[27]

In order to monitor the tip capacitance in the presence of a large background capacitance signal, a time dependant voltage signal (typically 5-15kHz) is applied to the sample and the signal resulting solely from the semiconductor-tip capacitance is detected by a lock-in amplifier.[26] Free carriers in the semiconductor are alternatively attracted and repelled from the region directly below the tip due to the alternating electric field produced by the AC bias applied to the semiconductor (Figure 2.2). This alternation between accumulation and depletion of carriers under the tip can be modeled as moving capacitor plates. The distance between the metallic tip and the depletion depth gives the capacitor plate separation distance. As the carriers are depleted from the surface directly below the tip, the "distance (d) between the plates" increases. Capacitance is related to the "plate" separation by the following equation:

$$C=\varepsilon A/d \qquad (2.2)$$

where C is capacitance, ε is the dielectric constant of the insulator between tip and sample, A represents the surface area of the plates, and d is the plate separation distance. Since capacitance is inversely proportional to the plate separation distance, this plate movement varies the capacitive load on the resonant sensor system, and therefore alters the resonance frequency of the capacitance sensor.

The depletion depth, which is modeled by the capacitor plate movement, depends on three quantities: the strength of the applied electric field, the quality and thickness of

the dielectric between the tip and semiconductor, and the free carrier concentration. Depletion occurs at deeper depths in the semiconductor with larger electric fields or lower carrier concentrations.[28] Measuring capacitance at a constant A.C. bias amplitude results in a change in the capacitance signal with doping density due to a change in the sample's depletion depth. More lightly doped regions produce a higher change in capacitance signal since they have a lower free carrier concentration and form deeper depletion layers. Changes in the depletion width, due to doping density variations, result in variation in the imaging volume and results in non-constant spatial resolution across an image.

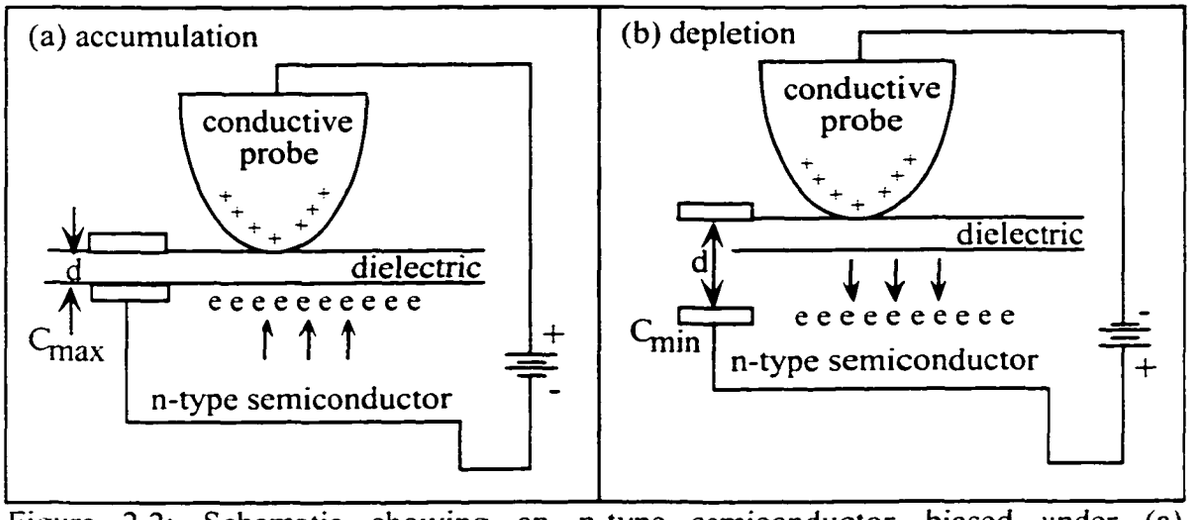


Figure 2.2: Schematic showing an n-type semiconductor biased under (a) accumulation (b) depletion conditions. Maximum capacitance (C_{max}) is produced under accumulation because carriers, represented by as "e", collect at the surface resulting in a minimum capacitor plate separation distance (d). Capacitance is minimized (C_{min}) under depletion where the plate separation distance (d) is large. The depletion depth increases with increased doping density.

Resolution variations can be overcome by operating the microscope in constant change in capacitance mode (dC/dV) resulting in a constant depletion depth and therefore constant spatial resolution across the resulting image.[29] Constant change in capacitance

mode (dC/dV) uses a feedback loop to adjust the amplitude of the oscillating voltage on the sample, thus keeping the change in capacitance and depletion depth constant when imaging any doping level in the sample.[29] A large bias voltage is required in heavily doped regions to produce the same depletion depth and change in capacitance that can be created at a lower voltage in a more lightly doped region. SCM operating in dC/dV mode has been used to profile semiconductor devices varying from 10^{13} to 10^{20} dopants/cm^3 with spatial resolution approaching 10 nm (increases with tip diameter).[27]

SCM is currently viewed as a promising method for two-dimensional dopant profiling since there is a possibility to make it a "turn key" quantitative technique. There are, however, several problems associated with using SCM for dopant profiling. The two nanometers of native oxide that grows on silicon exposed to air is too thin for SCM, since tunneling occurs through the native oxide.[29] This requires growth of a uniform thickness oxide layer on the silicon surface using low temperature methods in order to limit dopant diffusion during the oxide growth.[30] Oxide growth is time consuming and not always precise. Care must be taken in selecting an oxidation process since small variations in this oxide thickness can cause problems in the doping quantification process.[27] A poorly developed oxide can also contain large numbers of fixed charges which can affect the capacitance signal. In order to overcome the capacitance signal associated with a charging oxide, D.C. offset biases can be applied to the sample. Field induced oxidation (FIO) can occur if D.C. biases greater than 3 volts are used during SCM scanning.[31] FIO leads to decreased SCM sensitivity and, when it occurs in consecutive scans, it causes a local variation in oxide thickness, which change the capacitance signal and increase uncertainty in quantitative measurements.[31] Good sample preparation, however, can

often lower the D.C. offset needed. Low temperature annealing during UV exposure has been found to help minimize trapped charges in the oxide, thus reducing the D.C. offset voltage required for imaging.[31]

Junction delineation is the determination of the position of a pn-junction, and is crucial when determining the channel length of a device. A major problem with using SCM for junction delineation is that apparent junction positions have been observed to shift with the A.C. bias amplitude and D.C. offset bias applied to the sample.[32] This problem is associated with the fact that within the depletion region of the pn-junction both types of carriers (holes & electrons) are available. Applying a bias in this region via the SCM probe alters electric fields in the depletion region. This results in carriers flowing into the region under the tip from the adjacent n-type or p-type region depending on bias. Carrier movement can thus make n-type material look like p-type and vice-versa.[33] Changes in doping levels of both materials involved results in a different bias voltage dependence on junction position, since doping concentrations within the junction determine the depletion widths. It turns out that there is no ideal D.C. bias voltage which is best for imaging all semiconductor device structures. Imaging at the "mid-gap" or the bias midway between the n-type and p-type peaks in dC/dV plots gives the best approximation of the junction position. However, this is not easy and its accuracy is questionable.[33]

Additional problems arise when imaging samples containing metal or rough topography. Topographic features produce a false signal when the side of the tip interacts electrically with the side-wall of the topographic feature. Often it is difficult to determine if this signal results from a slight change in doping or if it is indeed a false

signal. Imaging metals in dC/dV mode produces a very large AC bias when the feedback loop attempts to keep the capacitance change constant with the rest of the sample. In metals charges always accumulate and spread out over the sample surface. Depletion never occurs into the bulk of the metal. This large AC bias often results in destruction of the cantilever via resistive heating resulting from shorting to the metal.[34]

2.6 Nano-Spreading Resistance Probe

Spreading resistance probe (SRP) is another of the recently proposed AFM based dopant profiling techniques.[26,35] The basic SRP principle of using single, two point, and four point probes to determine doping levels in semiconductors by measuring resistance and resistivity was developed years ago.[28] Nano-SRP or μ-SRP use a Tappingmode® AFM to scan a conductive diamond-like probe tip across a sample surface. The probing can be performed using a single or two-point probe as shown in Figure 2.3. A bias is applied between the probe and the base of sample for single point measurement or between the two probes for two-point measurement of the sample resistance. Resistance decreases with increased doping density since the carrier density increases with the number of donors and acceptors.[28] Calibration on a well characterized sample can provide a quantitative doping density map of the surface with 20% reliability.[13] Better doping quantification reliability should be attainable by improving the reproducibility of the probe-sample contact.

A major barrier to reliable SRP based doping level measurement is the need to supply an ohmic contact to the probe tips at the sample with a reproducible contact resistance. The probe points must be capable of repeated contact to the surface with no significant changes in the tip-sample resistance (R_s) or the measurement of the sample

spreading resistance (R_m) will not be reliable, since R_s and R_m add in series.[35] At each measurement point the tips must be able to break a fixed distance through the native surface oxide without deforming.

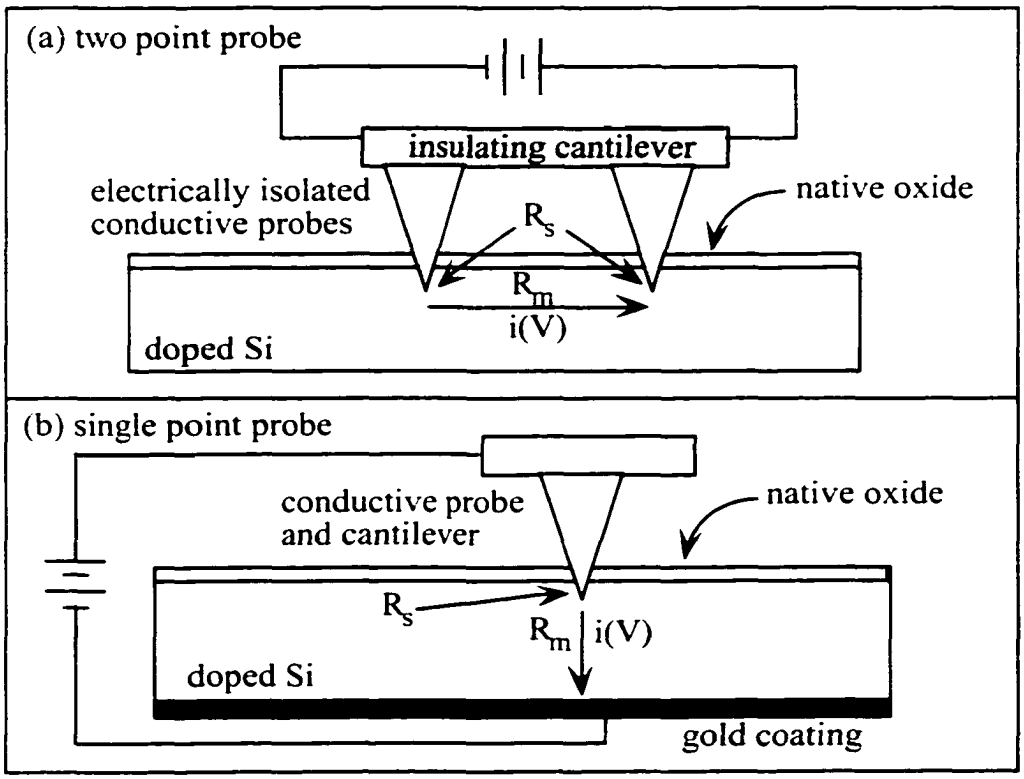


Figure 2.3: Point probes mounted on a cantilever for use in SRP imaging. With two point probes (a) bias is applied between the probes, thus inducing a current (i(V)) to flow. For single point probes (b) the bias is applied between the tip and sample base. Sample resistance (R_m) is measured in both cases. Changes in the depth of the physical penetration through the native oxide result in variations in the contact resistance (R_s), and can result in errors in the measurement of the spreading current.

Local variations in the oxide thickness can also result in inconsistent spreading resistance measurements since the contact resistance varies with oxide thickness. AFM force feedback is used to try to keep the contact resistance constant by applying a constant force to drive the probe a consistent distance through the oxide into the underlying semiconductor. The force applied to the tips (10^{-4} N to 10^{-6} N) at each point,

unfortunately, results in the destruction of the surface and prevents repeated scans of the same sample area.[36] Typically the surface roughness is increased by 1 nm during a SRP scan.[13] For optimal spatial resolution the force applied to the tips must be minimized so that the probe has a small contact area. When the SRP is operating at an optimal applied force the microscope is capable of 20 nm spatial resolution.[13]

In addition to its destructive nature and contact resistance problems, SRP often produces erroneous results when probing the low doping density side of a junction. Since electrons take the path of least resistance when traveling through a material, electronic carriers will jump across a junction into the more highly doped material, which has a lower resistance. This results in the measured resistivity appearing lower than it actually is, resulting in an erroneous dopant density measurement. As device sizes become smaller this problem increases since carriers do not have to travel as far to reach a medium of higher doping density or lower resistance.

2.7 Scanning Kelvin Probe

Scanning Kelvin force microscopy (SKFM) is based on the Kelvin method, which is the most common method used to measure contact potential differences between a reference electrode and a sample. Contact potential differences for a clean metal surface is given by the difference in work functions of the two materials.[37] Since the work function of a semiconductor depends on the doping level, maps of work function variations should translate into a doping profile.

Kelvin measurements are made by bringing a vibrating probe of known work function into close proximity of the surface to be measured. The system is treated as a

parallel plate capacitor and the contact potential between the two materials is given by the equation:

$$V_{CPD} = \frac{1}{e(\phi_2 - \phi_1)} \tag{2.3}$$

where Φ_2 and Φ_1 are the work functions of the conductors and e is the elementary charge.[37] An electric current results from the introduction of a periodic vibration between the two plates at a frequency ω with amplitude d_1:

$$i(t) = dC/dz(V_{bias}+V_{CPD})\omega d_1 \cos\omega t \tag{2.4}$$

where V_{CPD} is the contact potential difference, dC/dz is the derivative of the capacitance of this plate capacitor with respect to the tip sample separation distance, and V_{bias} is the bias voltage applied between the tip and sample.[37] The objective of the Kelvin method is to detect the zero point of current i(t) while the additional bias voltage is applied between the plates until the electric field between the plates disappears. At this point the contact potential difference is equal to the bias voltage.[37] The Kelvin method has high surface potential sensitivity but integrates over the entire surface.

Figure 2.4 shows a set of diagrams of a n-type silicon sample and metal reference electrode at electrical equilibrium and biased so that $V_{bias} = V_{CPD}$. Under equilibrium conditions (Figure 2.4a) the Fermi levels of the two materials equilibrate. This results in band bending in the semiconductor, resulting in depletion or accumulation of carriers at the surface, depending on the magnitude of the semiconductor's work function with respect to that of the reference electrode. Accumulation and depletion of the semiconductor surface results in flow of an electric current. At $V_{bias} = V_{CPD}$ the bands in the sample are flat and there is no electric field which can force mobile carriers out of

the surface region. Therefore, the current is zero. At this point all of the fixed charges in the space charge region are paired with mobile carriers. The mobile carriers act to shield the fixed charges. Thus, at $V_{bias} = V_{CPD}$, the capacitive force between the sample and probe is zero.

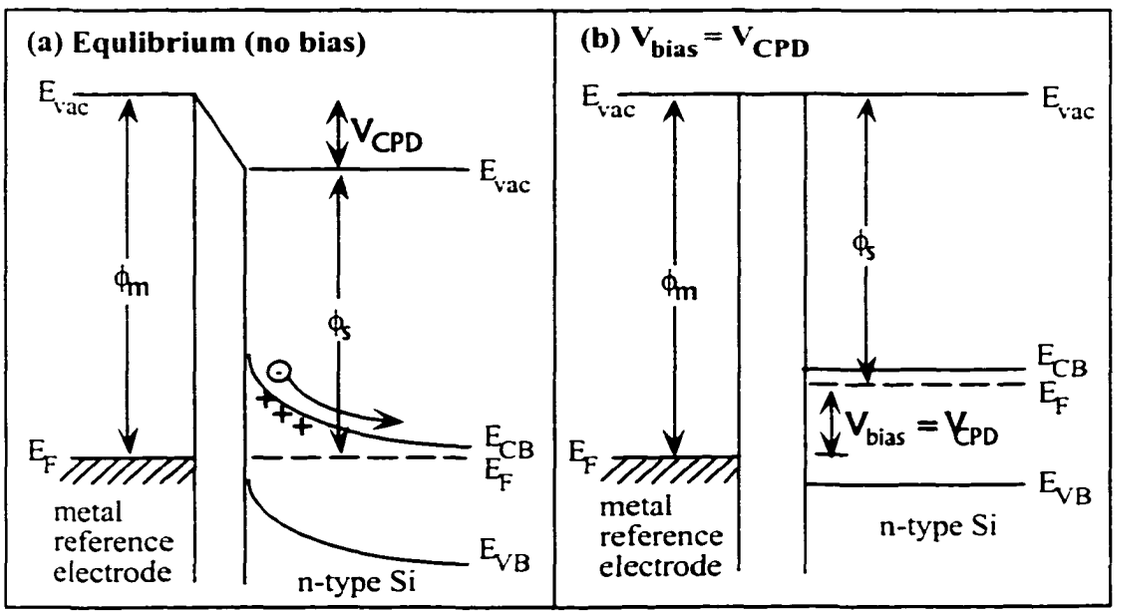


Figure 2.4: Energy band diagrams showing the measurement of the work function of a n-type semiconductor with respect to a known work function metal reference electrode. In electrical equilibrium (a) a contact potential difference(V_{CPD}) develops due to the equilibration of the Fermi levels of the two materials. Band bending in the semiconductor caused depletion of the surface, which results in flow of an electric current and a capacitive force between the plates due to the charging of the semiconductor surface. When a bias is applied to the semiconductor which equals V_{CPD}, the band bending is offset, resulting in flatband conditions in the semiconductor. The work function of the semiconductor is calculated as $\phi_m = \phi_s + V_{bias}$.

Scanning Kelvin force Microscopy (SKFM) takes the Kelvin method one step further by using an AFM setup to scan a small Kelvin probe (metal coated AFM tip) over a surface. An image is produced which reveals local variations in the work function instead of a work function value averaged over the entire surface. SKFM is capable of providing local surface potential information and topographic information

simultaneously. An AFM tip is positioned close to the surface using non-contact AFM and the cantilever sample separation is kept fixed at a constant value (~15 nm) while the Kelvin force measurements are taken. This is done by taking a topographic scan of a single line of the sample and then immediately retracing the topography from the first scan at an increased tip-sample separation. A feedback loop is used during the lift scan, which monitors the force on the cantilever and adjusts the tip-sample bias in order to keep the force constant throughout the scan. An image is produced showing the bias required to keep a constant tip sample force. Since the contact potential difference between the tip and sample are supposed to be equal when the current or force are zero the image plot indicates changes in contact potential of the sample with respect to that of the probe.

SKFM has been shown to have a spatial resolution of about 10 nm at a lift height of 15nm while scanning in ambient conditions.[38] Structures containing features of both n-type and p-type varying in concentration from 10^{14} cm^{-3} to 10^{20} cm^{-3} have been imaged using SKFM.[10] Operational sensitivity to doping concentration varies depending on the setting of the feedback loop and the sample structure being examined. In a study examining a bipolar structure, the sensitivity was determined to be ±12.5% for a doping density of 10^{16} cm^{-3} and ±17% for a doping density of 10^{19} cm^{-3}.[10] Currently SKFM imaging produces qualitative results. However, efforts are being made to make it a quantitative dopant profiling technique.

SKFM measurements have recently been performed in ultra high vacuum (UHV) in order to obtain higher spatial resolution and to assure that the work function of the probe remains stable since metal work functions are very sensitive to surface

contamination.[39] Unlike their ambient counterpart, UHV non-contact AFM systems are powerful enough to resolve features near atomic resolution.[40] The contact potential difference measurement, as described above, relies on the difference in work function between the surface of the tip and sample. Measuring the contact potential difference, and using the well-known work function of the tip, it is possible to determine the work function of the sample. Unfortunately, the work function of both the tip and sample vary with oxidation, surface contaminant adsorption, surface roughness, and tip shape. The work function of a semiconductor depends on doping density. However, errors in measuring doping density can result from changes in work function due to roughness, contamination and tip shape. Currently, SKFM has not sparked a major interest from the dopant profiling community due to its requirement for vacuum operation and difficulties with repeatability from tip to tip. If, however, these problems can be worked out, it may become a promising two-dimensional dopant profiling technique.

References

[1] *S. I. A. The National Technology Roadmap for Semiconductors, 4300 Stevens Creek Boulevard, San Jose, CA.*, (1997).

[2] A. W. Czanderna, *Methods of Surface Analysis*, Vol. 1 (Elsevier, New York, 1989).

[3] G. Binnig and H. Rohrer, Helvetica Physica Acta **55**, 762 (1982).

[4] G. Binnig, C. F. Quate, and C. Gerber, Phys. Rev. Lett. **56**, 930 (1986).

[5] Y. Martin, D. W. Abraham, and H. K. Wickramasinghe, Appl. Phys. Lett **52**, 1103 (1988).

[6] B. D. Terris, J. E. Stern, D. Rugar, and H. J. Mamin, Phys. Rev. Lett. **63**, 2669 (1989).

[7] G. Meyer and N. M. Amer, Appl. Phys. Lett. **57**, 2089 (1990).

[8] Y. Martin and H. K. Wickramasinghe, Appl. Phys. Lett. **50**, 1455 (1987).

[9] C. Schönberger, S. F. Alvarado, S. E. Lambert, and I. L. Sanders, J. Appl. Phys. **67**, 7278 (1990).

[10] T. Hochwitz, A. K. Henning, C. Levey, C. Daghlian, J. Slinkman, J. Never, P. Kaszuba, R. Gluck, R. Wells, J. Pekarik, and R. Finch, J. Vac. Sci. Technol. B **14**, 440 (1995).

[11] H. D. Liess, R. Mäckel, H. Baumgärtner, and J. Ren, Sensors and Actuators **13-14**, 739 (1993).

[12] P. D. Wolf, T. Clarysse, and W. Vandervorst, J. Vac. Sci. Tachnol. B. **16**, 320 (1998).

[13] W. Vandervorst, V. Privitera, V. Raineri, T. Clarysse, and M. Pawlik, J. Vac. Sci. Technol. B **12**, 276 (1994).

[14] C. Shafai, D. J. Thomson, and M. Simard-Normandin, *Two-dimensional Delineation Of Semiconductor Doping by Scanning Resistance Microscopy* (U.S Department of Commerce Gaithersburg, MD, 1994).

[15] R. Weisendanger, *Scanning Probe Microscopy and Spectroscopy: Methods and Applications* (Cambridge University Press, New York, 1994).

[16] D. R. Turner, J. Electrochem. Soc. **107**, 810 (1960).

[17] V. A. Ukaraintsev, R. McClothlin, M. A. Gribelyuk, and H. Edwards, J. Vac. Sci. Technol. B **16**, 476 (1998).

[18] E. S. Meieran and T. I. Kamins, Solid-State Electron. **16**, 545 (1973).

[19] W. Vandervorst, T. Clarysse, J. Vanhellemont, and A. Romano-Rodriguez, J. Vac. Sci. Technol. B **10**, 449 (1992).

[20] T. T. Sheng and R. B. Marcus, Solid-State Sci. Technol. **128**, 881 (1981).

[21] C. Spinella, V. Raineri, F. La Via, and S. U. Campisano, J. Vac. Sci. Technol. B **14**, 414 (1996).

[22] T. Takigami and M. Tanimoto, Appl. Phys. Lett. **58**, 2288 (1991).

[23] R. Alvis, B. Mantiply, and M. Young, J. Vac. Sci. Technol. B **14**, 452 (1996).

[24] V. Raineri, V. Privitera, W. Vantervorst, L. Hellemans, and J. Snauwaert, Appl. Phys. Lett. **64**, 354 (1994).

[25] M. Barrett, M. Dennis, D. Tiffin, Y. Li, and C. K. Shih, J. Vac. Sci. Technol. B **14**, 447 (1996).

[26] J. J. Kopanski, J. F. Marchiando, and J. R. Lowney, Materials Science and Engineering B **44**, 46 (1997).

[27] A. Born and R. Wiesendanger, Appl. Phys. A **66**, S421 (1998).

[28] S. M. Sze, *Semiconductor Devices: Physics and Technology* (John Wiley & Sons, New York, 1985).

[29] J. S. McMurray, J. Kim, and C. C. Williams, J. Vac. Sci. Technol. B **15**, 1011 (1997).

[30] Y. Huang, C. C. Williams, and M. A. Wendman, J. Vac. Sci. Technol. A **14**, 1168 (1996).

[31] V. V. Zavyalov, J. S. McMurray, and C. C. Williams, (**unpublished** Gaithersburg, MD (1998).

[32] R. N. Kleiman, M. L. O'Malley, F. Baumann, J. P. Garno, and G. L. Timp, *Exploring the Limits of Spatial Resolution in SCM Imaging* (Intel, Santa Clara, CA, 1998).

[33] M. L. O'Malley, G. L. Timp, S. V. Moccio, J. P. Garno, and R. N. Kleiman, Appl. Phys. Lett. **74**, 272 (1999).

[34] A. Erickson, (personal communication, 1999).

[35] R. B. Marcus, *Spreading Resistance Microprobe* (U.S Department of Commerce, Gaithersburg, MD, 1994).

[36] A. C. Diebold, M. R. Kimp, J. J. Kopanski, and D. G. Seiler, J. Vac. Sci. Technol. B **14**, 196 (1996).

[37] J. Lü, M. Guggisberg, R. Lüthi, M. Kubon, L. Scandella, C. Gerber, E. Meyer, and H.-J. Güntherrodt, Appl. Phys. A **66**, S273 (1997).

[38] H. O. Jacobs, P. Leuchtmann, O. J. Homan, and A. Stemmer, J. Appl. Phys. **84**, 1168 (1998).

[39] S. Kitamura and M. Iwatsuki, Appl. Phys. Lett. **72**, 3154 (1998).

[40] F.J. Giessibl, Science **267**, 68 (1995).

Chapter 3: Dopant detection in 2D Semiconductors with STM

3.1 Introduction to STM of Dopants in Layered Semiconductors

Due to its small probe size and measurement method scanning tunneling microscopy (STM) theoretically allows for direct measurement of the local electronic density of states and surface potentials induced by localized charges. For example, single dopant atoms, defect sites, or areas of differing doping type can be identified due to apparent height changes they induce in constant current STM images.[1-5] Additional measurements with scanning tunneling spectroscopy and current imaging tunneling spectroscopy (CITS) are capable of producing spatially resolved information on the local density of states (LDOS) in the sample.[6-8] This chapter will focus on the imaging mechanisms which cause STM imaged dopant atoms to appear as protrusions or depressions superimposed on the atomic lattice of semiconductor surfaces. Experiments on three-dimensional semiconductor materials such as Si or GaAs, have revealed that the apparent height or depth of the observed features lie in the range of about 0.1 nm, most likely resulting from local changes in electronic structure around the dopant atom.[4,9-11] Appearance of dopants in the topographic images of 3D semiconductors are often

attributed to a change in the occupied electronic density of states due to localized band bending or a change in the LDOS in the vicinity of the observed features.

Investigation of two-dimensional layered semiconductors, such as MoS_2 and WSe_2, by STM yields similar dopant related features, with tenfold larger protrusion/depression heights or depths than observed with the three-dimensional semiconductors mentioned above. These height changes cannot be explained solely by electronic effects since a 0.1 nm tip-sample distance corresponds to a change in the occupied density of states of about one order of magnitude. This means that the occupied density of states in the two-dimensional material would have to change ten orders of magnitude from the bulk semiconductor to the area around the dopant, which is nearly impossible. This suggests that forces between the tip and sample may have a role in the mechanism, leading to the observed features.

Earlier published results suggest that forces between the tip and sample are responsible for several STM imaging artifacts on layered materials. Examples of force interactions in STM imaging of layered materials include etching of layered materials[12,13], controlled nano-structuring via STM[14-17], pulling up of sample edges[18], and several other artifacts attributed to tip sample interactions.[19-21] The focus of the remainder of this chapter is the investigation of Coulomb force contributions to the STM imaging mechanism of dopant related features in layered semiconductors. In addition, the influence of electrostatic forces were studied by atomic force microscopy operated under an applied electric bias. Bias induced changes in topography and force calibration curves suggest a significant influence of Coulomb forces on the imaging process.

Additional measurements with CITS and STM suggest that topographic changes rather than localized electronic variations produce the protrusions and depressions.

3.2 STM Investigation of the Dopant Feature Formation

Transition metal dichalcogenides (TMD) are built up from repetition of a three-layered sandwich consisting of a top chalcogenide layer (S, Se, Te) a middle transition metal layer (e.g. Ti, Sn, Mo, W, Ta, etc.) and a bottom chalcogenide layer. The bonding within each sandwich layer is covalent, however adjacent sandwich layers are held together by weak van der Waals (VDW) forces (Figure 3.1). Layered materials can be easily cleaved, by peeling away the topmost layers, providing clean, atomically flat terraces of up to several thousand square Ångströms. The flat surfaces of the cleaved layered materials are ideally suited for atomic resolution SPM studies. Since the surfaces of layered materials contain virtually no dangling bonds,[22] a freshly cleaved surface can stay clean for a long period of time. This allows atomic resolution SPM studies even under ambient conditions. Electronically, the TMDs range from true metallic conductivity (NbS_2), through semi-metallic (WTe_2), semiconducting (MoS_2, WSe_2, SnS_2, $MoTe_2$), and insulating (HfS_2). STM can easily be applied to TMDs with conductivity ranging from metal to semiconductor. However, this work focuses on semiconducting TMDs.

As mentioned above, imaging of layered semiconductors with STM often yields protrusions superimposed on an otherwise flat atomic lattice of the surface.[23,24] These protrusions can be attributed to the dopant sites, as has been shown on WSe_2 crystals by correlating dopant densities counted in STM images with Hall measurements.[25] A strong bias dependence on the heights of the dopant features is observed in STM images

suggesting that a closer look into the forces interacting between the tip and surface is necessary. A set of experiments investigating sample-probe electrostatic force interactions were performed using a commercially available Nanoscope III® STM/AFM (Digital Instruments, Santa Barbara, CA) operating in ambient conditions. Dopant imaging was also performed in ultra high vacuum using an Omicron Variable Temperature STM (VT-STM®). Electrochemically etched tungsten tunneling tips (35-40 V AC in 1M KOH) were used in all STM experiments.

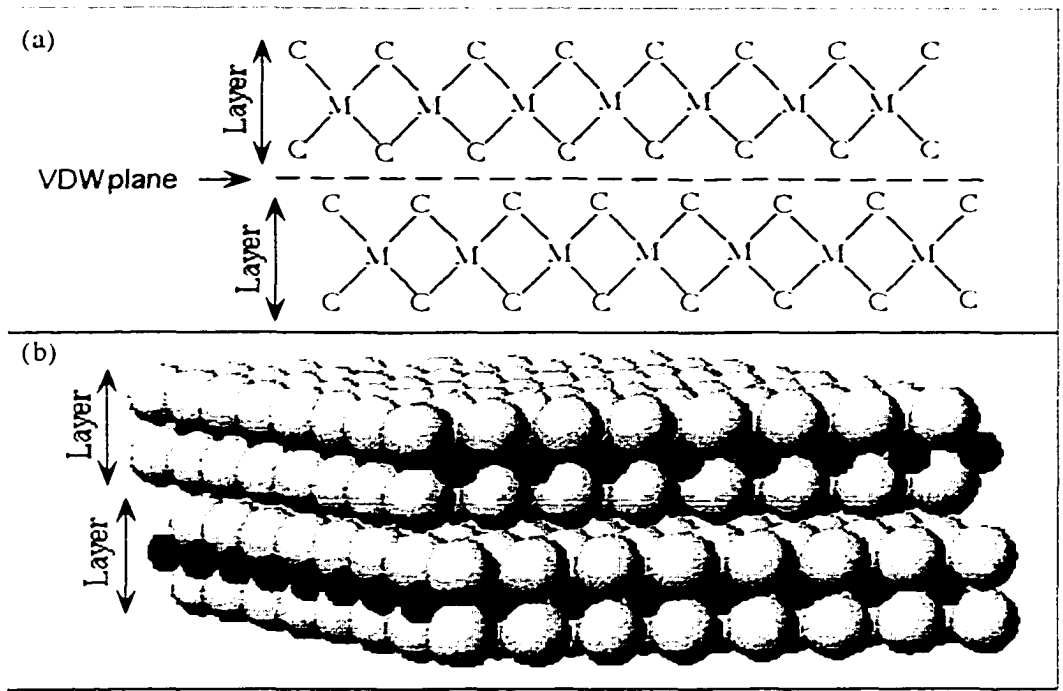


Figure 3.1: Simple bonding diagram (a) and space filling model of a typical transition metal dichalcogenide (TMD) crystal. The dark spheres and "M" atoms represent transition metals while light colored spheres and "C" atoms represent chalcogenide atoms.

Figure 3.2 shows a series of three STM images taken from the same area (60x60 nm^2) of a natural MoS_2 crystal at varied negative tunneling biases with respect to the grounded tip. From the shape of I(V) curves taken on the sample, the scanned area was determined to be p-doped (see Appendix A). The images at –500 and –1500 mV show

protrusion heights of about 1.2 nm while the same protrusions appears much smaller (about 0.2 nm) in the image taken at –70 mV. One additional protrusion becomes visible in the upper parts of the higher bias images that is not noticeable in the –70 mV picture. This is most likely a dopant residing in the second van der Waals layer of the MoS_2. The image taken at –1500 mV also shows multiple tip effects as revealed by the satellite protrusions appearing slightly above and to the left of the four main protrusions. All of these features point toward a bias-dependent morphology change of the surface. It is possible that the surface is pulled toward the tip in the vicinity of the dopant atoms at the higher biases, while the ring-like appearance of the of the protrusions at –70 mV indicates that the tip is pressing on the surface, instead of pulling.[2,26] The appearance of multiple tip features at –1500 mV indicates that the surface is pulled toward the tip more strongly, resulting in a bending of the surface around the tip apex, allowing more than one singular point on the tip to become a tunneling site.

A possible explanation for the observed result is that the applied bias results in a change in the surface potential of the sample. The negative sample bias, therefore, leads to a depletion of the surface leaving the dopant atoms ionized. An ionized dopant atom represents a localized charge on the order of one elemental charge. Presence of the dopant charge results in the induction of an image charge of opposite polarity in the tip. The Coulomb attraction of the charge and image charge results in a relatively strong force which exceeds the van der Waals force holding the layers of the MoS_2 together in the region around the dopant. These large electrostatic forces may explain the large protrusion since the feedback loop of the microscope retracts the tip in order to maintain a constant tunneling current as the surface is pulled toward the tip. At low biases, the

surface depletion is not as deep, and the charge on the dopants is more effectively screened by free charge carriers. This results in a weaker electrostatic force between tip and sample at lower bias, resulting in a lower protrusion height.

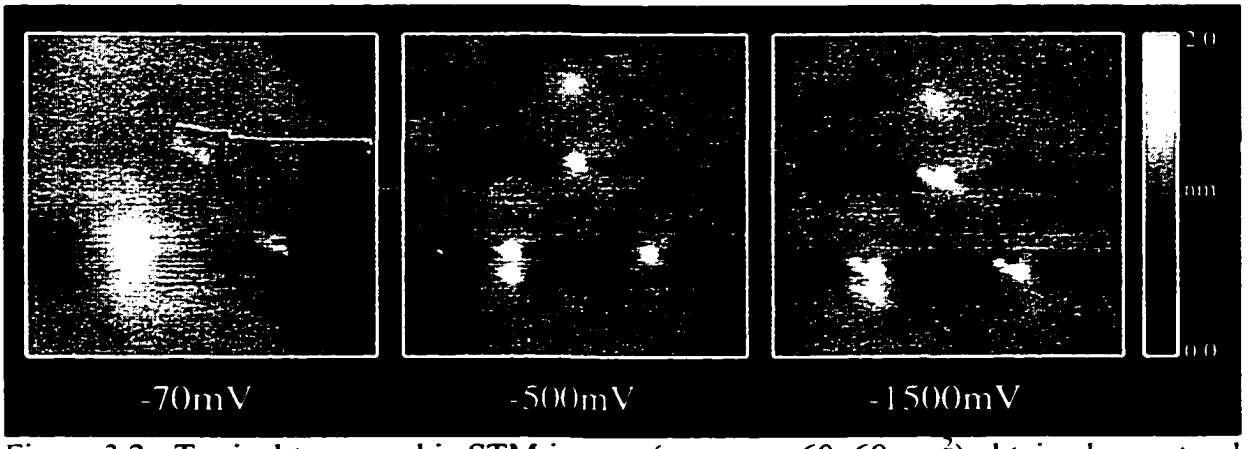


Figure 3.2: Typical topographic STM images (scan area 60x60 nm^2) obtained on natural MoS_2. The heights of the dopant related protrusions vary dramatically with the tunneling bias (given below the images). All images were taken at 100 pA setpoint current and at a scan rate of 5 lines/s.

A band diagram, shown in Figure 3.3(a), demonstrates how a negative sample bias leads to depletion of mobile charges from the surface, thus leaving surface dopant atoms ionized. At biases close to zero volts, the semiconductor is in or near flat band conditions. This results in an even distribution of the mobile charge density (ρ) throughout the semiconductor, as can be seen in Figure 3.3(b), which shows the mobile charge density (ρ) as a function of distance from the semiconductor surface. As the semiconductor becomes biased negatively, band bending is induced in the semiconductor. This causes an electric field to be setup in the material, resulting in the depletion of mobile charges (holes) from the surface region. Depletion of the surface region results in ionization of the dopants residing in the surface region. The ionized atoms can induce an image charge in the tunneling tip, thus creating an attractive electrostatic force between the tip and sample. The mobile charges that remain in the surface region screen the

ionized dopant, making the effective charge seen by the tip less than the elemental charge. As the bias on the semiconductor becomes more negative the surface is depleted to a greater depth. This increased depletion causes a decrease in the mobile charge density (ρ) at the surface, which decreases the screening of dopant charges by the mobile charges. This results in a larger effective charge seen by the tip, and a larger Coulomb force.

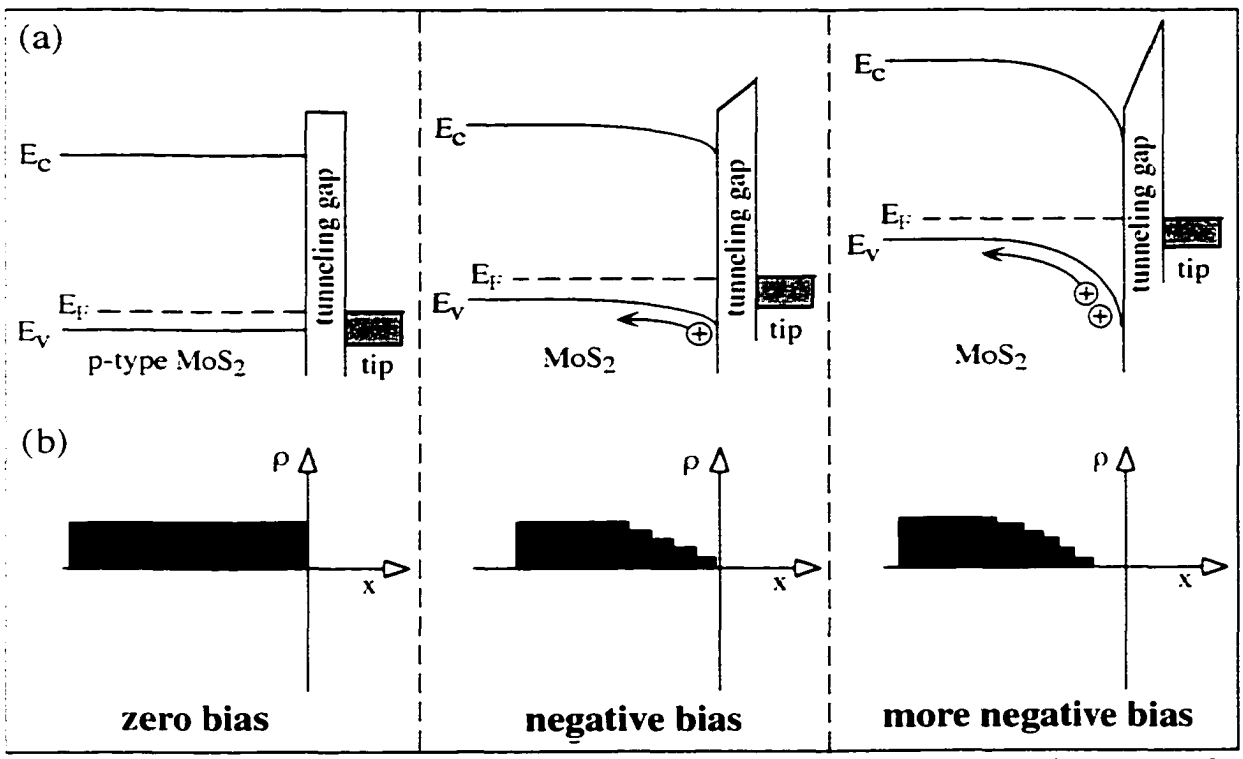


Figure 3.3: (a) Band diagrams and (b) mobile charge density, ρ, versus distance, x, for three biasing conditions. As the p-type MoS_2 sample is biased more negatively with respect to a grounded tunneling tip, the surface region becomes depleted of mobile charges. This decreases the screening of the ionized dopant cores, and increases the Coulomb force between charges in the tip and the dopant charges in the sample.

Figure 3.4 shows a schematic indicating how the surface is pulled toward the tip when the tip is in close lateral proximity to a dopant atom. Since the dopant atom is incorporated into the top van der Waals layer of the TMD, the pulling of the dopant toward the tip causes a distortion of the covalent lattice surrounding the ionized dopant.

This phenomenon causes the top van der Waals layer in the region around the ionized dopant to be pulled toward the tip. As the tip starts to scan a line containing a dopant atom there is virtually no Coulomb interaction between the tip and sample (Figure 3.4(a)). When the tip is within a certain distance of the dopant atom, Coulomb attraction between the tip and ionized dopant atom occurs (Figure 3.4(b)) provided that there is significant band bending in the sample. Only the area in the immediate vicinity of the dopant is pulled toward the tip. The tip retracts to keep the tunneling current equal to the setpoint value, which in turn keeps the tip sample separation constant. As the tip retracts, the surface is pulled up with the tip until it reaches the point at which the Coulomb force between the tip and sample is equal to the van der Waals force holding the layers together. As the tip scans past the dopant, the Coulomb force falls off with the increasing tip-dopant separation distance, and the lifted layer falls back to its original position (Figure 3.4(c)).

Often there are "quantitized" variations in the heights of features observed. This is attributed to the presence of dopants in both the first and second van der Waals layer of the material. Protrusions appearing with greater heights can be assumed to result from dopants in the topmost layer, while dopants in the second layer will not have as strong of a Coulomb force interaction with the tip, due to charge screening and increased tip dopant separation. Evidence for the height variations resulting from doping in different layers comes from an experiment in which the top layer was etched away during STM dopant observation.[25] Etching of TMDs during ambient STM experiments is well documented.[12,13] Although the exact mechanism has not been determined, etching has been attributed to electrostatic tip sample forces, electrochemical redox reactions, and

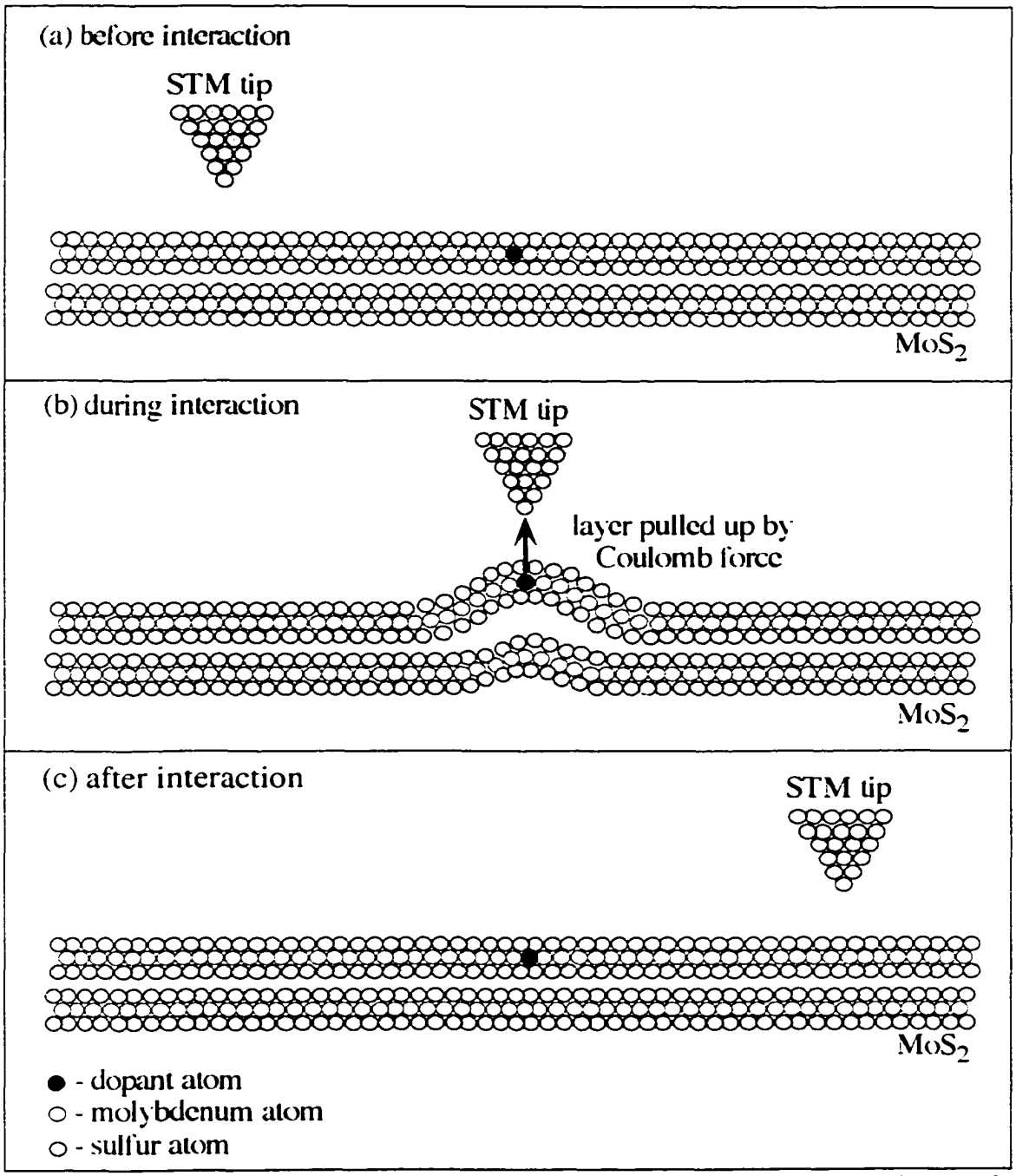


Figure 3.4: Diagram indicating the proposed mechanism by which van der Waals layers of a TMD are pulled apart when imaging in the vicinity of a dopant atom when imaged using STM. (a) As the tip starts, a scan line containing a dopant atom the top MoS_2 layer remains in contact with the underlying layer. (b) When the tip is inside a distance allowing for Coulomb attraction the top MoS_2 layer starts to be pulled toward the tip. The tip retracts under the control of the feedback loop until the Coulomb force equals the van der Waals interaction holding the MoS_2 layers together. (c) As the tip moves away from the dopant the Coulomb attraction falls off and the top layers fall back to their original position.

mechanical abrasion occurring during scanning. The true mechanism for this etching is most likely due to a combination of two or more of the above stated processes.

During a few experiments, dopant-related protrusions appearing at two discrete heights were observed. Etching of the top layer of the MoS_2 occurred on the same area in which dopant protrusions of two discrete heights were observed. After the top layer was removed by etching, the protrusion, which had appeared with the greater height, disappeared indicating that the observed dopant had been located in the top layer. The protrusion which previously appeared with a lower height reappeared in the post-etch scan with the same lateral position but with an increased protrusion height. The increase in apparent height resulted from the fact that the dopant was now in the top layer. This indicates that dopants can therefore be observed in layers below the topmost surface layer.

As was mentioned earlier, if the formation of features that are on the order of 1 nm in height were solely due to electronic effects, about ten orders of magnitude change in the local density of states (LDOS) in the vicinity of the dopant would be required. In order to examine the contributions of LDOS changes to the observed effects, we performed current imaging tunneling spectroscopy (CITS) (Appendix A) measurements on many samples. Figure 3.5 shows *(dI/dV)/(I/V)* transformed CITS spectra and the corresponding topograph obtained on a synthetic MoS_2 crystal. From the p-type rectifying shape of the original *I(V)* curves (similar curves represented in ref. 28), we determined the doping of the scanned area to be p-type. Similar *(dI/dV)/(I/V)* spectra obtained on p-type GaAs by Feenstra et al. (which has a similar band line up at the tip sample interface) are shifted by about 1 eV to positive biases.[27] This difference is

tentatively related to the different work functions of the tips used. Tungsten oxide (WO_3) has a considerably higher work function (5.8 eV)[28] than metallic tungsten (4.2-4.6 eV).[29] This might explain the 1 eV shift of our curves due to the higher negative voltage needed to tunnel from the conduction band of the sample into the oxidized tip. The image in Figure 3.5 shows four basic types of features on the MoS_2 surface: plane areas, protrusions, shallow depressions, and deep depressions. The depressions and protrusions can be related to the dopants (protrusions) and compensating dopants (depressions) located at different depths below the surface.[30] The CITS spectra, corresponding to the four basic features in the topograph, were transformed into *(dI/dV)/(I/V)* curves in order to determine the LDOS and compensate for bias-induced tip-sample distance changes. Each spectrum shown represents the average of about ten single *(dI/dV)/(I/V)* spectra selected from the region of the respective sites. It is obvious that the spectra do not differ much in their shape between the different areas. The rising parts of the curves on the left and right represent the LDOS of the valence band and conduction band, respectively. It is obvious that the valence band LDOS varies by about 50% of the values measured on the plane areas while the conduction band does not vary at all. In order to explain the protrusion and depression height changes measured on dopant sites, in the nm range, a much greater change in the LDOS should have been detected (several orders of magnitude) between the plane and dopant areas. The CITS measurements justify the assumption that the changes in the LDOS at dopant sites cannot be the sole explanation for their appearance as protrusions with such large heights.

The CITS data presented above was obtained with the following procedure: A full data set representing 64x64 single spectra over the entire scanned area was taken during

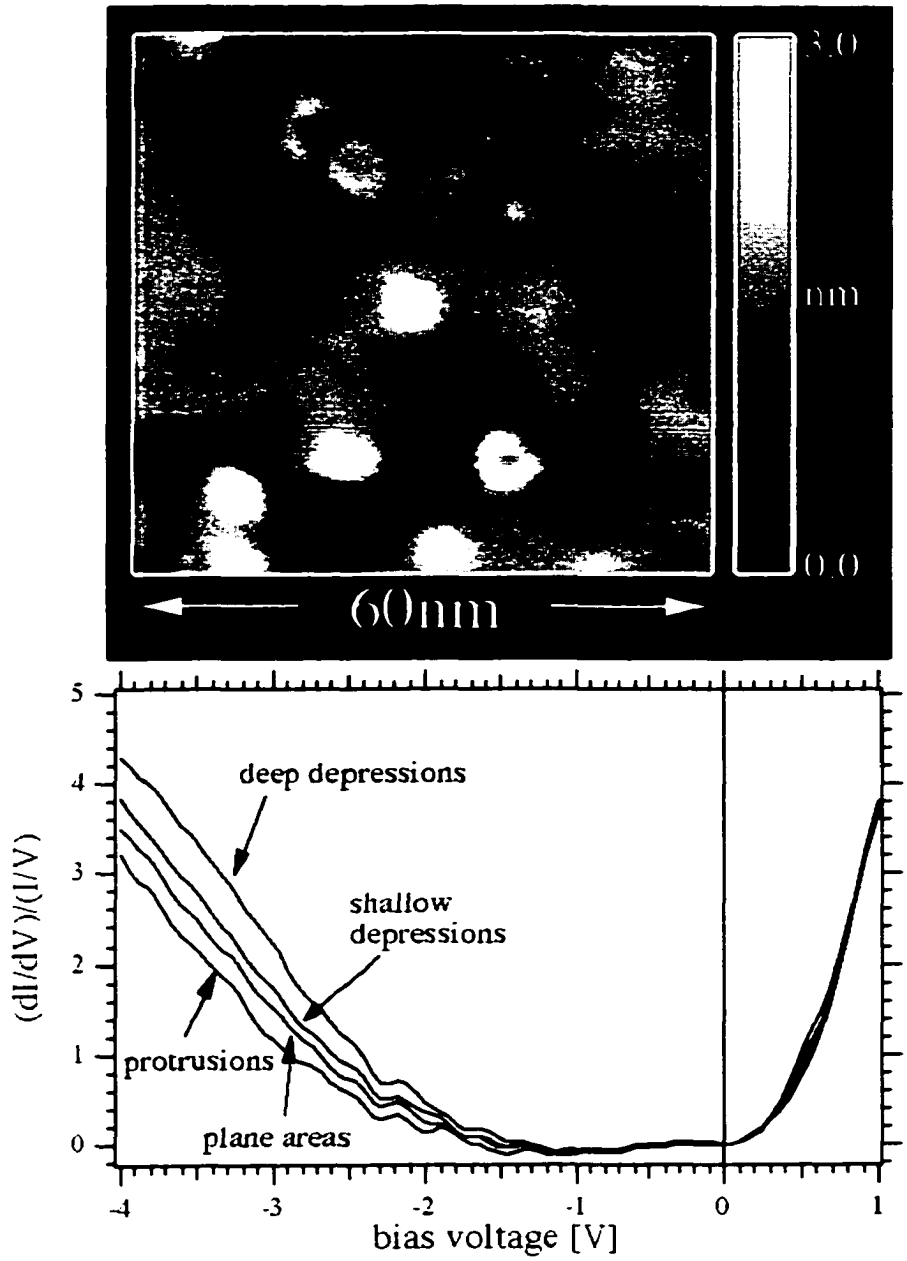


Figure 3.5: Current imaging tunneling spectroscopy (CITS) data obtained on a p-type MoS_2 crystal. The spectra in the lower part represent (dI/dV)/(I/V) curves reflecting the local density of states (LDOS) of different areas of the corresponding topograph (scanning conditions: -1500 mV bias; 100 pA setpoint current) in the upper part of the figure. The curves represent averages of about ten individual spectra selected on protrusions, plane surface, shallow depressions, and deep depressions.

one full scan. The data were then transferred into the data evaluation program IGOR® (Wavemetrics, Lake Oswego, OR). The curves were then selected from each of the different areas of interest from the image and *dI/dV* was calculated for each curve. Since the normalization to the tunneling gap resistance requires the division of the above differential by *I/V*, low currents in the gap region of the spectra induce strong noise in the resulting *(dI/dV)/(I/V)* spectra since *I* is very close to zero. In order to circumvent this problem, we added a small constant of 1 nA to the positive bias side and −1 nA to the negative bias side of the spectra. We want to emphasize that the constants were only applied in the normalization procedure and not in the *dI/dV* derivative. This resulted in the relatively noiseless spectra presented in this work. Experimenting with larger added constants shows that the band edge position remain in basically the same position. The only changes inflicted on the resulting spectra are moderate change in the absolute intensities. A constant of 1 nA did not result in more than a 5% changes in intensity, which is acceptable accuracy for this work.

3.3 AFM Investigation of Electrostatic Forces

A major problem associated with STM investigations of Coulomb force-related effects comes from the measurement process itself. Since the measurement basically results in the detection of a plane of constant local density of states (LDOS), any change in the LDOS will also change the resulting image. It is therefore difficult to sort out all involved effects and identify single contributions such as tip-sample distance changes, band bending, or deviations in the LDOS. In order to overcome these problems and to gain more insight into the Coulomb force contributions to the measurement process, AFM measurements supplementing the STM studies are valuable. The imaging process

with AFM depends only on the forces interacting between the cantilever and the sample. These images are therefore independent of variations in LDOS. The measurement of topographic features with and without an applied bias voltage allows for more insight into the effect of bias induced Coulomb forces.

In all of the following AFM experiments, silicon tapping mode cantilevers (Digital Instruments; Santa Barbara, CA, force constant 20-100 N/m) were used. For the investigation of Coulomb forces, the sample stage was disconnected from ground and connected to a power supply, allowing the application of a D.C. voltage to the sample. A 1 MΩ resistor was placed into the bias circuit in order to protect the cantilever and sample, in the case of point contact. The cantilever was grounded in all experiments. All of the following images were plane fit, and in some cases noise was eliminated by applying a low-pass filter.

Figure 3.6 shows two images obtained by tapping mode atomic force microscopy (TMAFM) with an applied bias applied on the p-type MoS_2 sample. TMAFM proves to be much more sensitive to bias induced effects than regular contact mode AFM. This increased sensitivity can be explained by the difference in the imaging process. In TMAFM changes in the amplitude of the cantilever vibration yield the height information. Thus, the cantilever is not touching the surface most of the time, making it very sensitive to force changes. In conventional contact mode AFM, the cantilever is touching the surface at all times and additional attractive forces do not result in significant tip-sample distance changes. Measurements in conventional contact mode, which monitors attractive forces with the tip, usually within a van der Waals force distance range from the surface, were not successful since the additional bias usually

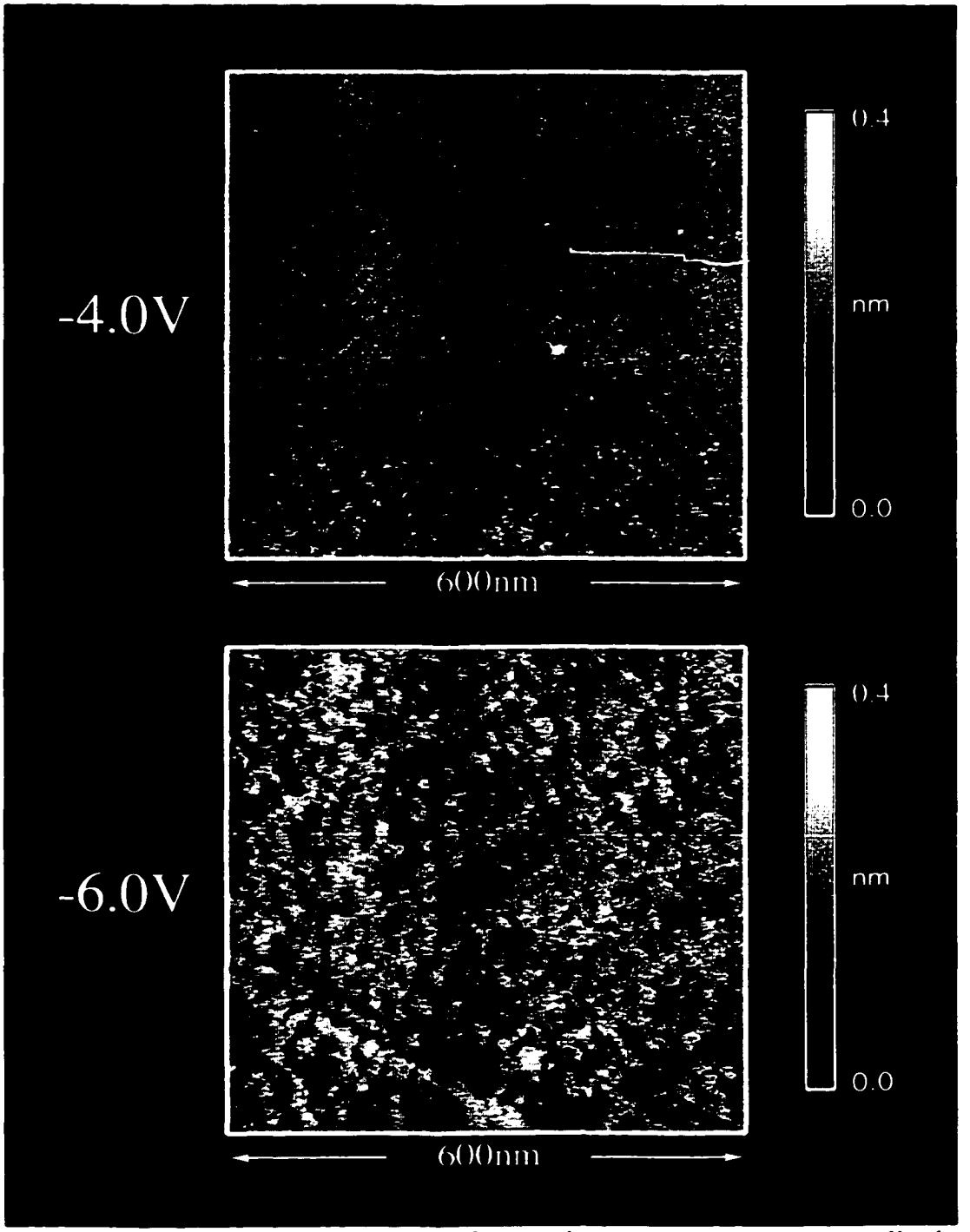


Figure 3.6: Tapping mode atomic force microscopy (TMAFM) amplitude images with additional bias voltage applied to a p-type MoS_2 sample. At –4.0 V bias, the scanned area appeared nearly flat and featureless. At –6.0 V, the whole area seems to be covered with dopant features.

resulted in disengagement of the cantilever. The TMAFM data shown in Figure 3.6 represents the amplitude changes of the cantilever vibration. At 0 V bias, the 600x600 nm^2 area appeared totally featureless and flat. The lower part of the image at –4.0 V shows the development of dopant like structures. At –6.0 V, the whole surface appears covered with these features having an average height of about 0.3 nm. From these measurements, we conclude that Coulomb forces play an important role in the imaging process of dopant sites.

In this context, we cannot rule out that in addition to the Coulomb forces, capillary forces from water adsorbed on the tip and sample may also be involved. Capillary forces may cause an amplification of the heights by adding to the Coulomb force attraction to the surface. This was the case for highly oriented pyrolytic graphite (HOPG) as was described by Mamin et al.[31] In order to investigate the effect of capillary forces adding to Coulomb forces during the local deformation of the surface in the vicinity of dopants, p-type MoS_2 was imaged with an Omicron VT-STM® operating at $5x10^{-10}$ torr. The surface water layer was removed by heating to 350 C° for 45 minutes prior to imaging. Figure 3.7 shows an ultra high vacuum (UHV) STM image (11x11 nm^2) of p-type MoS_2 taken at a tunneling bias of –1500 mV with a setpoint current of 100 pA. Atomic resolution in this image allows for a direct count (~20 atoms across) of the number of atoms affected in the dopant related surface distortion. The protrusion seen here is approximately 0.6 nm in height, while protrusions measured in an ambient environment under the same imaging conditions (bias and setpoint) measured 1.2 nm. This difference in height of 0.6 nm results from the severe reduction or absence of capillary forces in the UHV image. Under ambient conditions, the capillary forces add to

the Coulomb forces, allowing the surface to be pulled to a greater height. The reduced height in the UHV-STM image could also be due to variations in tip shapes. These variations could result in changing the magnitude of the electric field between the tip and sample. Variations in the electric field strength could result in a change in the apparent protrusion height.

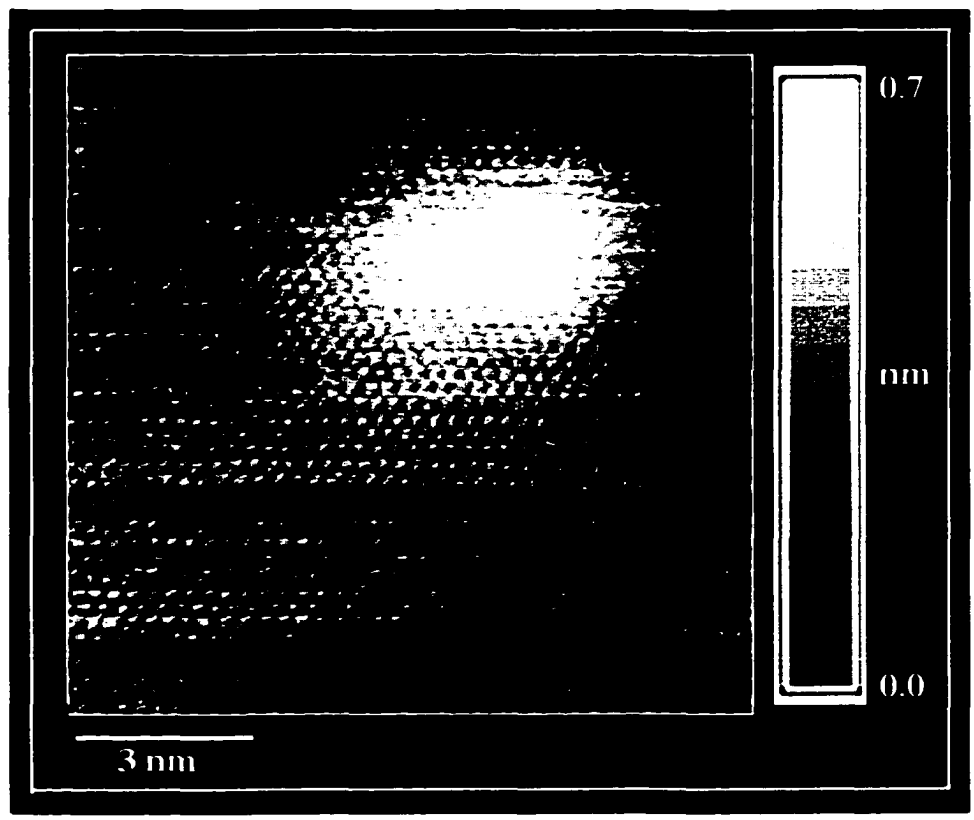


Figure 3.7 UHV STM image of synthetic MoS_2 crystal surface showing a protrusion at atomic resolution. The image was taken at -1500 mV with a setpoint current of 100 pA

A question arises as to whether the surface moves toward the cantilever or the cantilever moves toward the surface. From the data presented in Figures 3.2 and 3.6 it appears as though the surface is drawn toward the cantilever and tip, which is not surprising, since the single sandwich layers of the sample are only held together by relatively weak van der Waals forces. Investigating a much more rigid surface may test

this conclusion. In Figure 3.8, TMAFM images obtained on a gold-coated silicon wafer at different biases are presented. The gold layer consists of small clusters, which are distributed randomly over the surface. The image at 0 V bias shows that the roughness of the surface is about 15 nm. After applying increasingly negative biases, the features become less prominent (-3.0, -3.5 V) and finally disappear into the noise (-4.0, -5.0 V). Note the different height scales for the images at higher biases. It is evident that smaller protrusions disappear first while larger ones, like the feature in the upper right quadrant, are still visible at –4.0 V. This indicates that, in the case of the rigid surface, the cantilever moves away from the surface in order to compensate for the additional bending

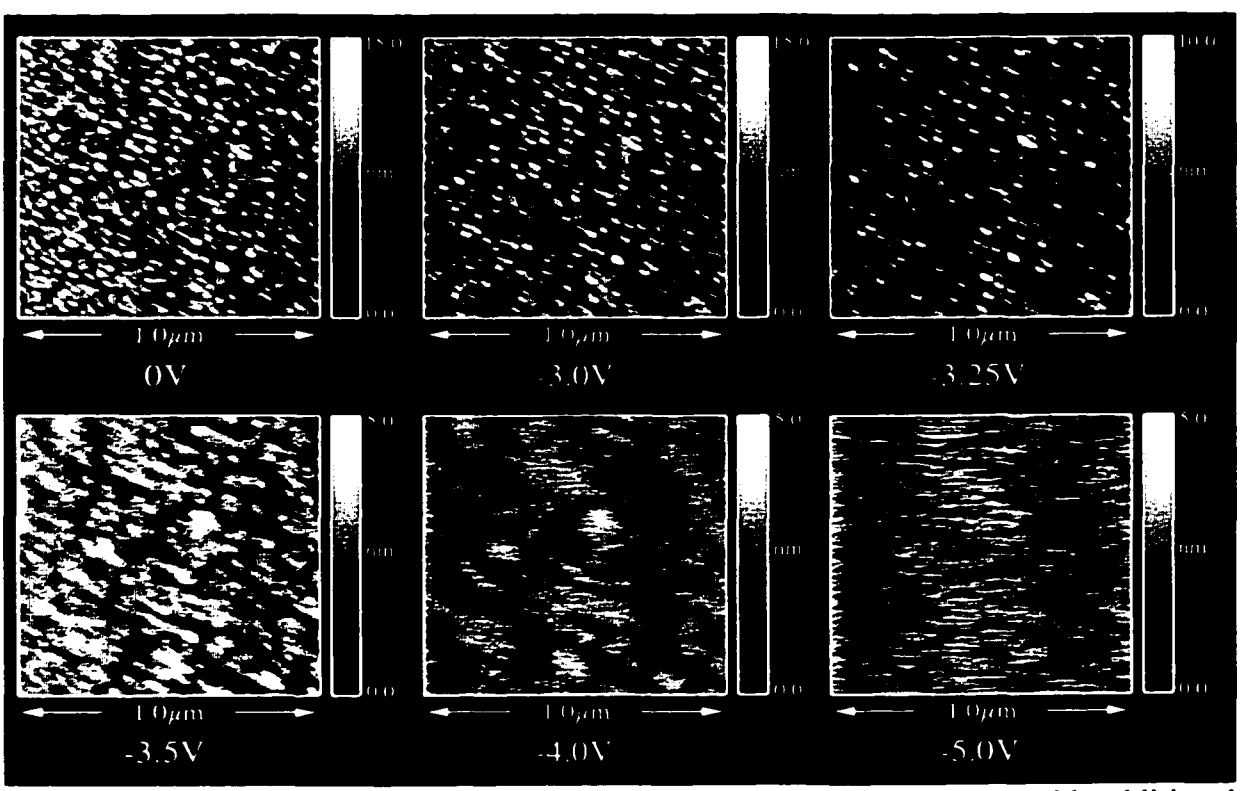


Figure 3.8: Tapping mode atomic force microscopy (TMAFM) images with additional bias voltage applied to a gold coated silicon wafer. At 0 V bias, the scanned area shows a clustered surface with a roughness of about 15 nm. At higher bias voltage (indicated below images), the features diminish and finally vanish into the noise. Note the different height scales of the images.

imposed by the Coulomb force while the surface does not move. The case of the layered material shown in Figure 3.6 opposes that of the metal since the features appear when applying the bias. This implies that the surface follows the retracting cantilever (until the tension in the substrate becomes large enough to compensate for the Coulomb force). The difference in charges between dopant areas and the plane surface then accounts for the height contrast observed.

The magnitude of the Coulomb forces can be determined by measuring force calibration curves (Appendix B) under bias. In Figure 3.9 several curves with and without bias are presented. The topmost curve represents the non-biased case. This curve was used to offset the axes for the correct (0,0) origin position. The curves lying below were taken at –2.0, -4.0, -6.0, -10.0, and –15.0 V bias voltage. All curves were measured while retracting the cantilever from the surface. The nonbiased curve therefore yields the magnitude of the attractive forces including capillary forces that have to be overcome to retract the cantilever from the surface. The determination of the magnitude of this force is accomplished by reading the force value from the curve right before the cantilever snaps back to its relaxed position. A value of about 650 nN is obtained as indicated by the dashed line. This allows an easy estimation of the additional Coulomb forces that act when the bias is applied by determining the additional force it takes to retract the cantilever to its relaxed position.

All of the presented force curves were calibrated using the slope of the contact region of the curves (negative part of the x-axis). The cantilever deflection curves (in nanometers) were then multiplied by the spring constant of the cantilever, yielding the applied force on the sample. The force constant we used was 60 N/m, which represents

the average of the minimum and maximum values stated by the manufacturer (20-100 N/m). The force and z-piezo axes were shifted in such a way that the (0,0) coordinate represents the point where the cantilever would have touched the sample surface without any interaction. This point was obtained by determining the intersection between the horizontal part of the non-biased curve, (far right side), and the contact area of the curve. Negative forces represent attractive forces while positive values indicate repulsive forces between the tip and sample.

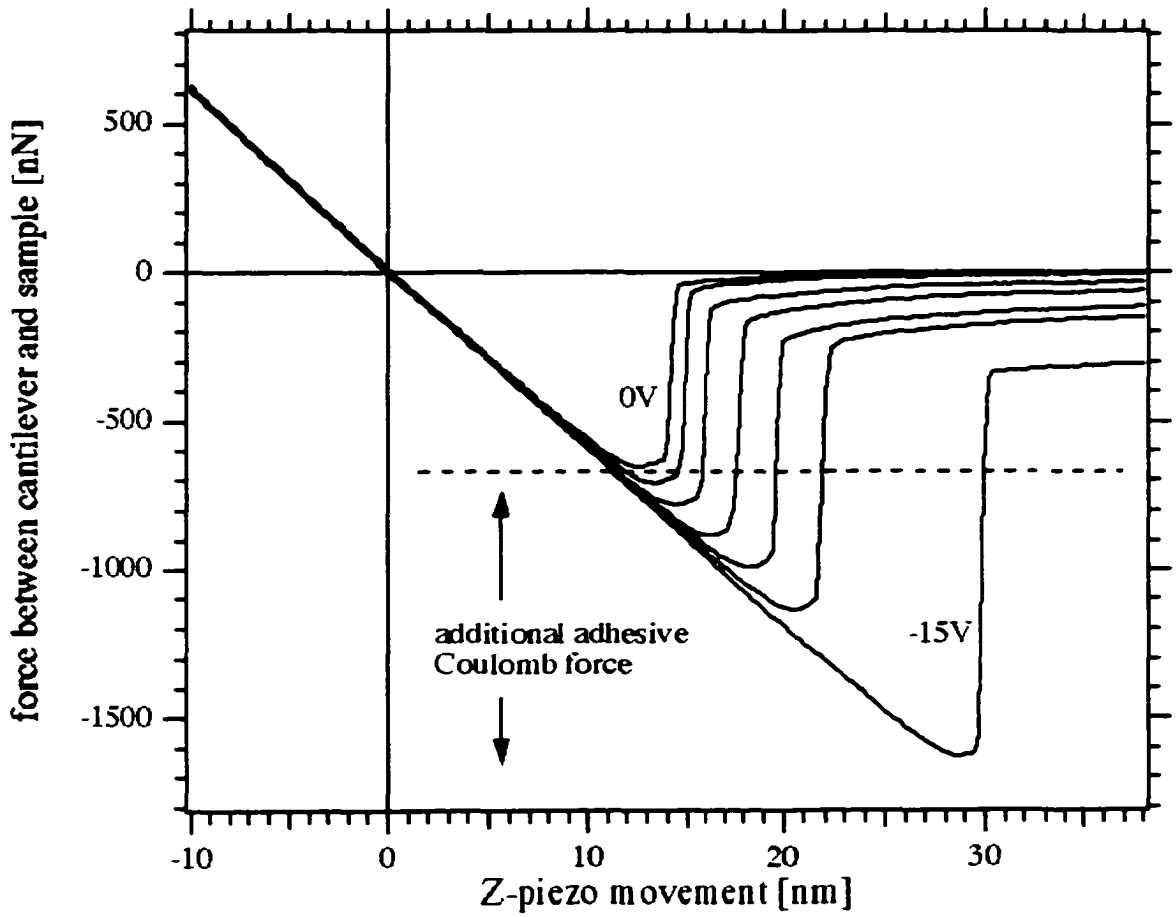


Figure 3.9: Force curves measured on the gold coated silicon wafer with additional bias (0, -2, -4, -6, -10, and −15 V) applied to the sample. The curves were measured while retracting the cantilever from the surface. The bias dependent changes of the force curves allow direct determination of the Coulomb forces between the sample and cantilever.

The results for the Coulomb forces are summarized in Table 3.1. The second column shows the Coulomb forces as determined from Figure 3.9. The third column contains the corresponding charge sitting on both cantilever and sample surface as calculated from the Coulomb equation for electrostatic force F between two point charges Q separated by d:

$$F = \frac{Q^2}{4\pi\varepsilon_0 \varepsilon d^2} \Leftrightarrow Q = \sqrt{4\pi\varepsilon_0 \varepsilon d^2 F} \tag{3.1}$$

A first-order approximation for the charge density on the cantilever apex and sample surface is in the contact area can be obtained from the equation for parallel plate capacitors. The relation between charge Q, bias V, area A, and distance d is given by:

$$A = \frac{Qd}{\varepsilon_0 \varepsilon V} \tag{3.2}$$

Estimates for the areas from this method are listed in the fourth column of Table 3.1. These values are equivalent to 10-14 nm diameter circular areas and are in good agreement with tip apex radii given by the cantilever manufacturer (5-10 nm). All values in Table 3.1 were obtained using dielectric constant of ε=1 (vacuum) and a distance of d=1.0 nm. We believe ε=1 is a fair approximation due to the similarity of our I(V) curves to those presented by Youngquist and Baldeschwieler[32] (which were measured under ultra high vacuum conditions), indicating that the tunneling still occurs through a vacuum gap despite the ambient conditions.

The value d represents the estimate for the distance needed to prevent significant conductivity across the cantilever-sample junction. In order to retract the tip, we believe that it first needs to overcome the capillary forces while the cantilever and sample are in electrical contact, thus preventing significant charge build up at the interface. When the

spring force in the cantilever exceeds the capillary forces, the tip and sample become separated, resulting in a capacitor like charging of the facing areas since the conductivity across the gap decreases several orders of magnitude. Tunneling distances in STM usually range from about 0.4-1.0 nm. Therefore, we estimate that the necessary separation between the tip and sample should be at least 1.0 nm in order to allow one-quarter of an elementary charge (e) to slightly below 1e (at –15 V).

Table 3.1: Coulomb forces depending on the applied bias voltage determined from the force calibration curves shown in Figure 3.9. Estimates for the charge [units: elementary charge (e)] occupying the cantilever, and the area of a hypothetical parallel plate capacitor having the same charge are also given.

Bias	Coulomb Force	Charge	Area (nm^2)
-2	70	17.4	157.6
-4	150	25.5	115.4
-6	230	31.6	95.2
-8	350	39.0	88.1
-10	480	45.5	82.5
-15	990	65.5	79.0

Since the dopant related localized charges should also be about 1e, it becomes clear that the presence of a dopant atom represents a significant disturbance to the charge density at such an interface, especially at low voltages (< 3 V) usually used in STM experiments. The change in the estimated areas at different biases probably reflects geometric effects and bias dependent changes in the cantilever-sample distance while still in close proximity. However, we believe that the assumptions made above are justified in order to obtain a qualitative estimate for the Coulomb forces involved.

3.4 Chapter Conclusions

We investigated the influence of Coulomb forces on STM imaging of dopant sites on layered semiconductor surfaces by performing STM measurements at different bias voltages. The images obtained show a dependence of the apparent height of the dopant sites on the tunneling bias. The heights of the dopant related protrusions is in the nm range, ruling out an explanation based solely on the local density of states and band bending effects. Results suggest that mechanical deformations of the surface due to bias-induced Coulomb forces are responsible for the observed effects. This conclusion is also supported by CITS measurements, which show only minor changes in the LDOS at dopant sites. In order to investigate the effects imposed by Coulomb forces, we performed additional measurements with TMAFM under applied D.C. biases. These measurements showed a strong influence of the applied bias on the measured topography. Measurement of force curves at varying biases allowed the bias-induced Coulomb forces to be measured. An estimation of the charge density on the cantilever and sample was determined from these force measurements. These results showed that the charge density is in the range of one elementary charge per/nm^2 or less, supporting the theory that the presence of localized dopant charges can strongly alter the forces between the tip and sample in STM measurements.

References:

[1] Z. F. Zheng, X. Liu, N. Newman, E. R. Weber, D. F. Ogletree, and M. Salmeron, Phys. Rev. Lett. **72,** 1490 (1994).

[2] S. N. Magonov, H.-J. Cantow, and M.-H. Whangbo, Surf. Sci. **318,** L1175 (1994).

[3] J. A. Stroscio, R. M. Feenstra, and A. P. Fein, Phys. Rev. Lett. **58,** 1668 (1987).

[4] G. Lengel, R. Wilkins, G. Brown, and M. Weimer, J. Vac. Sci. Technol. B **11,** 1472 (1993).

[5] S. Kordic, E. J. van Loenen, D. Dijkkamp, A. J. Hoeven, and H. K. Moraal, J. Vac. Sci. Technol. A **8,** 549 (1990).

[6] P. Avouris, Acc. Chem. Res. **27,** 159 (1994).

[7] M. F. Crommie, C. P. Lutz, and D. M. Eigler, Science **262,** 218 (1993).

[8] Y. Hasegawa and P. Avouris, Phys. Rev. Lett. **71,** 1071 (1993).

[9] A. R. Smith, K.-J. Chao, C. K. Shih, K. A. Anselm, A. Srinivasan, and B. G. Streetman, Appl. Phys. Lett. **69,** 1214 (1996).

[10] R. M. Feenstra, Semicond. Sci. Technol. **9,** 2157 (1994).

[11] Z. F. Zheng, M. B. Salmeron, and E. R. Weber, Appl. Phys. Lett. **64,** 1836 (1994).

[12] B. A. Parkinson, J. Am. Chem. Soc. **112,** 7498 (1990).

[13] E. Delawsky and B. A. Parkinson, J. Am. Chem. Soc. **114,** 1661 (1992).

[14] T. Schimmel, H. Fuchs, K. Graf, R. Sander, and M. Lux-Steiner, Phys. Status Solidi A **131,** 89 (1992).

[15] A. Sato and Y. Tsukamoto, Adv. Mater. **6,** 79 (1994).

[16] I.-W. Lyo and P. Avouris, Science **253,** 173 (1991).

[17] D. M. Eigler and E. K. Schweizer, Nature **344,** 524 (1990).

[18] R. Schlaf, P. G. Schroeder, M. W. Nelson, R. Stübner, S. Tiefenbacher, H. Jungblut, and B. A. Parkinson, Thin Solid Films **331,** 203 (1998).

[19] U. Dürig, J. K. Gimzewski, and D. W. Pohl, Phys. Rev. Lett. **57,** 2403 (1986).

[20] J. M. Soler, A. M. Baro, N. Garcia, and H. Rohrer, Phys. Rev. Lett. **57,** 444 (1986).

[21] M. Kuwabara, D. R. Clarke, and D. A. Smith, Applied Physics Letters **56,** 2396 (1990).

[22] R. Weisendanger, *Scanning Probe Microscopy and Spectroscopy: Methods and Applications* (Cambridge University Press, New York, 1994).

[23] T. W. Matthes, C. Sommerhalter, A. Rettenberger, P. Bruker, J. Boneberg, M. C. Lux-Steiner, and P. Leiderer, Appl. Phys. A **66,** S1007 (1998).

[24] R. Schlaf, D. Louder, M. W. Nelson, and B. A. Parkinson, J. Vac. Sci. Technol. **A15,** 1466 (1997).

[25] D. Louder, Ph.D. Thesis, Colorado State University, 1996.

[26] S. N. Magonov and M.-H. Whangbo, *Surface Analysis with STM and AFM: Experimental and Theoretical Aspects of Image Analysis* (VCH Verlagsgesellschaft GmbH, Weinheim, 1996).

[27] R. M. Feenstra, J. A. Stroscio, J. Tersoff, and A. P. Fein, Phys. Rev. Lett. **58,** 1192 (1987).

[28] H. Höchst and R. D. Bringans, Phys. Rev. B **26,** 1702 (1982).

[29] D. R. Lide, *CRC Handbook of Chemistry and Physics* (CRC Press, New York, 1998).

[30] D. Louder, R. Schlaf, and B. A. Parkinson, unpublished (1996).

[31] H. J. Mamin, E. Ganz, D. W. Abraham, R. E. Thomson, and J. Clarke, Phys. Rev. **34,** 9015 (1986).

[32] M. G. Youngquist and J. D. Baldeschwieler, J. Vac. Sci. Technol. B **9,** 1083 (1991).

Chapter 4: Bias Applied AFM for Dopant Profiling

4.1 BAAFM Introduction

In the previous chapter electrostatic forces were used to image individual dopant atoms located in layered semiconductors. Localized charging of the semiconductor's surface under an applied bias resulted in a Coulomb force between an STM tip and the investigated sample. This Coulomb force was greater than the Van der Waals force holding individual layers of the material together. As a result, the top layers were pulled toward the tip, allowing the dopants to be imaged as topographic STM features. These same electrostatic forces, occurring between a probe and biased semiconductor sample, may also be exploited to image doping in three-dimensional semiconductors such as Si, Ge, and GaAs. In this chapter we extend the use of applied bias scanning probe microscopy based dopant profiling to include patterned silicon semiconductor surfaces using a simple new atomic force microscopy-based dopant profiling method. This new method uses standard Tappingmode® atomic force microscopy (TMAFM) with an added D.C. bias applied between the tip and sample.[1,2] Monitoring the oscillating cantilever's phase signal in TMAFM, which is sensitive enough to detect bias induced Coulomb

forces, allows for resolution of minute surface charge variations between areas of different doping type. This, in turn, allows biased applied TMAFM to be used for mapping out surface doping density variations in the form of phase contrast images. In this chapter this new dopant profiling technique will be discussed in detail along with mechanistic aspects of the measurement process.

4.2 Electrostatic Force Imaging

As was mentioned above, we recently introduced a new TMAFM based technique capable of laterally resolving regions of varying dopant density and type in silicon substrates.[1,2] This method uses TMAFM with an additionally applied direct current (D.C.) bias between cantilever and sample. Depending on its polarity and magnitude, the applied bias voltage produces surface charging resulting in the introduction of Coulomb forces between the cantilever and the sample surface. The surface charge density and strength of these electrostatic forces have been found to vary as a function of doping density. Minute variations in the doping density dependent Coulomb forces can be monitored as a change in the phase of the cantilever oscillation relative to the cantilever driving frequency (TMAFM uses a cantilever excited into resonance with an oscillating piezoelectric driver for the imaging process).[3]

Phase imaging TMAFM has been a new development in the last couple of years.[4-6] This mode of imaging uses an oscillating cantilever as is used in obtaining a height image with traditional TMAFM. Height mode TMAFM imaging is performed using a cantilever oscillating near its resonance frequency.[7] The amplitude of the cantilever's oscillations is then monitored. As the cantilever interacts with the sample surface its amplitude is dampened. Scanning the oscillating cantilever over the sample surface

while adjusting the z-peizo displacement, via a feedback loop, in order to keep the cantilevers amplitude constant is used to produce a "height image." With phase imaging, instead of monitoring the amplitude of the cantilever, a phase signal is obtained by monitoring the angular difference between the cantilever "drive signal" and the detected "response signal." The change in the phase signal is recorded to produce an image as the cantilever is scanned across the sample.

Figure 4.1 demonstrates how a phase image is acquired using a typical TMAFM setup. A time-dependent drive signal is fed into a piezoelectric material, which is mechanically coupled to the cantilever. Oscillation of the piezo via a sinusoidal drive signal causes the cantilever to oscillate. The cantilever is tuned to oscillate near its resonance frequency and the cantilever's response is monitored by reflecting a laser beam off of the back of the cantilever into a position sensitive split photodiode detector. The amplitude response of this detected signal is used as a feedback signal, which is used to produce topographic images. The angular response of the detected signal is compared to the drive signal, using a lock-in amplifier, and the phase angle is measured. The relative change in the angular phase response is plotted as the cantilever is scanned across the sample surface and a phase image is produced.

Phase imaging was originally developed with the intent to map compositional variations in heterogeneous samples.[8-10] Early papers on phase imaging suggested that the phase contrast in images obtained on heterogeneous samples are related to surface stiffness variations or differences in the elastic modulus of the imaged materials.[8] This was suggested to explain the results of several papers in which the changes in hardness of the surface or varied chemical functionality resulted in imaging phase shifts. Recently,

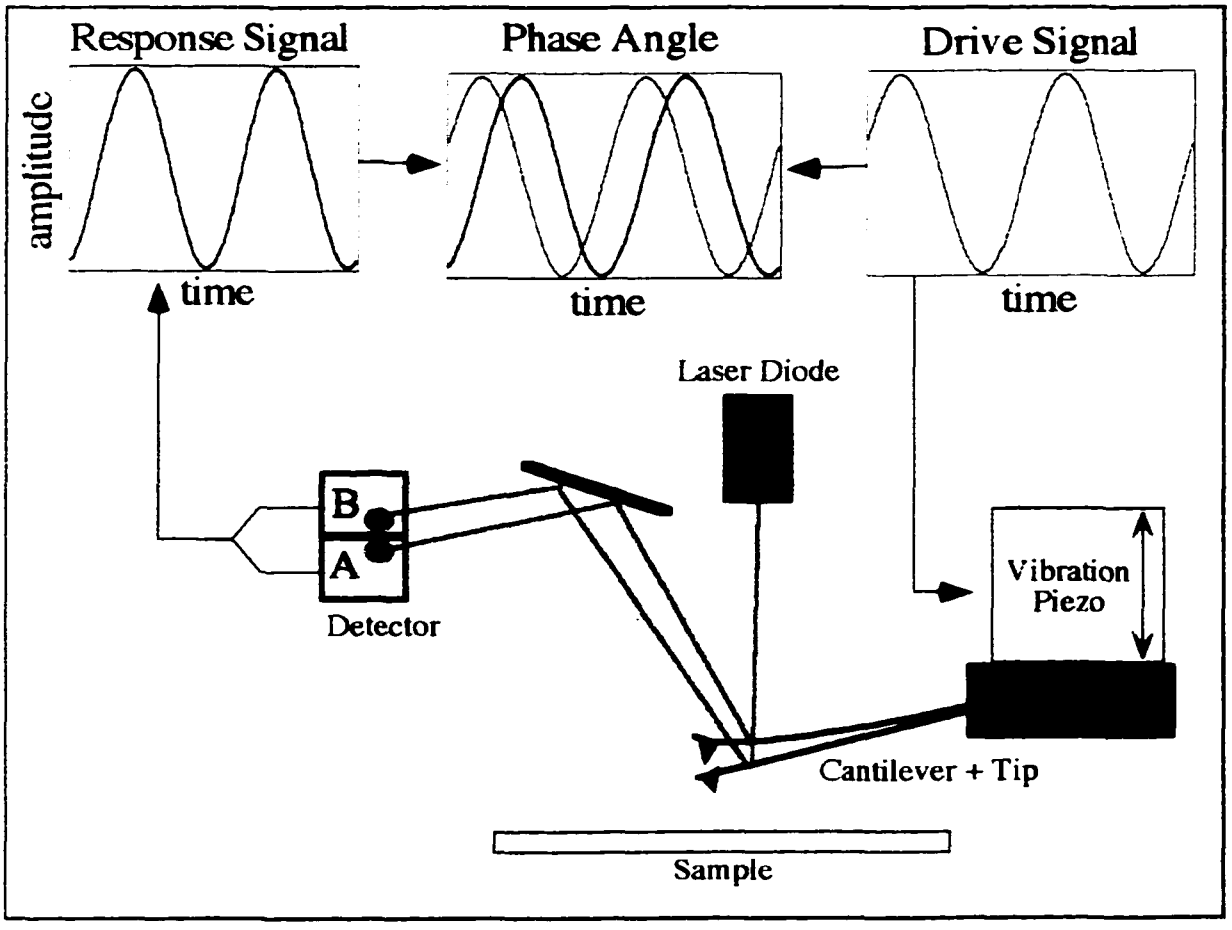


Figure 4.1: Diagram indicating how the vibration piezoelectric drive signal and the detector response signal in an AFM come together to produce a phase signal. The phase angle is the angular difference between the drive and response signals.

several groups have pointed out that phase of the oscillating cantilever can be related directly to the energy loss at the tip-sample junction.[6,11,12] This power dissipation mechanism suggests that phase imaging should be capable of detecting all types of tip-sample forces, which result in varied degrees of energy dissipation from the cantilever-sample system. In order to image variations in a sample's doping density we must rely on a phase shift which results from a Coulomb force attraction between a charged sample surface and an image charge produced in the tip of the oscillating cantilever.

The degree to which variations in the sample surface charge density affects the various properties of the oscillating cantilever can be demonstrated by looking at how the

capacitive force between the tip and an electrically biased metal sample alters the frequency and phase components of the cantilever. In Figure 4.2 the resonance peak shift (4.2(a)) and phase lag (4.2(b)) of a conductive "SESP" cantilever (Digital Instruments, Ni_2Si_3 coated Si, spring constant 2-7 N/m) are shown to depend on the bias applied between the cantilever and a gold-coated mica substrate. Biases were applied by disconnecting the sample stage from ground and connecting a D.C. power supply as well as a 10 MΩ resistor into the circuit. As the bias is increased the surface charge density at the sample's surface increases. As a result, an image charge is formed in the tip of the cantilever and a Coulomb attraction occurs between the surface charge and image charge.

During the measurement the cantilever and sample were held at a distance of approximately 100 nm from each other. The measurement was done using the Nanoscope's "cantilever tune" mode which is normally used in TMAFM to adjust the drive frequency to match the cantilever resonance frequency. In this mode the cantilever amplitude and phase angle are displayed as a function of the drive frequency which is swept across a predefined range. The 100 nm separation distance was adjusted by moving the cantilever closer to the surface until the resonance peak became completely dampened and then subsequently moving the cantilever up 100 nm by using the "step motor" command. The curves in Figure 4.2 were measured by ramping up the sample bias in 2 V increments and capturing the respective amplitude and phase curves using the "print screen" command.

As is obvious from 4.2(a), the resonance frequency shifts from 77.3 kHz at 0 V to 77.04 kHz at 16 V. This shift of the resonance frequency also causes the phase curves to shift as is shown in Figure 4.2(b). This demonstrates that under the actual

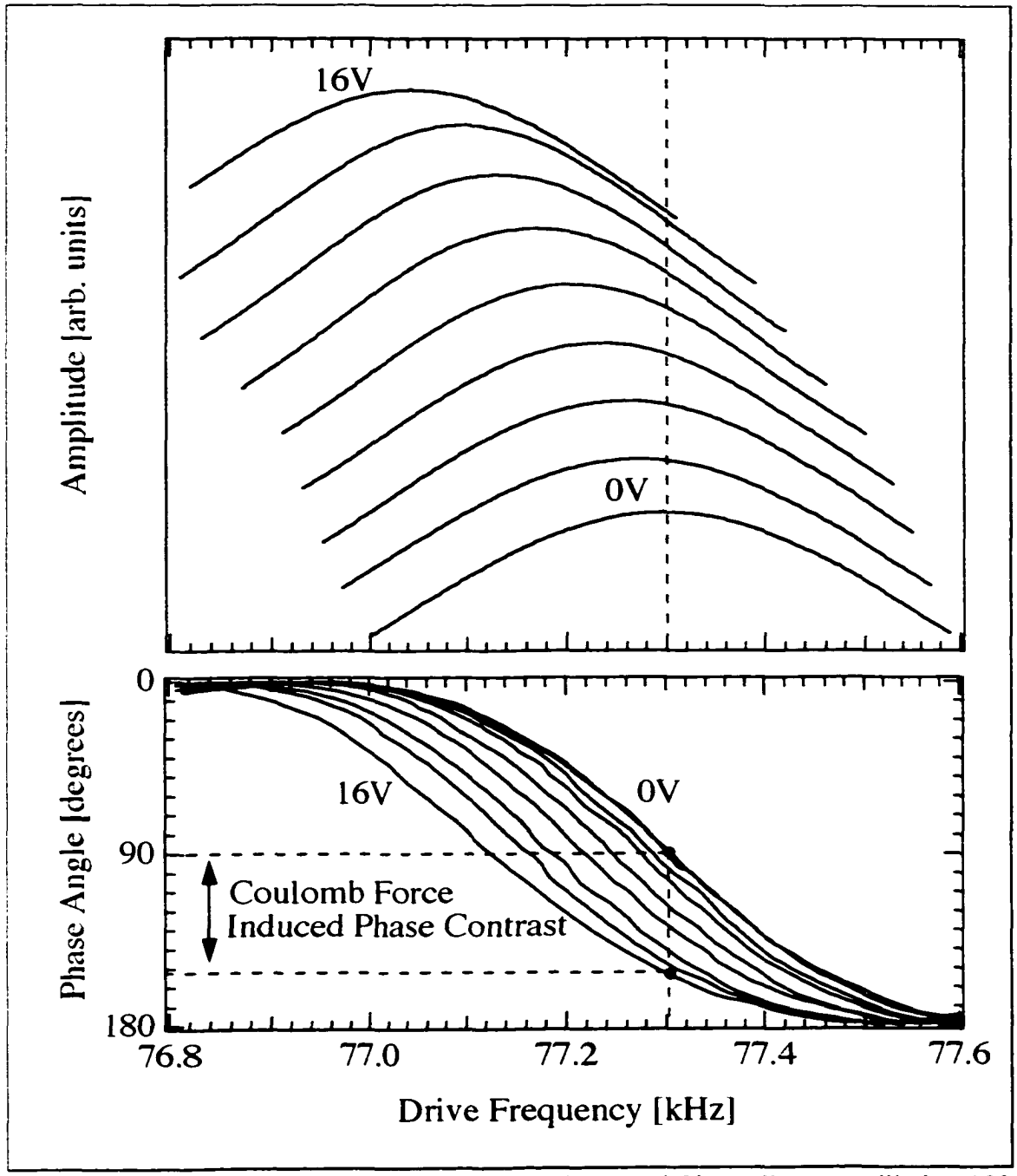


Figure 4.2: (a) Frequency response of a Ni_2Si_3 coated Si cantilever oscillating 100 nm over a gold coated sample biased in 2 volt increments. The curves are offset to improve clarity. (b) Phase angle response of the cantilever depending on the applied bias. Vertical dashed line represents the resonance frequency at zero bias.

measurement conditions in which the drive frequency is held constant near the maximum resonance frequency of the unbiased sample, the phase angle between the cantilever response and the drive signal will change depending on the applied bias. This is due to the resonance frequency decrease caused by the Coulomb force acting between the cantilever and sample. The magnitude of the shift is indicated by the dashed lines in the bottom graph which indicates that a shift of 65° is observed between 0 and 16 V bias.

Since phase imaging is sensitive to various types of tip-sample force interactions it is essential to limit the types of forces which produce the phase signal to only those of interest (i.e., electrostatic forces). Fortunately, phase signals resulting solely from electrostatic force interactions between surface charges and image charges induced in the tip of the cantilever can be monitored with confidence when using interleave lift mode imaging. Topographic and doping dependent phase images can be simultaneously acquired with this method by operating the microscope in the "interleave-lift-mode" in which each line in an image is scanned twice (Figure 4.3). The first scan produces a standard TMAFM height image where the amplitude of the cantilever oscillation is

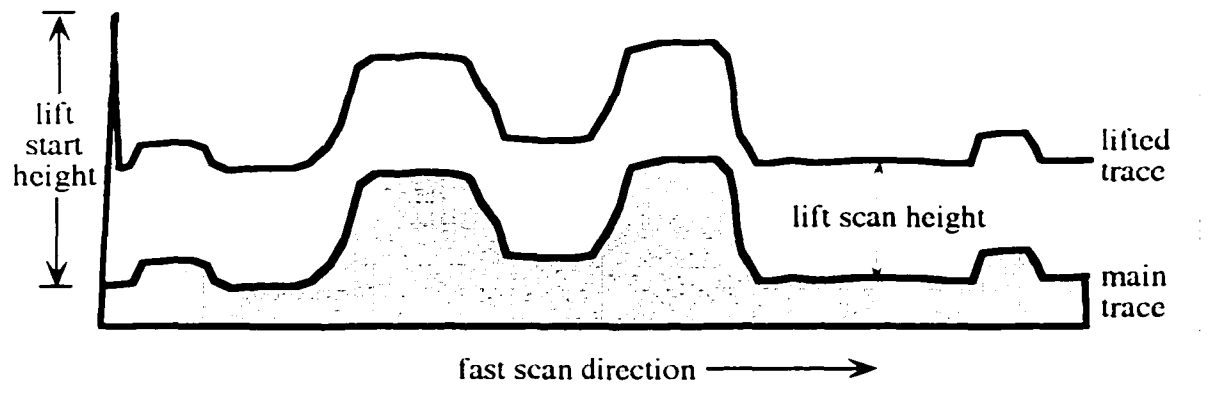


Figure 4.3: Scheme showing a side view of an interleave lift mode scan in which a line contributing to the topographic image taken in the "main trace," and a lift mode phase image scan line is obtained in the second trace. Note that the lifted trace exactly follows the contour of the main topographic trace.

monitored as a feedback signal. The second scan retraces the same line, at a user-defined height above the sample, which follows the previously determined topography profile. The result of this procedure is that the tip sample separation distance remains constant during the second scan when the doping-dependent phase image is measured. Additionally, the tip sample separation is great enough to limit direct contact with the surface, which may produce phase contrast related to factors such as surface stiffness. The interleave-lift-mode procedure, therefore, assures that the measured phase contrast depends exclusively on the sample surface charge density while phase contrast due to varying distance and short-range dispersion forces are avoided.

4.3 Test Structure Investigation

In order to determine if biased interleave lift-mode phase imaging is capable of detecting variations in doping, a silicon test sample was prepared using a photoresist mask and subsequent ion implantation.[13] The sample investigated was a p-doped silicon (100) wafer (boron, $1x10^{17}$ cm^{-3}) containing counter n-doped (arsenic, $5.9x10^{15}$ cm^{-3}) 1.3 μm diameter circular regions (prepared at Symbios Logic, Fort Collins, CO). After implantation the sample was annealed to activate the dopants. Prior to imaging, the sample was dipped into a dilute (pH 3.8) solution of HF to remove the surface oxide. This was followed by rinsing in distilled water and drying under a stream of nitrogen. Figure 4.4 shows a topographic image of a 3x3 μm^2 area of the p-doped silicon wafer which contains 1.3 μm diameter n-type circular regions prepared by ion implantation. The image was measured with TMAFM in height mode and shows one of the circular n-doped regions surrounded by a 5-6 nm high ring of ion-damaged photoresist. It is important to note for the following evaluation of the phase contrast data that the n-doped

areas inside photoresist ring and the p-doped areas outside the rings are of similar topographic height.

The circular pattern was then imaged again using lift mode phase imaging with a D.C. bias was applied to the sample. From this point on, the mode of imaging which uses interleave phase mode with the intent of obtaining a doping profile will be referred to as bias applied atomic force microscopy (BAAFM). All of the BAAFM images in this chapter were obtained using a Nanoscope III Multimode™ SPM (Digital Instruments, Santa Barbara, CA) in which a bias was fed into the sample stage by connecting its ground pin connector to a D.C. power supply. To avoid high currents between the cantilever and sample in the case of point contact, a 10 MΩ resistor was put in series in between the power supply and sample. A conductive cantilever (Digital Instruments,

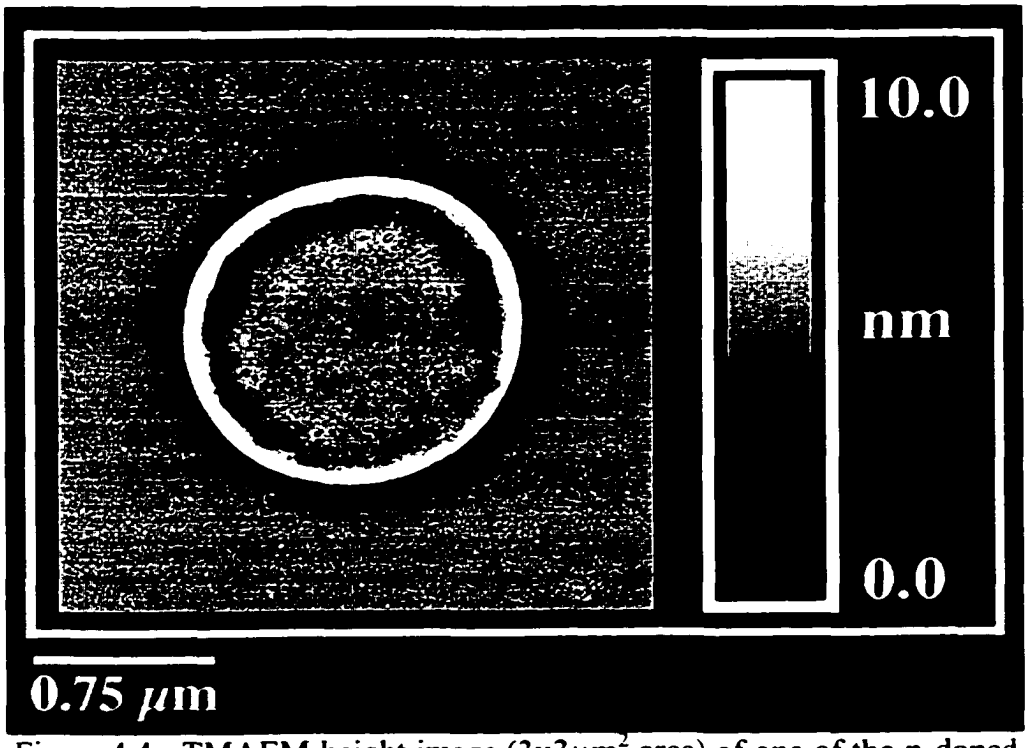


Figure 4.4: TMAFM height image (3x3μm^2 area) of one of the n-doped ion implanted circular areas. For details see text.

Santa Barbara, CA, SESP, Ni_2Si_3 coated Si, spring constant 2-7 N/m, tip radius 20-40 nm, quality factor ≈ 300) was used for all experiments in this chapter. For further details on instrument and software setup, consult Appendix C

The phase contrast images shown in Figure 4.5 are in the units given by the Nanoscope software and were only plane fit. In order to demonstrate the influence of the interleave height, two sets of data obtained at lift heights of 30 nm and 100 nm are shown on the left and right, respectively, in Figure 4.5. Both sets of images were measured at biases ranging from –20 V to +20 V. It is evident that the unbiased images (sample grounded) at both scan heights appear completely flat and featureless, demonstrating the absence of phase changes from influences other than the applied bias. It should be mentioned that the cantilever center-of-oscillation gradually moves toward the surface upon application of increasingly greater biases, potentially causing problems due to contact oscillation dampening with the surface. We will demonstrate below, using calibration curves measured on a gold film, that this effect can be ruled out at the lift heights used in our experiments.

Phase contrast between the n-type circle and p-type surrounding appears when the bias voltage is applied and a contrast reversal occurs upon switching the sign of the applied bias. The appearance of the residue ring in all the images, as an area of smaller phase lag, is most likely due to charging effects in the residual insulating photoresist material. It is evident that at both distances the contrast between n- and p-type regions first increases with the absolute value of the applied bias. Then after peaking at a certain voltage, the phase contrast decreases to the point where it almost completely vanishes. Comparison between the 30 nm and 100 nm lift height images shows that at 30 nm the

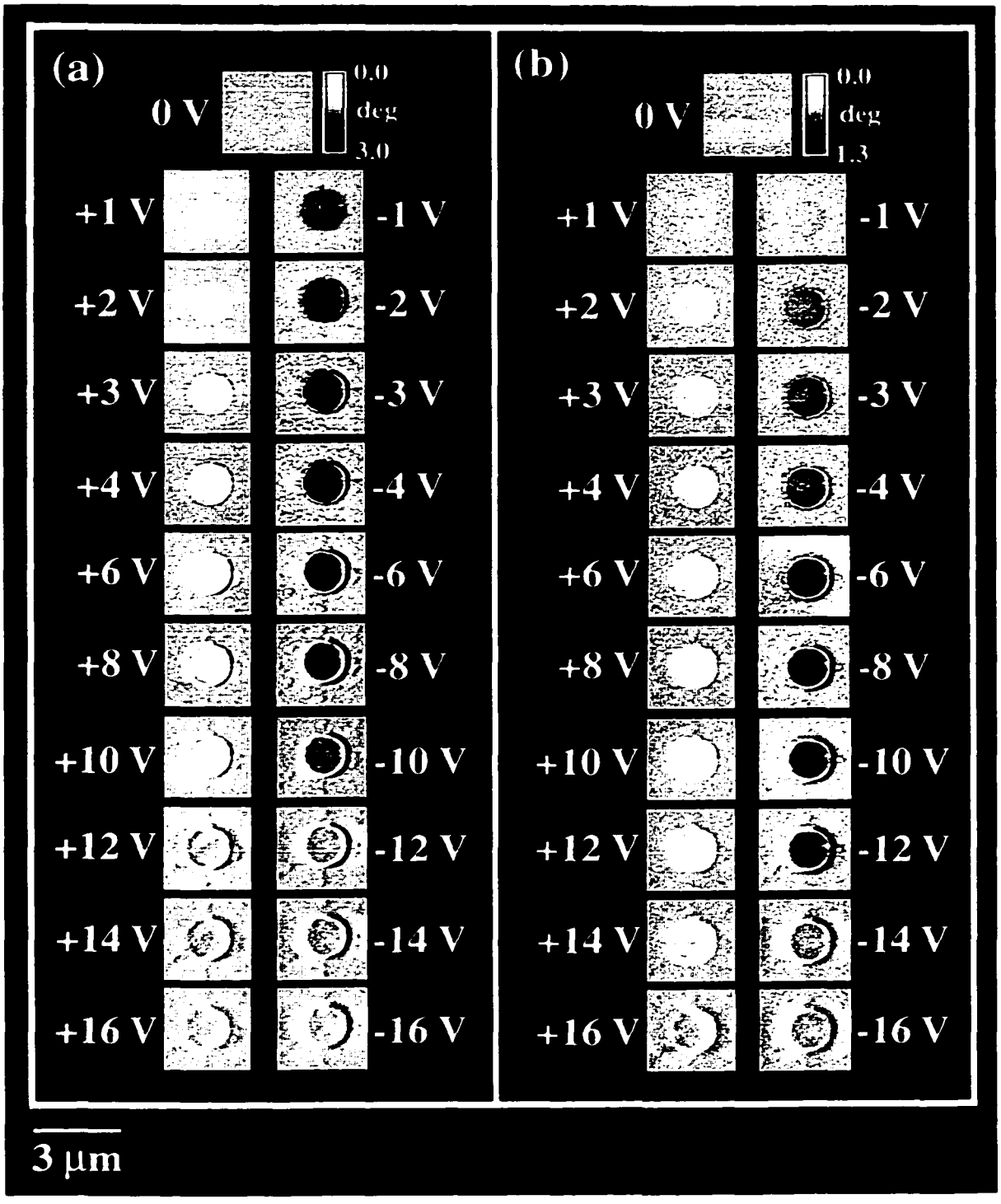


Figure 4.5: TMAFM phase images at different sample biases, on the same area of a silicon sample prepared by arsenic ion implantation of 1.3 μm circular regions ($5.9x10^{15}$ cm^{-3}) into a boron doped ($1x10^{17}$ cm^{-3}) silicon substrate. (a) Images on the left were taken at 30nm interleave lift height while (b) images on the right were taken at a lift height of 100nm.

maximum phase contrast is reached at lower biases and that the contrast diminishes faster than in the 100 nm case.

The main results presented in this chapter are the so-called "phase images." These images show the phase angle between the cantilever drive signal (alternating current) that drives the cantilever via oscillation of a piezo to which the cantilever is attached and the actual cantilever oscillation as displayed relative to the scan position. The above described interleave mode topography and phase angle can be measured simultaneously allowing a correlation between the topographic and phase images. Once a phase image is measured it is plane fit to remove scan artifacts. After completing this procedure the image contains only relative phase contrast variation between the different regions of the image (i.e. p- and n-type regions in our case). To obtain the average total phase angle (Figure 4.6(b)) of the image data relative to the drive signal, the plane fit parameters, which are stored in the header file, were used. Depending on the order of the polynomial used for the plane fit, up to three values are stored in the header after the plane fitting. The first of the three values corresponds to the zero order plane fit parameter that represents the average background phase angle of the entire image. This value was used to determine the absolute phase angle of the cantilever oscillation upon which the phase contrast values are superimposed.

The average phase angle values are assigned by the computer as "two complimentary sets of 16 bit numbers" which range from +32768 to – 32767. This range corresponds to phase angles between 0° and 180°. This assignment adheres to the usual convention used to describe the forced oscillator. In this convention, the driving frequency is fixed and the phase angle defines the phase of the oscillator relative to the

driver. It should be pointed out that in a forced elastic harmonic oscillator the oscillator frequency is always the same as the driver frequency, while amplitude and phase can change in response to changes in the driving frequency. Due to the relatively small amplitude at which the cantilevers are typically resonated, it is assumed that the oscillation is close enough to the harmonic case to use the terminology of the harmonic oscillator. In this convention a 0° phase angle corresponds to the extreme case of a driving frequency that is much lower than the resonance frequency resulting in oscillator movement which directly follows the driver. The 90° phase angle corresponds to the condition where the driving frequency is the same as the resonance frequency, hence the oscillator lags 90° behind the motion of the driver. A 180° phase angle corresponds the high frequency extreme case where the driving frequency is much greater than the resonance frequency, causing the oscillator to move opposite the driver. As was demonstrated in Section 4.2, an increase of the phase angle corresponds to an increase in the Coulomb force between the cantilever and the sample. This results from the attractive Coulomb force reducing the resonance frequency of the cantilever while the driving frequency remains constant. Thus, an attractive Coulomb force interaction corresponds to the situation where the driver is faster than the resonance frequency resulting in a phase angle between 90° and 180°.

The magnitude of the phase contrast for both sets of images and at positive and negative biases is shown in Figure 4.6. In this graph the phase contrast of the n-type region *relative* to the p-type region (which is defined as zero) is plotted with respect to the applied bias. See Appendix D for an explanation of phase contrast and absolute phase. As will be demonstrated below, a larger phase angle between cantilever and

driving frequency corresponds to a larger Coulomb force between cantilever and sample. Therefore, in Figure 4.6(a) a positive phase contrast value corresponds to a stronger force on the n-type area while a negative value indicates a weaker force *relative* to the p-type area. In this context, both of the curves demonstrate that upon bias reversal the force difference between n- and p-type areas also reverses. This is a first indication that the contrast mechanism is connected to band bending in the silicon surface induced by the applied bias. Since the band bending will be affected oppositely on p- and n-type regions at the same bias voltage, contrast reversal is to be expected when the bias is inverted. Figure 4.6(a) also shows that the phase contrast between p and n-regions reaches a maximum at about +5V and –3 V at 30 nm interleave height and at +7 V and –5 V at 100 nm cantilever sample separation.

The asymmetry between positive and negative bias maximum phase shifts in each series is probably related to the initial contact potential difference between cantilever tip and the sample surface due to the difference in work functions of up to 1 eV (work functions of the Ni_2Si[14] cantilever coating and Si (intrinsic)[15], Si (p-type 10^{16} cm^{-3})[15], and Si (n-type 10^{16} cm^{-3})[15] are 4.96 eV, 4.85 eV, 5.30 eV, and 4.40 eV respectively). The contact potential acts as a constant voltage offset superimposed upon the applied bias voltage. The wider peak widths of the 100 nm curve, when compared to the peaks of the 30 nm curve, are caused by the fact that changing the surface potential at larger tip-sample separation requires higher voltages to produce the same change as observed at smaller bias and smaller separation. The difference between the absolute contrast values in both series can be explained by the distance dependence of the Coulomb forces. We want to emphasize here that the contrast observed is the phase angle *difference between*

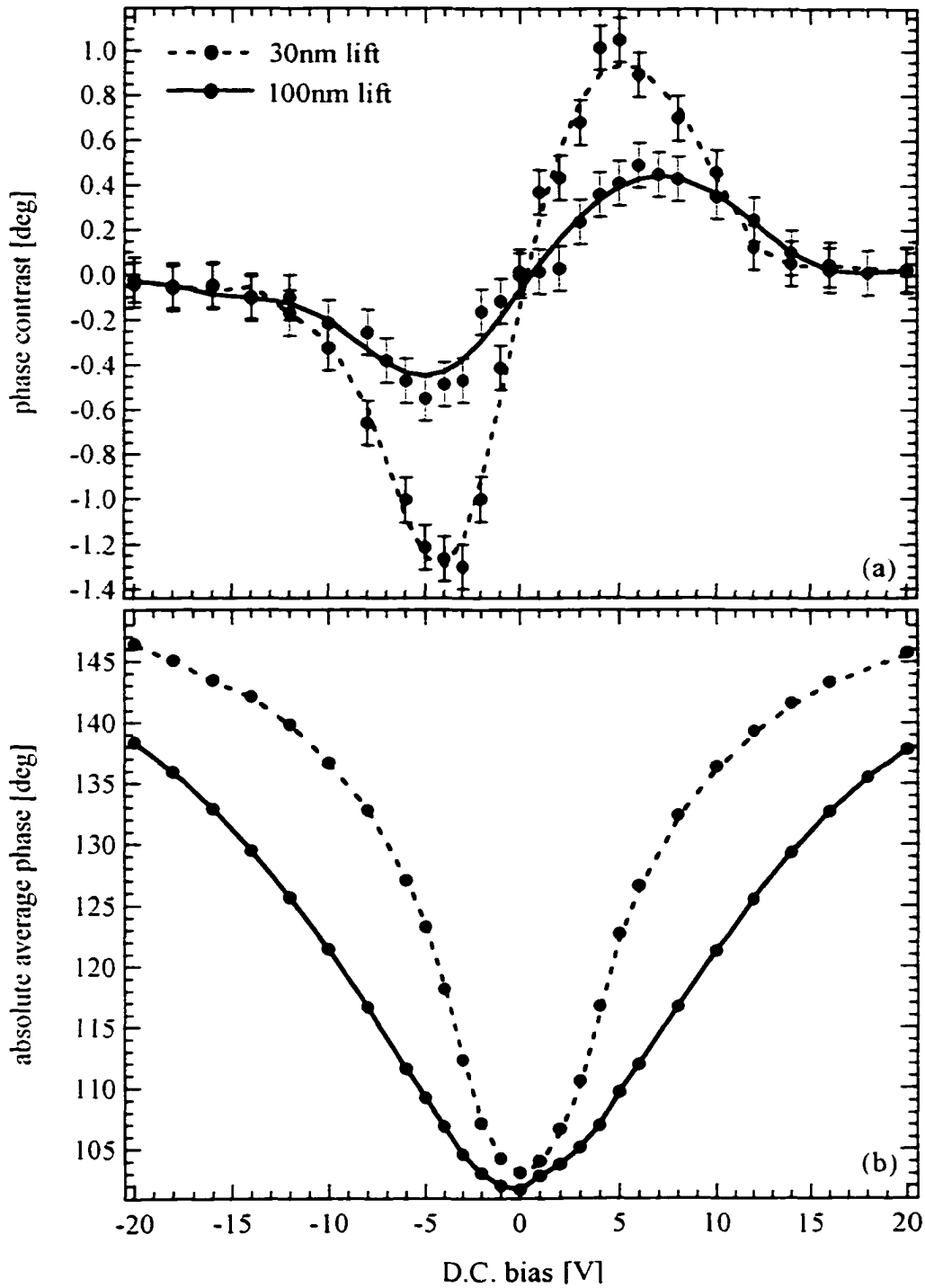


Figure 4.6: (a) TMAFM phase lag on the n-type region relative to the phase lag on the p-type region as a function of bias. (b) Absolute phase angle, averaged over each image, of the cantilever relative to the cantilever drive frequency plotted against the applied D.C. bias. Phase angle data in both images were taken from images presented in Figure 4.5. See Appendix D for an explanation of phase contrast and absolute phase.

p- and n-type regions. The dependence of the *absolute* phase angle change on the bias is shown in Figure 4.6(b). Comparison of the absolute phase angle values in this graph with the phase contrast values between p- and n-type areas demonstrates that the phase contrast between the differently doped areas is only a small fraction of the total phase change caused by the bias voltage. The absolute phase angle values were determined from the zero order plane fit coefficients as explained above. After the plane fit, the image shows the phase contrast between the different regions in the imaged area. This small contribution to the absolute phase angle change contains the information about the local doping density.

4.4 Imaging Mechanism

The overall shape of the contrast curves in Figure 4.6(a) is explained by a model where the surface potential induced at the sample surface by the applied bias is the mechanism responsible for the observed phase contrast between n- and p- doped areas. Figure 4.7 summarizes the changes in surface potential resulting from the applied bias, depending on the dopant type. The five scenarios shown in Figure 4.7 range from highly positive to highly negative sample biases. At 0 V only Coulomb forces caused by the contact potential (due to the work function difference between cantilever and sample areas) are acting between cantilever and the differently doped sample areas. The contact potential is relatively small when compared to the applied bias voltages and could be nearly the same for both areas if the cantilever work function is somewhere in between the work functions of the p- and n-type areas. Little or no contact potential difference results in no observable phase contrast. Positive or negative biases change the surface potentials on both of the doped areas oppositely with respect to each other. Moderate

negative biases increase the band bending on the p-type areas while the n-type area is driven towards the flat band condition. This results in more depletion on the p-type region which needs to be compensated by image charges on the cantilever while there is less depletion on the n-type region. This means that the Coulomb force between cantilever and p-type region will be larger than that between cantilever and the n-type region, resulting in a larger phase lag for the p-type than for the n-type region. At *moderate* positive biases the reverse happens, causing stronger band bending on the n-type area, and driving the bands more towards flat band conditions on the p-type area. This results in a larger phase lag on the n-type region than on the p-type region. At *high* positive or negative biases (top and bottom diagrams) the n-type area is driven into accumulation while the p-type region goes into inversion or vice versa. Both accumulation and inversion conditions are dramatically different from the depletion case since the Fermi level is driven into the bands. This means electronic states at the valence and conduction band edges become emptied or filled, which in both cases results in a high either positive or negative surface sheet charge density. Differences in doping no longer play a role in the phase contrast at high biases since the surface charge density under accumulation or inversion is primarily determined by the density of states at the band edges. Since both the n- and p-regions are either in accumulation or inversion at higher biases, the surface charge density in both regions becomes approximately equal, resulting in the disappearance of the phase contrast between these regions.

The relationship between applied bias, Coulomb force, and induced phase angle changes was investigated by measuring phase angle, amplitude, and force versus distance curves on a gold-coated mica substrate. The sample use for the calibration

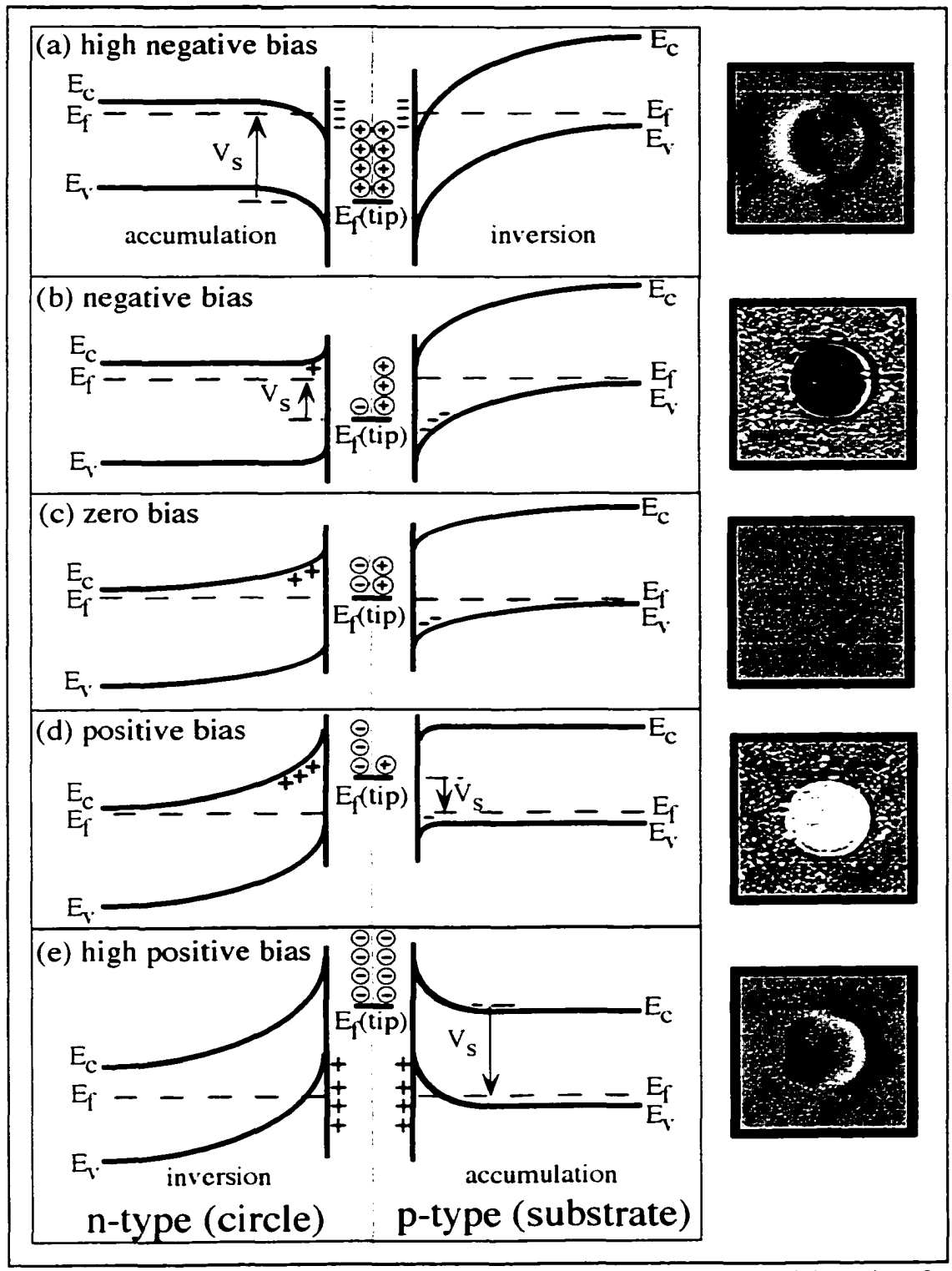


Figure 4.7: Energy level diagrams of the TMAFM-cantilever/air/Si junction for both the n-type area (circle) and p type substrate at positive, high positive, negative, high negative, and zero biases. E_c is the conduction band edge, E_v is the valence band maximum, E_f is the Fermi energy of the Si, E_f(tip) is the Fermi level of the cantilever, and V_s is the sample bias.

measurements was an epitaxial gold film grown on mica. The gold film showed flat terraces and islands of about 300 nm diameter. The curves were measured after the cantilever was positioned at the center of one of these islands. The data evaluation procedures used are discussed in Appendix E. The gold film was used since it provides a good reference due to the absence of any surface potential from surface charging or band bending found in semiconductors. Figure 4.8 shows cantilever phase (a) and amplitude vs. tip-sample separation (b) curves measured simultaneously in Tappingmode® (oscillating cantilever). The bottom graph (c) shows standard force calibration curves measured with a non-oscillating cantilever. All curves shown were measured as the tip approached the sample surface. The force calibration curves (c) directly demonstrate the influence of the bias induced Coulomb force on the cantilever deflection. The topmost curve corresponds to zero applied bias. The curve remains essentially unchanged as the cantilever approaches the surface until a tip-sample distance of about 17 nm. The cantilever then snaps towards the surface as the capillary forces acting on the cantilever exceed the spring constant of the cantilever. The curves beneath the zero bias curve were taken with applied bias. The bias was applied in 1 V steps up to 14 V. The graph shows that the cantilever start positions at the 120 nm distance already change dramatically upon application of the various biases. As the distance is reduced the cantilever bends even more toward the surface due to the increasing Coulomb force. The biased curves shape resemble a quadratic function expected from the influence of Coulomb forces. As the bias increases the "snap to contact" points shift to distances further away from the surface since the electrostatic force adds to the capillary forces when the bias is applied.

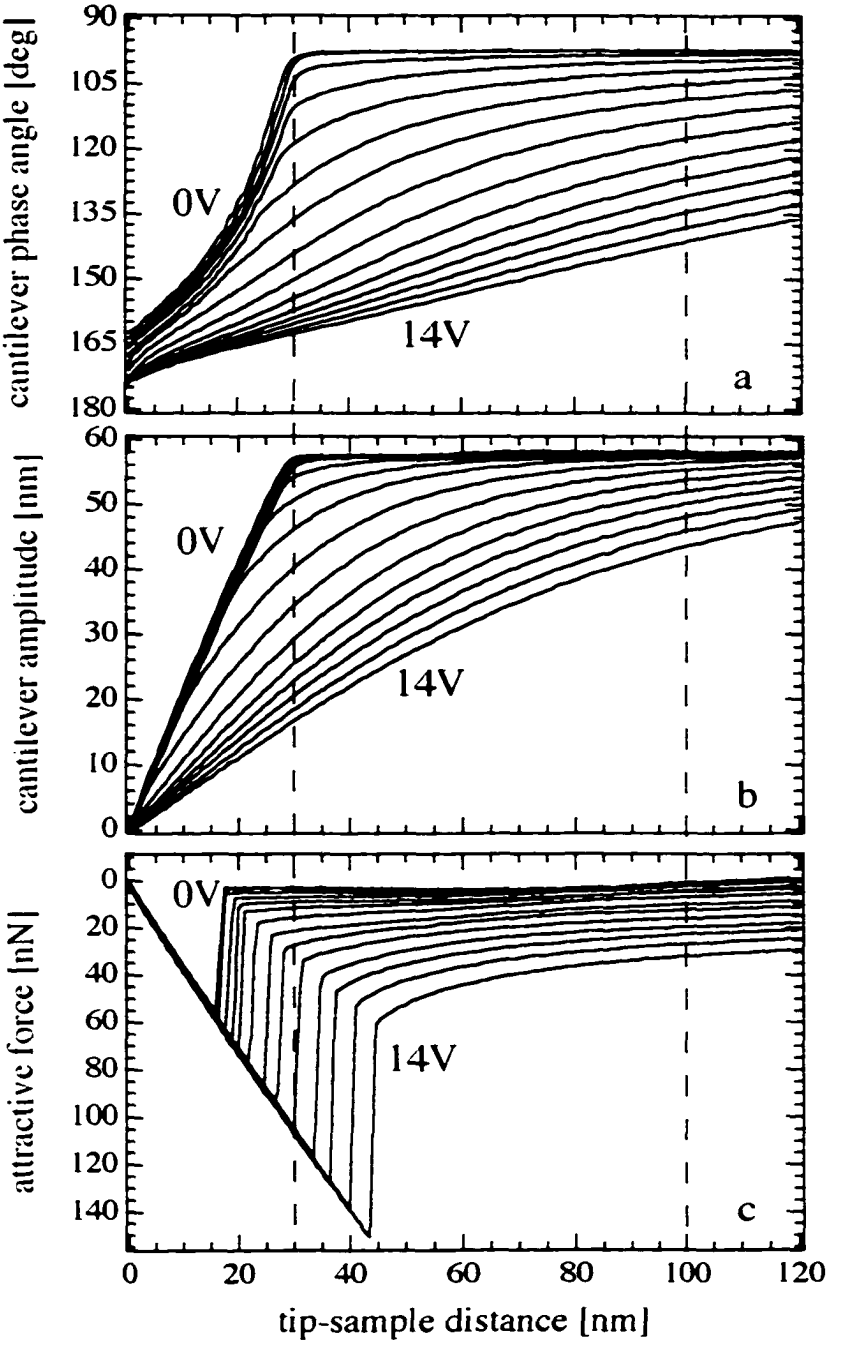


Figure 4.8: Cantilever calibration curves taken at various D.C. biases (1V increments between 1V and 14V) on a gold evaporated on mica sample: (a) TMAFM phase calibration, (b) TMAFM cantilever amplitude curves, (c) Contact mode force calibration curves (force units were obtained by multiplying the cantilever deflection data with the force constant of the cantilever).

The phase and amplitude calibration curves in Figures 4.8(a) and 4.8(b) demonstrate how the bias-induced Coulomb force affects the cantilever oscillation in TMAFM measurements. The curves presented in Figure 4.8(b) show the variation of the amplitude depending on bias and tip-sample separation. The top curve in this graph corresponds to the zero bias case. As the cantilever moves towards the surface (starting at 120 nm) the amplitude of about 58 nm remains unchanged at first, then as the initial contact occurs the amplitude of the oscillation becomes dampened until it is completely suppressed at zero tip-sample distance. As the bias increases the amplitude curves also bend down, similar to the force calibration curves in Figure 4.8(c). Comparison between amplitude and force calibration curves reveals rather peculiar behavior: the vertical line drawn at 30 nm allows for comparison of all curves at the point where the 0 V amplitude curve just starts to tap on the surface. When increasing the bias it should be expected that the cantilever would bend down toward the surface as is shown in the force calibration curves below. Therefore, the amplitude should be attenuated earlier (at larger tip sample separation) than in the unbiased case. However, the data show the opposite. As the bias is increased, the amplitude becomes attenuated smoothly as the tip approaches the surface and contact is only made after the curve intersects the 0 V curve. It is evident that at biases exceeding 8 V the cantilever hits the surface after the amplitude becomes zero. These curves do not intersect the contact region of the 0 V curve. This behavior can be explained by the fact that the application of the bias changes the resonance frequency of the cantilever so strongly that the amplitude is attenuated faster than the tip-sample distance shrinks. This is apparent in Figure 4.8(a) where the phase angle dependence on bias and tip-sample separation is plotted. As the bias is increased the resonance frequency

of the cantilever decreases which means the driving frequency "runs ahead," resulting in a phase angle increase. The data shown in Figure 4.8(a) clearly prove that the cantilever phase angle is extremely sensitive to the bias-induced Coulomb force. The 0V curve shows that the phase angle also increases upon contact between cantilever and surface. However, the bend of the phase curves is opposite in the contact case when compared to the Coulomb force case. This is especially evident in the 5 V, 4 V and 3 V curves where inflection points can be recognized at about 18 nm, 23 nm and 28 nm tip sample distances, respectively. I believe that these inflection points indicate the moment the tip starts touching the surface. These curves also show that at biases exceeding 5 V no contact occurs before the cantilever is at zero distance due to the amplitude attenuation mentioned above.

4.5 Distance Dependence

Since BAAFM is a lift-off technique where an increased tip-sample separation distance is required in order to reduce or eliminate short-range inter-atomic forces, the doping-dependent response signal should be examined as a function of separation distance. In Figure 4.9 phase contrast (a) and absolute phase (b) signals are plotted as a function of tip-sample separation at +5, 0, and –5 V. These phase contrast plots compare the phase signal between a substrate, doped p-type ($1x10^{17}$ cm^{-3}), and an ion implanted region which is doped n-type ($5.9x10^{15}$ cm^{-3}). In the phase contrast curves, (Figure 4.9 (a)) both curves are approximately symmetric about the zero volt curve especially at large tip-sample separations. The phase signal is large at small separation distances and falls off rapidly with increasing separation, indicating that for increased measurement sensitivity small tip-sample separations are desirable. Ideally, the contrast for the zero

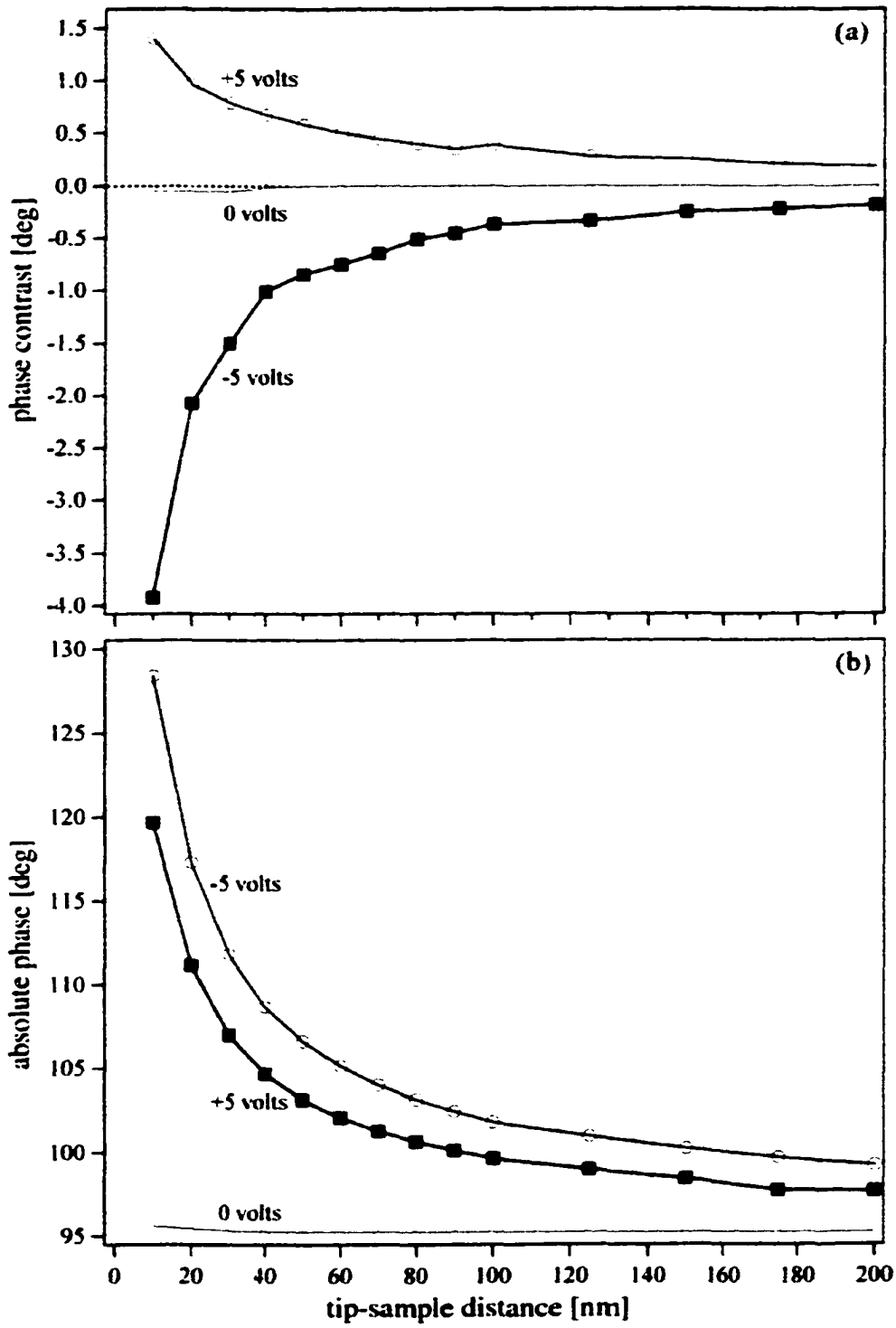


Figure 4.9: Phase contrast (a) and absolute phase (b) curves plotted as a function of the tip-sample separation distance over a substrate doped with p-type ($1x10^{17}$ cm^{-3}) regions and an implanted 1.3 μm diameter circular region which is doped n-type ($5.9x10^{15}$ cm^{-3}). Curves are plotted at +5 V, 0V, and –5 V with respect to the sample. See Appendix D for an explanation of phase contrast and absolute phase.

bias case should be nonexistent at all tip sample separations in which dopant dependent lift imaging is performed. This ensures that tip-sample interactions due to probe contact are eliminated. A slight phase contrast signal is observed in Figure 4.9(a) for separations up to about 30 nm. The signal observed at separations below 30 nm is small compared to the ±5 volt case indicating that it will not likely affect the doping profile. To be safe, all of the doping profiles presented in this text are recorded with tip sample separations of 30 nm or larger. If increased spatial resolution is desired it may be possible to decrease the separation distance to below 30 nm.

The absolute phase curves shown in Figure 4.9(b) indicate how the background signal varies with tip-sample separation distance. At large tip-sample separations the background signal falls off to a value close to 100 degrees. A phase signal of 90 degrees is expected for a forced oscillator being driven at its resonance frequency in the absence of external damping interactions. At large tip-sample separations the 10 degree shift in the phase from 90 degrees to about 100 degrees is most likely due to viscous damping as the cantilever moves through the air.[12] As the separation distance is decreased the Coulomb interaction, due to the attraction of surface charges to charges in the tip and diving board, starts to dominate the background phase signal. At separations of 10 nm the phase angle rapidly approaches 180 degrees. At such close distances one must make sure that the background signal never reaches 180 degrees. A background of 180 degrees may be reached if the Coulomb interaction becomes too great, either by decreased tip-sample separation or by increased surface charge density as a result of an increased surface potential. If the background does reach 180 degrees the detection will be at its maximum and it will be impossible to measure a phase contrast.

In order to determine if the mode of interaction between the tip and surface is force or energy based, phase versus distance curves were fit to the function

$$y = w_0 + \left(w_1 / (x - w_2)^{w_3}\right), \tag{4.1}$$

where the $w_0, w_1, w_2,$ and w_3 are fit coefficients. Figure 4.10 shows fits for the -5 V phase contrast and +5 V absolute phase versus distance curves in Figure 4.9(a & b). The blue curve (dashed) represents the case in which the w_3 coefficient was held constant at 2 during the fit while the red curve (solid) represents the case in which the w_3 coefficient was held at 1 during the fit. It is obvious that for both the phase contrast and absolute phase curves, the best curve fit results when the coefficient $w_3 = 1$. This indicates that the phase signal falls off as the inverse of the tip-sample separation (1/r).

Recently, it has been pointed out that phase contrast can be directly related to power dissipation between the tip and sample.[12,16-18] By taking advantage of the steady-state-motion of the cantilever, and calculating the power loss relative to the power put into the cantilever, it is possible to derive a relationship in which the phase of the oscillating cantilever is directly related to the power dissipated due to the tip-sample interactions.[12,17] At equilibrium the rate at which power is supplied to the cantilever must equal the rate at which power is lost from the cantilever. The power supplied to the cantilever (P_{ext}) is then equal to the sum of the power dissipated by viscous interactions with the environment (P_{air}) and the power dissipated in the tip sample interaction (P_{tip}),

$$P_{ext.} = P_{air} + P_{tip}. \tag{4.2}$$

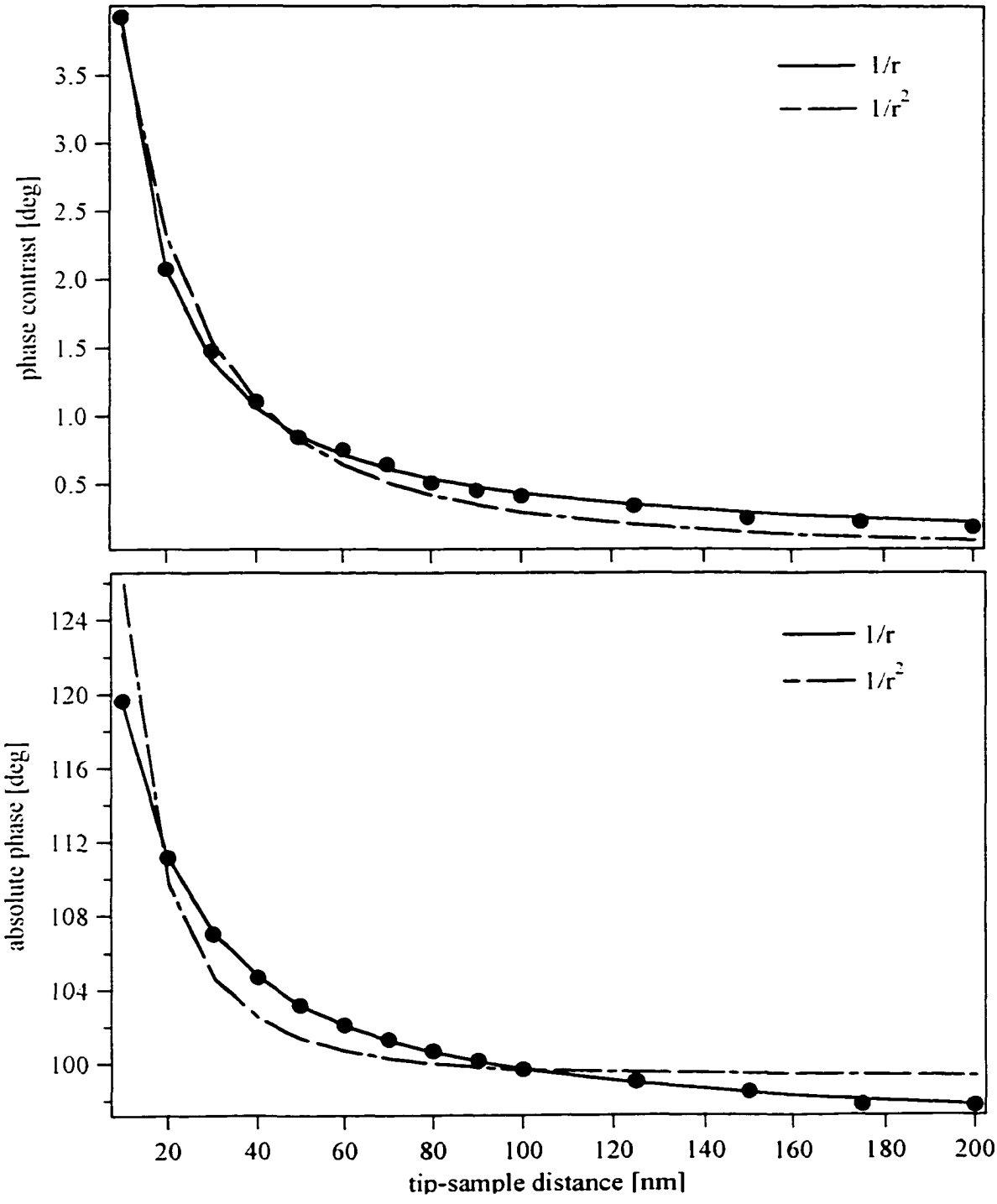


Figure 4.10: Negative 5 V phase contrast (a) and +5 V absolute phase (b) versus distance curves from Figure 4.9 forced to fit a $1/r$ and $1/r^2$ distance dependence. See Appendix D for an explanation of phase contrast and absolute phase.

From Equation 4.2, under the assumption of sinusoidal response an expression, which relates phase angle (ϕ) to the power dissipated by the tip sample interactions (P_{tip}) is deduced;[12]

$$\sin\phi = \frac{2P_{tip}}{k\omega z_1 z_d} + \frac{z_1\omega}{z_d\omega_0 Q_{cant}}, \tag{4.3}$$

where ω and ω_0 are the excitation and natural frequencies of the cantilever, respectively, while z_1 and z_d are the amplitude of the cantilever and the sinusoidal driving amplitude, respectively.

Energy or power dissipated in the tip-sample interaction results from resistive I^2R losses as charges in the semiconductor move away from and toward the surface as a result of the oscillation of the cantilever. Figure 4.11(a) shows the biased tip oscillating at an average distance of z_0 over a semiconductor sample. This system can be modeled as a capacitor between the tip and the sample and a parallel resistor and capacitor in the semiconductor, as shown in Figure 4.11(b). According to Kirchhoff's laws three equations can be written for this circuit. Kirchhoff's laws for left half of the circuit give,

$$V_B = \frac{Q_T}{C_T} + I_R R \tag{4.4}$$

where V_B is the bias at the tip, Q_T is the charge at the tip, C_T is the capacitance at the tip, I_R is the current through the resistive semiconductor, and R is the resistance of the semiconductor. Differentiating Equation (4.4) once with respect to time (t) gives,

$$0 = \frac{I_T}{C_T} - \frac{Q_T}{C_T^2}\frac{dC_T}{dt} + \frac{dI_R}{dt}R \quad . \tag{4.5}$$

where I_T is the current passing through the tip. Kirchhoff's laws for right half of the

circuit produce the following equation,

$$V_B = \frac{Q_T}{C_T} + \frac{Q_C}{C} \tag{4.6}$$

where C is the capacitance of the semiconductor and Q_C is the capacitive charge in the semiconductor. Differentiating Equation (4.6) once with respect to time (t) gives,

$$0 = \frac{I_T}{C_T} - \frac{Q_T}{C_T^2}\frac{dC_T}{dt} + \frac{I_C}{C} \tag{4.7}$$

where I_C is the oscillating current passing through the capacitor in the semiconductor. From conservation of current we can arrive at Equation 4.8.

$$I_T = I_R + I_C \tag{4.8}$$

Combining these equations for the circuit shown in Figure 4.11(b) one can derive the differential equation for I_R,

$$R(C + C_T)\frac{dI_R}{dt} + I_R\left(1 + R\frac{dC_T}{dt}\right) = +V_B\frac{dC_T}{dt}. \tag{4.9}$$

The term on the right hand side is the effective driving term for the oscillating charge.

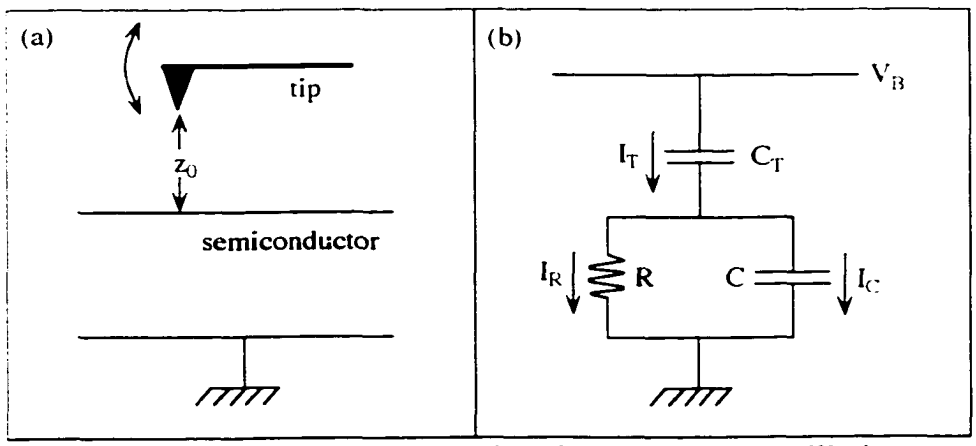


Figure 4.11: (a) Simple schematic of a cantilever oscillating at a distance of (z_0) over a semiconductor. (b) Circuit showing the capacitance (C_T) between the tip biased at V_B and the semiconductor sample, the current (I_R) through the semiconductor with resistance (R), and the current (I_C) through the semiconductor with capacitance (C).

The near exact equation of motion for the oscillating cantilever is given by,

$$z = z_0 + z_1 e^{i\omega t} \tag{4.10}$$

where z_0 is the average tip to sample separation, z_1 is the amplitude of oscillation, ω is the resonance frequency of the cantilever, and $i = \sqrt{-1}$. The oscillating cantilever makes C_T oscillate. An approximation of C_T is made by its first two Fourier coefficients

$$C_T \approx C_0 + C_1 e^{i\omega t} \tag{4.11}$$

where C_0, and C_1 are the average and variable capacitances, respectively. Making a similar approximation for the current gives

$$I_R = I_0 + I_1 e^{i\omega t}. \tag{4.12}$$

Putting equations 4.10, 4.11, and 4.12 into equation 4.9 gives

$$I_R = \frac{\omega C_1 V_B}{\sqrt{1+\omega^2 R^2 (C+C_0)^2}} \left(e^{i(\omega t-\beta)}\right) \tag{4.13}$$

where β can be determined from

$$\cos\beta = \frac{\omega R(C+C_0)}{\sqrt{1+\omega^2 R^2 (C+C_0)^2}} \tag{4.14}$$

and

$$\sin\beta = \frac{1}{\sqrt{1+\omega^2 R^2 (C+C_0)^2}}. \tag{4.15}$$

The average power lost in this resistor is given by

$$P_{ave} = \frac{1}{2} R\left[I^* I\right] = \frac{1}{2}\frac{V_B^2}{R}\left[\frac{\omega^2 R^2 C_1^2}{1+\omega^2 R^2 (C+C_0)^2}\right]. \tag{4.16}$$

To this point we have shown that the resistive power losses in the semiconductor are

related to the variable component of C_T, C_1^2. A further simplification can be made by assuming that $C_T << C_0$.

The real system consists of a metallic tip and a semiconductor surface; however, band bending in the semiconductor makes it difficult to model the tip sample capacitance C_T. As a convenient first approximation, I have chosen to model the tip as an insulating sphere of radius Δ, and the sample as a semi-infinite metal. The distance between the center of the sphere and the surface of the semi-infinite metal is z. Using the method of image charges to represent the effect of the semi-infinite metal, the electrostatic potential ϕ at a distance x from the semi-infinite metal is given by,

$$\phi = \frac{Q}{4\pi\varepsilon_0(z-x)} + \frac{-Q}{4\pi\varepsilon_0(z+x)} \tag{4.17}$$

where Q and $-Q$ are the charge of the two point charges. When x equals the difference between the surface to sphere center distance (z) and the sphere radius (Δ) or $(x = z - \Delta)$, ϕ is the potential across the capacitor, Q/C_T. Inserting $\Delta = z - x$ into equation and rearranging gives,

$$\phi = \frac{Q}{2\pi\varepsilon_0}\frac{z-\Delta}{\Delta(2z-\Delta)}. \tag{4.18}$$

The capacitance between the tip and the sample surface is thus

$$C_T = \frac{2\pi\varepsilon_0\Delta(2z-\Delta)}{z-\Delta}. \tag{4.19}$$

Equations 4.18 and 4.19 are valid for the case when the charge on a sphere is uniformly distributed around the surface of the sphere. Since the tip is actually conducting, the charge on the sphere will be pulled to one side of the sphere, creating an electric dipole. This is the case for the charged AFM tip. At large tip sample separation distances the

electric dipole correction will be small. However, Equations 4.18 and 4.19 fall apart when the tip-sample separation (z) is comparable with the tip radius (Δ).

The capacitance between the diving board portion of the cantilever and the surface is significant when comparing it to the capacitance between the tip and the surface. Even though the diving board sits about 10 μm above the surface it's surface area is large, making it's capacitance large. Therefore, capacitance contributions from the diving board must be added in parallel to the tip-sample capacitance as is given by

$$C_T = \frac{2\pi\varepsilon_0\Delta(2z-\Delta)}{z-\Delta} + \frac{\varepsilon_0 A}{d+z}. \tag{4.20}$$

In Equation 4.20 the first term comes from the tip and the second term consists of the diving board capacitance, where d is the distance between the diving board to sample distance and A is the area of the diving board. Inserting equations 4.10 and 4.11 into 4.18, and equating terms in e^0 and $e^{i\omega t}$ gives

$$C_0 = C_T(z_0) = \frac{2\pi\varepsilon_0\Delta(2z_0-\Delta)}{z_0-\Delta} + \frac{\varepsilon_0 A}{d+z}. \tag{4.21}$$

and

$$C_1 = -\varepsilon_0 z_1\left[\frac{A}{(d+z_0)^2} + \frac{2\pi\Delta^2}{(z_0-\Delta)^2}\right] \tag{4.22}$$

Using equation 4.16 for the power dissipation and the approximation that $C_0 \ll C$, one can show that the expression for the average power dissipated has the general form

$$P_{ave} \propto C_1^2 \propto \left[\frac{A^2}{(d+z_0)^4} + \frac{4\pi A\Delta^2}{(d+z_0)^2(z_0-\Delta)^2} + \frac{(2\pi)^2\Delta^4}{(z_0-\Delta)^4}\right]. \tag{4.23}$$

Finally, the phase angle (ϕ) as a function of the tip-sample separation (z) should vary with P_{ave}, since P_{ave} is inserted into Equation 4.3. Since the tip-sample separation

dependence in Equation 4.23 is not obvious, the equation was solved numerically and plotted in Figure 4.12. Phase contrast results from the difference in the signal between the large diving board and the small tip. The absolute phase signal is a sum of the contributions from the diving board and tip. In a metallic system C_1^2 should have the same distance dependence as the "absolute phase" and the C_1^2 cross term, namely $\left(\frac{4\pi A\Delta^2}{(d+z_0)^2(z_0-\Delta)^2}\right)$ from equation 4.23, should have the same distance dependence as the "phase contrast." Both the second and final terms in Equation 4.23 are involved in the phase contrast. However, the final term has a relatively small contribution to the phase contrast with respect to the second term (cross term) except when the tip sample separation distance is small with respect to the tip radius.

In Figure 4.12, the "absolute phase" is modeled by plotting C_1^2 as a function of the tip-sample separation (z_0), and the "phase contrast" distance dependence is modeled using the C_1^2 cross term plus final term, from Equation 4.23. In the modeled system a 5 V bias is applied to a cantilever with a diving board surface area (A) of 250 nm^2 (half of diving board area since one end is fixed), a tip radius (Δ) of 7.5 nm, and a tip length (d) of 10 μm. The C_1^2 and the C_1^2 cross term plus final term curves are fit to the function $C_1^2 = w0/z_0^{w1} + w2$ where $w0$,$w1$, and $w2$ are fit coefficients. During separate fits the w1 coefficient was held constant at $w1$=1 and $w1$=2 in order to see how the calculated data fit $1/z_0$ and $1/z_0^2$, respectively. The $w1$ coefficient was also allowed to float freely during an additional fit as is shown by the $w1$=free curves in Figure 4.12.

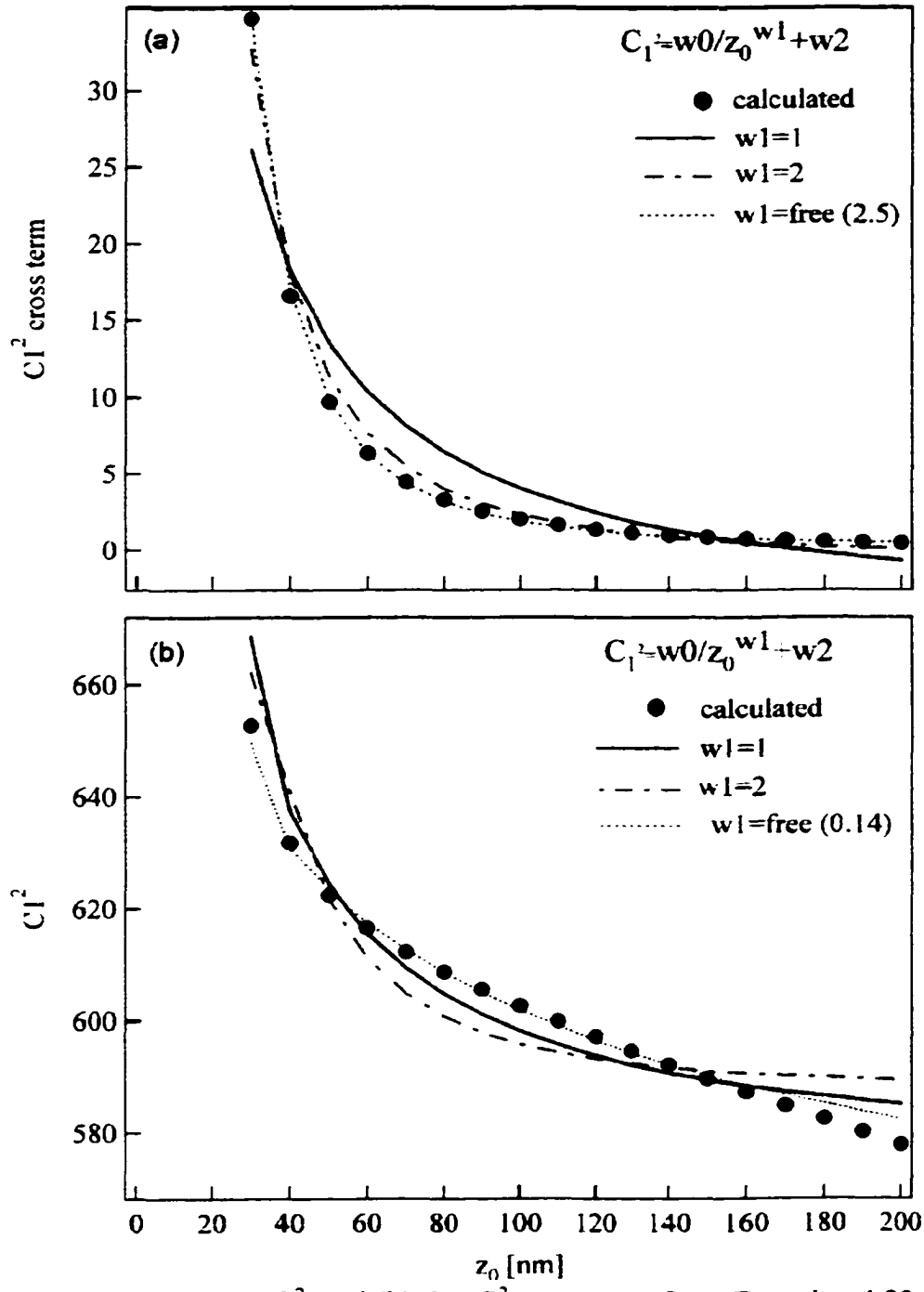


Figure 4.12: (a) C_1^2 and (b) the C_1^2 cross term from Equation 4.23 plotted as a function of the tip-sample separation (z_0). The system is modeled as a 5 V bias is applied to a cantilever with a diving board surface area (A) of 250 nm^2, a tip radius (Δ) of 7.5 nm, and a tip length (d) of 10 μm. Note that C_1^2 should have the same form as "absolute phase" and the C_1^2 cross term should have the same form as the "phase contrast."

The C_1^2 cross term plot in Figure 4.12(a) is fit better by a $1/z_0^2$ function than a similar function held constant a $1/z_0$. The C_1^2 function plotted in Figure 4.12(b) comes much closer to fitting a $1/z_0$ function. Fitting the phase versus distance data from Figure 4.10 gives a 1/distance fit for both the phase contrast curves and the absolute phase curves.

I have identified many possible causes for the discrepancy between the predicted and observed behavior. The first is that non-sinusoidal motion of the tip is significant, and approximations 4.10 and 4.11 are not valid. However, Cleveland et al. have shown that non-sinusoidal motion is not observed with all the tip-sample interactions they have studied.[12] Tamayo has also shown that equation 4.3 is generally valid even when one includes higher order Fourier components to the tip motion, with the only change being that the first term on the right side becomes a sum over the amplitudes of the Fourier components.[19] Non-sinusoidal behavior does not appear to be the source of the discrepancy.

The second reason for the discrepancy is that patch effect fields may cause significant deviations in the electric fields between the sample and the tip.[20-22] Patch effect fields arise when real surfaces with grain boundaries and finite resistances replace the ideal surfaces of my model, leading to electric fields parallel to surfaces. Patch effect fields are significant for tips on the order of hundreds of microns and tip to sample distances on the order of centimeters. In my experiments, grain boundaries will be more important due to the nanometers distances involved. Patch effect fields may be the cause of the discrepancy between my observation and the proposed model.

One of the more likely reasons for the discrepancy between measured observations and the model stems from assumptions made in the modeling process. The model assumes that the semiconductor behave as semi-infinite metallic sheet and the tip as a metallic sphere, while this is not actually the case. It is probably a good approximation to model the tip as a sphere and the semiconductor as a sheet. However, the sample is not a metal with a reservoir of mobile charges, as was assumed in the model. In reality, the semiconductor sample is biased, under depletion conditions, during the distance dependence experiment. The charge distribution in the depletion region of the semiconductor does not behave as a metallic sheet charge. Instead, there is some distribution of fixed charges in the depletion region and the mobile charges oscillate toward and away from the surface as the cantilever tip vibrates above the surface. Performing distance dependent phase angle measurements on a metal, as opposed to a semiconductor, may give some insight into whether the metallic surface approximation in the model is a poor assumption.

4.6 Surface Charge Density

Knowing the surface charge density (Q_s) of the semiconductor as a function of the surface potential (ψ) at various doping densities is important to understand the mechanism of electrostatic force-based dopant profiling. S.M. Sze derives an equation for determining the semiconductor surface charge density in a metal-insulator-semiconductor device in his text "*Physics of Semiconductor Devices*."[23] This derivation uses the one-dimensional Poisson equation to determine the electric field at the semiconductor surface. From Gauss' Law the space charge per unit area required to produce the field is determined to be:

$$Q_s = \sqrt{2\varepsilon_s kTp_{po}} \times \left[\left(e^{-\beta\psi} + \beta\psi - 1\right) + \frac{n_{po}}{p_{po}} \left(e^{\beta\psi} - \beta\psi - 1\right) \right]^{\frac{1}{2}} \qquad (4.24)$$

where p_{po}, and n_{po} are the equilibrium densities of the holes and electrons, respectively. The dielectric constant of the semiconductor, Boltzman's constant, and temperature are represented by ε_s, k, and T, respectively, and $\beta = 1/kT$. The equilibrium electron and hole densities can be approximated as: $p_{po} \approx N_A$ and

$$n_{po} \approx n_i^2 / N_A \qquad (4.25)$$

where n_i^2 and N_A are the intrinsic carrier densities and acceptor doping density. Substituting these terms into the above equation allows determination of the surface charge density as a function of surface potential at different p-type doping densities.

Figure 4.13 shows calculated surface charge densities (Q_s) at various n-type and p-type doping densities as a function of the surface potential (ψ) at 300 Kelvin. The zero potential in the above surface charge density equation represents the flat band situation for each doping density. In order to compare the charge density curves to each other, the Fermi levels at each doping density are offset and referenced to the intrinsic level which is shown in Figure 4.13 at zero potential. The p-type surface charge density curves were calculated and the n-type curves were plotted using the assumption that the absolute value of the surface charge is symmetric about the intrinsic level. Surface charge density values calculated in Figure 4.13 do not take into account the effects of surface states and bulk defects in the semiconductor. In the real semiconductor there is likely to be some charging under flat-band conditions due to surface states and bulk defects. This should result in an increase in the calculated surface charge near the flat-band potential, located

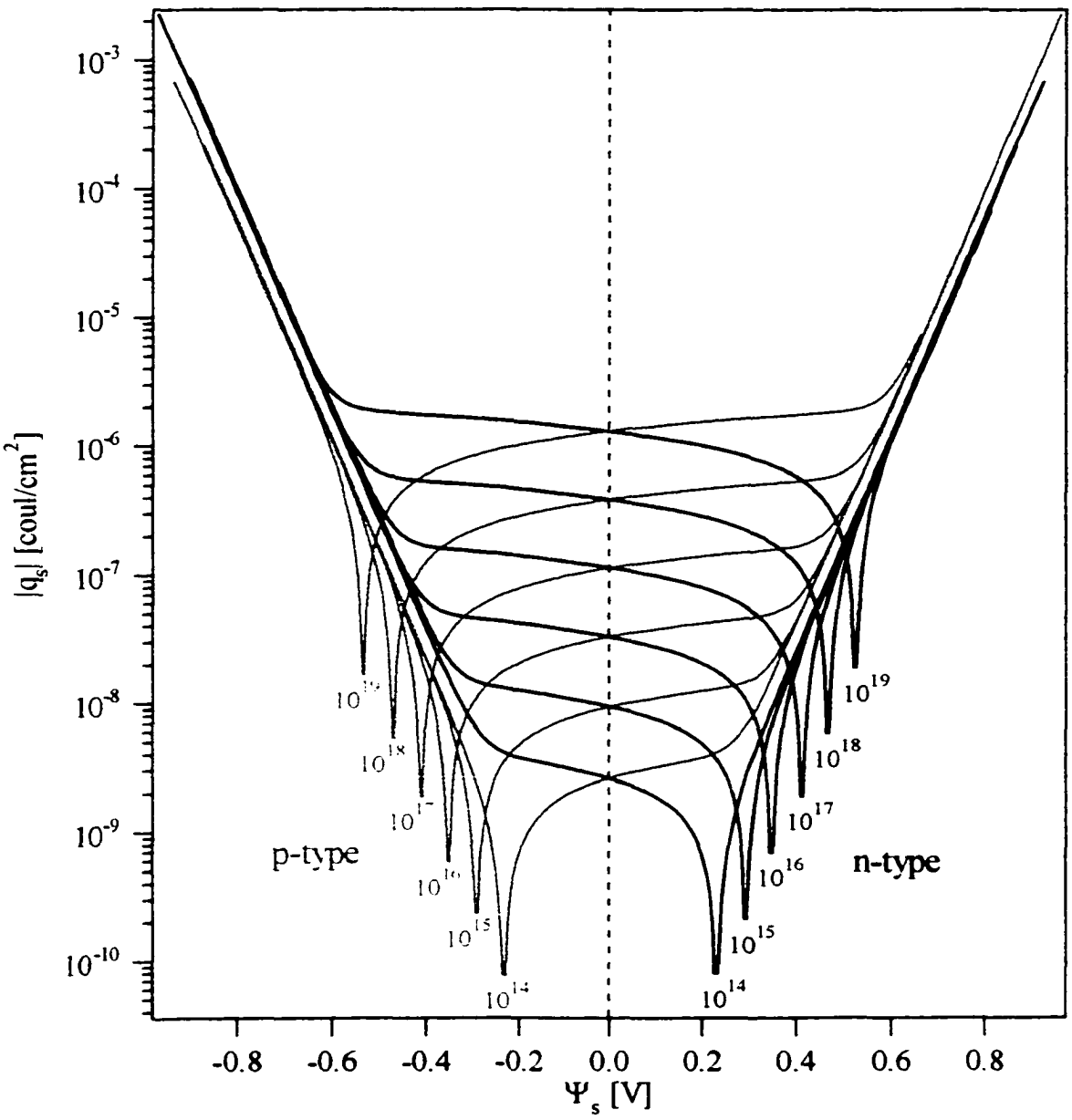


Figure 4.13: Calculated surface charge density (q_s) as a function of surface potential (Ψ_s) for various doping densities (dopants/cm^3) of silicon. Surface charge density calculations do not take into account defects and surface states.

at the lowest charge density region for each curve in Figure 4.13. If surface states due to dangling bonds and foreign atoms exist at the surface it is possible that the Fermi level of the semiconductor is pinned at the surface.[24,25] This would result in a surface charge density verses surface potential plot that is different from the one shown in Figure 4.13.

Surface charge densities in Figure 4.13 can be used to explain some of the phase behavior observed in BAAFM imaging of doped semiconductors. In BAAFM the difference in the surface charge density appears as a phase contrast between two areas of different doping density. Comparing the surface charge density for two curves of differing doping density in Figure 4.13 but at a given surface potential should indicate the relative magnitude and sign of the phase contrast that would be expected under ideal conditions. The absolute phase behavior that is seen in all of our BAAFM experiments is also explained by the surface charge density curves, as is discussed below.

A set of standard samples provided by the National Institute of Standards and Technology (NIST) were imaged using BAAFM in order to demonstrate how phase data relates to the surface charge density calculations presented in Figure 4.13. The NIST samples, which will be called N1, N2, N3, and N4, consist of 5 μm stripes implanted with 50 keV boron ions into a substrate doped n-type to 1 Ωcm (~5×10^{15} cm^{-3}).[26] The boron ion dose for the N1, N2, N3, and N4 samples were 10^{13}, 10^{14}, 10^{15}, and 10^{15} cm^{-2}, respectively. The implants were activated at 900 °C for 10 minutes. Samples N1, N2, and N3 were dipped into a 2% HF solution prior to imaging. However, a 10 nm-thick oxide was grown over the implanted regions of the sample N4 and a 20 nm oxide was grown over the unimplanted regions.

Figure 4.14 shows phase contrast curves as a function of sample bias for the N1, N2, and N3 NIST samples. These samples consist of an n-type substrate (~$5x10^{16}$ cm^{-3}) which is dosed with 50 keV boron ions. The doses are not large enough to compensate for the number of n-type dopants present, therefore, regions of the samples implanted with boron remain n-type. The low ion doses result in the doping density in the implanted region effectively being lower doped than the n-type substrate since each ion cancels the effect of a n-type dopant. Figure 4.14 indicates that distinguishing between two doping densities that are similar, such as the implanted region versus the substrate in the N1 sample, requires an increased applied D.C. bias. Variations in the doping density between the substrate and implanted region in the N3 sample produce a larger difference in the surface charge density which results in a larger difference in the phase angle recorded at a given bias. This in turn results in a phase contrast that can be detected at lower applied biases.

In order to determine whether Figure 4.13 properly reflects changes in phase contrast between areas of differing doping density will require the development of a good set of doping standards. A standard containing all of the doping densities presented in Figure 4.13, both n- and p-type in one sample, would be beneficial. Absolute phase contrast difference between all of the various regions could be used to validate the use of Figure 4.13 for interpretation of the BAAFM results. Since the surface charge density scale in Figure 4.13 is logarithmic there may be some doping density in which the surface charge density is too low to induce a phase shift. Differentiating between two doping densities below this threshold doping density would therefore be impossible. A good doping standard could be used to determine the lower detection limit.

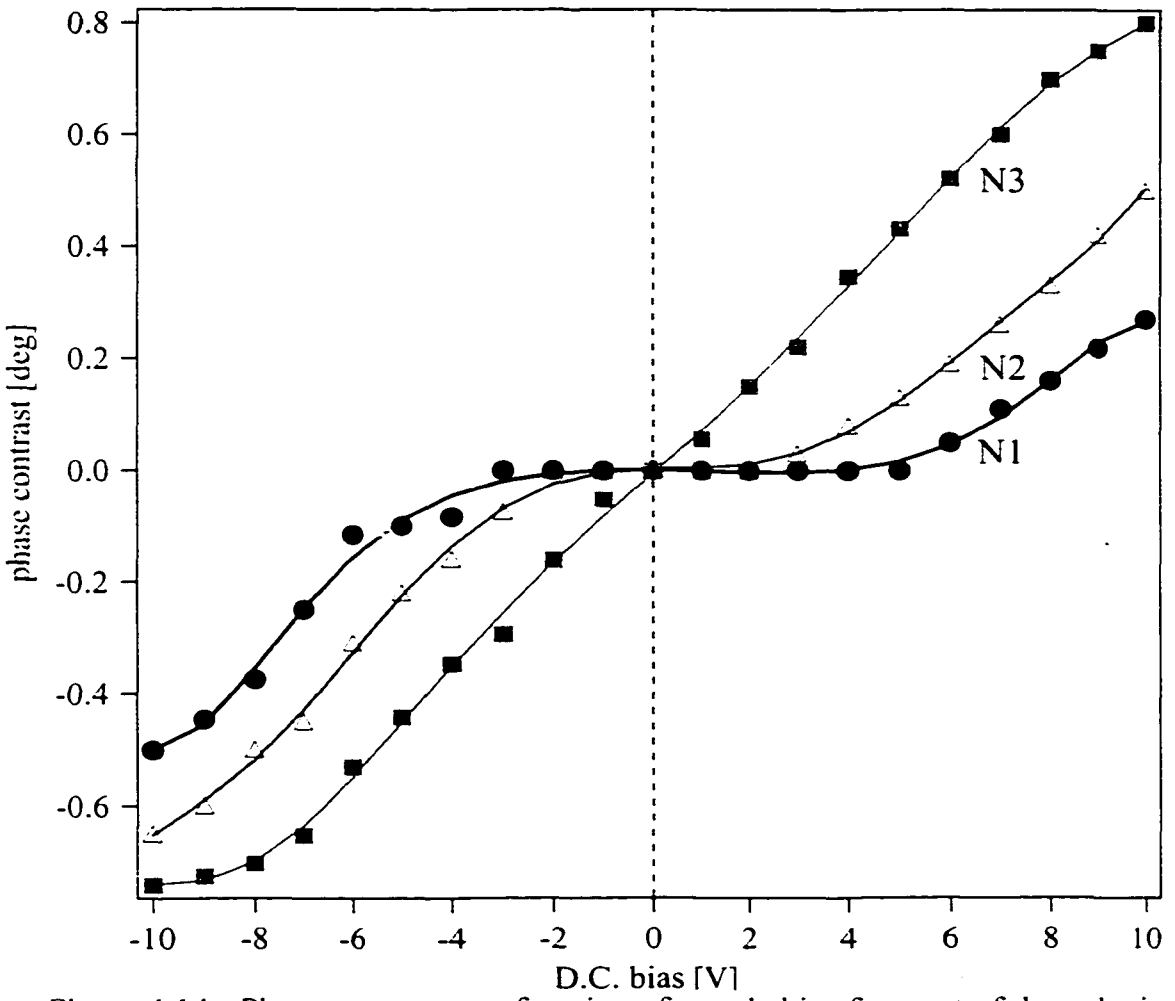


Figure 4.14: Phase contrast as a function of sample bias for a set of three doping standards provided by NIST. The substrate is doped n-type 1 Ωcm with a 50 keV boron implantation dose for samples N1, N2, and N3 being 10^{13}, 10^{14}, and 10^{15} cm^{-2} respectively. See the text for a detailed description of the samples.

Better doping standards could also be used to determine if the non-linearity of the phase response that is predicted by the charge density plot (Figure 4.13) truly exists. Since the charge density axis in the graph shown in Figure 4.13 is logarithmic it is expected that the charge density and therefore the phase signal may not be linear with the doping density. Variations in the doping density at higher doping densities may be more easily detected than at lower doping densities where the difference in the surface charge is smaller.

The loss of phase contrast at either highly positive or highly negative potentials, observed in all phase-based dopant profiling images, can be explained by the merging of all of the charge density curves (n-type and p-type) at sufficiently large surface potentials. At large biases the semiconductor is forced into accumulation or inversion, as is demonstrated in Figure 4.7. Under these conditions, the surface becomes essentially metallic with a sheet of charged carriers, resulting in a surface charge density that is nearly identical for all doping densities. With the surface charge density being nearly identical for all doping densities there is no longer a difference in surface charge density for regions of varied doping density and the phase contrast disappears. This can be seen in Figure 4.13 where the calculated surface charge density curves at all doping densities converge at high surface potentials. From Figure 4.13 it would also be expected that the loss of contrast should occur at lower biases for more lightly doped regions. This stems from the convergence of the lower doping density surface charge density curves at lower surface potentials.

The absolute phase in Figure 4.15 is observed to increase with increasing positive and negative bias. Absolute phase results from the total surface charge density appearing across the imaged region, not the difference in surface charge density between two regions of different doping density. The absolute phase as a function of sample bias should reflect the total charge density across a scan as a function of the surface potential as is seen in Figure 4.13. The general shape of the curves in Figure 4.12 shows that the total surface charge density at all doping densities increases with bias in the same way that the curve in Figure 4.6(b) does. This indicates that the calculated surface charge in the accumulation and inversion regions properly models the absolute phase background signal.

Figure 4.15 shows the absolute phase behavior of the NIST samples as a function of the D.C. bias applied to the sample. Note that in Figure 4.15 the N4 curve has a much larger absolute phase signal at a given bias than the other three curves. This overall increase in phase signal is due to charging of the oxide, which is 10 nm thick over the implanted region and 20 nm thick over the unimplanted region. Charge densities in oxides are almost always larger than the charge density in silicon at a given bias.[13] When imaging samples containing both oxides and exposed doped silicon, the oxides will show up as bright areas in the phase image relative to the silicon at both positive and negative biases. Bias dependent phase contrast reversal should occur for two areas of silicon with different doping densities since the surfaces charge differently as a result of different fixed charge densities. One will have a larger surface charge density at a given sign of the bias while the other region has a larger surface charge density at the opposite sign. The larger surface charge density region will always appear lighter than the lower charge density region. The oxide always seems to appear lighter than the silicon since the charge density in the oxide is always greater than in the silicon. Since oxides tend to charge more than the doped silicon, the absolute phase curve should be expected to be offset toward higher phase angles when oxide-containing samples are imaged. This results from the greater charge density that produces a larger Coulomb interaction and therefore a larger phase shift. Oxides and silicon are often present in the same sample when cross sections are imaged. An example of a BAAFM cross-section of a commercially available integrated circuit produced at Hewlett Packard will be presented in Chapter 5.

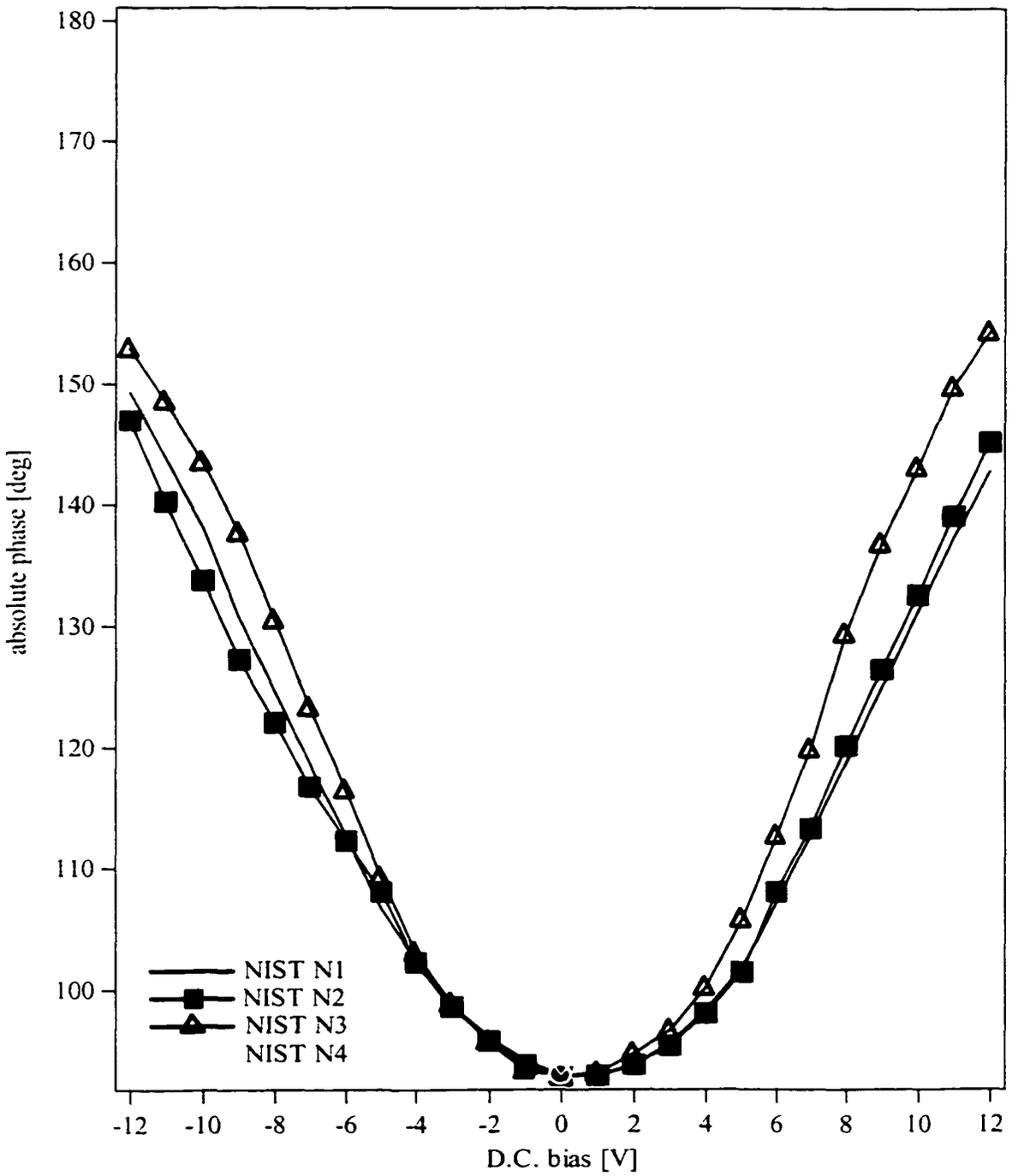


Figure 4.15: Absolute phase as a function of sample bias for a set of four doping standards provided by NIST. Substrate is doped n-type 1 Ωcm with the 50 keV boron implantation dose for samples N1, N2, N3, and N4 being 10^{13}, 10^{14}, 10^{15}, and 10^{15} cm^{-2} respectively. Samples N1,N2, and N3 were dipped in HF in order to remove any surface oxide while the N4 sample had a 10 nm thick oxide covering the implanted region and a 20 nm thick oxide grown over the unimplanted substrate. See the text for a detailed description of the samples.

4.7 Humidity

Low humidity is essential for dopant profiling with electrostatic force microscopy. As a general rule imaging dopants using BAAFM is no longer effective after reaching relative humilities larger than about 15%. With increasing humidity the water layer on the tip and sample increases in thickness as a result of surface wetting processes.[27] The potential drop across the tip-sample separation gap decreases as the dielectric constant of the liquid surface coating increases. This results in a larger surface potential at any given bias applied to the sample. Since water has a larger dielectric constant than that of air, as the water layer thickness increases the surface potential also increases.

Even though the surface potential and therefore the surface charge density is greater at a given bias when there is water in the tip-sample gap, the phase contrast signal disappears. The lack of phase contrast with water layers can be explained by dielectric screening. Figure 4.16 shows a simplified schematic drawing of a charged sample with an image charge on the tip and a condensed water layer between the tip and sample. This situation can also be modeled as an electrostatic dipole, containing two point charges, as shown to the right in Figure 4.16, where there is a frequency component associated with the electrostatic potential due to the cantilever's oscillation near 70 kHz. When the tip oscillation frequency is less than the plasmon frequency we can assume the static case, where the charges are fixed in space. This is the case for the BAAFM where the resonance frequency is about 70 kHz while the plasmon frequency for silicon is about 10^6 Hz.[28] Assuming a static condition, the Coulomb potential between the image charge at the tip and a point charge in the sample is given by:

$$E = \frac{1}{4\pi\varepsilon_0\varepsilon}\frac{q_{tip}q_{sample}}{r} \tag{4.26}$$

where q_{tip}, q_{sample}, ε_0, ε, and r are the charge of the tip, charge of the sample, permittivity of free space, dielectric constant of the material between the tip and semiconductor, and the distance between the charges respectively.[29]

If we consider how the Coulomb potential is affected by changing the value of the dielectric constant of the medium between the tip and sample it becomes obvious why the signal disappears as the humidity increases. Replacing the dielectric constant of air (1.0) with that of water (80.4), in Equation 4.26, results in a dramatic decrease in the Coulomb interaction potential.[30] The phase contrast between two regions of different doping density on a semiconductor is usually less than 3 to 5 degrees when imaging in low humidity. A dramatic drop in the Coulomb potential due to an increase in dielectric

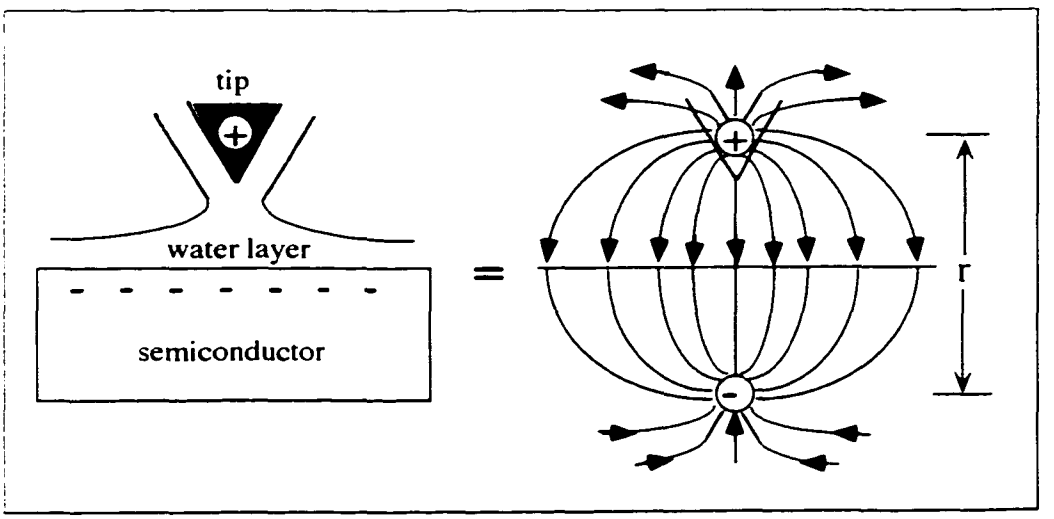


Figure 4.16: Diagram indicating the location of the water layer between the tip and sample. The charging of the surface is shown on the left to be equal to the case on the right in which all charges are represented by an electrostatic dipole.

constant would result in an eighty-fold decrease in the phase signal. This would cause a loss of signal large enough to not allow distinguishing between the phase contrast between two areas of different doping densities.

4.8 Cantilevers

Properties of cantilever probes should be investigated since the sensitivity and repeatability of dopant-dependent phase signals may be related to the nature of the cantilever probe used. Several "conductive" and "non-conductive" cantilevers are commercially available. Each of these cantilevers varies in one or more of the following properties: coating material, force constant, resonance frequency, tip size, cantilever dimensions, etc. Physical properties of several cantilevers are provided in Table 4.1. In this section the effect of the cantilever's force constant and coating materials are investigated. Selecting cantilevers that are identical except for one physical property (i.e., surface coating or force constant) makes it possible to monitor the dependence of the doping-dependent phase signal on that property. In this section, each cantilever's phase signal is monitored as a function of sample bias voltage while imaging a sample consisting of a p-type substrate ($1x10^{17}$ cm^{-3}) and a n-type ion implanted region ($5.9x10^{15}$ cm^{-3}).

When investigating the influence of the cantilever's coating material on the dopant-dependant phase signal, a set of four Digital Instruments cantilevers were used. All of these cantilevers are identical to the Digital Instruments standard non-coated"FESP" silicon cantilever (resonance frequency ≈ 65 kHz) in all properties except for surface coating material. The exposed probe surfaces of the FESP, SESP, EFM-50, and MESP cantilevers consist of lightly doped silicon, nickel silicide, platinum-iridium,

Table 4.1: Manufacturer specifications for several commercially available cantilevers. Suppliers of the cantilevers are Digital Instruments (Santa Barbara, CA), NT-MTD (Moscow, Russia), Olympus (Tokyo, Japan), and Thermomicroscopes (Sunnyvale, CA) are denoted as DI, NT-MTD, Olympus, and Thermo respectively. The * denotes that these cantilevers are also available with W_2C, TiN, TiO, or Si_3N_4 coatings.

supplier	cantilever	surface material	force const. (N/m)	res. freq. (kHz)	radius of curvature (μm)	length (μm)	width (μm)
DI	EFM-50	Pt/Ir	1-5	60-100	5-10	225	50
DI	DNISP	diamond	100-300	35-65	< 25	356	50
DI	FESP	Si	1-5	60-100	5-10	225	50
DI	IBMSC	Si	50-200	275-450	4-10	125	50
DI	MESP	Co/Cr	1-5	60-100	25-50	225	50
DI	SDNP (a)	Cr	0.58	N/A	20-60	200	50
DI	SDNP (b)	Cr	0.32	N/A	20-60	100	50
DI	SDNP (c)	Cr	0.12	N/A	20-60	200	50
DI	SDNP (d)	Cr	0.06	N/A	20-60	100	50
DI	SESP	Ni_2Si_3	1-5	60-100	5-10	225	50
DI	TESPW	Si	20-100	200-400	15-20	125	50
NT-MTD	c-SC11 (a)	Si*	0.35	20	N/A	200	40
NT-MTD	c-SC11 (b)	Si*	6.0	180	N/A	90	60
NT-MTD	nc-SC11 (a)	Si*	3.0	40	N/A	200	40
NT-MTD	nc-SC11 (b)	Si*	48	360	N/A	90	60
NT-MTD	c-SC12 (a)	Si*	0.7	90	N/A	110	35
NT-MTD	c-SC12 (b)	Si*	1.7	160	N/A	90	35
NT-MTD	c-SC12 (c)	Si*	0.6	75	N/A	130	35
NT-MTD	c-SC12 (d)	Si*	0.05	15	N/A	300	35
NT-MTD	c-SC12 (e)	Si*	0.03	10	N/A	350	35
NT-MTD	c-SC12 (f)	Si*	0.08	20	N/A	250	35
NT-MTD	nc-SC12 (a)	Si*	6	180	N/A	110	35
NT-MTD	nc-SC12 (b)	Si*	14	320	N/A	90	35
NT-MTD	nc-SC12 (c)	Si*	4.5	150	N/A	130	35
NT-MTD	nc-SC12 (d)	Si*	0.35	30	N/A	300	35
NT-MTD	nc-SC12 (e)	Si*	0.25	20	N/A	350	35
NT-MTD	nc-SC12 (f)	Si*	0.65	40	N/A	250	35
NT-MTD	c-SC21 (a)	Si*	0.12	10	N/A	290	40
NT-MTD	c-SC21 (b)	Si*	2.0	110	N/A	110	40
NT-MTD	nc-SC21 (a)	Si*	0.9	20	N/A	290	40
NT-MTD	nc-SC21 (b)	Si*	17	260	N/A	110	40
Olympus	OTESPA	Si (0.001 Ωcm)	~42	200-400	< 10	160	N/A
Thermo	nc-ultralever (a)	Si (0.001 Ωcm)	2.1	80	12	180	25
Thermo	nc-ultralever (b)	Si (0.001 Ωcm)	3.2	90	12	180	38
Thermo	nc-ultralever (c)	Si (0.001 Ωcm)	13	280	12	85	18
Thermo	nc-ultralever (d)	Si (0.001 Ωcm)	17	320	12	85	28
Thermo	c-ultralever (a)	Si (0.001 Ωcm)	0.26	40	12	180	25
Thermo	c-ultralever (b)	Si (0.001 Ωcm)	0.4	45	12	180	38
Thermo	c-ultralever (c)	Si (0.001 Ωcm)	1.6	140	12	85	18
Thermo	c-ultralever (d)	Si (0.001 Ωcm)	2.1	160	12	85	28
Thermo	mfm-ultralever (a)	Co	2.1	80	12	180	25
Thermo	mfm-ultralever (b)	Co	3.2	90	12	180	38
Thermo	mfm-ultralever (c)	Co	13	280	12	85	18
Thermo	mfm-ultralever (d)	Co	17	320	12	85	28

and cobalt-chrome, respectively. Figure 4.17 shows the bias voltage dependence of the phase contrast signal observed between the substrate (p-type: $1x10^{17}$ cm^{-3}) and implanted region (n-type: $5.9x10^{15}$ cm^{-3}) of a test sample taken at a lift height of 100 nm for all four of these Digital Instruments cantilevers. Note that all four of the curves in Figure 4.17 are identical to within experimental error (±0.05 degrees). It is somewhat surprising that the doping dependent phase signal is independent of the cantilever's conductive coating material.

Cantilever coating:

Data from Section 4.4 suggested that variations in the work function of the cantilevers with respect to the sample may result in the appearance of contrast at low tip-sample biases due to a contact potential difference between the tip and sample. This should result in an offset in the surface potential that could induce band bending and surface charging. We would also expect that variations in work function from one tip coating to another tip coating could result in variations in the phase contrast observed from tip to tip since each material should have a different work function. Work functions of several cantilevers, including those shown in Figure 4.17, were measured using a Physical Electronics Quantum 2000 x-ray photoelectron spectrometer (XPS).

Electron spectroscopy is an ideal method for measuring work functions.[31] One only needs to measure the position of the Fermi level binding energy (E_F), which should be zero for a calibrated spectrometer, and the position of the secondary electron cutoff (E_{SE}), located at low kinetic energy. To prevent measuring the work function of the spectrometer, the sample was biased at –8 volts. Should the work function of the spectrometer be greater than the work function of the sample, the secondary electron

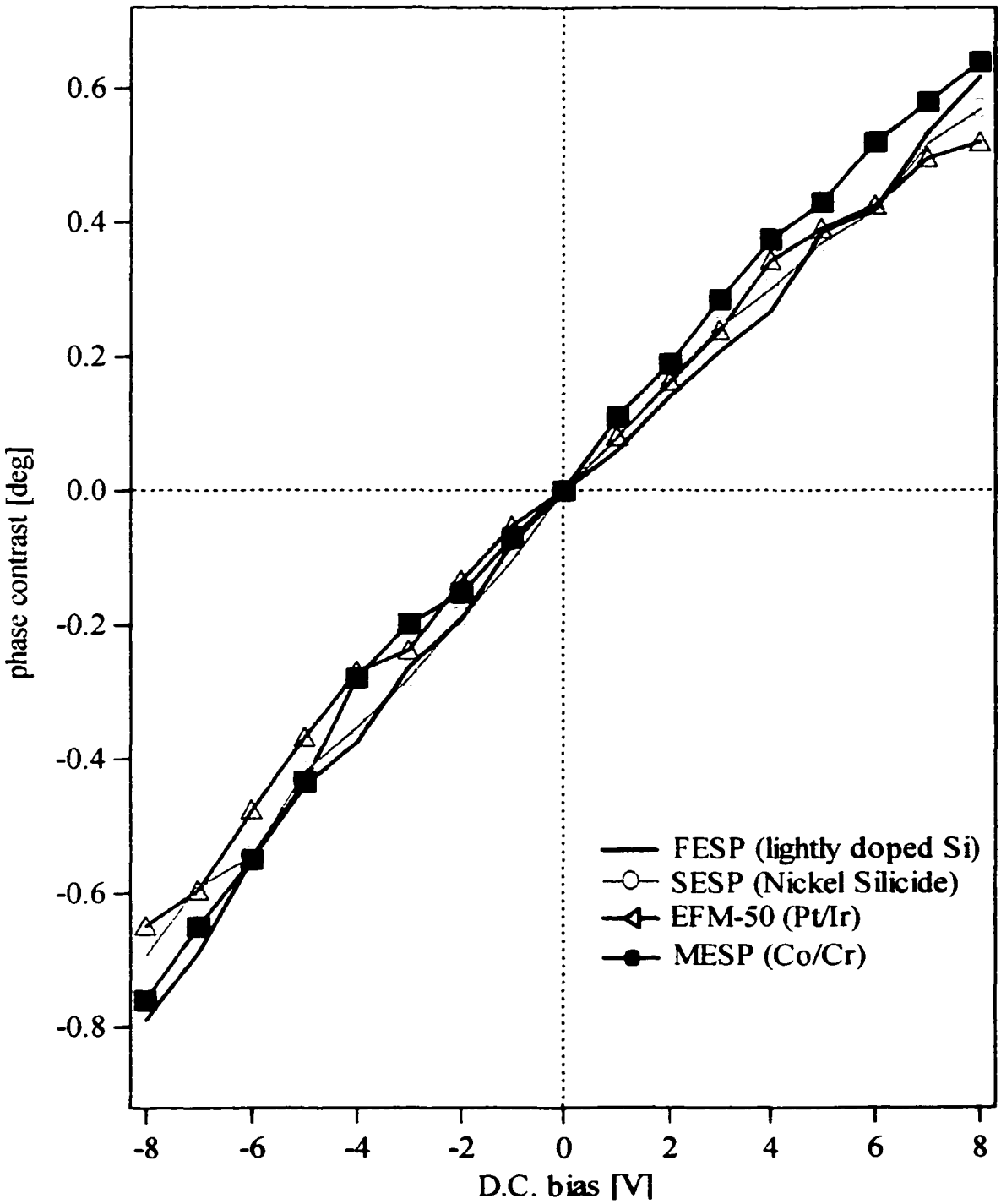


Figure 4.17: Phase contrast as a function of bias voltage for four cantilevers which are identical except for their surface coating. The "phase contrast" in this plot is produced by measuring the difference in phase signal between the substrate (B: $1x10^{17}$ cm^{-3}) and implanted region (As: $5.9x10^{15}$ cm^{-3}) of an ion implanted "dopant circle" sample . See Appendix D for an explanation of phase contrast.

cutoff will represent the work function of the spectrometer and not the sample. When the work function of the sample is dominant as a result of biasing the sample, its work function (Φ_{sample}) is simply given by:

$$\Phi_{sample}=h\nu-(E_{SE}-E_F), \tag{4.27}$$

where $h\nu$ is the incident photon energy.[31]

XPS-measured secondary electron edge positions and their corresponding work functions are presented in Figure 4.18 and Table 4.2 for several cantilevers. The data indicate that all of the work functions lie between 4.12 and 4.25 eV with the exception of the nickel silicide-coated SESP cantilever and the Pt/Ir coated EFM-50 cantilever which have work functions of 3.95 eV and 3.68 eV, respectively. All of the cantilevers measured were loaded into the XPS vacuum chamber in the same oxidized/air contaminated condition that they are in when used for BAAFM dopant profiling. It is well known that work functions are very sensitive to surface contamination, which tends to lower the work function of a given material dramatically.[31] Note that after Ar^+ sputter cleaning, the Pt/Ir cantilever's work function climbed from 3.67 eV to 5.72 eV. Note that the work functions reported here can only be used to approximate the actual work

Table 4.2: XPS determined work functions of various cantilevers with different types surfaces at the probe tip.

cantilever	supplier	surface material	ϕ (eV)
EFM-50	Digital Instruments	Pt/Ir	3.67
SESP	Digital Instruments	Nickel Silicide	3.95
SDNP	Digital Instruments	Chrome	4.12
SCS11	NT-MTD	Si_3N_4	4.20
MESP	Digital Instruments	Cr/Co	4.12
TESDP	Digital Instruments	hardened Si	4.20
Ultralever	Thermo Microscopes	highly doped Si	4.16
Ultralever	Thermo Microscopes	Si	4.25
DNP	Digital Instruments	Si_3N_4	4.25

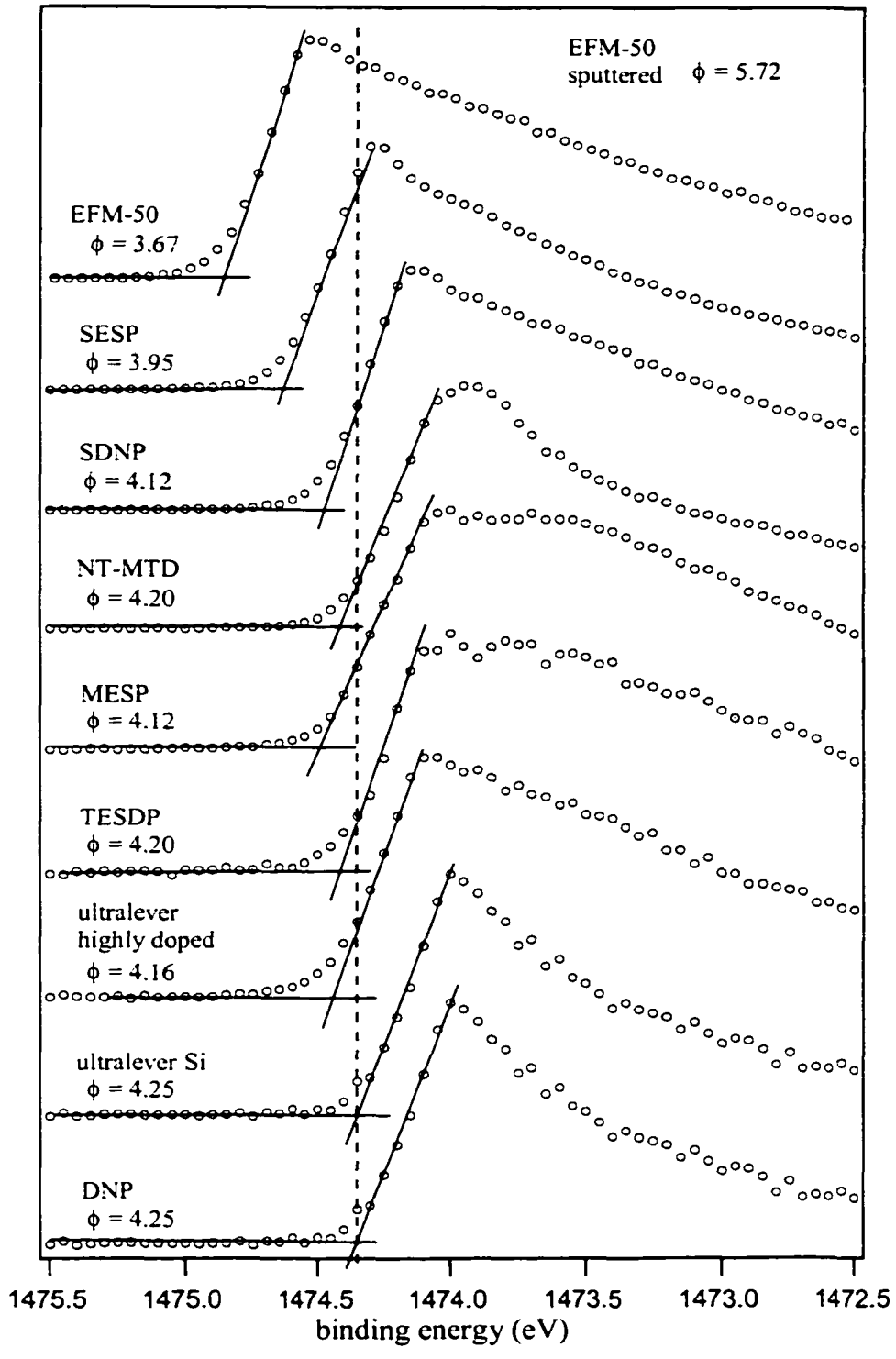


Figure 4.18: XPS data showing the secondary edge spectra from which the work functions of the cantilevers in Table 4.2 were calculated.

function of the tip of the cantilever, since work functions are also known to depend on the surface roughness.[31] A decrease in the work function is typically observed at the apex when a pointed structure such as a cantilever is measured.[31]

Surface contamination and a high degree of curvature at the apex of a cantilever result in lowering the cantilever's work function. The uniformity of the phase contrast curves observed in Figure 4.17 can be explained by the converging value of the cantilever work functions to approximately the same value due to surface contamination and similarities in tip shape. Influences due to work function variations between various cantilevers are further decreased in the BAAFM experiment since the potential difference felt at the surface is decreased due to an increased tip-sample separation distance. This lowering of the contact potential difference is due to a tip-sample separation dependence on the potential felt across the separation gap. Even though clean cantilevers in vacuum can have large differences in their work functions, surface contamination and a distance dependent potential drop result in the measured doping dependent phase signal being virtually independent of the cantilever's surface coating. Work function differences are relatively small compared to the bias applied to the sample. Therefore, effects may only be seen in the 0 to ±1 V range. Contrast differences below ±1, however, are difficult to distinguish from noise.

Cantilever force constant:

The spring constant of a cantilever should have some influence on the sensitivity of the dopant profiling measurement. Figure 4.19 shows phase contrast (a) and absolute phase (b) curves taken 100 nm above a standard sample consisting of 1.3 μm diameter ion implanted circles (n-type: 5.9×10^{15} cm^{-3}) on a 1×10^{17} cm^{-3} p-type substrate. The

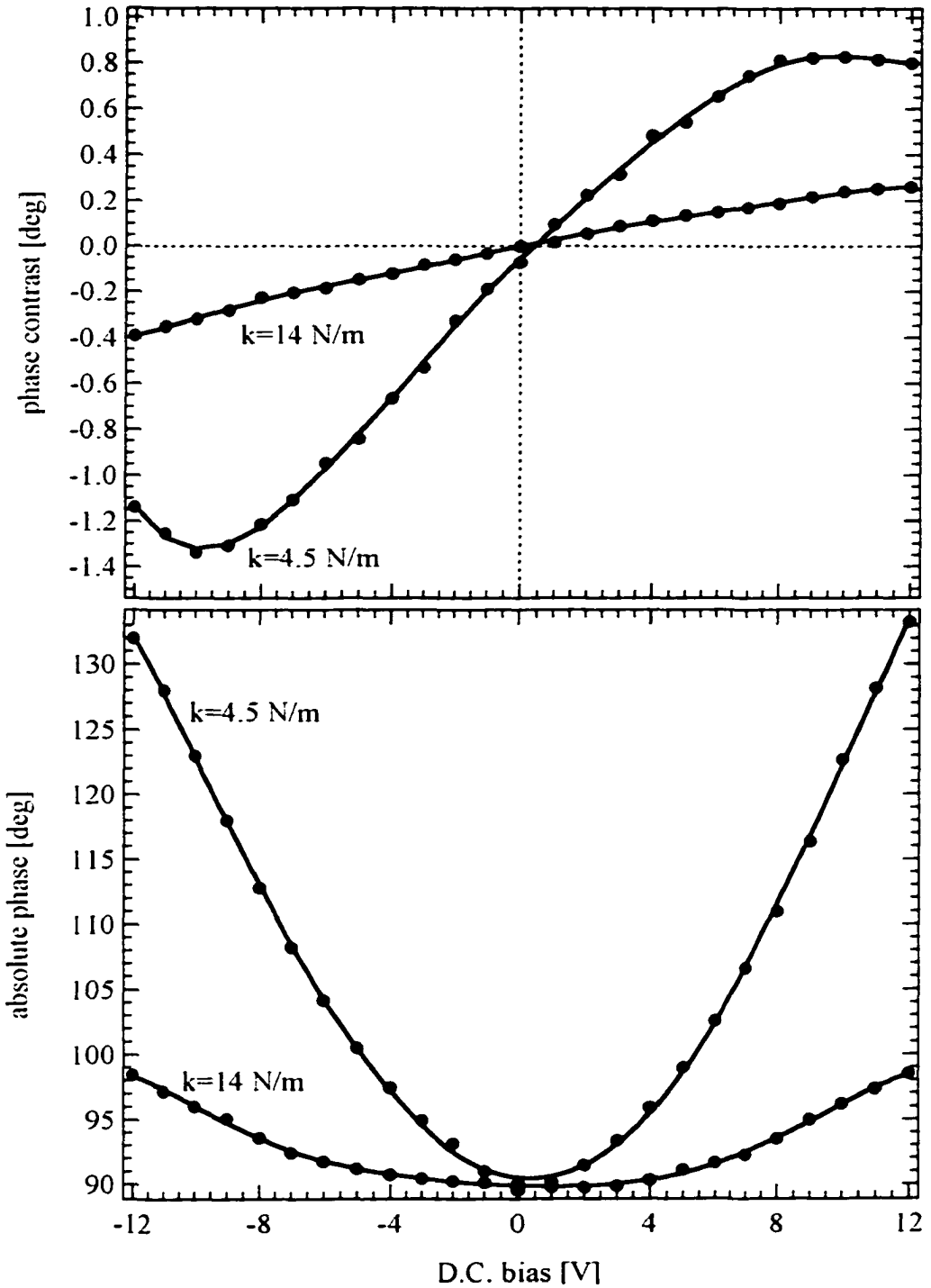


Figure 4.19: Phase contrast as a function of bias for two cantilevers which are identical in all physical properties except for their spring constants. The "phase contrast" in this plot represents the difference in phase signal measured between the substrate (B: $1x10^{17}$ cm^{-3}) and implanted region (As: $5.9x10^{15}$ cm^{-3}) of an ion implanted "dopant circle" sample. See Appendix D for an explanation of phase contrast and absolute phase.

substrate was imaged using two NSCS12 cantilevers (B and C) attached to a common NSCS12 substrate (NT-MTD Co., Moscow, Russia). Both cantilevers are attached to the same silicon substrate in order to ensure that all physical properties are identical except for the force constant and resonance frequency. Cantilevers B and C have force constants of 14 N/m and 4.5 N/m, respectively. Note that both the phase contrast (a) and absolute phase (b) signals observed in Figure 4.19 with the lower force constant cantilever are larger than the signals recorded for the higher force constant cantilever. The same type of cantilever force constant dependence, in which a larger phase signal is observed for lower force constant cantilevers, was observed when measuring the phase signal with Ultralever cantilevers (ThermoMicroscopes, Sunnyvale, CA). Lower force constant cantilevers are therefore more desirable for dopant profiling since they can be perturbed by smaller Coulomb interactions than a cantilever that has a larger force constant. Cantilevers with force constants that are too low, however, may be perturbed too much by the Coulomb interaction. This may result in an absolute phase (background signal) which is so large that it is impossible to record a doping dependent phase contrast signal. Digital instruments MESP (Co/Cr coated) and EFM-50 (Pt/Ir coated) cantilevers, both with force constants of about 3 N/m, seem to work well for dopant profiling. A Pt/Ir coating is somewhat more durable than a Co/Cr coating, therefore the EFM-50 cantilevers do not suffer from loss of resolution due to tip wear as quickly as the MESP cantilevers do.

4.9 Chapter Conclusions

We performed TMAFM measurements on patterned silicon wafer samples with an additional direct current (D.C.) bias applied between cantilever and sample. Our results show that areas of different doping density cause a bias-dependent variation in the

phase angle between the cantilever oscillation and its driving frequency. This phase angle contrast was used to characterize the lateral doping profile of silicon wafers patterned by photolithography with subsequent ion implantation.

The imaging contrast mechanism was investigated by calibrating the force measurements and amplitude/phase relationships with applied biases on a gold-coated mica substrate. The impact of the Coulomb force, induced by the applied bias on the cantilever oscillation, was then determined. The distance dependence of the phase signal was shown to fall off with the inverse of the distance (z). A model based on energy dissipation, via I^2R losses in the semiconductor, was developed to try to explain the $1/z$ dependence of the phase signal.

A model was developed in which the bias-induced surface potential (band bending) variation between differently doped areas of the surface is identified to be the cause for the image contrast observed in the phase images. The model relates surface charge density variations, produced by biasing the sample, into phase contrast observed between areas of varied doping density. Surface charges densities as a function of the sample's surface potential were calculated for different doping densities. These curves were used to validate many of the assumptions made in the mechanistic model. The effects of humidity and cantilever properties were also investigated.

References

[1] M. W. Nelson, P. G. Schroeder, R. Schlaf, and B. A. Parkinson, J. of Vac. Sci. Technol. B **17,** 1354 (1999).

[2] M. W. Nelson, P. G. Schroeder, R. Schlaf, B. A. Parkinson, C. W. Almgren, and A. N. Erickson, Appl. Phys. Lett. **74,** 1421 (1998).

[3] R. Schlaf, P. G. Schroeder, M. W. Nelson, R. Stübner, S. Tiefenbacher, H. Jungblut, and B. A. Parkinson, Thin Solid Films **331,** 203 (1998).

[4] R. Brandsch, G. Bar, and M.-H. Whangbo, Langmuir **13,** 6349 (1997).

[5] R. Garcia, J. Tamayo, M. Calleja, and F. Garcia, Appl. Phys. A **66,** S309 (1998).

[6] J. Tamayo and R. Garcia, Appl. Phys. Lett. **71,** 2394 (1997).

[7] R. Weisendanger, *Scanning Probe Microscopy and Spectroscopy: Methods and Applications* (Cambridge University Press, New York, 1994).

[8] S. N. Magonov, V. Elings, and M.-H. Whangbo, Surf. Sci. **375,** L385 (1997).

[9] R. S. McLean and B. B. Sauer, Macromolecules **30,** 8314 (1997).

[10] M. H. Whangbo, G. Bar, and R. Brandsch, Appl. Phys. A **66,** S1267 (1998).

[11] J. Tamayo and R. Garcia, Appl. Phys. Lett. **73,** 2926 (1998).

[12] J. P. Cleveland, B. Anczykowski, A. E. Schmid, and V. B. Elings, Appl. Phys. Lett. **72,** 2613 (1998).

[13] J. W. Mayer and S. S. Lau, *Electronic Materials Science: For Integrated Circuits in Si and GaAs* (Macmillan Publishing Company, New York, 1990).

[14] E. Bucher, S. Schulz, M. C. Lux-Steiner, P. Munz, U. Gubler, and F. Greuter, Appl. Phys. A **40,** 71 (1986).

[15] D. R. Lide, *CRC Handbook of Chemistry and Physics* (CRC Press, New York, 1998).

[16] W. Denk and D. W. Pohl, Appl. Phys. Lett. **59,** 2171 (1991).

[17] R. Garcia, J. Tamayo, and A. San Paulo, Surf. Interf. Anal. **27,** 312 (1999).

[18] P. Grütter, Y. Liu, P. LeBlanc, and U. Dürig, Appl. Phys. Lett. **71,** 279 (1997).

[19] J. Tamayo, Appl. Phys. Lett. **75,** 3569 (1999).

[20] G. S. LaRue, Ph.D. Thesis, Stanford University, 1978.

[21] A. F. Hebard, Ph.D. Thesis, Stanford University, 1970.

[22] J. D. Phillips, Ph.D. Thesis, Stanford University, 1983.

[23] S. M. Sze, *Semiconductor Devices: Physics and Technology* (John Wiley & Sons, New York, 1985).

[24] R. Schlaf, R. Hinogami, M. Fujitani, S. Yae and Y. Nakato, J. Vac. Sci. Technol. A **17**, 164 (1999).

[25] L. J. Huang and W. M. Lau, J. Vac. Sci. Technol. **A10**, 812 (1992).

[26] A. S. Grove, *Physics and Technology of Semiconductor Devices* (John Wiley and Sons, Inc., New York, 1967).

[27] A. W. Adamson, *Physical Chemistry of Surfaces*, 5th ed. (John Wiley & Sons, New York, 1990).

[28] C. Kittel, *Introduction to Solid State Physics*, 5th ed. (John Wiley & Sons, New York, 1976).

[29] J. R. Reitz, F. J. Milford, and R. W. Christy, *Foundations of Electromagnetic Theory* (Addison-Wesley, Reading, 1980).

[30] D. Halliday, R. Resnick, and J. Walker, *Fundamentals of Physics* (John Wiley & Sons, New York, 1993).

[31] S. Hüfner, *Photoelectron Spectroscopy: Principles and Applications*, Second Edition ed. (Springer, New York, 1996).

Chapter 5: Profiling of Commercial Devices

5.1 Introduction to Dopant Profiling on Real Devices

High-resolution two-dimensional dopant profiling methods are used to support the development of new processes for integrated circuit (IC) production and for investigation of doping-related problems in commercial ICs.[1] Previous chapters dealt with development of BAAFM for dopant profiling. In this chapter we will look at qualitative doping profiles of commercial semiconductor devices using the BAAFM technique. Sample preparation for both top-down and cross-sectioned samples will be addressed since sample preparation has been one of the issues that have plagued many of the 2D dopant profiling methods.

5.2 Top Down Sample Preparation

Depending on the processing step in which a wafer is pulled from the production line, "deprocessing" may be needed in order to produce a top down doping picture of the device of interest. Deprocessing involves removal of metal, insulating dielectrics, and other materials lying on top of the doped semiconductor.[2] Since phase contrasts in BAAFM can arise from bias induced charging of overlying metals and oxide,[1] it is

essential to remove overlying oxides, dielectrics, polymers, and metal layers. Removal of these layers can be accomplished using a few chemical steps.

Stripping back most of today's processed wafers to bare silicon for top down BAAFM and SCM analysis can be accomplished by using the following procedure. Passivation is removed with a reactive ion etch (Technics RIE Etcher: P_{O_2} 20 sccm, P_{CF_4} 40 sccm, RF 300 watts, time 15 min., chamber partial pressure 50 mTorr). Reactive ion etching is followed by an "aluminum etch" (16:1:1:1, phosphoric acid, nitric acid, acetic acid, DI water) which is used to remove metal. Heat the aluminum etch solution to 60 C° and etch 2 minutes. Rinse in DI water and dry in a stream of N_2. Etch in HF (49% vol./vol.) for 30 seconds in the dark in order to remove oxides and other dielectrics. Reactions performed in illuminated conditions may result in photo-electrochemical dopant delineation etching of areas of varied doping density.[3-5] Rinse and repeat the "aluminum etch." After the second "aluminum etch" the sample is rinsed in hydrogen peroxide 30% (60C°) for 1 minute. This should remove tungsten from the surface. The sample is washed in DI, dried in a stream of dry N_2, and examined under a high power microscope to see if all of the poly-silicon has been removed. In some cases it may be desirable to etch the wafer for a shorter period of time so that the poly-silicon remains for the analysis. Altering reaction times may be necessary in order to strip back areas produced with differing processes.

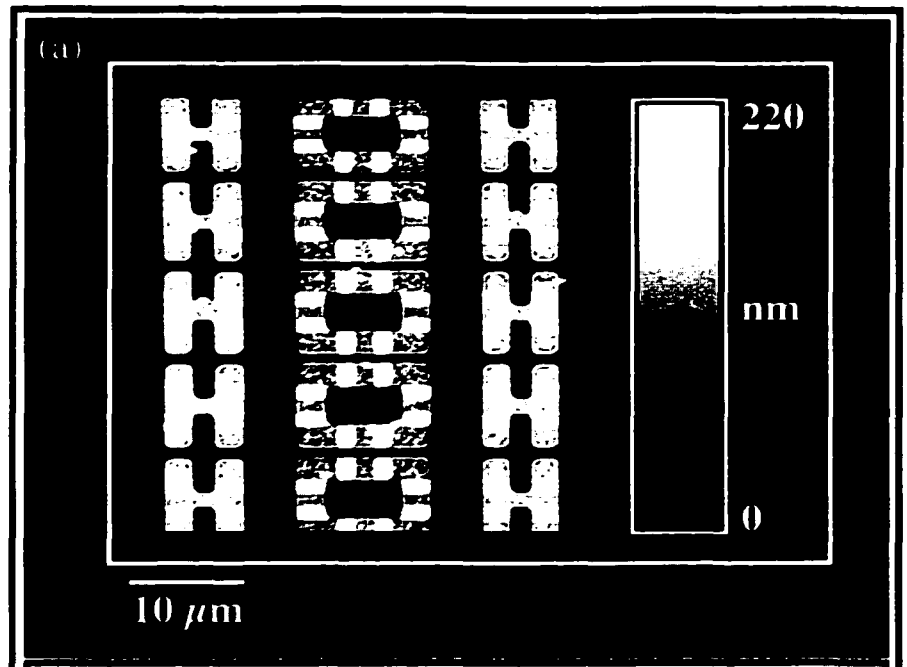
5.3 Top-Down Investigation of a SRAM device

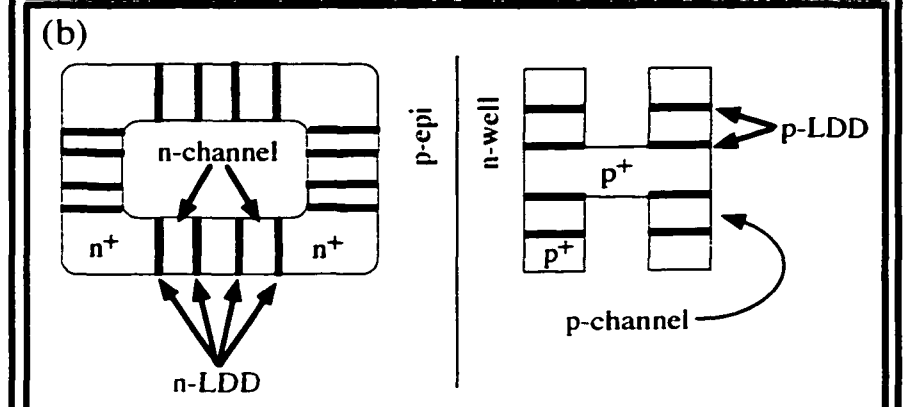
In this section, we tested the BAAFM method on static random access memory (SRAM) integrated circuit devices produced at Sandia National Laboratories using 0.6 μm CMOS technology.[6] All experiments were carried out using a Nanoscope® III

Multimode scanning probe microscope (Digital Instruments, Santa Barbara, CA) in which the sample stage was disconnected from ground and connected to a D.C. power supply, thus allowing application of a bias to the sample. A 10 MΩ resistor was put in series between the power supply and sample in order to protect the cantilever and the sample in case of electrical point contact. A conductive "SESP" cantilever (Digital Instruments, Santa Barbara, CA, Ni_2Si_3-coated Si, spring constant 2-7 N/m) was used in all of the experiments in this section. The SRAM's passivation, metal gate electrodes, oxide and silicide layers were removed using the method discussed above in Section 5.2. Prior to imaging the sample was dipped into a 2% HF solution in order to hydrogen passivate the surface and thus greatly reduce the number of surface states in the band gap.[7,8]

A topographic TMAFM image of the investigated sample is shown in Figure 5.1(a). The image shows a 40x40 μm^2 area. From the z-scale bar it is evident that the height of the features in the imaged area vary by about 200 nm. Features in this image result only from changes in the topographic height of the sample, not from doping-related effects. Figure 5.1(b) shows a top down doping schematic, which correlates the topographic features shown in Figure 5.1(a) with doping type and density. The schematic indicates that the SRAM contains p-type areas ranging in doping density from $2x10^{16}cm^{-3}$ to $4x10^{19}cm^{-3}$ (p-epi, "n-channel", p-LDD, and P+) and n-type features whose doping densities range from $1x10^{17}cm^{-3}$ to $2x10^{20}cm^{-3}$ ("p-channel", n-well, n-LDD and N+). Note that the p-epi and n-well can not be differentiated in the topographic image.

Imaging top-down samples with large topographic features can result in erroneous measurements near the edges of large features due to the side of the tip interacting with





	doping type	conc. (cm^{-3})	species	dose (cm^{-2})	Energy (keV)
p-epi	p	$2x10^{16}$			
n-channel	p	$2x10^{17}$	BF_2^+	$2x10^{12}$	30
n-LDD	n	$5x10^{18}$	P+	$6x10^{13}$	40
n^+	n	$2x10^{20}$	As^+	$4x10^{15}$	100
n-well	n	$2x10^{17}$	P+	$1.7x10^{13}$	900
p-channel	n	$1x10^{17}$	P+	$1x10^{12}$	175
p-LDD	p	$3x10^{18}$	BF_2^+	$6x10^{13}$	40
p^+	p	$4x10^{19}$	BF_2^+	$2x10^{15}$	45

Figure 5.1: (a) TMAFM height image of a SRAM device produced at Sandia National Laboratories. (b) Design schematic of the features shown in the topographic image. The structures contain p- and n-type areas of doping densities ranging between $10^{16}cm^{-3}$ and $10^{20}cm^{-3}$.

the steep side wall of the sample. Side wall interactions show up as scan direction dependent features. Often this problem can be fixed by decreasing the scan rate, thus allowing the tip to pull up at a rate which is fast enough to eliminate much of the side wall influence. It may also be necessary to increase the interleave tip-sample separation height. Using the computer-applied bias voltage only during the lift also aids in elimination of side wall-related effects. Biasing during both the main and interleave scans can result in blurring of the main scan which results in a poor trace for the interleave scan. This bad interleave trace can result in close probe proximity near high features, thus resulting in a greatly increased signal at the edge of features.

In Figure 5.2, a series of seven TMAFM phase images taken from a portion of the same area in Figure 5.1 are shown at sample biases ranging from –6V to +6V at a 100 nm interleave lift height. The image at the top represents the zero bias case. This image shows a weak phase contrast, which is most likely related to the contact potential difference between the tip and sample. The contact potential difference by itself already induces a weak Coulomb interaction between cantilever and sample. Topography related effects can be excluded since the edges of the protruding features would be expected to show scan direction dependent contrast which is not seen in the measurements. Applying a D.C. bias between the tip and sample changes the phase contrast between the differently doped areas depending on the magnitude and sign of the applied bias. Similar to the results presented in Chapter 4, contrast reversal occurs upon switching the sign of the sample bias. Also, biases higher than 6V resulted in an attenuation of the phase contrast due to the strong accumulation of carriers at the surface on both n- and p-type areas. This is also in accordance with the results presented in Chapter 4 where the model proposed

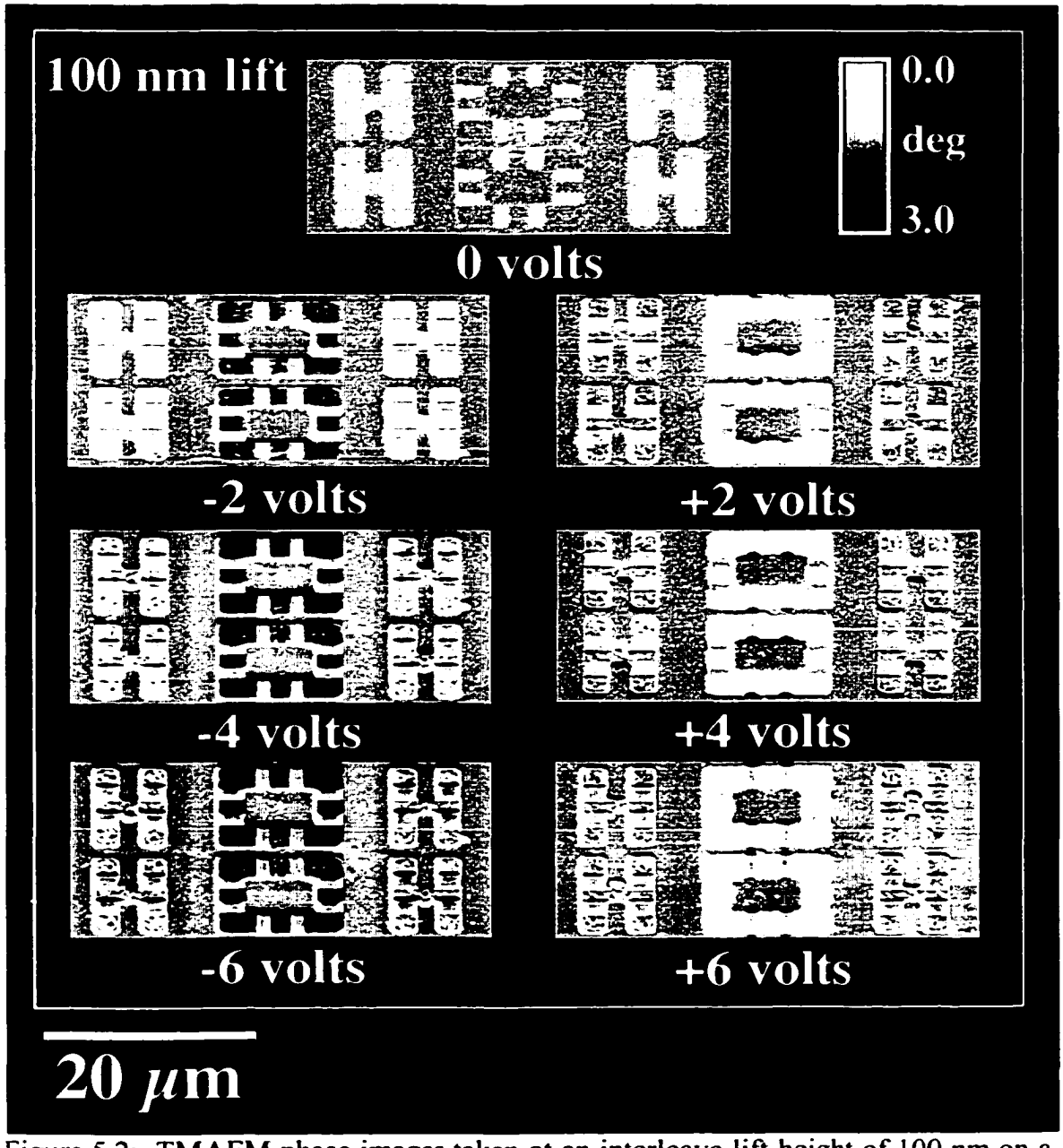


Figure 5.2: TMAFM phase images taken at an interleave lift height of 100 nm on a 16 x40 μm area of the SRAM sample. The two rows of images towards the right side were taken at different sample biases. The top row was measured at positive biases while the bottom row was measured at corresponding negative biases. Phase contrast reversal occurs when the bias is switched. The image on the left was obtained by grounding the sample.

for the contrast mechanism is discussed in detail. Figure 5.2 clearly demonstrates that the differently doped n-type areas yield clearly distinguishable phase signals.

Figure 5.3 shows the -4V phase image from Figure 5.2 with the z-scale magnified by a factor of 4. At this magnification the phase contrast differences between the "n-well" and "p-epi" areas are easily distinguishable. At negative biases the n-type areas appear "dark" with the highest doping density (N+) appearing the darkest (highest phase contrast). The n-well, which has the second highest n-type doping density, appears dark and has the second highest phase contrast. Regions doped p-type appear as "light" at negative bias. However, it is not as easy to determine the true shade of the P+ regions due micro surface roughness and possible silicide contamination in these regions. The overall doping density contrast range of the images allows us to estimate that the maximum doping density dynamic range might well exceed 10^{16} to $10^{20}cm^{-3}$. The p-channel (n-type $1x10^{17}$ cm^{-3}) and n-well (n-type $2x10^{17}$ cm^{-3}) show a dramatic phase

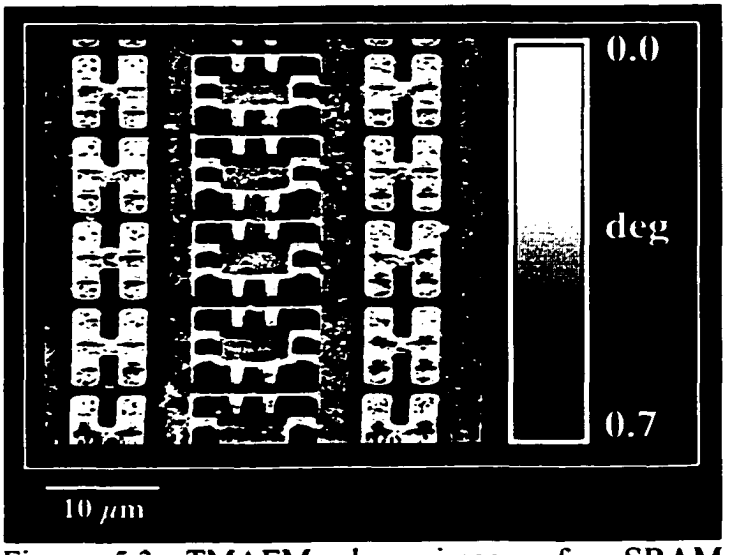


Figure 5.3: TMAFM phase image of a SRAM devices taken with a lift height of 100 nm with –4 V applied to the sample. Image is from the same scan as the –4 V image in Figure 5.2, however, the z-scale has been collapsed in order to emphasize contrast between the n-well and p-epi structures.

contrast difference, indicating that at 10^{17} dopants/cm^3 the BAAFM technique is capable of distinguishing between 1/10 an order of magnitude in doping density.

In Figure 5.2 the highest doping density region (n+ 10^{20} dopants/cm^3) seems to stand out at both positive and negative biases. It seems much more difficult to obtain contrast for regions of much lower doping density. Increased sensitivity at higher doping density is most likely due to the logarithmic dependence of the surface charge density on doping density.[9] This non-linearity in the doping-dependent phase signal was discussed in Chapter 4 section 4.6

We investigated a commercial SRAM integrated circuit pattern with areas of known doping density using TMAFM with an additionally applied D.C. bias. Our results demonstrate that the dynamic range of the technique for the detection of doping densities during one scan spans a range from at least from 10^{16}cm^3 to 10^{20}cm^{-3}. Furthermore, there is a general trend in the measurement indicating that at negative bias the increasingly higher n-type dopant concentrations appear successively darker and higher p-type doping concentrations appear successively lighter. The opposite is true at positive biases. Our experiments also indicate that the lateral resolution of the method is below 100 nm. These results further suggest that this method has potential to become a valuable non-destructive method for two-dimensional dopant profiling of semiconductor devices and satisfy the proposed specifications of the National Technology Roadmap for Semiconductors.[10]

5.4 Cross-Section Sample Preparation

Many of the problems associated with two-dimensional dopant profiling of cross-sectioned samples involve sample preparation methods.[11] The final state of the polished

semiconductor surface seems to be important. Oxide thickness, oxide charging, the degrees of hydrogen termination, adsorbed particulates, micro-cracking, and good electrical contact to the sample are just a few of the issues which arise when attempting to prepare a cross-section that will produce a high quality image of the doping in the sample. The remainder of this section includes a discussion of two procedures I have used for preparation of cross-section samples used for producing reliable BAAFM cross-sectional doping profiles. The two methods which will hereafter be referred to as the "Allied method" and "Buehler method" are fairly similar. However, the Allied method is faster and is more reliable when trying to section through small features. The Allied method also decreases the amount of doping density dependent topographic features which can often result in artifacts when imaging with most dopant profiling methods.

Buehler Method:

With the Buehler Method, three cleaved IC parts are glued together using a conductive silver epoxy (Figure 5.4(a)). Glass slides are then attached to the outside surfaces of the three ICs using silver epoxy on the back surface and clear epoxy on the front glass-IC interface. This creates a laminated sandwich structure which then is potted in a clear epoxy. Potting produces a rigid structure, reducing stresses that could otherwise be transferred to the silicon when one of the sandwich edges is being polished. Polishing with successively smaller polishing particles eventually results in a surface which can be easily imaged with AFM.

Typically, the samples should be cleaved approximately about 500-1000 μm from the area of interest, making it possible to polish down to the area of interest. Cleaves should be made so that each glass or silicon part is successively smaller. This aids in alignment

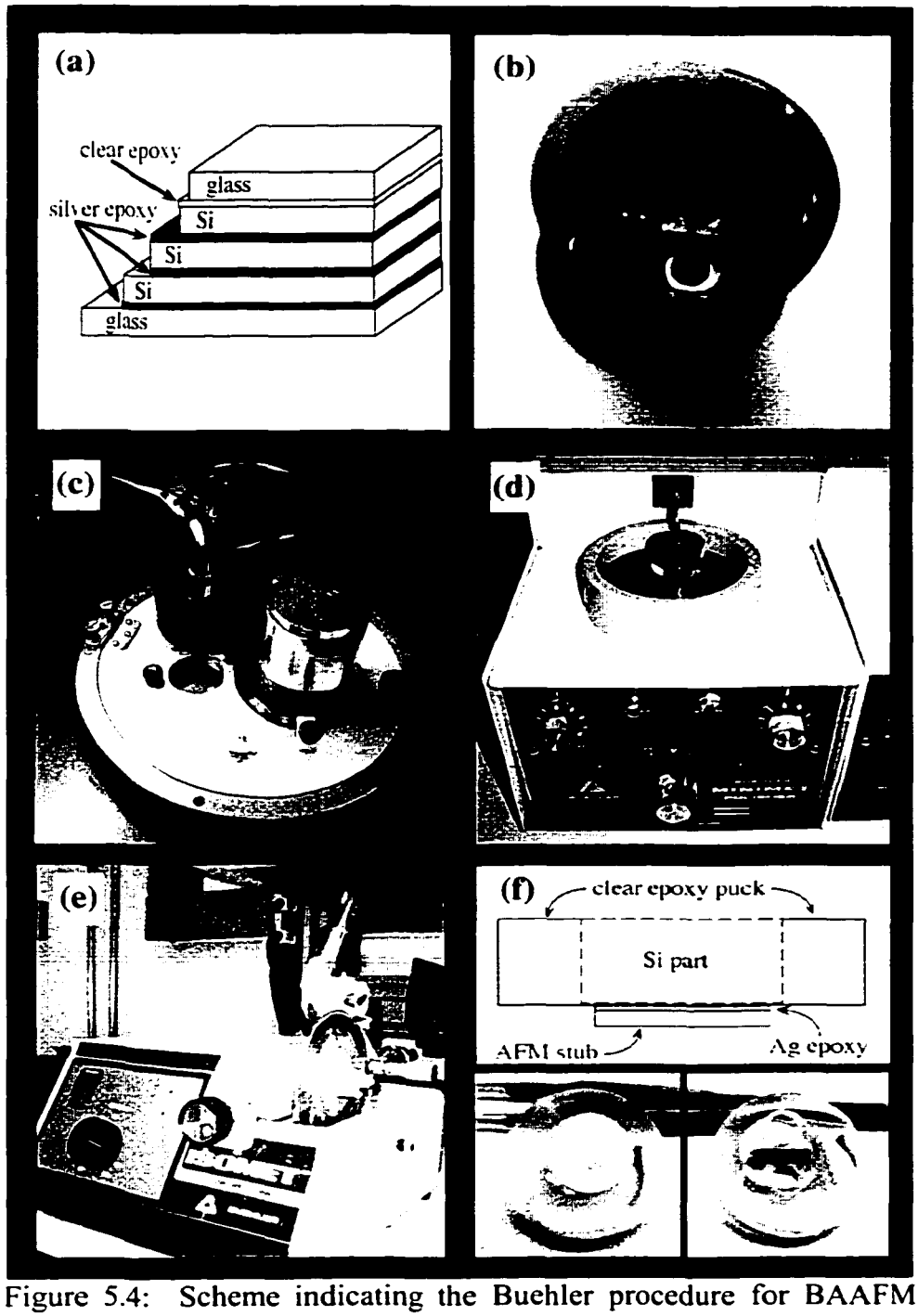


Figure 5.4: Scheme indicating the Buehler procedure for BAAFM cross-section sample preparation. Images above demonstrate the following sample preparation steps: (a) lamination of IC parts between glass slides, (b) potting the Si part in Buehler Epoxide® epoxy, (c) grinding both surfaces of the epoxy puck flat and parallel using a Buehler wet lapidary wheel, (d) polishing using a Buehler Minimet® polisher and either diamond slurry or colloidal silica, (e) cutting off the sample with a diamond saw (f) mounting the sample on a steel AFM puck.

of the front "polishing" edge of the parts when gluing the sandwich stack together. Glass slides are cleaved and prepared so that they fit on the top and bottom of a sandwich stack of wafers. Before adhering the ICs together with epoxy, the entire backside of each part should be scratched lightly with a scribe in order to remove any oxide, thus improving electrical contact. Alternatively the oxide may be removed from the back of the wafer by dropping a small bead of 40% HF on the back of the part. When working with the HF it is important to make sure that the bead of solution does not get wicked to the front side of the part or damage may occur to the area of interest. Stack, align, and epoxy the parts together as shown in Figure 5.4(a). Be sure to place the target area of each part at the even or flat edge of the sandwich. The glass slides then are glued to the outside of the wafer stack in order to prevent rounding of the edges of the silicon parts during the polishing. If using a heat cure epoxy be sure not to thermally shock the part. This could result in microcracking which could eliminate or greatly degrade the electrical contact to the target region.

After the epoxy holding the layers together has set, the sandwich is potted in a 32 mm diameter mold using Buehler Epoxide® epoxy (Figure 5.4(b)). Prior to placing the part in the mold and pouring epoxy it is essential to spray-coat the walls of the mold with mold release. The face of the sandwich to be polished is placed at the bottom of the mold so that it is not buried in the epoxy. Immediately after pouring the mold, the epoxy is out-gassed in a vacuum desiccator for about 20 minutes. The sample is then removed from the desiccator and set aside 24 hours for the epoxy to set. The epoxy puck containing the sample is removed from the mold and the backside is ground flat using a 45 μm grit Buehler wet lapidary wheel (shown in Figure 5.4(c)). The front surface of the puck is

then ground with a 15 μm grit wheel using a wet lapidary grinding wheel until the surface of the silicon sandwich is flat. The remainder of the polishing is carried out with a Buehler Minimet® polisher (fig 5.4(d)). Minimet® polishing cycles of about 30 minutes are performed with each of the following grit polycrystalline diamond suspensions (Allied Tech Products Inc.; Rancho Dominguez, CA): 45 μm, 30 μm, 15 μm, 9 μm, 3 μm, and 1 μm. For each of the diamond suspension steps a new Buehler Textmet® polishing cloth is used. The final polishing steps involve a 5 minute Minimet® cycle using 0.3 μm and 0.05 μm Gamma Micropolish II Alumina (Buehler LTD.; Lake Bluff, IL) diluted 1:4 with deionized water followed by a 10 minute polish using a 0.05 μm colloidal silica suspension (Allied Tech Products Inc.). A 40-7212 Microcloth® (Buehler Ltd.) is used as the polishing surface for both the alumina and silica polishing steps. If rounding of edges or pitting of the sample surface is observed by AFM analysis the alumina polishing step should be eliminated. Reaching the target region depth during the colloidal silica polishing step is difficult but can be done with practice.

Viewing the sample at any magnification under an optical microscope after rinsing the sample with water and drying with nitrogen should yield optically flat silicon surfaces under bright field and a uniform looking surface under dark field. At this point the surface looks optically flat. However, particles from the colloidal silica-polishing step cover the sample's surface. These particles are not easily removed since the particle's small radius of curvature results in fairly strong adsorption to the silicon surface. Figure 5.5 shows AFM images of a cross-section after the colloidal silica polish at various stages of surface cleaning. After a light water rinse followed by drying under high pressure N_2 (Figure 5.5(a)), the sample surface remains littered with multiple layers

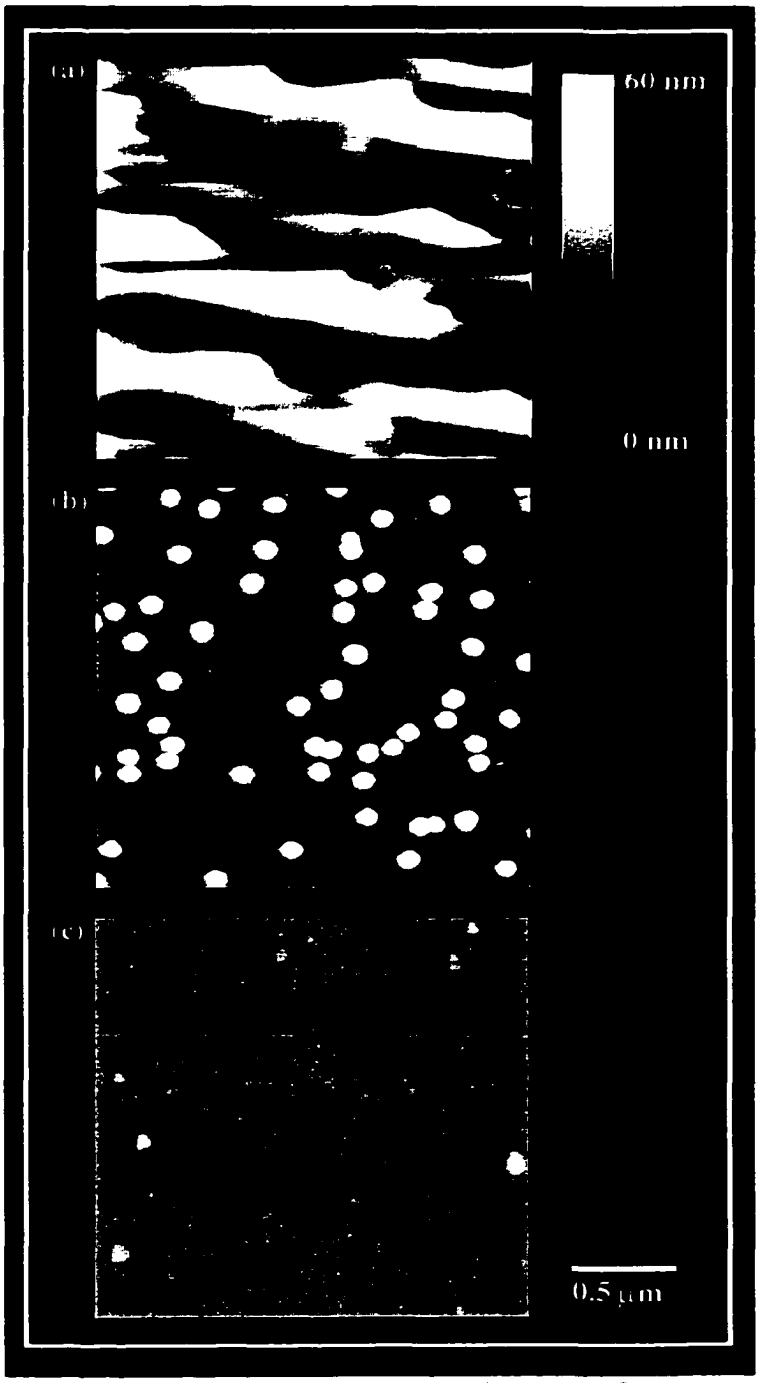


Figure 5.5: Topographic AFM images of a cross-sectioned silicon sample after the final 0.05 μm colloidal silica polishing step. The sample surface was then cleaned by (a) rinsing with low pressure water followed by drying with high pressure N_2, (b) rinsing with high pressure water for 5 min. followed by drying with high pressure water, and (c) ultrasonic cleaning for 10 min. in isopropyl alcohol followed by 20 min. in deionized water.

of colloidal silica particles. A high pressure water rinse followed by drying under high pressure dry nitrogen results in removal of most of the colloidal silica from the surface. Individual silica particles (50nm – 75nm diameter) sitting on a flat silicon surface can be easily seen in the AFM image (Figure 5.5(b)). The silica particles charge in BAAFM, thus, producing bright spots in the lifted phase image. The particles are therefore undesirable since they obscure details in the doping profile. Ultrasonic cleaning in a isopropyl solution for 20 minutes followed by 10 minutes in deionized water removes practically all of the particulates on the silicon surface (Figure 5.5(c)). Cleaning for a longer period of time may be needed for high resolution imaging.

After the final polishing step with colloidal silica the sample puck should be cut to its final thickness and mounted on a steel AFM sample mounting stub. The epoxy puck containing the embedded sample is cut off parallel to the polished face using a Isomet® (Buehler LTD.) low speed saw with a Buehler diamond wafering blade (Figure 5.4(e)). The cut should be made so that it goes through the silicon sandwich embedded in the epoxy. This results in silicon being exposed on the back site of the sample puck, allowing for good electrical contact to be made to the sample from the backside. During the cut-off step care should be taken to ensure that water being used as the cutting coolant does not make its way onto the polished surface of the sample. If this occurs the final colloidal silica polishing step and ultrasonic cleaning should be repeated.

The oxide should be removed from the silicon on the back side of the sample puck using HF (47%). After the sample is "cut-off," it is mounted on a metal AFM sample mounting stub. The AFM stub should be cleaned using a Scotchbrite® pad. Silver epoxy is used to attach the sample puck to the AFM stub after both surfaces are clean and

free of oxide (Figure 5.5(f)). Be sure to attach the AFM stub to the part of the sample puck which contains the exposed silicon sample. Good electrical contact is essential at this point since BAAFM relies on the bias applied between the tip and sample. After mounting the sample on the AFM stub the sample is ready to be imaged in a Digital Instruments Dimension Series AFM.

Allied Method:

As was mentioned above, the Allied method is similar to that of the Buehler method. However, it has the advantage of being faster and more reliable when trying to section through small features. When using the Allied method, the first step is to mount the sample and a cover glass directly to a Allied chuck (Figure 5.6(a)). Silver epoxy is used at the silicon-chuck interface in order to make good electrical contact. It is also essential to remove any oxide from the backside of the silicon part with 47% HF prior to attaching the silicon to the chuck. The glass slide is then adhered to the top of the silicon part using a clear two-part epoxy. Glass is used to keep the edge of the part from rounding during the polishing. It is also possible to see through the glass-epoxy layer, thus allowing direct monitoring of the position reached in the sample during the polishing by optical microscopy. A picture of the glass-silicon-chuck is shown in Figure 5.6(b). Note that the green wire is attached to the Allied chuck using silver epoxy and will later be used to make an electrical connection to the sample holder. This is essential since the Allied chuck is to tall to fit into the either of the Dimension 5000's magnetic sample chucks.

The Allied chuck can be mounted in the Allied Techprep8®, lapidary grinder/polisher shown in Figure 5.6(c). Once mounted in the Techprep8®, the sample is polished for approximately 1-5 minutes with each of the following Imperial™ Diamond

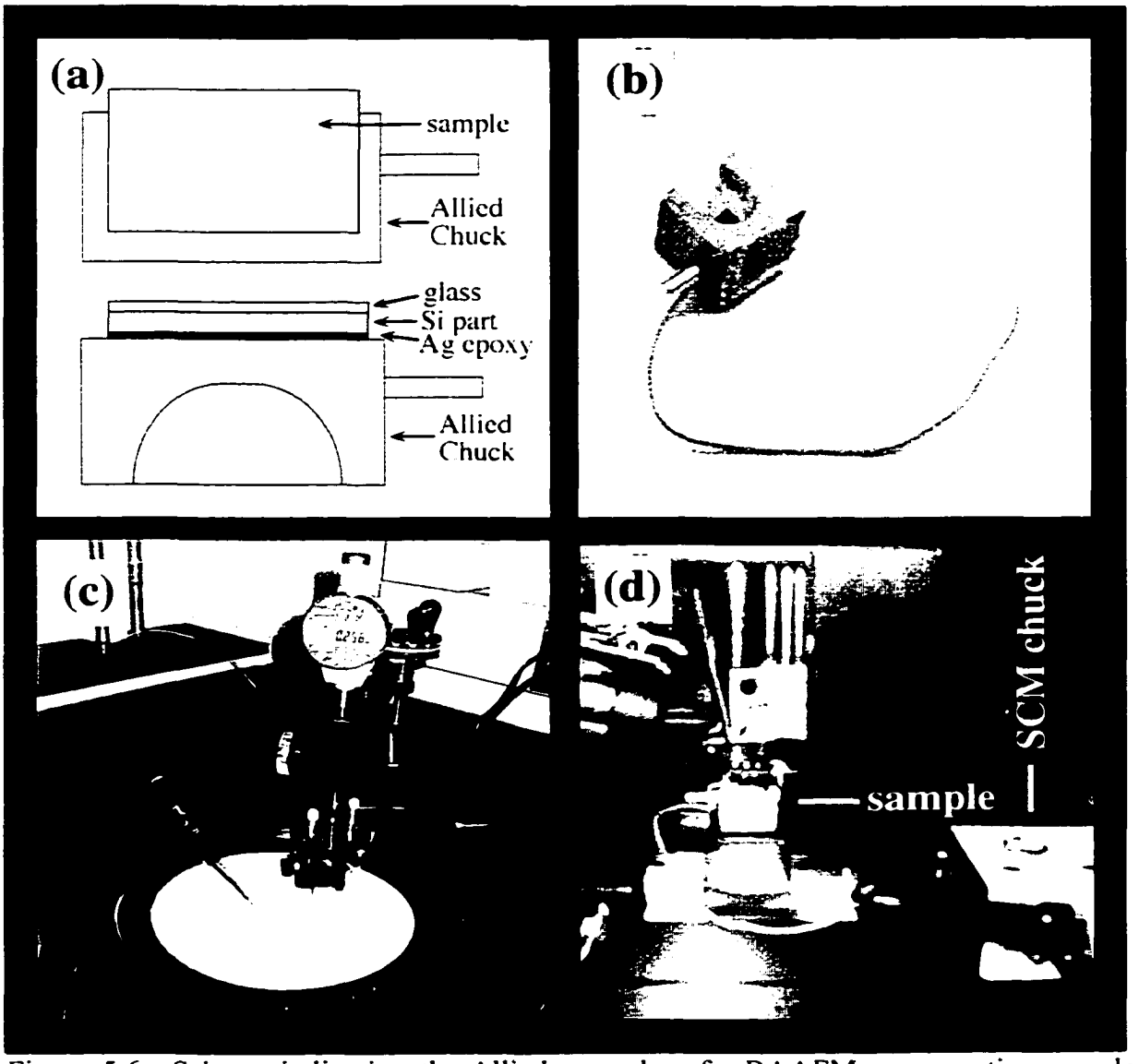


Figure 5.6: Scheme indicating the Allied procedure for BAAFM cross-section sample preparation. Images above demonstrate the following sample preparation steps: (a) mounting of IC parts with a glass cover to an Allied sample mounting fixture, Allied sample mounting fixture with a Si IC part mounted to it and a green bias transition wire attached to the back side of the sample with silver paste, (c) polishing using an Allied techprep8™ polisher and either Allied Imperial™ diamond lapping film disc or Allied colloidal silica, (d) mounting of the sample on an Allied sample mounting fixture in the D5000 AFM with a bias transmission line (green wire) running between the back of the sample and the magnetic AFM chuck.

Lapping Films suspensions (Allied Tech Products Inc.; Rancho Dominguez, CA): 30 μm, 15 μm, 9 μm, 6 μm, 3 μm, 1 μm, 0.5 μm, and 0.1 μm. Prior to switching to the next grit film the sample should be inspected by optical microscopy under both light and dark field, to ensure that large scratches from the previous grit film have been removed. The progress toward the target area on the IC should also be tracked by optical microscopy during each polishing step. If the target is a long way off it may be necessary to polish for an extended time with larger grit films in order to remove more material. After the 0.1 μm diamond films has been used and the target has been reached, the lapidary wheel is switched to a wheel containing a Red Final C polishing cloth (Allied Tech Products Inc.) and the sample is polished with 0.05 μm Gamma Micropolish II Alumina (Buehler LTD.; Lake Bluff, IL) diluted 1:4 with deionized water. This step is followed by polishing for 2-5 minutes using a 0.05 μm colloidal silica suspension (Allied Tech Products Inc.). The sample is then cleaned in the ultrasonic cleaner as was done in the Buehler method. The sample can now be mounted in the AFM on the allied chuck.

One of the advantages of using the Allied method is that the sample can be mounted in the AFM on the polishing chuck. This allows for future polishing of the sample if the target area has not quite been reached or if the surface is too rough to acquire a good image. The Allied chuck is too tall to be mounted on the magnetic sample chuck mount located on the Dimension's sample stage so a slight modification must be made to the sample stage in order to accommodate the Allied chuck and sample. A depression or groove that is 6 mm deep and 22 mm wide exists in the center of the Dimension's vacuum sample chuck. Adhering a magnet at the bottom of the groove allows one to mount tall samples (up to ~ 20 mm in height) on the sample stage as long

as the samples are adhered to an AFM sample puck. Attaching a wire between the sample and the AFM's pre-existing magnetic chuck allows the application of a bias to the sample.

As a final step in the cross-section preparation, using both polishing methods, the sample surface must be cleaned by rinsing in high pressure water followed by ultrasonic cleaning in isopropanol and deionized water. At this point a decision must be made as to whether the sample will be imaged using BAAFM or SCM. Imaging with BAAFM requires dipping the sample in a 2% HF solution for 1 minute followed by rinsing in deionized water. SCM imaging, however, requires growth of SiO_2 over the area to be imaged. Oxide growth on the SCM sample is accomplished by heating the sample at 300 C° for 45 minutes in air (30 % RH). While heating the sample is irradiated with 315 nm ultraviolet light from a 30 watt Cole Parmer 9815 series ultraviolet lamp. Ultraviolet irradiation is said to reduce the trapped charge density in the oxide and aide in the annealing of surface damage due to polishing.[11] Oxidation under these conditions is currently the standard method for preparation of silicon surfaces for SCM analysis.

In order to determine the relative difference in the two surface treatment methods samples for SCM and BAAFM analysis were investigated by x-ray photoelectron spectroscopy (XPS). Figure 5.7 shows a XPS depth profile for both samples in which the O1s, Si2p, and C1s peaks were monitored to determine the percent concentration of these elements as the sample surface was sputtered away with 1 keV argon ions. This ion energy is chosen in order to minimize sample damage and to maximize sputtering rate.[12] After their respective surface treatments the BAAFM (Figure 5.7(a)) and the SCM (Figure 5.7(b)) were allowed to sit in the laboratory for 1 hour prior to introduction into

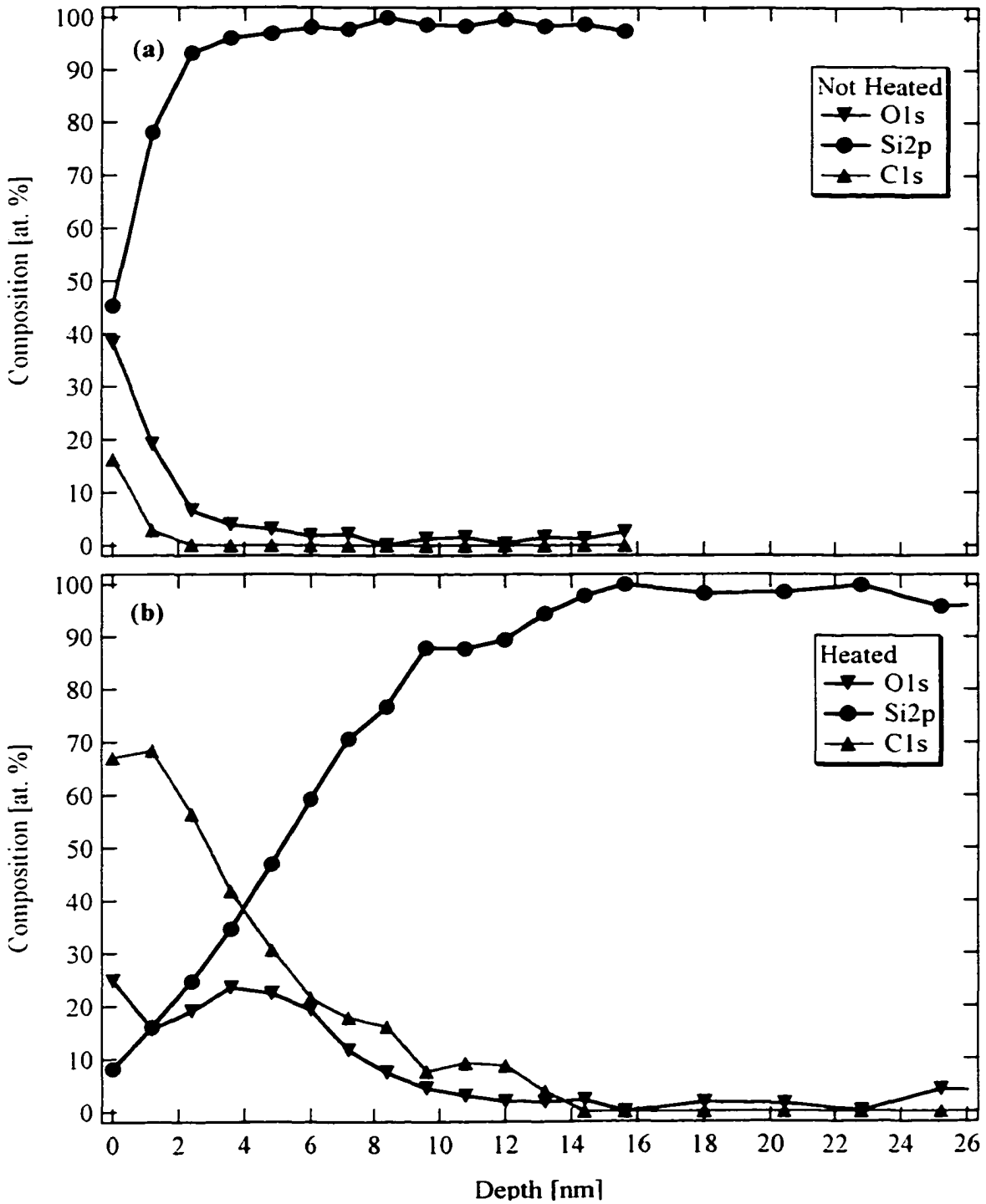


Figure 5.7: X-ray photoelectron spectroscopy (XPS) Ar^{+} sputtered depth profile of a (a) sample prepared for BAAFM analysis by dipping in a 2% HF solution for 1 minute, and a (b) sample prepared for SCM cross-sectional analysis by thermally growing (300 C° for 45 min.) an oxide on the sample surface under a 315 nm ultraviolet lamp.

the XPS vacuum chamber in order to simulate the condition that the samples would be imaged. Oxygen and carbon in the BAAFM sample were shown to only penetrate into the first 2 nm of the sample's surface. The approximate depth of the oxide grown for SCM was, however, on the order of 8 to 10 nm. Carbon, most likely from the epoxy used in sample preparation, also penetrates about 12 to 14 nm into the sample.

5.5 Investigation of a Cross-Sectioned Bipolar Device

A Hewlett Packard bipolar transistor was chosen for cross-sectional dopant profiling using BAAFM and SCM. A computer simulated iso-contour doping drawing of this transistor is shown in Figure 5.8. A bipolar transistor consists of an emitter, base, and collector, as was described in Section 1.5 of Chapter 1. This device is a npn-transistor with the emitter, poly-silicon base and collector doped with arsenic, boron, and phosphorous, respectively. The emitter, which was added in the drawing after the simulation, is highly n-doped at about 10^{21} dopants/cm^{-3}. Doping in the p-type base and n-type collector is on the order of 10^{18} cm^{-3} and 10^{16} cm^{-3}. The buried layer is arsenic doped (n-type) with the concentration of the center of the buried region being about 10^{19} cm^{-3}. An isolation layer or channel stop is shown on the right side of the drawing. This region is doped heavily p-type and is used to electrically isolate neighboring transistors.

Figure 5.9 shows a set of cross-sectional BAAFM images of one of Hewlett Packard's bipolar transistors. The topmost image shows a topographic cross-section of the device after polishing using the "Buehler" method. The surface washed with high-pressure water and then ultrasonically cleaned in isopropanol followed by water in order to remove colloidal silica, remaining after the final polishing step. A few silica spheres can still be seen on the sample surface in the topographic image. Prior to imaging, the

sample was dipped in a 2% HF solution in order to remove surface oxide and to passivate the surface.[8] The metal contacts in the device appear as the highest features in the cross-section image. Note that in the topographic image the buried layer and collector can be faintly seen as slight depressions. This slight contrast variation from that of the bulk silicon is due to doping density-dependent differential etching silicon during polishing or dipping in HF prior to imaging.[3] Field oxide appears as a much deeper depression due to a faster etching rate during the polishing and HF passivation steps.[8]

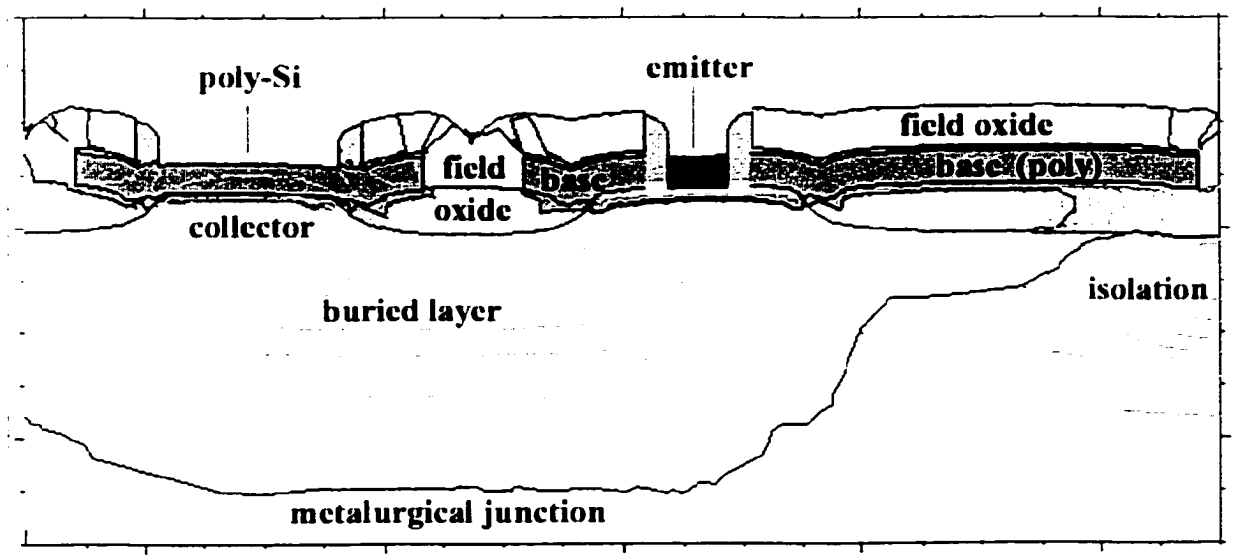


Figure 5.8: Calculated doping scheme of a Hewlett Packard bipolar transistor shown in cross-section. Hewlett Packard has several bipolar devices, which vary in their dimensions. Relative dimensions of device parts in this scheme are not necessarily the same as the device that was measured by BAAFM and SCM.

The three lower images in Figure 5.9 were taken of the same area as the topographic image however using BAAFM with a lift height of 75 nm and biases of –2, 0, and +2 volts. The 0 V image shows a lack of phase contrast indicating that topography related effects have been eliminated and the tip-sample separation is constant throughout the lifted scan. Bias-dependent contrast reversal can be seen in some areas of the +2 and –2 V images, with the buried layer showing the most obvious

reversal. Note that at both positive and negative biases the field oxide appears bright. This results from the oxide charging to about the same charge density at both positive and negative biases.[1] We have observed that oxide and metals appear bright no matter what the sign of the bias. The poly base appears to stay at a mid-tone contrast value at both signs of the bias. The highly doped emitter is most likely showing up as a very dark band lying below the metal above the right edge of the buried layer in the –2 V image. In the +2V case the emitter shows contrast that is bright enough to make it indistinguishable from the metal contact. In the –2 V image the transition from mid-tone gray to a lighter shade may be outlining the position of the metallurgical junction which is shown in the calculations in Figure 5.8. The channel stop cannot be seen since it is out of the field of view in this set of images.

Figure 5.10 shows a scanning capacitance microscopy (SCM) dC/dV image, which was taken above the same structure as is shown in the BAAFM images in Figure 5.9. Scanning capacitance microscopy was described in Section 2.5. After imaging with BAAFM, a thermal oxide was grown over the same sample surface by heating on a hotplate at 300 C° for 45 min with simultaneous 315 nm UV exposure. Variations in the between the BAAFM and SCM surface compositions can be observed in the XPS depth profile of these surfaces shown in Figure 5.7. It is quite obvious at first glance that scanning capacitance reveals much of the same information as is acquired using the BAAFM method.

The SCM image contains streaks which can be attributed to the coating material on the Co/Cr tip (MESP Digital Instruments, Santa Barbara, CA) being worn off as the tip is dragged across the surface. Tip wear in SCM imaging is common and has been

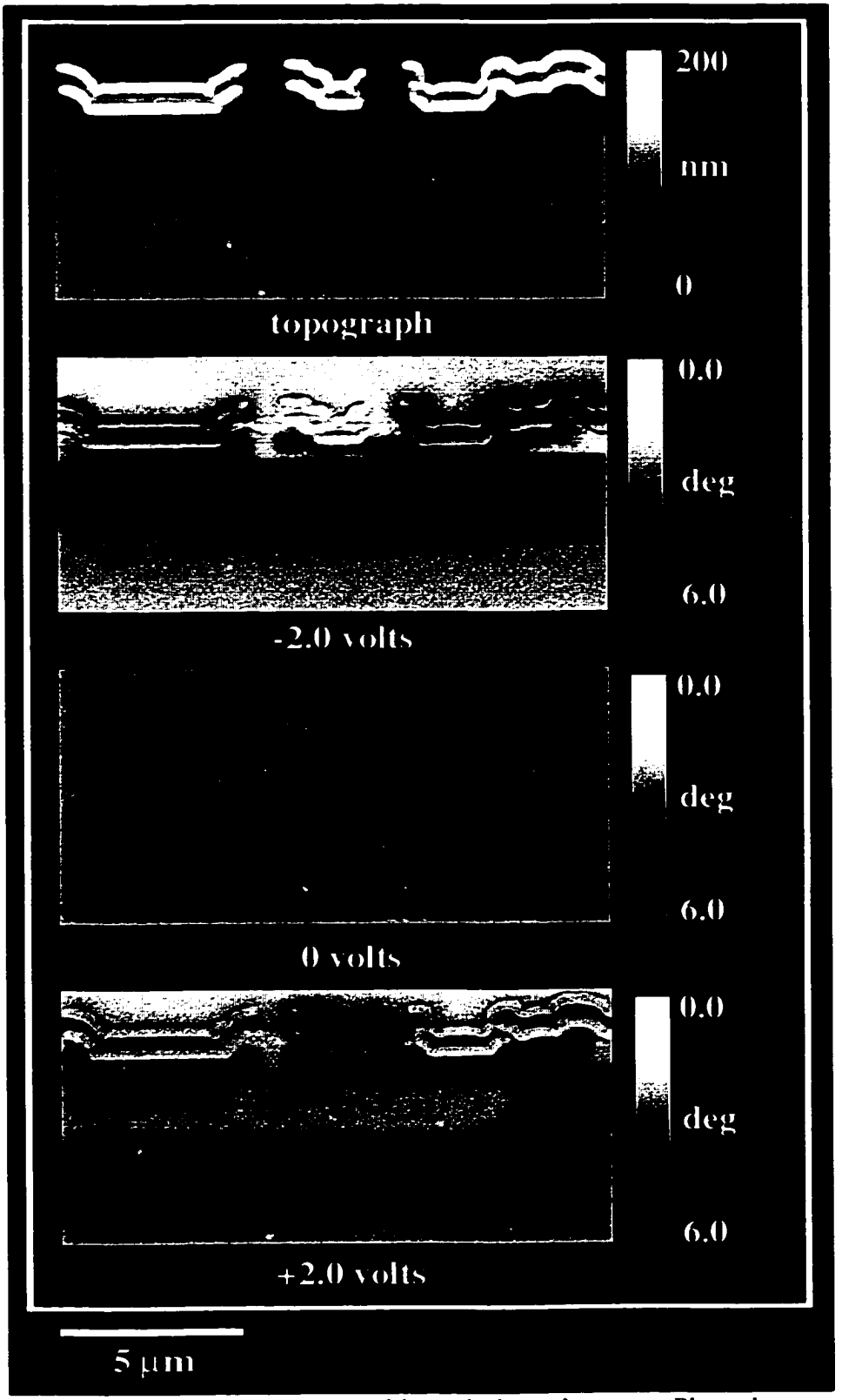


Figure 5.9: TMAFM topographic and phase images. Phase images were acquired at an interleave lift height of 75 nm at –2, 0, and +2 volts.

attributed to the fact that soft metal coatings used in combination with contact mode imaging can result in loss of tip material on the sample surface.[11] Often, tip wear is revealed by a loss of spatial resolution as the sharp tip wears flat or by streaking or noise spikes. Loss of resolution can occur within the first few scan lines after engagement of the cantilever. This can make it impossible to scan a wide area and then zoom in and focus on analyzing a gate structure. Since BAAFM operates in tapping mode, tip wear is much less problematic. Often, tips may be used for many days without any loss of resolution.

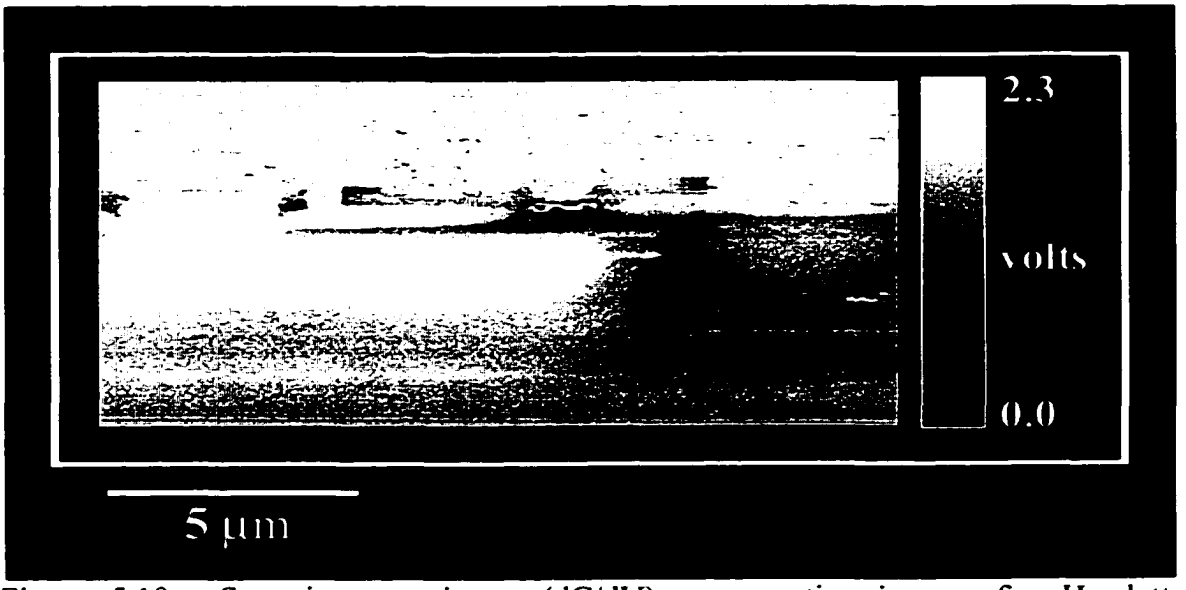


Figure 5.10: Scanning capacitance (dC/dV) cross-section image of a Hewlett Packard bipolar transistor.

Although not all of our SCM images show streaking, the image presented in Figure 5.10 represents a typical SCM image. Many of the images of the cross-sectioned sample resulted in indistinguishable dopant profiles since the image quality was too poor to observe features. Most of the BAAFM samples, however, produced images like those shown in Figure 5.9. Problems with sample preparation are often blamed for poor image quality in dopant profiles. Cross-sectioning samples produces the possibility that surface

defects and bulk defects created during the cleaving and polishing of the sample can result in surface charges. Charges that result from the growth of the surface oxide can also appear in samples prepared for SCM. These surface charges result in either a change in capacitance or a change in the phase signal. If charging is not consistent across the sample surface, contrast in both SCM and BAAFM images will not represent a dopant profile and the surface damage will dominate. Much attention must be placed on development of sample preparation techniques for SCM and BAAFM.

BAAFM and SCM should be considered complementary techniques. Based on the imaging mechanism, sensitivities for BAAFM should increase with increasing doping density since the phase signal is sensitive to surface charge density that increases with doping density. SCM measures a change in the derivative of the capacitance with respect to the amplitude of an applied A.C. voltage. The depletion width is larger for lower doping densities than for higher doping density. As the capacitance changes due to the changing value of the A.C. bias applied to the sample, charge carriers in the sample move between the surface and the maximum depletion depth determined by the A.C. voltage. The change in depletion width is largest for lower doping densities making capacitance most sensitive to low doping density samples. Therefore, depending on the composition of the sample to be imaged, one method may be preferred over the other.

References:

[1] J. W. Mayer and S. S. Lau, *Electronic Materials Science: For Integrated Circuits in Si and GaAs* (Macmillan Publishing Company, New York, 1990).

[2] A. S. Grove, *Physics and Technology of Semiconductor Devices* (John Wiley and Sons, Inc., New York, 1967).

[3] M. Barrett, M. Dennis, D. Tiffin, Y. Li and C. K. Shih, J. Vac. Sci. Technol. B **14,** 447 (1996).

[4] D. Gräf, M. Grundner and R. Schulz, J. Vac. Sci. Technol. **A7,** 808 (1990).

[5] V. A. Ukaraintsev, R. McClothlin, M. A. Gribelyuk and H. Edwards, J. Vac. Sci. Technol. B. **16,** 476 (1998).

[6] R. F. Pierret, *Field Effect Devices*, Vol. IV, 2nd ed. (Addison-Wesley Publishing Company, Reading, 1990).

[7] R. Schlaf, R. Hinogami, M. Fujitani, S. Yae and Y. Nakato, J. Vac. Sci. Technol. A **17,** 164 (1999).

[8] L. J. Huang and W. M. Lau, J. Vac. Sci. Technol. **A10,** 812 (1992).

[9] S. M. Sze, *Semiconductor Devices: Physics and Technology* (John Wiley & Sons, New York, 1985).

[10] *S. I. A. The National Technology Roadmap for Semiconductors, 4300 Stevens Creek Boulevard, San Jose, CA.,.*

[11] V. V. Zavyalov, J. S. McMurray and C. C. Williams, Rev. Sci. Instrum. **70,** 158 (1999).

[12] D. Briggs and M. P. Seah, *Practical Surface Analysis: Auger and X-ray Photoelectron Spectroscopy*, Vol. 1 (John Wiley & Sons, Chinchester, 1996).

Conclusions:

We investigated the influence of Coulomb forces on scanning tunneling microscopy (STM) images of dopant sites at layered semiconductor surfaces. The images show a tunneling bias dependence on the apparent height of the dopants. Observed protrusion heights are in the 1 nm range, ruling out an explanation based solely on the local density of states (LDOS) and band bending effects. Results suggest that mechanical deformations of the surface due to bias-induced Coulomb forces are responsible for the observed effects. This conclusion is supported by current imaging tunneling spectroscopy measurements, which show only minor changes in the LDOS of the dopant sites with respect to the MoS_2 lattice.

To investigate the effects resulting from Coulomb forces, we performed additional measurements with Tappingmode® atomic force microscopy (TMAFM) under applied D.C. biases. These measurements showed a strong influence of the applied bias on the measured topography. Measurement of force curves at varying biases allowed the bias-induced Coulomb forces to be measured. An estimation of the charge density on the cantilever and sample was then determined from these force measurements. These results showed that the charge density is in the range of one elementary charge per/nm^2 or less, supporting the theory that the presence of localized dopant charges can strongly alter the forces between the tip and sample in STM measurements.

Doping in three-dimensional semiconductors was investigated using TMAFM with an additional D.C. bias applied between cantilever and sample. Areas of different

doping density cause bias-dependent variations in the Tappingmode® phase angle. This phase angle contrast was used to characterize the lateral doping profile of doped silicon test patterns. The imaging contrast mechanism was investigated by calibrating the force measurements and amplitude/phase relationships with applied biases on a gold-coated mica substrate. The impact of the Coulomb force induced by the applied bias on the cantilever oscillation was then determined.

A model was developed showing how variations in band bending between differently doped areas of the surface produces the contrast observed in the phase images. Surface charge density variations, produced by biasing the sample, are related to phase contrast observed between areas of varied doping density. Surface charges density as a function of the sample's surface potential was calculated for different doping densities. These curves were used to validate many of the assumptions made in the mechanistic model. Humidity was determined to adversely affect the measurement process, since a water layer at the semiconductor surface results in dielectric screening of the doping charge from the tip. The nature of the cantilever coating was determined to not influence the measurement process, however, lower spring constant cantilevers prove to increase signal-to-noise.

Dopant profiling of commercial silicon based devices was performed from the top-down and in cross-section. From measurements presented here and from measurements on proprietary devices we were able to distinguish between p- and n- type areas ranging from 10^{16} to 10^{20} dopants/cm^3. The newly developed technique also showed the ability to distinguish between two areas differing by 1/10 an order of magnitude at 10^{17} dopants/cm^3. The detection limit may, however, be lower than this.

Future Directions:

Although the BAAFM technique has been tested on commercially available devices in one of Hewlett Packard's IC analysis facilities, a great deal of work still must be done in order to make BAAFM a "turn key" quantitative technique. The spatial resolution achievable with BAAFM at various lift heights must be known in order to determine if BAAFM can be used for imaging doping levels in future generations of increasingly smaller integrated circuits. Quantification of dopant profiles is also essential to the development of BAAFM. Some of the key issues in quantification will involve determining sensitivity, finding the maximum and minimum detectable doping range, and finding a way to quantitatively reference all doping levels to each other in the phase contrast image. Currently, the type of standard samples that exist for characterizing doping profiling methods limits the development of the techniques themselves. Development of new standards will most likely require industrial partnerships since small scale lithographic patterning followed ion implantation is a costly endeavor not likely be attainable in a university laboratory.

Determining the spatial resolution ultimately attainable with BAAFM at various lift heights will require the creation of a set of resolution test standards. These standards should consist of repeated ion implanted curricular or hexagonal features. Each sample should have consecutively smaller diameter features. By imaging these samples the resolution of the technique can be determined when it is not possible to resolve these features. Since AFM is capable of resolving very small features, it will be necessary to

create a doping mask in which these features reach a size of a couple nanometers in diameter. Masuda and Satoh have shown that it is possible to produce such a mask in experiments where they created gold "nanodot" arrays on a silicon substrate.[1] The arrays were fabricated by using an anodic porous aluminum membrane as a vacuum evaporation mask. Anodic porous aluminum membranes, which are formed by anodizing aluminum in an acidic electrolyte, have a unique honeycomb structure that is uniform in the diameter and spacing of holes. Removal of the bottom part of the anodic porous alumina layer by an appropriate etching process can yield porous membranes, which can be traversed by an ion traveling in a straight path. The same masks that Masuda and Satoh used could be used for an ion implantation mask to produce implanted features of various nanometer scale sizes.

BAAFM research still needs to catch up to SCM in the development of quantitative models that allow calculations of dopant profiles from contrast in an image. Development of a set of quantitative doping standards will be key to the advancement of BAAFM and other two-dimensional dopant profiling techniques. There is a great need for production of a doping standard that contains patterns of various doping densities and type in an area of less than 50x50 μm on a flat silicon (100) surface. Doping standards are currently available from the National Institute of Standards and Technology (NIST) as well at the Interuniversity Microelectronics Center (IMEC) located in Leuven, Belgium. NIST provides a set of eight silicon samples, four with the same background doping density p-type substrate and the other four with the same n-type background doping density. A pattern is implanted into each substrate with a different dose of 50 keV boron ions. Unfortunately the exact surface doping density is not known and

comparison between areas of different doping density is difficult since they are not contained in the same sample. IMEC offers a set of two doping standards, one being n-type and the other being p-type.[2] The samples consist of a silicon (100) wafer on top of which are grown multiple epitaxial layers, with each 5 μm thick layer increasing in doping density by an order of magnitude. The IMEC standards offers the ability to image a large number of doping densities of the same type. However, a difficult cross-sectioning process must be performed in order to image the standards. Since the surface preparation of cross-sections is not dependable, it is difficult to use cross sections as a quantitative standard. The IMEC standards also tend to be somewhat fragile and the layers can often delaminate during cross-sectioning.

Development of a new set of standards with many areas of both doping density and type patterned into a small (50 x50 μm) area on a silicon (100) surface will allow for BAAFM doping density quantification studies. It should then be possible to determine the minimum and maximum detectable doping densities as well as the sensitivity to various doping densities. It is suspected, but not known at this point that BAAFM is more sensitive to higher doping densities than lower doping densities. It should be possible to calculate how the phase signal relates to the doping density from first principle calculations. An in-depth study of a new set of doping standards should allow for the development and validation of a mathematical model which could be used for quantitative or semi-quantitative profiling of commercial semiconductor devices.

While investigating BAAFM, humidity was found to eliminate doping related phase contrast. It was suggested that the phase contrast disappeared as a result of dielectric screening in the presence of an adsorbed water layer on the sample's surface. It

would be interesting to test this dielectric screening theory by attempting to image a sample under a variety of solvents having different dielectric constants. Fluid cells for performing TMAFM experiments in a liquid environment exist, however, use of such devices often proves quite difficult. Theoretically it should be possible to image a doped sample under a low dielectric constant solvent. If dielectric screening is the reason for the loss of contrast in humid conditions, as the dielectric constant is increased the doping related contrast should diminish.

BAAFM has been shown to be a promising new method for high spatial resolution qualitative, and possibly quantitative, dopant profiling of semiconductor devices. BAAFM has been discussed as being a technique, which is complimentary to Scanning Capacitance Microscopy. Currently, however, nano-Spreading Resistance Probe Microscopy (nano-SRP) may be the most promising technique for truly quantitative two-dimensional dopant profiling.[2-4] In order for nano-SRP to work a conductive diamond-like cantilever must be indented through the samples native oxide layer to a reproducible depth and a resistance measurement is made. The nano-indentation of the surface creates a surface that is too rough for repeated imaging making it nearly impossible to scan a large area with nano-SRP and then to zoom in and scan another area with increased resolution, since the surface would be damaged before the second scan.

One way of using nano-SRP is to first scan the sample with some other dopant profiling method in order to determine the probe's location. After the desired imaging area is located the imaging method can be switched to nano-SRP and a single quantitative scan can be taken. SCM is one obvious choice of techniques that may be used for nano-

SRP probe location. However, the requirement of having an oxide covered sample would be detrimental to the SRP method since the probe would have to be forced through the oxide layer. BAAFM would actually be ideal for determining the probe location in nano-SRP imaging. The oxide is removed by etching in a dilute HF solution prior to BAAFM imaging. This would result in minimizing the probe depth needed after switching to nano-SRP. An additional advantage to using BAAFM would be that the probe could be used in tapping mode for both BAAFM and nano-SRP. Only the tapping-mode setpoint value would have to be adjusted in order start the nano-indentation of the surface. In the near future BAAFM should be used as a prelude to nano-SRP imaging.

References:

[1] H. Masuda and M. Satoh, Jpn. J. Appl. Phys. **35,** L126 (1996).

[2] T. Clarysse, M. Caymax, P. D. Wolf, T. Trenkler, W. Vandervorst, J. S. McMurray, J. Kim, C. C. Williams, and G. N. J.G. Clark, J. Vac. Sci. Technol. B. **16,** 394 (1998).

[3] R. B. Marcus, *Spreading Resistance Microprobe* (U.S Department of Commerce, Gaithersburg, MD, 1994).

[4] P. D. Wolf, T. Clarysse, and W. Vandervorst, J. Vac. Sci. Tachnol. B. **16,** 320 (1998).

Appendix A: Current Imaging Tunneling Spectroscopy

Studying the bias dependence of the tunneling current using STM allows for the extraction of various types of spectrographic information on a spatial scale which ultimately approaches the atomic level. This spectroscopic capability combined with high spatial resolution is one of the most powerful features of STM.[1,2] Spectroscopic information from bias-dependent STM measurements can be obtained in a number of ways based on various experimental techniques. One of the most common of these techniques is Current Imaging Tunneling Spectroscopy (CITS).

CITS involves the local measurement of I-V curves at constant tip-sample separation. This particular spectroscopy method uses a sample and hold amplifier to alternatively switch the feedback loop on and off. When the feedback is switched on, a constant stabilization voltage is applied between the tip and sample and the tip height is adjusted to maintain a constant tunneling current. When the feedback loop is deactivated, the voltage applied between tip and sample is ramped linearly between two user-defined values and an I-V curve is measured at a fixed tip height. The voltage is then set back to the stabilization bias and the feedback loop is switched back on, after the I-V curve is measured. Acquiring I-V curves rapidly while scanning at a slow speed allows for simultaneous generation of a constant current topographic image and spatially resolved I-V curves.

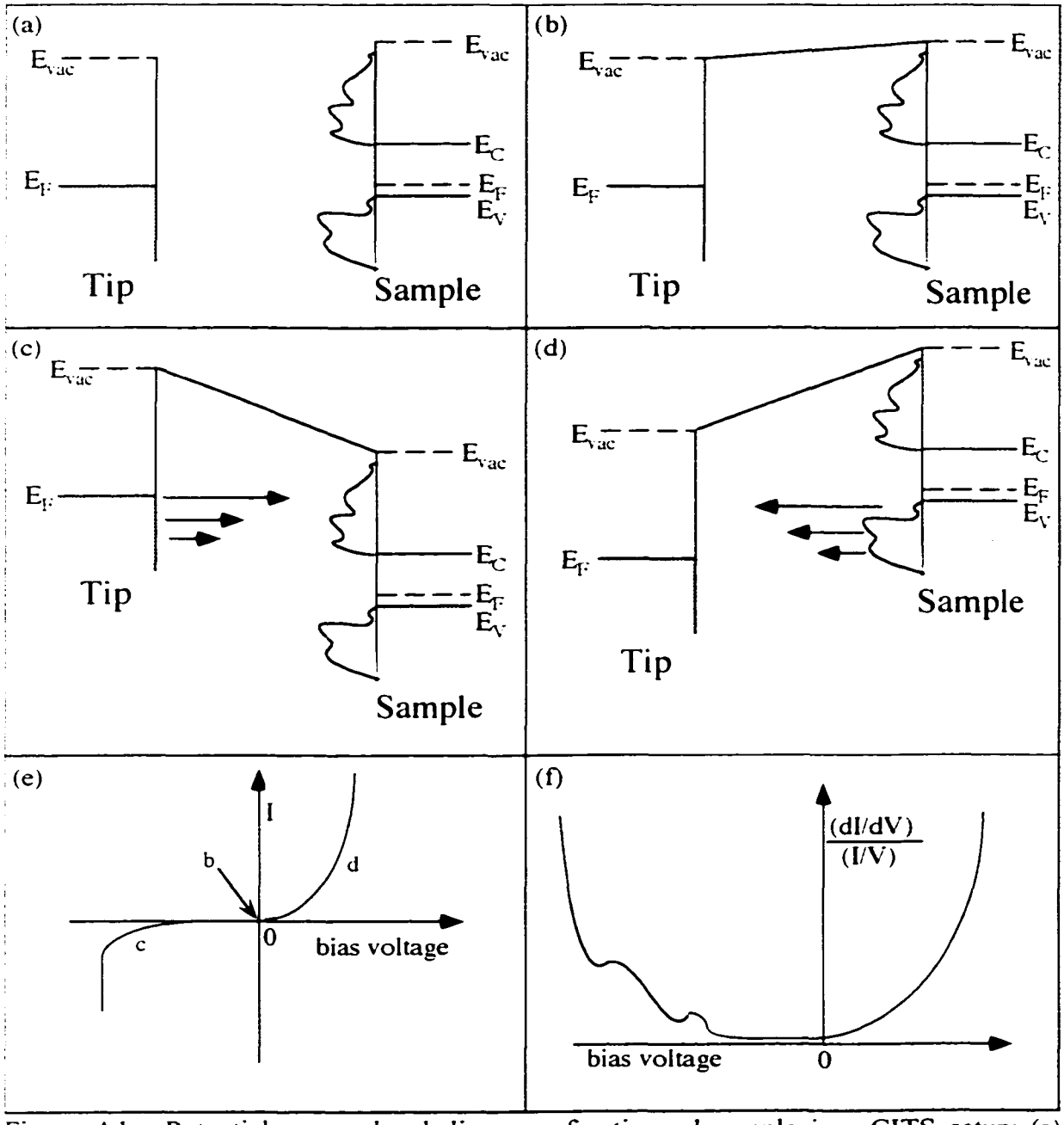


Figure A1: Potential energy level diagrams for tip and sample in a CITS setup: (a) independent tip and sample, (b) sample and tip at equilibrium, (c) positive tunneling bias with electrons tunneling from tip to sample, and (d) negative sample bias with electrons tunneling from sample to tip. (e) Scheme indicating a CITS I(V) curve typical for a p-type semiconductor. (f) Scheme showing a differentiated I(V) curve.

To demonstrate what is occurring during a tunneling I-V experiment we will consider a simple potential energy diagram for a system consisting of a tunneling tip and semiconductor sample which are separated by a small vacuum gap (Figure A1(a)). Under zero applied bias the Fermi levels of the tip and sample are at equilibrium (Figure A1(b)). Application of a bias to the sample results in a shift of the energy levels (upward or downward) by an amount |eV| depending on whether the polarity of the bias is positive (Figure A1(c)) or negative (Figure A1(d)). For positive biases the net tunneling current arises from electrons that tunnel from the occupied states in the tip into unoccupied states in the sample. With negative biases the tunneling occurs from occupied states in the sample into unoccupied states in the tip. Polarity of the bias therefore determines whether the occupied or unoccupied electronic states in the sample are probed.

By varying the bias it is possible to select the electronic states that contribute to the tunneling current and, in principle, measure the local electronic density of states. If the density of states (DOS) being tunneled into or out of is larger, then the tunneling current will increase. As the bias voltage is ramped electrons will tunnel out of **all** energy levels which are greater in energy than the Fermi level. This means that as the bias increases the density of states being probed are not only the ones at the current bias, but rather all states with lower biases than the current voltage. Figure A1 (e) shows a drawing indicating the shape of an I(V) curve taken on a p-type semiconductor. Points b, c, and d on the curve indicate bias conditions represented in the potential energy diagrams shown in Figure A1 (b, c, and d).

Since the bias is ramped, in an I-V experiment, the first derivative (dI/dV (V)) represents the electronic density of states to a first approximation (Figure A1(f)). Tunneling from the tip into the sample mainly probes the sample's empty states with

negligible influence from the tip's occupied states. Tunneling from the sample to the tip, however, is much more sensitive to the electronic structure of the tip's empty states.[3] The I-V curve depends on the DOS of both the tip and sample. If the electronic DOS of the tip varies much this will be seen in the resulting I-V curve.

References:

[1] R. J. Hamers, Ann. Rev. Phys. Chem. **40**, 531 (1989).

[2] R. M. Tromp, J. Phys.: Condens. Matter. **1**, 10211 (1989).

[3] T. Klitsner, R. S. Becker, and J. S. Vickers, Phys. Rev. B **41**, 3837 (1990).

Appendix B: AFM Force Curves

Recording a cantilever's deflection as a function of its position above a sample produces force calibration curves in Atomic Force Microscopy (AFM). Knowing the cantilevers force constant it is then possible to determine the force acting upon the cantilever as a function of the tip-sample separation. Changes in the shape of the force curves can be used to determine if long range forces are involved or whether the surface is hard or soft. Relative amounts of tip-sample adhesion can also be determined my comparing the hysteresis measured between the extension and retraction of the piezo.

Figure B1 shows a hand drawn representation of a typical contact mode AFM force calibration curve. Several points on the force curve are indicated by numbers. Point 1, lying on the red curve, indicates the where the z-piezo starts to extend and pushes the tip toward the surface. At this point there is no force interaction with the surface. As the piezo continues to extent toward the surface eventually the tip is pulled toward the surface rapidly (point 2). This point is referred to as the snap to contact.

As the tip presses into the surface the cantilever starts to bend upward. This is represented by linear region sitting between points 2 and 3. For a hard surface the cantilever is bending at a rate which is equal to the rate of z-piezo movement the slope of the contact region should theoretically be unity. Raw data usually gives slopes that are

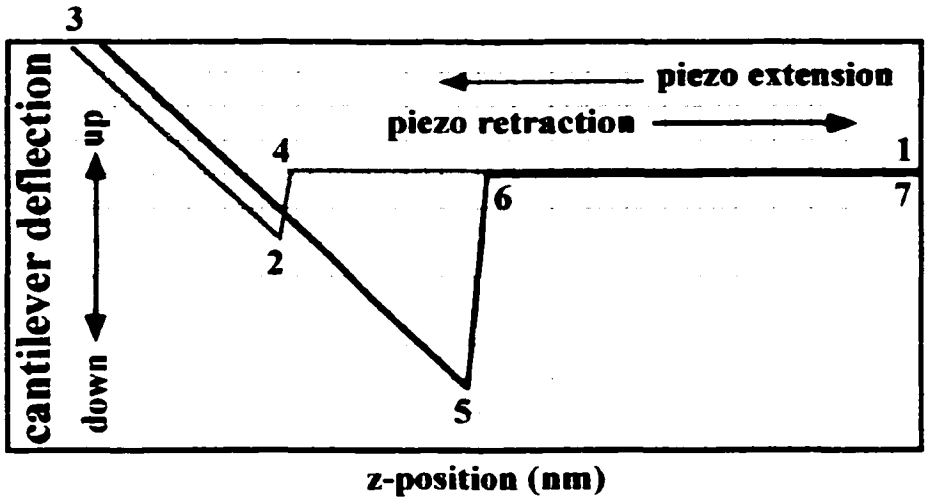


Figure B1: Drawing representing a typical force calibration curve recorded by contact mode atomic force microscopy.

not unity since the actual cantilever deflection is difficult to calibrate. Forcing the slope of the contact region to one after acquiring the curve is usually how the deflection component of force curves is calibrated.

If the tip continues to press into the surface it will reach a point where it can no longer bend with out breaking. The direction of the z-piezo is then reversed (point 3) and the piezo starts to retract the cantilever away form the surface. As the cantilever is retracted the cantilever relaxes downward until the tip forces are in equilibrium with the surface forces (point 4). After point 4 the piezo continues to retract and the cantilever bends downward as the surface attraction holds onto the tip (up to point 5). As the piezo continues to retract the tip finally breaks free of the surface attraction forces and the cantilever bound sharply upward (point 6). As the piezo continues to retract the tip continues to ascend. There is no further contact with the surface during the cycle.

Appendix C: BAAFM Instrument & Software Setup

In order to apply a bias between the sample and tip it is necessary to make some modifications to both the software and hardware of a Digital Instruments Multimode™ or Dimension™ 5000 SPM. When applying a bias it is possible to use either the Nanoscope's internal power supply or an external D.C. power supply. Applying a bias internally, through the computer, allows application of any bias between −12 V and +12 V. The major advantage to internally applying a bias is that it is possible to apply the bias during the lift scan and turn it off during the main topographic trace. Images with higher spatial resolution are usually obtained when the bias is applied only during the lift scan since the main topographic trace tends to smear out when the surface is electrically charged. A blurry topographic image will add error to the tip-sample separation distance in the lifted image since the trajectory of the lift image is dependent on the lifted image. An external power supply is typically used only when biases greater that ±12 V are required.

Jumper pins, which are located in the left side electronics panel in the Dimension 5000 and in the base of the Multimode™ SPM, must be wired as indicated in Figure C1 to use the microscope with internal or external biasing. The position of the jumper pin board in the Dimension™ 5000 SPM is indicated by an arrow in Figure C1(a). Inside the side panel of the extender box a set of pins set in the configuration

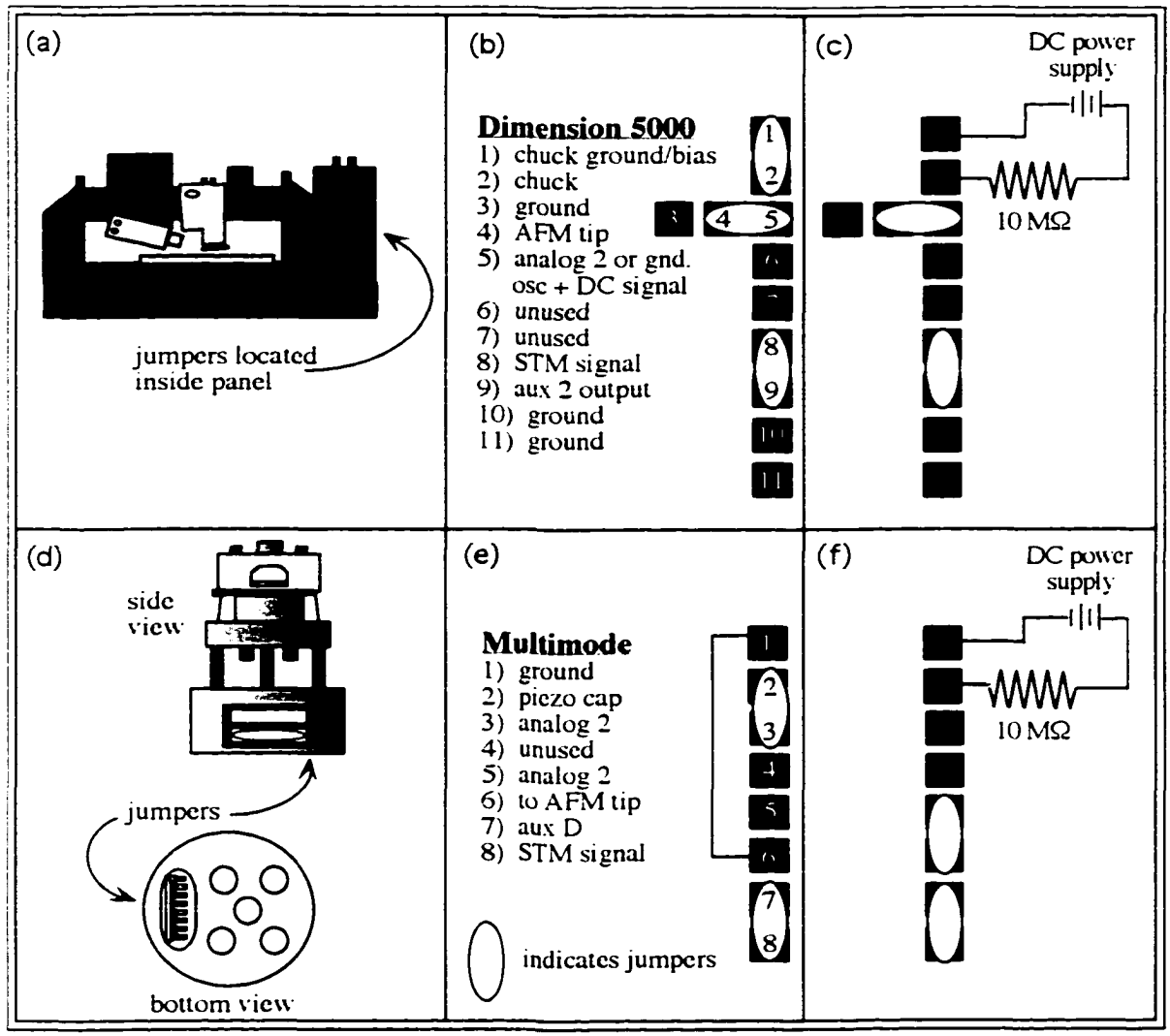


Figure C1: Diagrams indicating the (a) location of jumper pins for applying biases with in a Dimension™ 5000 SPM, (b) Jumper connections for applying an internal bias through the computer on a Dimension™ 5000 SPM, (c) Jumper connections for applying an external bias on a Dimension™ 5000 SPM (d) location of the jumper pins in a Multimode™ SPM, (e) Jumper connections for applying an internal bias through the computer on a Multimode™ SPM, and (f)) Jumper connections for applying an external bias on a Multimode™ SPM.

shown in Figure C1(b) and (c) can be found. The white ovals in these drawings represent wires or shorting caps which are used to electrically connect the pins. In Figure C1(b) the chuck ground (pin 1) is connected to the sample chuck (pin 2) so that the sample will be at ground. Analog 2 (pin 5) is connected to the tip (pin 4). This results in the bias voltage value entered into the analog 2 window in the Nanoscope software to be applied to the tip. Figure C1(c) shows how the jumper pins should be connected in order to apply an external D.C. bias. In this case the tip is connected to ground and a bias is applied to the sample stage by connecting an external power supply between pins 1 and 2. To avoid high currents between the sample and cantilever in case of point contact, a 10 MΩ resistor is put in series between the power supply and sample.

Setting the jumper pins in a Digital Instruments Multimode™ SPM is similar to setting up a Dimension 5000. Figure C1(d) indicates the location of the jumpers, which are located on the bottom of the Multimode's base. For internal biasing the analog 2 (pin 3) bias is applied to the sample chuck (pin 2) and the AFM tip (pin 6) is grounded (pin 1). External biasing is performed by connecting a power supply between ground (pin 1) and the sample stage (pin 2) with a 10 MΩ resistor in line, as was done with the Dimension™ 5000. The AFM tip (pin 6) is connected to the analog 2 pin which is set to ground potential by the computer. Note that in all cases, except for the internal biasing of the Dimension™ 5000, the bias is applied to the sample.

Several parameters in the Nanoscope software must be setup properly for the AFM to operate properly and apply biases and measure phase angles while in lift mode. First of all the AFM should be set up for tapping mode. To apply biases internally the analog 2 feature in the software must be enabled. This is done by switching the analog 2

control from "attenuation switch" to "user defined." In Nanoscope software versions 4.32 and later this is done by selecting "microscope select" under the "DI" menu bar. When the microscope select window comes up select "edit" then "advanced." In the advanced window the analog 2 option is then switched to used defined. On the microscope's extender (phase) box the switch on the left must be set to analog 2 for the computer defined analog 2 bias voltage to reach the AFM. The switch on the right is always set to the normal operation position.

Interleave lift mode is set up in the software by selecting "interleave" under the "panels" menu bar. The "interleave mode" selection should be set to "lift." "Lift start height" is used to set a height to pull the tip away from the sample. It is intended to pull the tip far enough from the sample surface to break it free from the influence of the water layer. The "lift scan height" is then used to set the tip-sample separation distance for the remainder of the lifted scan. If applying the bias via the computer the analog 2 selection should appear in the interleave control window. If the circle in front of the interleave control box's "analog 2" selection appears gray, the bias entered in the main control box's analog 2 selection will be applied to both the main ant lifted scan. If, however, the circle in front of the analog 2 is filled in with green the bias in the interleave control analog 2 box will be applied to the lifted trace and the main control analog 2 selection will be applied to the main topographic trace. In all of the experiments in this dissertation in which the bias was applied via the computer, the interleave analog 2 control was operating independently of the main analog 2 control which was set to zero volts bias. When operating in interleave mode it is also essential to set the "line direction" for the topographic and lifted data channels to retrace. This ensures that the z-

piezo has reached the full tip sample separation height by the time the lift scan starts. As a final step to setting up the software the "scan line" option for the topographic image should be set to "main" and for the lifted image it should be set to "interleave."

Appendix D: Phase contrast and Absolute Phase

Phase contrast represents the contrast observed from comparing two regions in a phase image. In many of the figures presented in Chapter 4, phase contrast is plotted against either D.C. bias or the tip sample separation distance. These plots contain phase data taken from several images. A physical parameter is changed while acquiring each image. The phase contrast in the images is attained by digitally cross-sectioning the image through a specific feature. Figure D1 shows a cross section of such an image. The phase contrast is basically a measurement of the difference in height of one feature in the image with respect to another feature. Frequently, in Chapter 4, the phase angle of an ion-implanted test pattern (circle) is compared to the phase angle of the surrounding (substrate) region.

In flattened atomic force microscopy (AFM) images the background is removed. This produces a plot in which all of the data in the image is referenced with respect to the data point in the image that has the lowest value. In Digital Instruments data files the magnitude of the background signal, shown if Figure D1, can be determined from looking at the plane fit parameters in the file's header. The background (average phase angle) values are assigned by the computer upon flattening the image as "two sets of 16 bit numbers," which range from +32768 to – 32767. In the case of phase images this

range corresponds to phase angles between 0° and 180°. Knowing where the background signal falls in the computers A/D range (+32768 to – 32767) allows one to calculate the "absolute phase angle" of the background. When measuring dopant profiles with BAAFM the magnitude of the "absolute phase angle" represents the surface charge integrated over the surface of the entire imaged area. As the D.C. bias is increased, the surface charge increases over the entire imaged area. This results in an increase in the absolute phase contrast (background) signal.

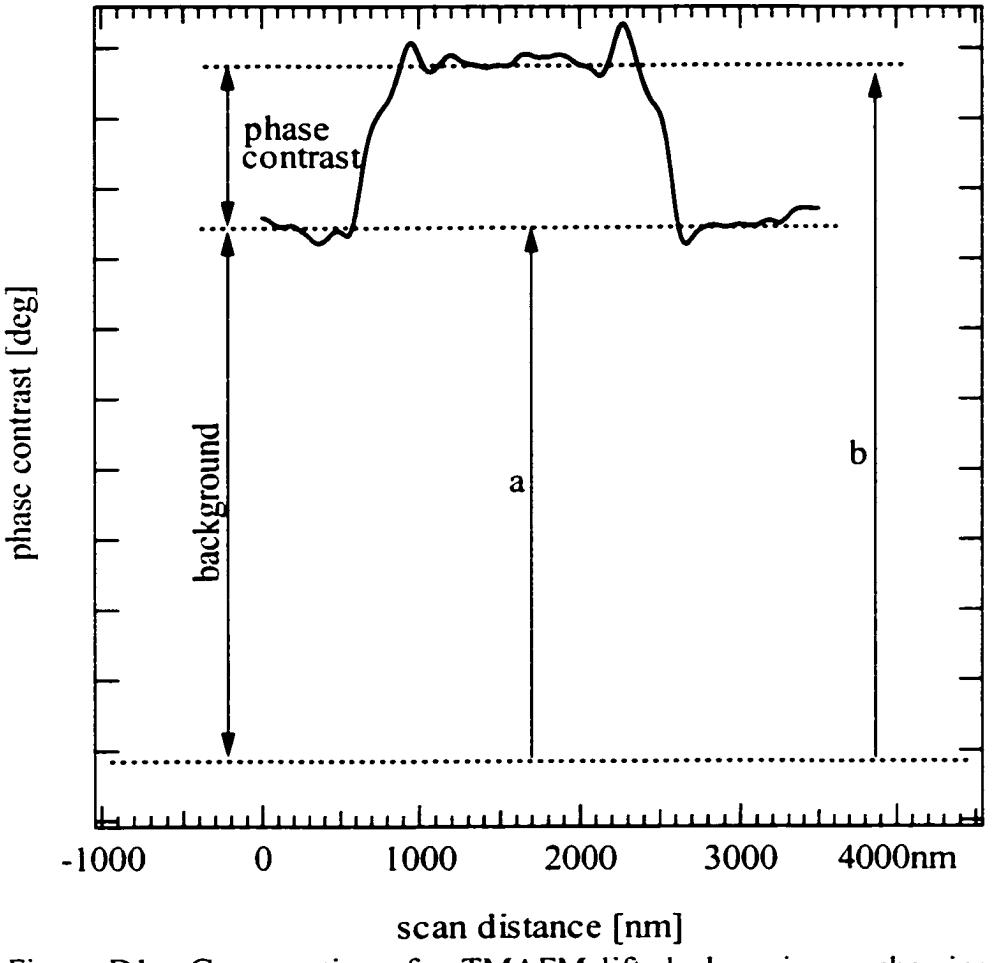


Figure D1: Cross-section of a TMAFM lifted phase image showing how phase contrast and absolute phase are measured. Phase contrast compares the contrast seen in the image with the background subtracted (phase contrast = b-a). Absolute contrast is attained by averaging the background phase signal over the entire sample.

Appendix E: Data Evaluation of Calibration Curves

Any force calibration curves (except the phase lag versus distance curves) taken with Nanoscope equipment is given in relative Z-piezo displacement units and in "twos complement 16 bit" data that correspond to the deflection of the cantilever. Relative Z-piezo displacement means that the absolute position of the cantilever is not given by the software but has to be determined from the curves themselves. Therefore, to convert these data into real deflection, force or amplitude values, the curves were calibrated as follows:

Standard force calibration curves were obtained from cantilever deflection versus A-piezo displacement curves by transforming the original 16 bit deflection data into true deflection values in nanometers. This was accomplished by multiplying the spectra with a constant that yielded a slope of 1 for the contact region of the curves (the left part of the curve where the cantilever is in contact with the sample surface, where its deflection directly equals the cantilever displacement). The obtained deflection values were then multiplied by the force constant of the cantilever resulting in true force values. The spring constant of 4.5 N/m, used in the force calculations, represents the average of the minimum and maximum values stated by the manufacturer (2-7 N/m). The x axis of these curves was calibrated to obtain true tip-sample distances by determining the intersection of the non-contact region of the zero bias curve (Right part of the curve

which is nearly horizontal; the weak wavy superimposed shape is an artifact which is always present in these curves measured on both our Nanoscope III setups) with the contact region of the curve. This intersection point was also defined as the origin of the force scale since it represents the point where the cantilever would have touched the sample surface had in not been attracted by capillary/Coulomb forces before it touched. The origin of the amplitude graphs was determined similarly by using the point where the contact region (left part of the graphs) bends off into a horizontal line corresponding to complete suppression of the vibration due to direct contact with the sample surface. The amplitude scale was calibrated by multiplying the curves with a constant so that the contact region would exhibit a slope of 2. This value was chosen to account for the fact that amplitudes correspond to the distance between both elongation maxima. This only approximates the real measurement conditions, since the driving force of the cantilever will accelerate the cantilever after being dampened due to surface contact during its upward motion away from the surface. Therefore, the true slope can be expected to be slightly different from 2. Phase calibration curves were measured simultaneously with the amplitude curves. Therefore, the x-axis calibration of the amplitude curve was used for the phase graphs. The phase scale shows the values as given by the Nanoscope software.

Appendix F: Hydrodesulfurization Catalysis Research Proposal

Observation of Individual Catalytic Hydrodesulfurization Reactions with a Variable Temperature, Ultrahigh Vacuum Scanning Tunneling Microscope

A Proposal Submitted as a Degree Requirement for Chemistry Doctoral Work at Colorado State University

Mark W. Nelson
Department of Chemistry
Colorado State University
Fort Collins, CO 80523

Abstract

Hydrodesulfurization (HDS) involves the catalytic removal of organic sulfur from petroleum feedstocks. Current environmental regulations require a substantial decrease in the sulfur content of raw petroleum during processing. Additionally, sulfur in raw petroleum poisons catalysts used in many steps of the refinement process, thus adding substantial costs to the finished products. Removal of a majority of the sulfur from feedstocks during early stages of the refinement process is crucial for petroleum processing. Improving HDS efficiency is becoming increasingly important as petroleum feedstocks with a low sulfur content are depleted, forcing the use of feedstocks containing high levels of sulfur. Understanding the mechanism of catalysis and the nature of the binding sites for the HDS process is essential for designing catalysts with superior efficiency and resistance to poisoning. We propose to create and study individual catalytic sites on the surface of MoS_2 (the most common industrial HDS catalyst) by the selective removal of sulfur surface atoms via the application of voltage pulses from the tip of a scanning tunneling microscope (STM). These potential catalytic sites will then be reacted with model compounds; their reactivity as a function of time, exposure to H_2 and temperature will also be probed. New insights into the mechanism of the important hydrodesulfurization reaction will be obtained.

Background

As low sulfur content petroleum reserves are depleted, feedstocks high in sulfur will be required to meet future energy demands. Removal of sulfur is currently required by environmental regulations, which attempt to decrease SO_x emissions from petroleum combustion. Additional problems arise from organic sulfur in petroleum feedstocks, which readily poison catalysts used in the petroleum refinement process. The high cost associated with catalyst replacement makes it desirable to remove as much sulfur as possible early on in the refinement process. Hydrodesulfurization (HDS) is the catalytic removal of organic sulfur from petroleum and coal feedstocks. The sulfur may be in the form of aromatic ring structures, as thiols, or thioethers. A generic reaction for the process can be written as follows:

$$C_xH_yS + nH_2 \text{ -----> hydrocarbons} + H_2S \qquad (F1)$$

The most common industrial catalyst used for HDS is supported MoS_2 or WS_2, with Co or Ni added as an activity promoter. The exact mechanism of catalysis is still a matter of much debate. Several recent review articles have summarized the current thinking on the possible mechanisms for this important reaction [1, 2, 3, 4, 5]. HDS is known to occur on many transition metals; in the case of MoS_2, the active site is known to be an exposed Mo atom. It is unknown, however, if the catalytic metal site exists at a Van der Waals edge site or at a missing sulfur defect on the basal plane of the MoS_2. HDS catalysis may occur at both basal vacancies or edge sites; however, reactivities may differ greatly. Insights into the relative activities of these two types of sites would aid in improving the design of HDS catalysts.

The first models of solid state HDS catalysis assumed that the active site was an exposed Mo edge site [6, 7]. In this work sites on the basal planes were considered

inactive since basal plane sulfur anions are more strongly bonded than sulfur at edge sites, making creation of a basal plane vacancy more difficult. Exposure of the basal plane of single crystal MoS_2 to sulfur containing reactants in a vacuum showed no reaction, suggesting that edge sites play the major role in HDS chemistry [8]. Active basal plane sites, however, are not likely to be created on MoS_2 at low pressures. Since, single crystal MoS_2 with large areas of well-defined edge sites is almost impossible make, HDS activity at edges could not be confirmed.

More recently, there has been support suggesting that basal plane vacancies may be present in large numbers at high hydrogen pressures typically used in the industrial process. The basal plane activity was initially tested by sputtering a single crystal MoS_2 sample with Ar^+ in a vacuum to create exposed metal sites of various size on the basal plane. The sputtered basal plane showed greatly increased activity over that of the unsputtered surface, indicating that basal plain defects may be active toward HDS [9]. Sputtering produces defect sites of various size and shape. This raises questions as to which types of basal plane surface vacancies are catalytically active: large ones, small ones, or defects of a specific shape.

Recently there has been evidence that HDS active edge sites can be produced on the basal plane as etch pits one van der Waals layer thick. In an STM investigation of the gasification of MoS_2, Chu and Schmidt saw the surface of single crystal MoS_2 etch under positive hydrogen pressures and elevated temperatures (conditions similar to those used in industrial HDS) [10]. Under hydrogen partial pressures between 10 and 100 torr, and at temperatures of 450°C, they saw triangular etch pits one van der Waals layer thick and about 500nm to a side, form on the basal plane of MoS_2. At higher temperature (600°C) and hydrogen pressure these pits grew larger (3 um on a side) and started to coalesce to

give a large area of exposed edge sites. Addition of thoiphene to the gas mixture resulted in a slowing down of the etching process. Chu and Schmidt suggest that thiophene is most likely reacting at edge sites where it deposits its sulfur, thereby counteracting the etching process. This study shows that the active sites are continually changing during the catalytic process.

Promoters such as Ni and Co may also increase activity by increasing the number of nucleation sites for etch pit formation, thus drastically increasing the area of edge sites [5.10]. Chu and Schmidt also looked at the gasification process of MoS_2 containing Ni and Co. In the Ni containing samples the size of etch pits was greatly enhanced while the Co samples produced small and very dense pits compared to those observed the pure MoS_2. Adjustment of the edge site to rim ratio of the pits is attributed to changing the HDS to dehydrogenation (HDY) reaction rates with respect to one another [11]. Since size, depth, and density of surface features seems to greatly alter the activity of the catalyst, having the ability to look at incorporation of reactants into controlled size and shape basal plane vacancies and edge sites is important to improving the understanding of HDS.

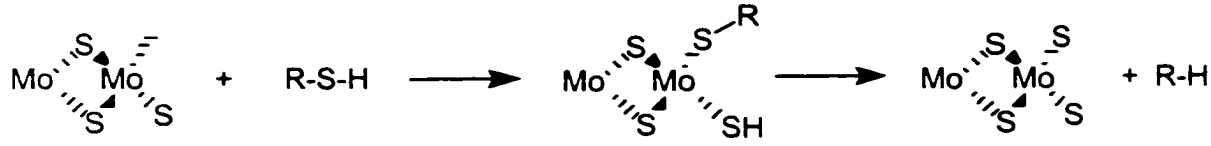
With the advent of scanning probe techniques new insight into events occurring at the scale of individual molecules can be obtained. Several papers have described the study of catalytic reactions on a molecular level using scanning tunneling microscopy (STM) [11, 12]. These studies have been limited by the use of metal single crystals, which have a multitude of catalytic sites, and so the individual catalytic events were not discernible. In this proposal we will create and study the reactivity of individual catalytic sites, which are active for the hydrodesulfurization of thiol molecules.

We choose to focus on a single step in the HDS reaction. Many of the catalytic studies using reactors have used thiophene, benzothiophene or dibenzothiophene as

models for aromatic sulfur containing species, which may be found in high sulfur content feedstocks. Reactions with these model compounds consist of multiple steps where first the aromatic ring is hydrogenated followed by hydrogenolysis of an aliphatic C-S bond. Since observation of individual catalytic events is a frontier experiment we propose to focus on this final, but crucial, step in the HDS reaction as indicated in the following reaction:

$$R\text{-}S\text{-}H + H_2 \text{ ----> } R\text{-}H + H_2S \qquad (F2)$$

The hydrogenolysis reaction may also proceed in a stoichiometric manner without the release of H_2S, resulting in a filled sulfur vacancy:



We intend to examine which of these reactions occurs on carefully created individual catalytic sites on the MoS_2 surface.

Creation of Individual Catalytic Sites

Workers at Hitachi have reported removal of individual sulfur atoms from basal plane surfaces of MoS_2 by applying voltage pulses to the tip of a STM operating in ultra high vacuum (UHV) [13]. Precise control over sulfur atom removal allowed for the message "peace '91 HCRL" to be written by atom removal on a MoS_2 surface, as shown in Figure F1. Unlike the earlier work of Eigler, who wrote "IBM" in individual Xe atoms on a Ni surface, the atomic scale letters produced by the Hitachi group are indefinitely stable at room temperature. Our research group has experience producing controlled size and shape structures on MoS_2 and other 2D metal chalcogenides [14, 15]. We also have a

long history of preparation and growth of single crystals of these 2D metal chalcogenide compounds.

As previously discussed, some HDS mechanisms implicate basal plane vacancies as the active sites for HDS since these defects result in coordinately unsaturated metal atoms which can be the organic sulfur binding site. To prepare individual catalytic sites, we plan to remove sulfur atoms from the basal plane of MoS_2 or other layered dichalcogenides, which will expose coordinately unsaturated metal atoms on the layer below. Atom removal will be accomplished with STM voltage pulses as was previously described, which will allow controlled production of reaction sites of virtually every size and shape. Atoms will be removed individually, or in groups where a specific geometry of the exposed metal sites can be prepared. Three fold coordinately unsaturated Mo atoms are attributed to be active sites for HDY, and may be involved in HDS as well [16]. Figure F2 shows the possible geometries for removing three sulfur atoms from the MoS_2 surface, which result in different arrangements of the underlying exposed metal. In one geometry (Figure F2(a)) the three sulfur atoms above a single Mo atom have been removed, allowing for possible adsorption of the reactant's sulfur to the d_{z2} orbital of the centrally exposed Mo atom. During the adsorption process the reactant eliminates hydrogen which should associate with a nearby unsaturated sulfur atom at the edge of the vacancy forming an S-H termination. In the case of the other threefold vacancy (Figure F2(b)), the thiol's sulfur atom would be bound in a threefold Mo site with the H atom again bound as a terminal SH moiety. There may be significant reactivity differences between these two types of threefold sites. We intend to investigate differences in threefold site reactivity as well as reactivity of larger vacancy sites.

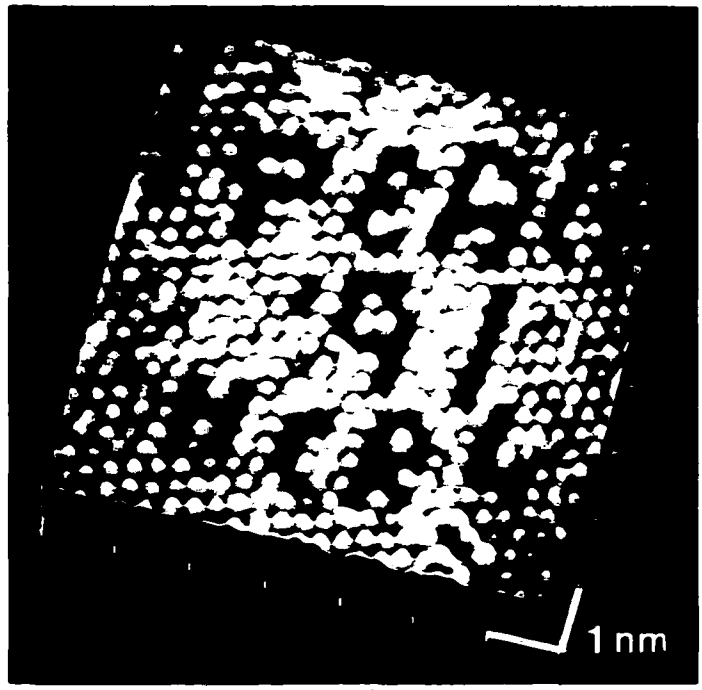


Figure F1: STM image of MoS_2 in which Hitachi researchers removed individual sulfur atoms from the surface by applying voltage pulses with an UHV-STM. A nanostructured message "Peace '91 HCRL" has been written as a result of the atom removal.

After the catalytic site is prepared via STM, the vacuum chamber will be dosed with volatile thiol molecules using a leak valve attached to the UHV-STM chamber. We may require use of the low temperature facility of our variable temperature STM (VT-STM) to cool the substrate, thus facilitating adsorption of the thiol molecules to the catalyst surface. After initial adsorption thiol molecule migration will be monitored by the STM, and individual thiol molecules can then be imaged as they migrate to, and chemically adsorb into, the holes left by sulfur atom removal. We will first try 1-pentanethiol since it has a combination of high vapor pressure at room temperature and a long enough hydrocarbon chain to make it obvious in the STM image. Once a thiol is bound to the STM created catalytic site, small doses of H_2 will be admitted to the chamber. Two possible routes for the hydrogenolysis reaction are then possible.

Hydrogen can dissociatively adsorb to dangling sulfur bonds created by removal of a sulfur atom from the surface, a mechanism favored by some HDS models [5,17]. Alternatively the H_2 can dissociate on the STM tip if a Pt or Pt-Ir tip is used. Hydrogenation of carbonaceous deposits do to tip hydrogenation has been previously reported [12]. Either way the activated hydrogen will then be available to continue the HDS of the adsorbed thiol, resulting in the desorption of alkane and H_2S, or the plugging of the catalytic site by the sulfur left behind when the alkane desorbs (as was summarized in the reaction schemes above). The true mechanism will be easily determined from the STM images.

The reaction's hydrogen source can be further investigated by switching to a W tip which does not adsorb hydrogen dissociatively. If no reaction is observed, as was the case with the previously reported "nanocatalysis", using a Au tip, then we can surmise that the H atoms from the Pt or Pt-Ir tip were the source of the necessary hydrogen in the reaction. It may be that the activation barrier required for the catalytic reaction will not be surmounted at the low temperatures where thiol adsorption is done. The temperature of the substrate can then be increased gradually and the temperature required for the reaction to occur can then be observed. Alternatively, the temperature can be kept constant at a point where the reaction rate is in a convenient time scale for the STM (tens of seconds to minutes) and significant number of individual catalytic or stoichiometric reactions can be observed, then a good idea of the activation barrier for the observed reactions can be obtained. Reaction the reactions rate constant can be measured for the individual events. If a statistically statistics will be accessed by creation of an array of identical reaction sites and monitoring them simultaneously with the STM while the reaction proceeds.

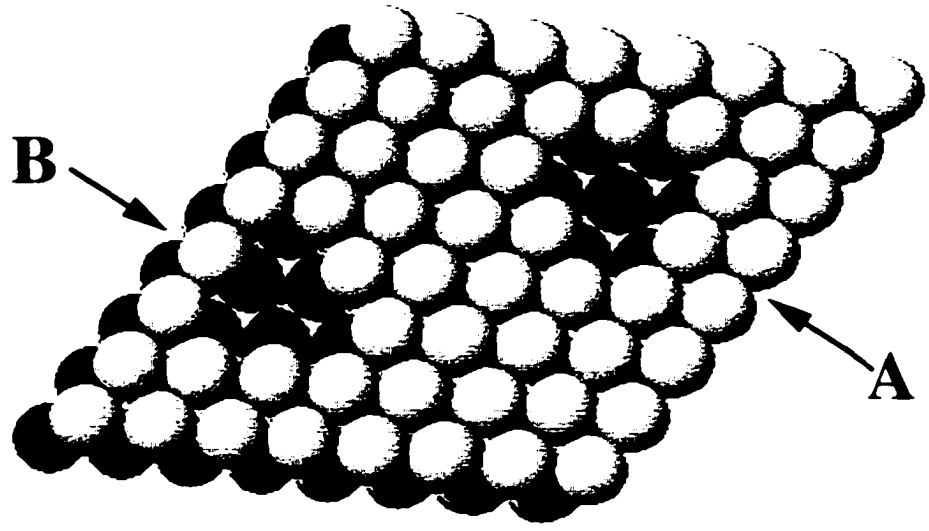


Figure F2: Possible types of threefold sulfur vacancies, which may be created in the surface of MoS_2.

Edge planes have also been implicated as active sites for HDS catalysis. HDS studies at edges are not widely reported since it is nearly impossible to create single crystal MoS_2 with a sizable and well defined edge surface. We can, however, prepare oriented edge planes on a very small scale by creating small defect sites on the MoS_2 surface using the Hitachi technique mentioned above. Rastering a STM tip over these defect sites at high bias results in layer-by-layer removal of material around the defect [13]. We have seen this etching phenomenon in STM experiments, under ambient conditions, in which one Van der Waals layer thick etch pits are created within the basal plane. We have also seen evidence for this occurring under ultra high vacuum. Controlled size triangular surface pits are created with well-defined edges (Figure F3), which look similar to the pits mentioned in the above Chu and Schmidt experiment. These edge sites are high in surface free energy providing a place where adsorbates collect. Exposed metals at the edge site thus have a high probability of interacting with

an adsorbed reactant's sulfur, making edges attractive sites to study for their HDS reactivity. We will investigate the edge reactivity the same way as the basal plane will be studied. First we will create etch pits of a controlled size. This will be followed by dosing the surface with a thiol and observation of the reaction as well molecular orientation relative to the edge using the variable temperature STM.

Additional Studies

Once we can show proficiency at observing single catalytic events using the simplified alkane thiol model, work should be extended to many of the more common model reactants such as thiophene, benzothiophene, and dibenzo-thiophene. These systems will be more difficult to study however since they involve many more steps than the alkane thiols. Debate exists over the size of the binding site required for catalysis of these large molecules. Questions also arise at to whether the binding of the reactant thiol occurs through one or more bonds with the surface. Varying the size and shape of the surface vacancies with the STM will allow us to look at how these more bulky molecules fit into different types of active sites. We will also look at other substrates. WSe_2 crystals are available in our laboratory and should be amenable to Se atom removal in a similar fashion to S atom removal from MoS_2. The removal of Se atoms from WSe_2 will result in larger "holes" (lattice constant for WSe_2 a = 3.28 Å while for MoS_2 a = 3.16 Å) with more space for substrate molecules to bind. $ReSe_2$ has a more complex, lower-symmetry layered structure than the Mo and W dichalcogenides and has been demonstrated to have a higher catalytic activity for HDS than unpromoted MoS_2. The four different Se surface sites and the four different Re sites would provide a slightly different chemical environments for our studies.

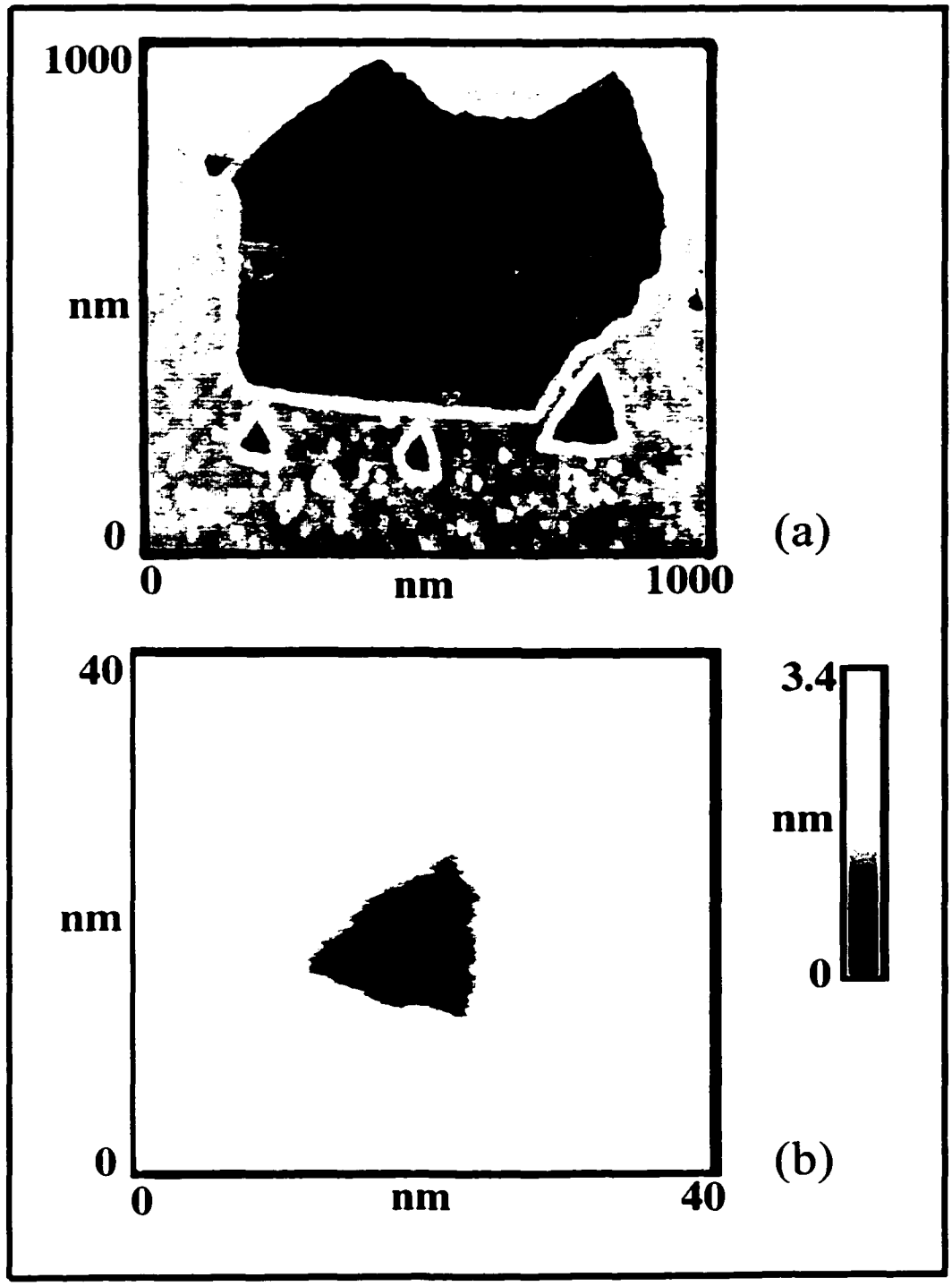


Figure F3: STM images of triangular etch pits formed by rastering a STM tip over the scan area. (a) Large etch pits formed in two different Van der Waals layers. Pits in the top layer have coalesced into a large pit. (b) Etch pit at early stages of formation.

Industrial HDS catalysts contain Co and or Ni as activity promoters. The actual role of these promoters in the catalysis is unknown. As observed by Chu and Schmidt the promoted activity may result from the ability of Ni and Co to act as nucleation sites forthe formation of layer thick etch pits on the basal plane of the catalyst [10]. Our research group has shown that it is possible to spatially locate impurity atoms in the first couple of surface layers of 2D metal chalcogenides such as MoS_2 by STM imaging at slightly increased biases [18]. Combining the ability to locate impurities with etching capabilities of the STM will allow us to monitor etch rates in the vicinity of a promoter atom. Formation of etch pits primarily in the immediate vicinity of the promoter would point to the nucleation site mechanism for the Ni and Co promotion effect.

Available Facilities

The Fort Collins flood on July 28th, 1997 submerged our ultra high vacuum (UHV) laboratory, destroying all of our UHV surface analysis capabilities. Fortunately we were able to replace this equipment with a new Omicron Multiprobe UHV surface analysis system. At the heart of this new system is a variable temperature STM which is capable of atomic resolution at temperatures between 50K and 1100K. A sapphire leak valve will allow delivery of controlled dosages of thiols to the STM sample stage, making it possible to introduce reactants to the active sites created by the STM. Varying the microscope's temperature will allow us to control reaction rates so the reactants may be monitored on the time scale of STM scanning. The integrated Omicron system allows UHV sample transfer between a number of different techniques including photoelectron spectroscopies (XPS and UPS) which allow the determination of the atomic composition as well as chemical and electronic state of the surface region of the substrates. Secondary ion mass spectroscopy (SIMS) will allow determination of the concentration of

impurities in our substrates. SIMS will be necessary for examining the Ni and Co concentrations when they are used for promotion of the HDS catalysis. The chamber also includes an isolated molecular beam deposition chamber for the controlled deposition of metals and organics onto substrates under ultra clean conditions.

References

1. Chianeli, R.R., "Fundametal studies of transition metal sulfide hydrodesulfurization catalysts", *Catal. Rev. - Sci. Eng.* , **26**, 361 (1984).

2. Prins, R., Beer, V. H. J. D. and Somorjai, G. A., "Structure and function of the catalyst and the promoter in Co-Mo hydrodesufurization catalysts", *Catal. Rev. -Sci. Eng.* , **31**, 1-41 (1989).

3. Startsev, A., "The mechanism of HDS catalysis", *Catal. Rev. -Sci. Eng.* , **37**, 353 (1995).

4. Topsøe, H. and Clausen, B. S., "Importance of Co-Mo-S type structures in hydrodesulfurization", *Catal. Rev.* , **26**, 395 (1984).

5. Delmon, B., "Selectivity in HDS, HDN, HDO and hydrocracking combustion of remote control and other new concepts", *Bull. Soc. Chim. Belg* , **104**, 173 (1995).

6. Farragher, A. L., "Surface vacancies in close-packed crystal structures", *Adv. Colloid Interface Sci.* , **11**, 3 (1979).

7. Voorhoeve, R. J. H. and Stuiver, J. C. M., "Kinetics of Hydrogenation on Supported and Bulk Nickel-Tungsten Sulfide Catalysis", *J. Catal.* , **23**, 228 (1971).

8. Salmeron, M., Somorjai, G. A., Wold, A., Chianelli, R. and Liang, K. S., "The adsorption and binding of thiophene, butene and H_2S on the basal plane of MoS_2 single crystals", *Chem. Phys. Rev.* , **90**, 105 (1982).

9. Farias, M. H., Gellman, A. J., Somorjai, G. A., Chianelli, R. R. and Liang, K. S., "The CO coadsorption and reactions of sulfur, hydrogen, and oxygen on clean and sulfided Mo(100) and on MoS/sub 2/(0001) crystal faces", *Surface Science* , **140**, 181 (1984).

10. Chu, X. and Schmidt, L. D., "Processes in MoS_2 gasification", *J. Catal.* , **144**, 77 (1993).

11. Cremer, P. S., McIntyre, B. J., Salmeron, M., Shen, Y.-R. and Somorjai, G. A., "Monitoring surfaces on the molecular level during catalytic reactions at high pressures by sum frequency generation vibrational spectroscopy and scanning tunneling microscopy", *Catal. Lett.* , **34**, 11 (1995).

12. McIntyre, B. J., Salmeron, M. and Somorjai, G. A., "Nanocatalysis by the tip of a scanning tunneling microscope operating inside a reactor cell", *Science* , **265**, 1415 (1994).

13. Hosoki, S., Hosaka, S. and Hasegawa, T., "Surface modification of MoS_2 using an STM", *Appl. Surf. Sci.* , **60/61**, 643 (1992).

14. Parkinson, B. A., "Layer-by-layer nanometer scale etching of two-dimensional substrates using the scanning tunneling microscope", *J. Am. Chem. Soc.* , **112**, 7498 (1990).

15. Delawski, E. and Parkinson, B. A., "Layer-by-layer etching of two-dimentional metal chalcogenides with the atomic force microscope", *J. Am. Chem. Soc.* , **114**, 1661 (1992).

16. Kasztelan, S., Toulhoat, H., Grimblot, J., and Bonnelle, J.P., "A geometrical model of hydrotreating catalysis - prediction of catalytic activity variations with composition", *Bull. Soc. Chim. Belg.*, **93**, 831, (1984).

17. Sundberg, P., Moyes, R. B. and Tomkinson, J., "Inelastic neutron scattering spectrocsopy of hydrogen adsorbed on powdered-MoS_2, MoS_2-alumina, and nickel-promoted MoS_2", *Bull. Soc. Chim. Belg.* , **100**, 967 (1991).

18. Schlaf, R., Lauder, D., Nelson, M. W. and Parkinson, B. A., "Influence of electrostatic forces on the investigation of dopant atoms in layered semiconductors by scanning tunneling microscopy/spectroscopy and atomic force microscopy", *J. Vac. Sci. Technol. A* , **15**, 1466 (1997).