LEABHARLANN CHOLÁISTE NA TRÍONÓIDE, BAILE ÁTHA CLIATH | TRINITY COLLEGE LIBRARY DUBLIN
Ollscoil Átha Cliath | The University of Dublin

Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Trinity College Dublin

Combined Scanning Tunneling Microscopy (STM) & Surface Stress Measurements (SSM) of Chemical Reactions on Semiconductor Surfaces – Oxidation of the Si(111)–(7 × 7) Surface

A thesis presented to the University of Dublin,
Trinity College for the degree of Doctor of Philosophy

David E. Meehan

Under the supervision of
Prof. John J. Boland
June 1st, 2012

Thesis 9677

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Dedicated to my parents, who made my education possible.

Abstract

A novel combined scanning tunneling microscopy (STM) and surface stress measurement (SSM) system has been developed to allow absorbate-induced changes in surface stress to be measured and related to the structural and electronic changes causing them *in situ* and with atomic resolution. Here, the system is used to investigate the oxidation of the Si(111)–(7×7) surface at room temperature. The site-specific evolution of surface stress is measured and associated with relief of the intrinsic tensile stress of the (7×7) unit cell. It is shown that stress relief is greatest in the faulted half of the unit cell, consistent with the known larger tensile stress of that half. It is further shown that metastable species proposed to exist on the surface at room temperature do not have significant stress signatures. A widely accepted reaction model is also extended to account for the known preference of oxygen for reacting with the faulted half of the unit cell. Finally, the validity of three reaction mechanisms proposed in the literature is assessed in terms of the above findings. The implications for the room-temperature reaction mechanism are discussed.

Acknowledgements

In the five years that have passed since I first started my Ph.D., I have encountered many people to whom I owe no small debt of gratitude. First among these is my research supervisor, Prof. John Boland, without whose constant support, encouragement and patience the work described in this thesis would not have been possible. I am indebted to him for affording me the great privilege of working under his supervision. I could not have asked for a better supervisor.

Few battles are fought alone, and the same can be said of a Ph.D. In that regard, I owe particular thanks to my former colleague and close friend, Dr. Niall Kinahan. Without his knowledge, patience and determination, the work described in this thesis would not have been possible, nor would my Ph.D. have been as much fun or as rewarding an experience as it was. It was a privilege to work with him, and it remains a privilege to call him my friend. I would also like to thank Dr. Tetsuya Narushima and Dr. Steffen Sachert for their enormous contributions to the work. It could not have been done without their efforts.

I would like to thank the members of the Boland research group, both past and present, for their support, advice and friendship over the years. You made work fun, enjoyable and something to look forward to each day. I wish you the very best in your respective futures. To the many incredible friends that I have made along the way – you know who you are – I cannot begin to express my gratitude. You have gifted me with, and continue to gift me with, some of the best years of my life. I will forever be in your debt.

Lastly, I would like to thank my parents, Claire and Denis, to whom I have dedicated this thesis. Without their efforts, my education simply would not have been possible and I would not have the many wonderful opportunities that I have today. Their constant support, endless patience and innumerable sacrifices have made these opportunities possible. I am and will always be especially grateful to them.

The work presented in this thesis was made possible by the generous financial support of Science Foundation Ireland under Grant No. 06/IN.1/I106. It is gratefully acknowledged.

Contents

Declaration i

Dedication iii

Abstract v

Acknowledgements vii

Introduction 1

References 2

1 Scanning Tunneling Microscopy 3

1.1 Theory 4

1.1.1 Elastic tunneling through a one-dimensional potential energy barrier of rectangular shape 4

1.1.2 Elastic tunneling through a one-dimensional potential energy barrier of arbitrary shape 6

1.1.3 Elastic tunneling in planar metal-insulator-metal (MIM) tunnel junctions 6

1.2 Instrumentation 10

1.2.1 Coarse positioner 11

1.2.2 Scanner 12

1.2.3 Control electronics 13

1.2.4 Vibration isolation 14

1.2.5 Tip preparation 16

1.3 Operation 18

1.3.1 Constant-current mode 18

References 20

2 Surface Stress 22

2.1 Definition 23

2.2 Origins 23

2.2.1 Intrinsic surface stress 23

2.2.2 Adsorbate-induced surface stress 26

2.3 Measurement 28
2.3.1 The Stoney Equation 28
References 33

3 The Si(111)–(7 × 7) Surface **35**
3.1 Physical structure 36
3.2 Electronic properties 38
3.3 Mechanical properties 38
3.4 Chemical properties 42
References 43

4 Development of a Combined Scanning Tunneling Microscopy and Surface Stress Measurement System **45**
4.1 Design 47
4.1.1 Sample holder 47
4.1.2 Surface stress measurement 50
4.1.3 Sample heater 52
4.1.4 Sample pusher 55
4.2 Operational considerations 56
4.2.1 Parasitic capacitance 57
4.2.2 Sample oscillation 71
References 74

5 The Oxidation of the Si(111)–(7 × 7) Surface at Room Temperature **75**
5.1 Background 76
5.2 Experimental 78
5.3 Results & Discussion 79
5.4 Conclusions 83
References 85

Conclusions **87**
References 88

Introduction

The growth of silicon dioxide is a critical step in the production of virtually all silicon-based electronic devices. As the size of these devices approaches the atomic scale, the use of ultra-thin oxide films will become integral to sustaining device performance and reliability. In order to grow these films, a deeper understanding of the chemistry underlying the oxidation process is required. Substantial progress has been made in this regard, most notably in relation to the oxidation of Si(100) and Si(111) surfaces [1–8]. In spite of this, however, many questions remain unanswered. While numerous attempts have been made to address these questions, few [9, 10] have attempted to do so in the context of *surface stress*. Surface stress refers to the mechanical forces acting at a surface due to the configuration of the chemical bonds holding it together. The interaction of adsorbates with a surface can induce changes in surface stress. In the case of silicon surfaces, these changes can be explained in terms of the chemical nature of the adsorbate, the bonding topology of the surface, and atomic size mismatch between the surface and the adsorbate [11]. They can also be measured. This enables the observation of mechanical forces acting at a surface during a chemical reaction. Since these forces arise due to changes in chemical bonding, their observation provides useful information about the underlying reaction mechanism. When considered alone, however, the utility of this information is limited. Surface stress is a *macroscopic* quantity, while its origins lie in structural and electronic changes occurring on the *atomic* scale. In order to maximize its value as a probe of chemical reactivity, it must be directly relatable to these changes. To this end, a combined scanning tunneling microscopy (STM) and surface stress measurement (SSM) system has been developed [12]. This system allows adsorbate-induced changes in surface stress to be measured and related to the structural and electronic changes causing them *in situ* and with atomic resolution. Here, it is used to investigate the oxidation of the Si(111)–(7×7) surface at room temperature. This reaction was selected because of the long-standing controversy surrounding its mechanism, and because of the well-known properties of the Si(111)–(7×7) surface. It provides an excellent demonstration of the capabilities and usefulness of the combined system. This thesis consists of five chapters. Chapters 1, 2 and 3 are reference chapters discussing STM, surface stress and the Si(111)–(7×7) surface respectively. Chapter 4 discusses the design and construction of the combined STM/SSM system. Finally, Chapter 5, building upon the preceeding chapters, discusses the application of the combined system to the oxidation of the Si(111)–(7×7) surface at room temperature.

References

[1] E. A. Irene, Crit. Rev. Solid State Mater. Sci. **14**, 175 (1988).

[2] T. Engel, Surf. Sci. Rep. **18**, 93 (1993).

[3] H. N. Waltenburg and J. T. Yates, Chem. Rev. **95**, 1589 (1995).

[4] R. J. Hamers and Y. Wang, Chem. Rev. **96**, 1261 (1996).

[5] A. Stesmans and B. Nouwen, Phys. Rev. B **61**, 16068 (2000).

[6] B. Nouwen and A. Stesmans, Mat. Sci. Eng. A **288**, 239 (2000).

[7] A. Stesmans and V. V. Afanas'ev, Appl. Surf. Sci. **168**, 324 (2000).

[8] W. Futako, T. Umeda, M. Nishizawa, T. Yasuda, J. Isoya, and S. Yamasaki, J. Non-Cryst. Solids **299-302**, 575 (2002).

[9] D. Sander and H. Ibach, Phys. Rev. B **43**, 4263 (1991).

[10] N. T. Kinahan, D. E. Meehan, T. Narushima, S. Sachert, K. Miki, and J. J. Boland, Phys. Rev. Lett. **104**, 146101 (2010).

[11] R. Meade and D. Vanderbilt, Phys. Rev. Lett. **63**, 1404 (1989).

[12] T. Narushima, N. T. Kinahan, and J. J. Boland, Rev. Sci. Instrum. **78**, 053903 (2007).

CHAPTER 1

Scanning Tunneling Microscopy

Scanning tunneling microscopy is a microscopy technique that uses *electron tunneling* to image electrically conductive surfaces in real space with atomic resolution. The basic operation of a scanning tunneling microscope (STM) involves positioning an atomically sharp metal tip within several Ångströms of a surface and scanning the tip across the surface while a small potential difference is applied between the two. This is accomplished using piezoelectric actuators that allow precision movement of the tip in three dimensions (x, y and z) with a lateral (x–y) resolution of 1 Å and a vertical (z) resolution of 0.01 Å [1]. When the wave functions of the foremost tip atom and the nearest underlying surface atom overlap, the applied potential difference causes electrons to flow from one to the other. This process is referred to as electron tunneling, while the resulting flow of electrons is referred to as the *tunneling current*. The magnitude of the tunneling current is exponentially dependent on the distance between the tip and the surface, and also depends on their respective local density of states (LDOS). The LDOS sampled depends on the sign and magnitude of the energy window defined by the product of the elementary charge, e, and the applied potential difference, $\pm V$. There are two imaging modes. In *constant-current mode*, the tip is scanned across the surface while the tunneling current is held constant by a feedback loop. At the same time, the vertical position of the tip is recorded as a function of lateral tip position and translated into a topographic map of the surface. This mode is mainly used to image atomically rough surfaces where the risk of a tip-surface collision occurring is high. In *constant-height mode*, the tip is scanned across the surface while the vertical position of the tip is held constant by disabling the feedback loop. At the same time, the tunneling current is recorded as a function of lateral tip position and translated into a charge density map of the surface. This mode is mainly used to image atomically smooth surfaces where the risk of a tip-surface collision occurring is low. Scanning tunneling spectroscopy (STS) can also be performed, allowing LDOS and other useful physical quantities to be measured with atomic resolution. Regardless of the imaging mode or spectroscopic technique used, it is important to note that all data measured represent a convolution of both topographic *and* electronic features – proper interpretation of this data is seldom trivial and always requires careful analysis. This chapter presents an overview of the theory, instrumentation and operation of a typical STM.

1.1 Theory

The operating principle of the STM is based upon electron tunneling, a concept arising from elementary quantum mechanics. This section provides an introductory-level theoretical description of electron tunneling and explains its role in the operation of the STM. For the sake of simplicity, only *elastic tunneling*, where the energy of tunneling electrons is conserved, will be considered. For a discussion of *inelastic tunneling*, where the energy of tunneling electrons changes due to interaction with other physical systems, several excellent sources are recommended [2, 3].

1.1.1 Elastic tunneling through a one-dimensional potential energy barrier of rectangular shape

Classical mechanics describes the interaction of an electron with a one-dimensional potential energy barrier of rectangular shape by the expression

$$E = \frac{p^2}{2m} + V(x) \tag{1.1}$$

where E is the energy of the electron, p is the momentum of the electron, m is the mass of the electron, and $V(x)$ is the height of the barrier at any position x. Here, the electron is treated as a particle. Rearranging Eq. (1.1), the momentum of the electron is given by

$$p = \sqrt{2m\left[E - V(x)\right]}. \tag{1.2}$$

Eq. (1.2) has real solutions in regions of the barrier where $E > V(x)$. These are referred to as *classically allowed* regions. The electron can exist in classically allowed regions. Eq. (1.2) has no real solutions in regions of the barrier where $E < V(x)$. These are referred to as *classically forbidden* regions. The electron cannot exist in classically forbidden regions. Quantum mechanics describes the same system by the expression

$$-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2}\psi(x) + \left[V(x) - E\right]\psi(x) = 0 \tag{1.3}$$

where $\hbar$ is the reduced Planck constant, m is the mass of the electron, $\psi(x)$ is the electron wave function at any position x, $V(x)$ is the barrier height at x, and E is the energy of the electron. Here, the electron is treated as a particle with wave properties. The solutions to this equation for a barrier of height V_0 and width s are

$$\psi(x) = \begin{cases} Ae^{-ikx} + Be^{ikx} & \text{for } x < 0 \\ Ce^{-\kappa x} + De^{\kappa x} & \text{for } 0 < x < s \\ Fe^{ikx} & \text{for } x > s \end{cases} \tag{1.4}$$

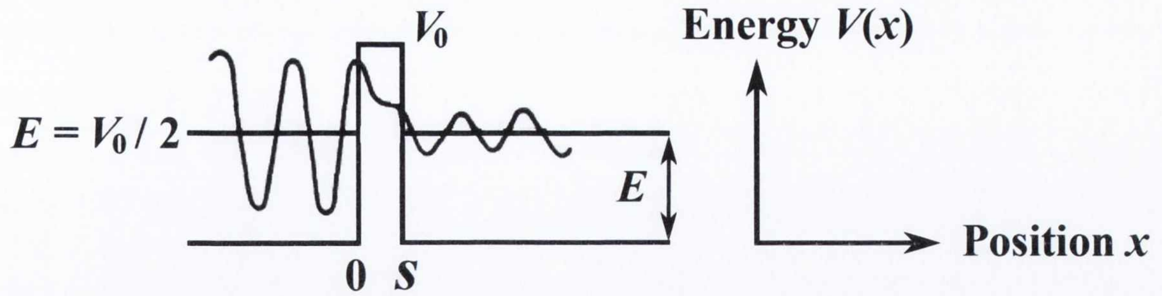


Fig. 1.1. Wave function $\psi(x)$ for an electron of energy $E = \frac{1}{2}V_0$ interacting with a potential energy barrier of height V_0 and width s. Left of the barrier ($x < 0$), the wave function is oscillatory. Within the classically forbidden region ($0 < x < s$), the wave function decays exponentially with a decay length of κ^{-1}. Right of the barrier ($x > s$), the wave function becomes oscillatory once again [4].

where the wave vector k and decay constant κ are given by

$$k = \frac{\sqrt{2mE}}{\hbar}, \qquad \kappa = \frac{\sqrt{2m[V_0 - E]}}{\hbar}.$$

Fig. 1.1 illustrates these solutions for an electron of energy $E = \frac{1}{2}V_0$. Classical mechanics predicts that the electron cannot exist in the barrier region defined by $0 < x < s$. Quantum mechanics, on the other hand, predicts that the electron *can* exist in this region, and further predicts that the electron wave function will decay exponentially across the region with a probability of appearing on the other side of the barrier. In doing so, the electron is said to have *tunneled* through the barrier. The decay length of the wave function of an electron at the Fermi level of an electrode with a work function φ is given by

$$\frac{1}{\kappa} = \frac{\hbar}{\sqrt{2m[V_0 - E]}} = \frac{\hbar}{\sqrt{2m\varphi}}. \tag{1.5}$$

The ratio of the tunneling current density transmitted from an electrode on one side of the barrier to the tunneling current density received by an electrode on the other side of the barrier is described by the *transmission coefficient* T, where

$$T = \left|\frac{F}{A}\right|^2. \tag{1.6}$$

Eq. (1.4) can be solved to obtain coefficients F and A by matching $\psi(x)$ and $\frac{\mathrm{d}}{\mathrm{d}x}\psi(x)$ where $V(x)$ is discontinuous at $x = 0$ and $x = s$. Eq. (1.6) can then be solved to yield the transmission coefficient, which evaluates to

$$T = \frac{4k^2\kappa^2 \sinh^2(\kappa s)}{1 + (k^2 + \kappa^2)^2}. \tag{1.7}$$

For a barrier of width s significantly larger than the wave function decay length κ^{-1}, the transmission coefficient approximates to

$$T \approx \frac{16\kappa^2 k^2}{(\kappa^2 + k^2)^2} e^{-2\kappa s}. \tag{1.8}$$

Eq. (1.8) shows that the probability of an electron tunneling through the barrier from one electrode to another is *exponentially* dependent on (*i*) the barrier width s and (*ii*) the square root of the effective barrier height $[V_0 - E]$ through the decay constant κ. The ability of the STM to observe changes in surface topography and electronic structure with atomic resolution is due to the exponential dependence of the transmission coefficient on these two parameters. For example, a 1 Å change in the barrier width results in the transmission coefficient changing by approximately one order of magnitude [4].

1.1.2 Elastic tunneling through a one-dimensional potential energy barrier of arbitrary shape

The description of tunneling outlined in Section 1.1.1 can be extended to one-dimensional barriers of arbitrary shape using the *Wentzel-Kramers-Brillouin (WKB) approximation* [2]. Using the approximation, the probability $D(E)$ of an electron tunneling through a barrier of arbitrary shape is given by

$$D(E) = \exp\left(\frac{-2}{\hbar}\int_{s_1}^{s_2} \sqrt{2m\,[V(x) - E]}\,\mathrm{d}x\right) = \exp\left(-2\int_{s_1}^{s_2} \kappa(x,E)\,\mathrm{d}x\right) \tag{1.9}$$

where E is the energy of the electron, $\hbar$ is the reduced Planck constant, s_1 and s_2 are the turning points of the barrier, $\Delta s = (s_2 - s_1)$ is the barrier width, m is the mass of the electron, $V(x)$ is the barrier height at any position x, and κ is the decay constant. Note that use of the WKB approximation is appropriate only if the value of $V(x)$ is significantly greater than E and the slope of the barrier either side of s_1 and s_2 is gentle.

1.1.3 Elastic tunneling in planar metal-insulator-metal (MIM) tunnel junctions

The description of tunneling outlined in Section 1.1.2 can be used to describe elastic tunneling in planar metal-insulator-metal (MIM) junctions as shown in Fig. 1.2. In this case, it is assumed that both electrodes are in thermal equilibrium and exhibit free-electron behaviour. From Eq. (1.9), the probability $D(E)$ of an electron tunneling from one electrode to another through a barrier of arbitrary shape is given by

$$D(E) = \exp\left(\frac{-2\sqrt{2m}}{\hbar}\int_{s_1}^{s_2} \sqrt{[\eta + \phi(x) - E]}\,\mathrm{d}x\right) \tag{1.10}$$

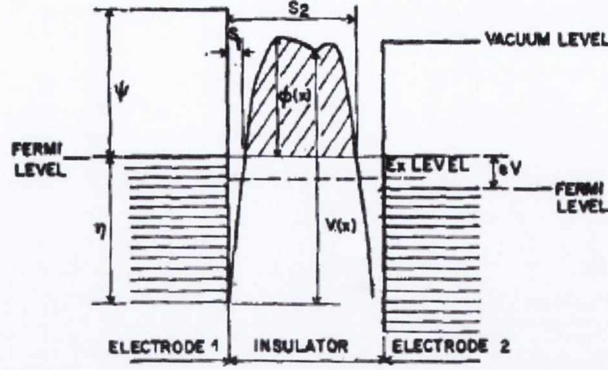


Fig. 1.2. Potential energy barrier of arbitrary shape between two metal electrodes separated by an insulator [5].

where E is the energy of the electron, $\hbar$ is the reduced Planck constant, s_1 and s_2 are the distances from the surface of electrode 1 to the points where the barrier crosses the Fermi level of electrode 1, Δs is the barrier width at the Fermi level of electrode 1, m is the mass of the electron, η is the Fermi energy of electrode 1, and $\phi(x)$ is the barrier height above the Fermi level of electrode 1 at any position x. By introducing a mean barrier height term, $\overline{\phi}$, the net tunneling current density J at 0 K can be expressed as [5]

$$J = \underbrace{J_0 \overline{\phi} \exp\left(-A\Delta s\sqrt{\overline{\phi}}\right)}_{J_1} - \underbrace{J_0\left(\overline{\phi}+eV\right)\exp\left(-A\Delta s\sqrt{\overline{\phi}+eV}\right)}_{J_2} \tag{1.11}$$

where

$$J_0 = \frac{e}{4\pi^2\hbar\,(\beta\Delta s)^2}, \qquad \overline{\phi} = \frac{1}{\Delta s}\int_{s_1}^{s_2}\phi(x)\,\mathrm{d}x, \qquad A = \frac{2\beta\sqrt{2m}}{\hbar}$$

and β is a correction factor approximately equal to one. V is the potential difference between the electrodes. Eq. (1.11) can be interpreted as a current density J_1 flowing from electrode 1 to electrode 2 and a current density J_2 flowing from electrode 2 to electrode 1, with the difference between them equal to the net current density J. If no voltage is applied, the current density is zero. If a voltage V is applied, the barrier is perturbed, an energy window of unoccupied states eV wide opens and the current density becomes non-zero. Further expressions for the current density have been derived for rectangular barriers across a number of voltage ranges [5]. These include:

(i) Low voltages

When $eV \approx 0$, $\overline{\phi} = \phi_0$ and $\Delta s = s$, the current density can be expressed as

$$J = \left(\frac{e^2}{4\pi^2\hbar^2}\right)\left(\frac{V}{s}\right)\left(\sqrt{2m\phi_0}\right)\exp\left(-As\sqrt{\phi_0}\right). \tag{1.12}$$

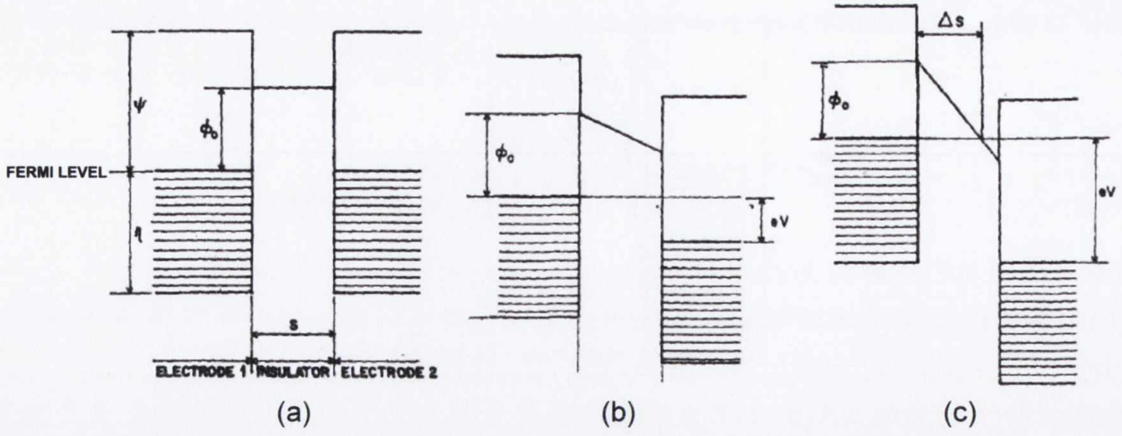


Fig. 1.3. Rectangular potential energy barrier between two metal electrodes separated by an insulator under (a) low voltage conditions; (b) intermediate voltage conditions; and (c) high voltage conditions [5].

Eq. (1.12) shows that the current density is exponentially dependent on (*i*) the barrier width s and (*ii*) the square root of the mean barrier height ϕ_0. Note that this is in agreement with Eq. (1.8) for the transmission coefficient presented on page 6. Eq. (1.12) also shows that the current density is linearly dependent on the applied voltage, implying that the tunnel junction exhibits ohmic behaviour at low voltages. Note that the STM operates in the low voltage regime. The barrier under these conditions is shown in Fig. 1.3a.

(ii) Intermediate voltages

When $eV < \phi_0$, $\overline{\phi} = \left(\phi_0 - \frac{eV}{2}\right)$ and $\Delta s = s$, the current density can be expressed as

$$J = J_0 \left[\phi_0 - \frac{eV}{2}\right] \exp\left(-As\sqrt{\phi_0 - \frac{eV}{2}}\right) - J_0 \left[\phi_0 + \frac{eV}{2}\right] \exp\left(-As\sqrt{\phi_0 + \frac{eV}{2}}\right). \quad (1.13)$$

The barrier under these conditions is shown in Fig. 1.3b.

(iii) High voltages

When $eV > \phi_0$, $\overline{\phi} = \frac{\phi_0}{2}$ and $\Delta s = \frac{\phi_0 s}{eV}$, the current density can be expressed as

$$\begin{aligned} J = &\left[\frac{e^3 (F/\beta)^2}{8\pi^2 \hbar \phi_0}\right] \exp\left(-\frac{4\pi\beta\sqrt{m}}{eF}\sqrt[3]{\phi_0}\right) \\ &- \left[\frac{e^3 (F/\beta)^2}{8\pi^2 \hbar \phi_0}\right]\left[1 + \frac{2eV}{\phi_0}\right] \exp\left(\frac{4\pi\beta\sqrt{m}}{eF}\sqrt[3]{\phi_0}\sqrt{1 + \frac{2eV}{\phi_0}}\right) \end{aligned} \quad (1.14)$$

where $F = V/s$ is the electric field strength in the barrier. The barrier under these conditions is shown in Fig. 1.3c. When $eV > (\eta + \phi_0)$, the Fermi level of electrode 2 lies below the conduction band of electrode 1. Under these conditions, electrons cannot tunnel from

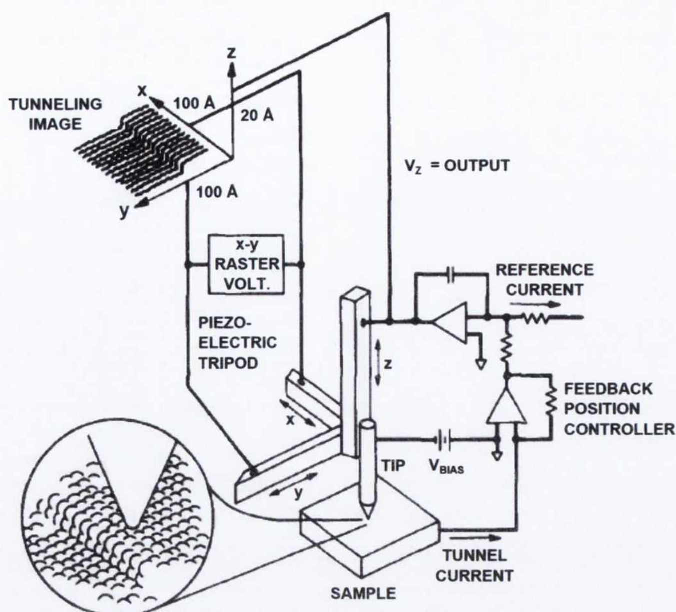


Fig. 1.4. Schematic of a typical scanning tunneling microscope. A sharp metal tip is mounted on a tripod scanner comprising three mutually orthogonal piezoelectric actuators responsible for moving the tip in three dimensions (x, y and z) parallel (x-y) and perpendicular (z) to the surface of an electrically conductive sample. The x and y piezos are driven by a raster scan generator, while the z piezo is driven by a feedback controller designed to maintain the tunneling current at a pre-set value by adjusting the distance between the tip and the surface. By recording the feedback voltage applied to the z piezo while scanning the tip across the surface, a topographic map of the surface can be obtained [7].

electrode 2 to electrode 1 since there are no unoccupied states in electrode 1. Conversely, more electrons can tunnel from electrode 1 to electrode 2 since there are more unoccupied states in electrode 2. In this case, the second term of Eq. (1.11) becomes zero and Eq. (1.14) reduces to an expression for field emission [6] given by

$$J = \left[\frac{e^3 (F/\beta)^2}{8\pi^2 \hbar \phi_0}\right] \exp\left(-\frac{4\pi\beta\sqrt{m}}{eF}\sqrt[3]{\phi_0}\right). \tag{1.15}$$

The first experimental observation of low voltage tunneling was performed using an early precursor to the STM called the *topografiner* [8, 9]. Fig. 1.4 shows a schematic of the instrument in the configuration later adopted by the STM. The basic operation of the topografiner involved positioning a sharp tungsten tip within several Ångströms of an electrically conductive surface and raster scanning the tip across the surface while a potential difference was applied between the two. This was accomplished using piezoelectric actuators that allowed precision movement of the tip in three dimensions (x, y and z) with a lateral (x–y) resolution of ~ 400 nm and a vertical (z) resolution of ~ 3 nm. A servo mechanism was used to maintain the position of the tip at a constant distance above the surface. By recording the feedback voltage applied to the z piezo during scanning, a topographic map of the surface could be obtained. This could be achieved in both

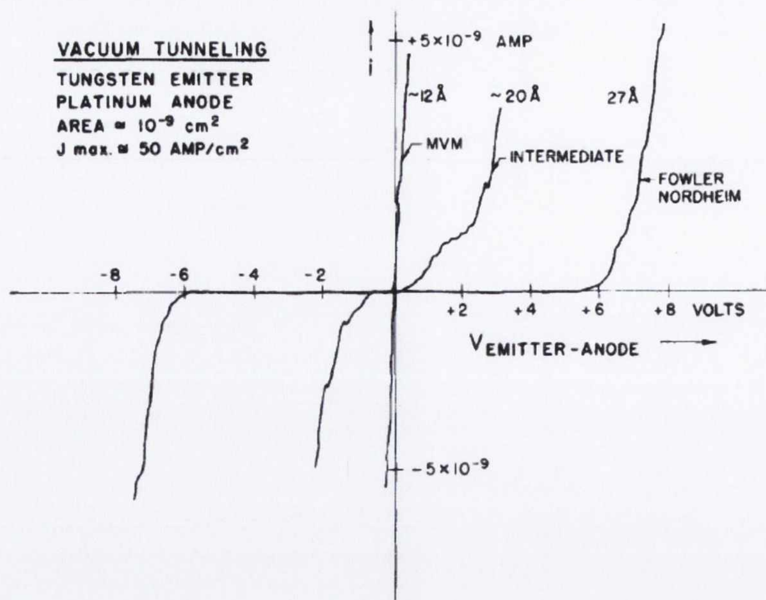


Fig. 1.5. Current-voltage (I-V) characteristics for three different tip-surface separation distances. In the low voltage regime ($\sim$ 12 Å), the current density changes linearly with junction voltage. In the intermediate voltage regime ($\sim$ 20 Å), the current density no longer changes linearly with junction voltage. In the field emission voltage regime ($\sim$ 27 Å), the current density is negligible until the tip-surface potential exceeds the relevant work function and the tunneling barrier has been perturbed substantially [8].

intermediate and field emission voltage regimes. Low voltage tunneling could be achieved by disconnecting the servo loop, disabling the x and y actuators and driving the z actuator manually. This was necessary due to the presence of mechanical noise and deficiencies in the stability of the servo mechanism. The current density observed in the low, intermediate and field emission voltage regimes is shown in Fig. 1.5. In spite of its limitations, the topografiner provided the first experimental observation of low voltage tunneling and confirmed the linear dependence of the tunneling current on the junction voltage predicted by Eq. (1.12). These advances would lay the foundations upon which the development [10–13] and earliest applications [14, 15] of the STM would later take place.

1.2 Instrumentation

The STM consists of four major components. These include (*i*) a coarse positioner to position the sample within tunneling range of the tip, (*ii*) a scanner to scan the tip across the sample surface, (*iii*) control electronics to regulate the tip-surface separation distance, and (*iv*) a vibration isolation system to reduce vibrational noise to a level sufficient to enable stable imaging with atomic resolution. This section provides an overview of these components and their role in the operation of the STM. Tip preparation will also be discussed. For a detailed overview of the design and construction of a functional STM, several excellent sources are recommended [3, 16–20].

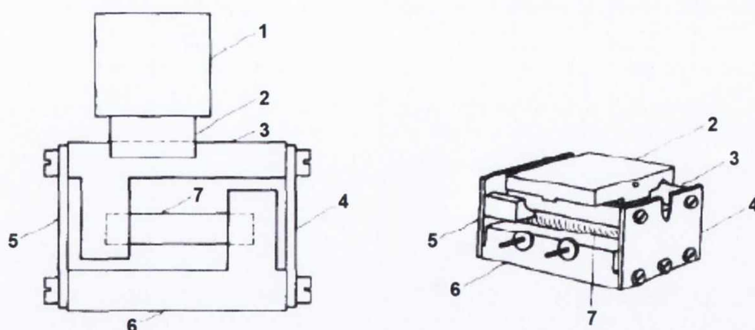


Fig. 1.6. Schematic of a typical inertial slider. The components are labeled as follows: (1) load, (2) stage, (3) rod, (4 and 5) flexible metal plates, (6) base (7) piezoelectric element [21].

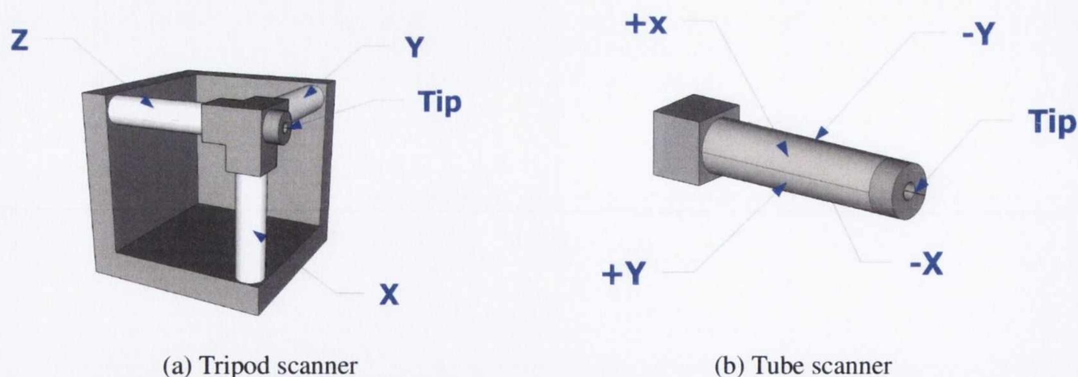
1.2.1 Coarse positioner

The coarse positioner is responsible for positioning the sample within tunneling range of the tip. This involves step-wise movement of the sample to within less than one micron of the tip over a travel range of several millimetres. In order to prevent a tip-sample collision, the minimum step size must be smaller than the maximum travel range of the scanner in the plane perpendicular to the surface. Sample movement must also be possible in the plane parallel to the surface.

The most common implementation of coarse positioning involves the of use of an *inertial slider* [21]. The operation of an inertial slider is based upon the controlled motion of a sliding mass along a surface using inertia. Fig. 1.6 shows a typical inertial slider consisting of a stage, rod and piezoelectric element. The stage is clamped to the rod, while the rod is connected to the piezoelectric element. When a slow-rising, fast-falling sawtooth voltage is applied to the piezo, the stage steps along the rod at the frequency of the applied voltage.

This process occurs in two phases. In the first phase, the slow-rising voltage causes slow elongation of the piezo. This is accompanied by slow movement of the stage-rod assembly in the direction of elongation. In the second phase, the fast-falling voltage causes rapid contraction of the piezo. This is accompanied by rapid movement of the stage-rod assembly in the direction of contraction. During this time, the inertia of the stage rises sufficiently to overcome the friction between it and the rod. This causes the stage to decouple from the rod. The stage re-couples to the rod once the rod has returned to its original position. This results in one step of the stage along the rod. The step direction can be reversed by applying a fast-rising, slow-falling sawtooth voltage to the piezo.

By mounting a sample on the stage, it can be positioned within tunneling range of the tip in the plane perpendicular to the surface. Multiple inertial sliders can be combined to extend positioning to the plane parallel to the surface. Advantages of this implementation include simplicity, stability, reliability and the ability to perform nanometre-sized steps over a large travel range. Alternative implementations include the use of 'louse'- [11], 'inchworm'- [22], and magnetic-type [23, 24] walkers. Fully mechanical techniques can also be used [16, 25].



(a) Tripod scanner (b) Tube scanner

Fig. 1.7. Schematic of two common STM scanner types showing (a) a tripod scanner consisting of three mutually orthogonal piezoceramic tubes (x, y and z) with a tip mounted at the scanner apex, and (b) a tube scanner consisting of a single piezoceramic tube with metalized inner and outer walls. The outer wall is evenly divided into four isolated quadrants. A tip is mounted at the free end of the tube.

1.2.2 Scanner

The scanner is responsible for scanning the sample surface. In order to image the surface with atomic resolution, it must be possible to move the tip in three dimensions (x, y and z) with a lateral (x–y) resolution of 1 Å and a vertical (z) resolution better than 0.01 Å [1]. This is achieved using a scanner assembly constructed from bar- or tube-shaped piezoceramic elements. For a piezoceramic bar of length l and thickness h, application of a voltage V across the thickness of the bar results in a change in length Δl given by

$$\Delta l = d_{31}\left(\frac{l}{h}\right)V \tag{1.16}$$

where d_{31} is the relevant piezoelectric coefficient. The value of the piezoelectric coefficient depends on the piezoceramic used and varies with temperature. Length changes typically range from 1-6 Å V^{-1} up to a maximum extension of several microns. Tube-shaped elements offer higher sensitivity than bar-shaped elements due to their reduced wall thickness. For a piezoceramic tube of length l and thickness h, application of a voltage V between the inner and outer walls of the tube results in a change in length Δl given by

$$\Delta l = d_{31}\left(\frac{l}{h}\right)V. \tag{1.17}$$

The two most common types of scanner are the tripod scanner [18] and the tube scanner [26–28]. Fig. 1.7a shows a tripod scanner consisting of three mutually orthogonal piezoceramic tubes joined together by vacuum-compatible epoxy [29]. By applying voltages to each of the tubes, a tip mounted at the apex of the scanner can be scanned across a surface with atomic resolution. Tripod scanners have low resonance frequencies ($\sim$ 1-5 kHz flexing and stretching) due to their composite construction, rendering them unsuitable for applications requiring high scanning speeds. They are also affected by mechanical cross-talk between tubes, whereby the motion of one tube influences the motion

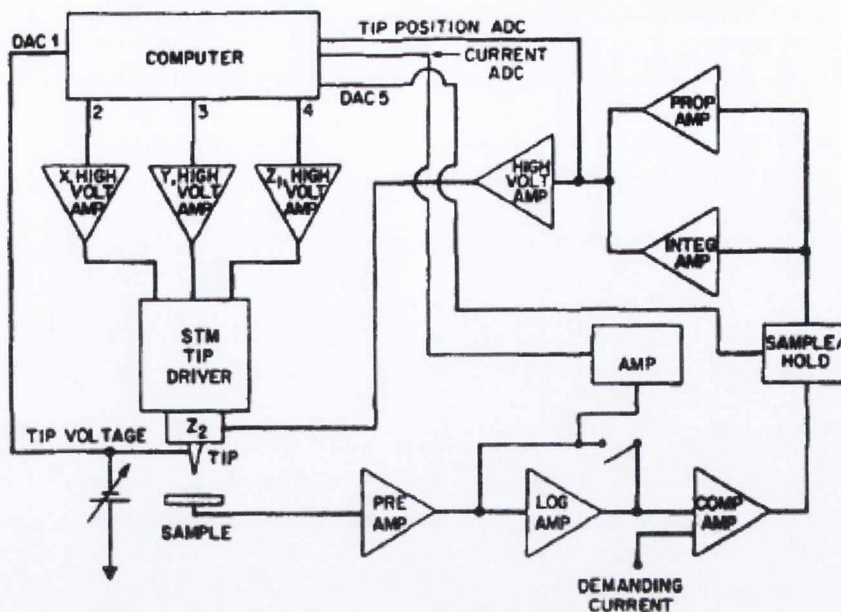


Fig. 1.8. Block diagram of a typical negative feedback circuit used to regulate the tip-surface separation distance [4, 20].

of another. For these reasons, many STM designs incorporate tube scanners instead of tripod scanners. Fig. 1.7b shows a tube scanner consisting of a single piezoelectric tube with metalized inner and outer walls. The outer wall is evenly divided into four electrically isolated quadrants, while the inner wall is connected to ground. By applying opposite voltages of equal magnitude to any two opposing quadrants while the others are connected to ground, the tube will bend in the direction perpendicular to the applied electric field. This allows the tube to move in two dimensions. By applying the same voltage to all four quadrants, the tube will change length. This extends movement of the tube to three dimensions. Like tripod scanners, a tip can be centre-mounted on one end of the tube and scanned across a surface with atomic resolution. Tube scanners offer substantially higher resonance frequencies ($\sim$ 8 kHz flexing, $\sim$ 40-100 kHz stretching) than tripod scanners, making them more suitable for applications requiring high scanning speeds. Like tripod scanners, however, they also suffer from mechanical cross-talk issues.

1.2.3 Control electronics

The control electronics are responsible for regulating the tip-surface separation distance. Fig. 1.8 shows a typical negative feedback circuit used to achieve this. Once tunneling has been established, the tunneling current is converted to a voltage and amplified by a high-gain transimpedance amplifier. The amplifier gain is normally 10^6-10^9 V A^{-1}. High gain is necessary due to the extremely small size of the tunneling current, which typically ranges from 0.01-50 nA [3]. Following amplification, the voltage is linearized by a logarithmic amplifier. Linearization is necessary due to the exponential dependence of the tunneling current on the tunneling barrier width predicted by Eq. (1.12). Following linearization, the voltage is compared with a reference voltage representing the tunneling current requested by

the user. The difference between these two voltages forms the input to a voltage integrator with a variable time constant. A proportional amplifier may also be included at this stage in order to shorten the settling time of the circuit at the expense of introducing a small error. In practice, however, most feedback circuit designs do not include a proportional amplifier.

The output of the integrator forms the input to a high-voltage amplifier, the output of which is connected to the z piezo of the scanner. If the tunneling current measured is *greater* than the tunneling current requested, the feedback circuit applies a voltage to the z piezo in order to move the tip *away* from the surface. If the tunneling current measured is *less* than the tunneling current requested, the feedback circuit applies a voltage to the z piezo in order to move the tip *towards* the surface. This process is repeated until an equilibrium tip-surface separation is achieved. The x and y piezos of the scanner are also connected to high-voltage amplifiers driven by a computer-controlled digital-to-analog converter (DAC).

Most feedback circuit designs also include a sample-and-hold amplifier inserted before the integrator. This allows the integrator input to be set to zero, thereby disabling the feedback loop and allowing the tip-surface separation distance to be held constant for a short period of time. During this time, the tunneling current can be measured solely as a function of the applied sample voltage. This forms the basis of tunneling spectroscopy.

1.2.4 Vibration isolation

The role of vibration isolation is to reduce vibrational noise to a level sufficient to enable stable imaging with atomic resolution. Atomically clean surfaces typically exhibit corrugation amplitudes of 0.1 Å or less. In order to resolve these corrugations, the tip-surface separation distance must be stable to within 0.01 Å or better. This is achieved by limiting the response of the STM to internal and external vibration sources. The extent to which this may be achieved depends on the structural rigidity of the STM, the nature of the vibration sources, and the characteristics of the vibration isolation system used.

The amplitude response of a system to an external vibration is described by the *amplitude transfer function*. This is defined as the ratio of the amplitude of the external vibration to the amplitude of vibration of the system. For an external vibration with an amplitude of 1 μm, the amplitude transfer function must be 10^{-6} or less in order for the amplitude transferred to the system to be 0.01 Å or less. This can be achieved by suspending the system from a spring. The extension ΔL of a spring of stiffness κ supporting a mass m is given by

$$\Delta L = \frac{mg}{\kappa} \tag{1.18}$$

where g is the acceleration due to gravity. The fundamental frequency f_0 is given by

$$f_0 = 2\pi\sqrt{\frac{g}{\Delta L}}. \tag{1.19}$$

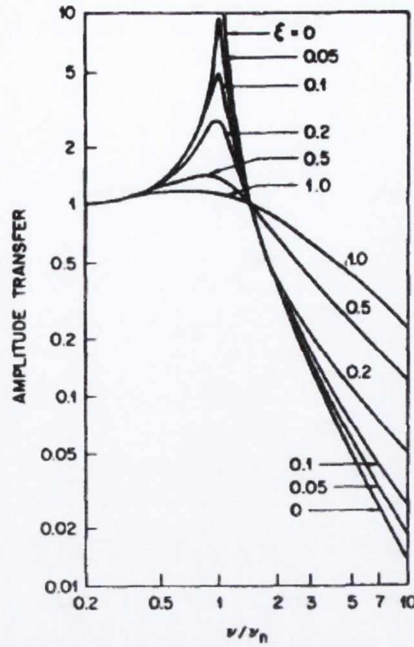


Fig. 1.9. Graph of the amplitude transferred from an external vibration of frequency ν to a spring of fundamental frequency ν_η plotted as a function of normalized frequency ν/ν_η and damping ratio ξ [20].

The amplitude transfer function for the spring is given by

$$T_s(f_v) = \left[\frac{1 + \left(\frac{2\xi f_v}{f_s}\right)^2}{\left(\frac{1 - f_v{}^2}{f_s{}^2}\right)^2 + \left(\frac{2\xi f_v}{f_s}\right)^2} \right]^{\frac{1}{2}} \tag{1.20}$$

where f_v is the frequency of the external vibration, f_s is the fundamental frequency of the spring, and ξ is the damping ratio of the spring. Fig. 1.9 shows a graph of the amplitude transfer function plotted as a function of normalized frequency and damping ratio. It can be seen from the graph that vibrations below the fundamental frequency result in complete transfer, vibrations at the fundamental frequency result in amplified transfer, and vibrations above the fundamental frequency result in reduced transfer. It can also be seen that increasing damping reduces transfer at the fundamental frequency, but amplifies transfer above the fundamental frequency.

Two conclusions may be drawn from these observations. First, effective vibration isolation systems have low fundamental frequencies. Second, balanced performance across a wide frequency range requires a compromise between between damping and vibration isolation. Single-stage spring suspension systems are rarely used alone in practice. This is due to the fact that a spring length of at least 25 cm would be required to reduce the fundamental frequency of the system to 1 Hz. Two-stage spring suspension systems [18, 19] are more common. These systems typically consist of an outer set of springs supporting an inner set of springs, which in turn support the mass requiring vibration

isolation. In the case of an STM, this is usually achieved by suspending the microscope from a single-stage spring suspension system and mounting the STM housing on a vibration isolation table. Rigid plate stacks and elastomers [30] can also be employed.

Regardless of the suspension system used, the most important factor in limiting the response of an STM to internal and external noise sources is the structure of the STM itself. The amplitude transfer function for the STM is given by [17]

$$T_m(f_v) = \frac{\left(\frac{f_v}{f_m}\right)^2}{\left[\left(\frac{1-f_v^2}{f_m^2}\right)^2 + \left(\frac{f_v}{f_m Q}\right)^2\right]^{\frac{1}{2}}} \tag{1.21}$$

where f_v is the frequency of the external vibration, f_m is the fundamental frequency of the STM, and Q is the quality factor of the tip-sample junction [4]. The total amplitude transfer function for the combined STM and spring suspension system is given by

$$T_{s+m}(f_v) = T_s(f_v)T_m(f_v) \tag{1.22}$$

Eq. (1.22) shows how the amplitude transfer function of the STM can be compensated for by the amplitude transfer function of a well-designed spring suspension system. The total amplitude transfer function in the frequency range of $f_m \geq f_v \geq f_s$ approximates to [3]

$$T(f_v) = \left(\frac{f_s}{f_m}\right)^2. \tag{1.23}$$

Eq. (1.23) provides a simple means of assessing the effectiveness of a given spring-microscope combination. For example, if f_s is 2.5 Hz and f_m is 25 kHz, the amplitude transfer function will be 10^{-8}. This is sufficient to guarantee the stability of the tip-surface separation distance to within 0.01 Å as required.

1.2.5 Tip preparation

The size, shape and chemical composition of the tip are key factors in determining the quality of data measured by the STM. The primary objective of tip preparation is to produce sharp, high-quality tips with low aspect ratios that minimize tip vibration [31]. Tips are usually formed from hard materials such as tungsten and platinum-iridium alloy. Tungsten is typically used for UHV applications, while platinum-iridium alloy is used for in-air and corrosive gas applications. This subsection will focus on the preparation of tungsten tips.

Tungsten tips can be formed using a variety of methods [3]. The most common of these is the *DC drop-off method* [31]. This involves electrochemically etching a section of tungsten wire in sodium hydroxide solution until a sharp tip is formed. The steps of the method are as follows. A 10 mm long piece of 0.25 mm diameter tungsten wire is mounted lengthwise on a vertical manipulator and positioned above the centre of a small beaker containing a

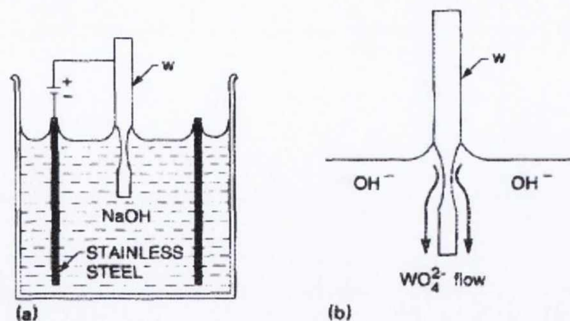
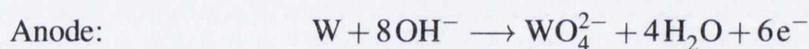


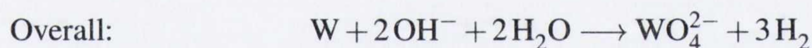
Fig. 1.10. (a) Schematic of an electrochemical cell showing a tungsten wire being etched in a sodium hydroxide solution. The tungsten wire serves as the anode, while a ring of stainless steel wire serves as the cathode. (b) Schematic of the tip etching mechanism showing the formation of WO_4^{2-} upon reaction of the immersed tungsten wire with hydroxyl anions [31].

2 M sodium hydroxide solution. A ring of platinum wire is immersed in the solution at the periphery of the beaker. This forms the cathode. The tungsten wire is lowered 1-3 mm below the surface of the solution. This forms the anode. A 4-12 V potential difference is applied to the tungsten wire to initiate the following electrochemical reaction [31]:

Cathode: $$6H_2O + 6e^- \longrightarrow 3H_2 + 6OH^-$$



Anode: $$W + 8OH^- \longrightarrow WO_4^{2-} + 4H_2O + 6e^-$$



Overall: $$W + 2OH^- + 2H_2O \longrightarrow WO_4^{2-} + 3H_2$$

Fig. 1.10 shows a schematic of the etching process. Etching takes place primarily at the air-solution interface. When the weight of the immersed section of tungsten wire overcomes the tensile strength of the etched section, the etched section fractures and *drops off*. This leaves a sharp tip at the end of the remaining wire. An electronic circuit monitoring the etching current immediately cuts off the potential difference and the tip is removed from the manipulator. Residual sodium hydroxide is removed by rinsing the tip in deionized water and high purity alcohol followed by drying with nitrogen gas. The tip is then annealed for several hours at 1000 K to remove the tungsten oxide layer and any contaminants remaining from the etching process [3]. Once cool, the tip may be stored for future use.

Several parameters can be adjusted in order to control the size and shape of the tip [31]. These include the magnitude of the potential difference applied to the tungsten wire, the wire immersion depth, the sodium hydroxide concentration, the shape of the meniscus and the potential difference cut-off time. The most important of these is the potential difference cut-off time; in general, the shorter the cut-off time, the sharper the tip. For more detailed information on tip preparation and alternative tip preparation methods, several excellent sources are recommended [3, 4].

1.3 Operation

The STM can be operated in either *constant-current* or *constant-height* imaging modes. In constant-current mode, the tip is scanned across the sample surface while the tunneling current is held constant by a feedback loop described in Section 1.2.3. At the same time, the feedback voltage applied to the z piezoelectric actuator is recorded as a function of lateral tip position and translated into a contour map of the surface. This map is representative of the surface topography *only* if the local density of states (LDOS) sampled is constant over the entire surface. In constant-height mode, the tip is scanned across the surface while the vertical position of the tip is held constant by disabling the feedback loop. At the same time, the tunneling current is measured as a function of lateral tip position and translated into a current density map of the surface. Constant-current mode is the more widely used of the two modes. This section will focus on constant-current mode. For a detailed discussion of STM imaging modes and related information on scanning tunneling spectroscopy (STS), several excellent sources are recommended [2–4, 32].

1.3.1 Constant-current mode

Interpretation of the contour map obtained in constant-current mode is non-trivial. This is because the tunneling current varies not only as a function of the distance between the tip and the surface, but also as a function of their respective LDOS at a given energy. In order to interpret the contour map correctly, it is first necessary to evaluate the electronic contributions of the tip and the surface to the tunneling current. This is achieved using the *Tersoff-Hamann approximation* [33–35]. The Tersoff-Hamann approximation provides a simplified three-dimensional description of tunneling between the tip and the surface based upon first-order perturbation theory [36] expressed in the *Transfer-Hamiltonian formalism* [37]. In this formalism, the tunneling current I flowing between the tip and the surface in *one dimension* is expressed as [34]

$$I = \frac{2\pi e^2}{\hbar} V \sum_{\mu\nu} |M_{\mu\nu}|^2 \delta\left(E_\nu - E_F\right) \delta\left(E_\mu - E_F\right) \tag{1.24}$$

where $\hbar$ is the reduced Planck constant, V is the voltage applied to the sample, $M_{\mu\nu}$ is the tunneling matrix element between the unperturbed tip and surface wave functions ψ_μ and ψ_ν, E_μ and E_ν are energy eigenvalues corresponding to ψ_μ and ψ_ν respectively, and E_F is the Fermi energy. Note that this expression is valid only for small voltages and temperature [34]. Eq. (1.24) can be solved by evaluating $M_{\mu\nu}$, given by [34]

$$M_{\mu\nu} = -\frac{\hbar^2}{2m} \int \left(\psi_\mu^* \nabla \psi_\nu - \psi_\nu \nabla \psi_\mu^*\right) \mathrm{d}\mathbf{S} \tag{1.25}$$

where the integral is evaluated over an imaginary surface lying between the tip and the surface, and $\nabla \psi_\mu$ and $\nabla \psi_\nu$ are the derivatives of ψ_μ and ψ_ν respectively.

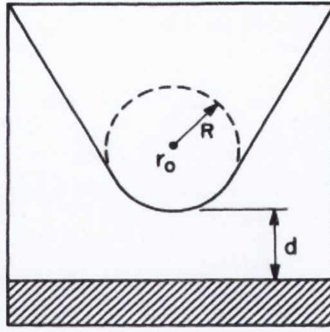


Fig. 1.11. Schematic of an STM tip as described in the Tersoff-Hamann approximation. The tip is spherically symmetrical with a radius of curvature R centred at $\mathbf{r}_0$, while d is the distance between the tip and a nearby surface. [34].

In order to solve Eq. (1.25), exact expressions for ψ_μ and ψ_ν are required. An exact expression cannot be obtained for ψ_μ, however, because the atomic structure of the tip is unknown. The Tersoff-Hamann approximation circumvents this problem by assuming that the structure of the tip is spherically symmetric with a radius of curvature R centred at a point $\mathbf{r}_0$ as shown in Fig. 1.11. This allows ψ_μ to be described by an s-type wave function, thus allowing Eq. (1.25) to be evaluated to [34]

$$M_{\mu\nu} = \frac{4\pi\hbar^2}{2m} \cdot \frac{R}{\sqrt{\Omega_t}} \cdot e^{\kappa R} \cdot \psi_\nu(\mathbf{r}_0) \tag{1.26}$$

where κ is the decay constant and Ω_t is the tip volume. Eq. (1.24) then becomes [34]

$$I = \left[\left(\frac{32\pi^3 e^2}{\hbar}\right)\left(\frac{\phi^2 R^2}{\kappa^4}\right)\right] \cdot V \cdot \rho_t(E_F) \cdot e^{2\kappa R} \cdot \sum_\nu |\psi_\nu(\mathbf{r}_0)|^2 \delta(E_\nu - E_F) \tag{1.27}$$

where ϕ is the tunneling barrier height, $\rho_t(E_F)$ is the density of states per unit volume of the tip at E_F, and the summation term is the surface LDOS at E_F, $\rho_s(E_F)$, evaluated at $\mathbf{r}_0$.

Eq. (1.27) shows that $\rho_s(E_F, \mathbf{r})$ is proportional to I. Assuming the s-type tip wave function approximation is valid, this implies that images obtained in constant-current mode are contour maps of constant $\rho_s(E_F)$ evaluated at $\mathbf{r}_0$. Again, this interpretation is valid only for small voltages and temperatures. Since ψ_μ and ψ_ν decay exponentially into the tunneling barrier in the z direction perpendicular to the surface, $\psi_\nu(\mathbf{r}) \propto e^{-\kappa z}$, and therefore $|\psi_\nu(\mathbf{r}_0)|^2 \propto e^{-2\kappa(R+d)}$, where d is the distance between the tip and the surface. This implies that $I \propto e^{-2\kappa d}$, which is in agreement with Eq. (1.12) for the tunneling current density presented on page 7. There are limitations to the applicability of the Tersoff-Hamann approximation. If the LDOS varies across the surface, $\rho_s(E_F, \mathbf{r})$ will no longer be proportional to I, and the contour map obtained in constant-current mode will no longer be representative of the surface topography. Furthermore, if the tip wave function cannot validly be described by an s-type wave function, or if the voltage applied to the sample is too large, the approximation cannot be applied.

References

[1] C. Bai, *Scanning Tunneling Microscopy and its Applications* (Springer, 2000), 2nd ed.

[2] R. Wiesendanger, *Scanning Probe Microscopy and Spectroscopy: Methods and Applications* (Cambridge University Press, 1994).

[3] C. J. Chen, *Introduction to Scanning Tunneling Microscopy* (Oxford University Press, 2008), 2nd ed.

[4] J. A. Kubby and J. J. Boland, Surf. Sci. Rep. **26**, 61 (1996).

[5] J. G. Simmons, J. Appl. Phys. **34**, 1793 (1963).

[6] R. Fowler and L. Nordheim, Proc. R. Soc. Lond. A **119**, 173 (1928).

[7] J. A. Golovchenko, Science **232**, 48 (1986).

[8] R. Young, J. Ward, and F. Scire, Phys. Rev. Lett. **27**, 922 (1971).

[9] R. Young, J. Ward, and F. Scire, Rev. Sci. Instrum. **43**, 999 (1972).

[10] G. Binnig, H. Rohrer, C. Gerber, and E. Weibel, Appl. Phys. Lett. **40**, 178 (1982).

[11] G. Binnig and H. Rohrer, Helv. Phys. Acta. **55**, 726 (1982).

[12] G. Binnig, H. Rohrer, C. Gerber, and E. Weibel, Phys. Rev. Lett. **49**, 57 (1982).

[13] G. Binnig and H. Rohrer, Surf. Sci. **126**, 236 (1983).

[14] G. Binnig, H. Rohrer, C. Gerber, and E. Weibel, Phys. Rev. Lett. **50**, 120 (1983).

[15] G. Binnig, H. Rohrer, F. Salvan, C. Gerber, and A. Baro, Surf. Sci. **157**, L373 (1985).

[16] J. E. Demuth, R. J. Hamers, R. M. Tromp, and M. E. Welland, IBM J. Res. Develop. **30**, 396 (1986).

[17] D. W. Pohl, IBM J. Res. Develop. **30**, 417 (1986).

[18] S. Park and C. F. Quate, Rev. Sci. Instrum. **58**, 2010 (1987).

[19] M. Okano, K. Kajimura, S. Wakiyama, F. Sakai, W. Mizutani, and M. Ono, J. Vac. Sci. Technol. A. **5**, 3313 (1987).

[20] Y. Kuk and P. J. Silverman, Rev. Sci. Instrum. **60**, 165 (1989).

[21] D. W. Pohl, Rev. Sci. Instrum. **58**, 54 (1987).

[22] W. G. May, Jr., *Piezoelectric electromechanical translation apparatus*, U. S. Patent 3902084 (1975).

[23] D. P. E. Smith and S. A. Elrod, Rev. Sci. Instrum. **56**, 1970 (1985).

[24] B. W. Corb, M. Ringger, and H.-J. Guntherodt, J. Appl. Phys. **58**, 3947 (1985).

[25] J. E. Demuth, R. J. Hamers, R. M. Tromp, and M. E. Welland, J. Vac. Sci. Technol. A. **4**, 1320 (1986).

[26] C. P. Germano, IRE T. Audio **7**, 96 (1959).

[27] G. Binnig and D. P. E. Smith, Rev. Sci. Instrum. **57**, 1688 (1986).

[28] K. Besocke, Surf. Sci. **181**, 145 (1986).

[29] EPO-TEK H21D, Epoxy Technology, Inc., Billerica, MA.

[30] C. Gerber, G. Binnig, H. Fuchs, O. Marti, and H. Rohrer, Rev. Sci. Instrum. **57**, 221 (1986).

[31] J. P. Ibe, P. P. Bey, S. L. Brandow, R. A. Brizzolara, N. A. Burnham, D. P. DiLella, K. P. Lee, C. R. K. Marrian, and R. J. Colton, J. Vac. Sci. Technol. A. **8**, 3570 (1990).

[32] J. A. Stroscio and W. J. Kaiser, *Scanning Tunneling Microscopy*, vol. 27 of *Methods of Experimental Physics* (Academic Press, 1993).

[33] J. Tersoff and D. R. Hamann, Phys. Rev. Lett. **50**, 1998 (1983).

[34] J. Tersoff and D. R. Hamann, Phys. Rev. B **31**, 805 (1985).

[35] P. K. Hansma and J. Tersoff, J. Appl. Phys. **61**, R1 (1987).

[36] J. Bardeen, Phys. Rev. Lett. **6**, 57 (1961).

[37] M. H. Cohen, L. M. Falicov, and J. C. Phillips, Phys. Rev. Lett. **8**, 316 (1962).

CHAPTER 2

Surface Stress

As materials are reduced in size to the atomic scale, the role of surfaces in determining their properties becomes critically important. One of the most important factors determining these properties is *surface stress*. Surface stress is a thermodynamic quantity defined as the reversible work per unit area required to stretch a surface elastically [1]. In simpler terms, it refers to the mechanical forces acting at a surface due to the configuration of the chemical bonds holding it together. It can be intrinsic to a surface, induced by adsorbate-surface interactions, or applied externally. The importance of surface stress has been demonstrated across a wide range of topics including, amongst others, surface reconstruction [2–5], thin film growth [6–10], electronic structure [11] and chemical reactivity [12–20]. In the case of the Si(100)–(2×1) surface, for example, externally applied strain has been shown to alter the (1×2) and (2×1) domain populations observed following annealing [3–5]. In the case of As and Ge adsorption on the same surface, surface stress has been shown to strongly influence the film growth process [6–8]. In the further case of O_2 adsorption on the Si(111)–(7×7) surface, reacting oxygen species have been shown to generate unique, measurable stress signatures directly related to the underlying reaction mechanism [18]. The effects of surface stress can be investigated in two ways. First, changes in surface stress can be measured and related to the processes causing them with the aid of theoretical calculations and complementary experimental results. These measurements are generally performed using the capacitive [12] and optical [21] cantilever bending techniques described later. Recent advances in instrumentation [22] have combined these techniques *in situ* with scanning tunneling microscopy (STM), allowing changes in surface stress to be related to the processes causing them using atomic resolution imaging [18]. Second, surface stress can be applied externally and the effects observed using STM [4, 5, 22], second-harmonic generation spectroscopy (SHG) [20] and other experimental techniques. These effects have been investigated bond-specifically [20] and with atomic resolution [4, 5]. This chapter presents an overview of surface stress in the context of crystalline solids, comprising a precise definition of surface stress, a discussion of its origins and a description of the techniques currently available to measure it.

2.1 Definition

Surface stress can be defined in several equivalent ways depending on the context within which it is discussed. The most general of these describes it as a change in the bulk *stress tensor* near a surface or interface [23]. A stress tensor is a matrix of normal and shear stress components describing the stress at any point in a material within a given coordinate system. For the purposes of this discussion, the tensor coordinate system is defined in three dimensions (x, y and z) by axes parallel (x and y) and perpendicular (z) to the surface, as shown in Fig. 2.1. Tensors are indexed in the x and y directions by indices i and j respectively. The surface stress can then be described by the expression [23]

$$\tau_{ij}^{s} = \int_{-\infty}^{+\infty} \left[\tau_{ij}(z) - \tau_{ij}^{b} \right] \mathrm{d}z \tag{2.1}$$

where τ_{ij}^{s} is the surface stress tensor at ij, $\tau_{ij}(z)$ is the bulk stress tensor at ij evaluated at z, and τ_{ij}^{b} is the bulk stress tensor at ij near the surface. This expression is valid for isolated surfaces and for interfaces formed between two solids, two liquids, or a solid and a liquid, assuming the bulk stress components are the same for each phase. The sign of τ_{ij}^{s} is positive when the surface stress is *tensile* and negative when the surface stress is *compressive*. Tensile stress refers to stress due to stretching of the surface bonds, while compressive stress refers to stress due to compression of the surface bonds. The dimension of τ_{ij}^{s} is force per unit length (N m^{-1}), although this is often expressed thermodynamically as the work per unit area (eV m^{-2}) normalized to the area of a unit cell. For reference, 1 N m^{-1} is equivalent to 1 J m^{-2}, 1 eV is equivalent to 1.602×10^{-19} J, and 1 N m^{-1} is equivalent to 6.242×10^{18} eV m^{-2}. Note that Eq. (2.1) can be combined with elasticity theory to model surface stress where the dimensions and mechanical properties of a material are already known [23].

2.2 Origins

The origins of surface stress can be discussed in terms of intrinsic surface stress and surface stress induced by adsorbate-surface interactions. This section discusses each case in the context of metal and semiconductor surfaces.

2.2.1 Intrinsic surface stress

The coordination of atoms at the surface of a crystal differs from that of atoms in the bulk. Each surface atom (or "adatom") possesses excess charge arising from reduced coordination. In the case of metal surfaces, this charge is transferred to the lateral bonds between the adatoms and to the backbonds connecting them to the underlying atomic layer. These bonds become shorter and stronger as a result. Two consequences arise from this. First, there is a reduction in the distance between the surface and the underlying atomic layer. This occurs due to shortening of the backbonds. Many metal surfaces exhibit this

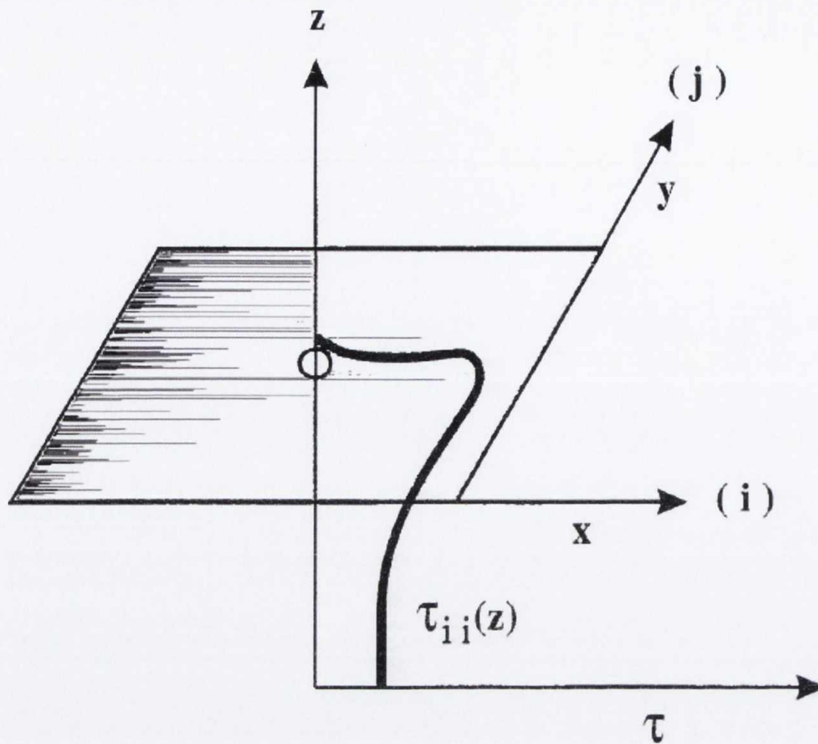


Fig. 2.1. Illustration of the variation of the bulk stress tensor $\tau_{ij}(z)$ near a surface. The tensor coordinate system is defined in three dimensions (x, y and z) by axes parallel (x and y) and perpendicular (z) to the surface. Tensors are indexed in the x and y directions by indices i and j respectively [23].

behaviour [23, 24], with the effect most pronounced in the case of surfaces with open structures [23]. Some exceptions include [23] the nearly-free electron Mg(0001) [25], Al(111) [26] and Al(100) [27] surfaces, none of which exhibit simple contraction of the first interlayer. Second, the adatoms experience forces pulling them in the directions of the lateral bonds. This occurs due to shortening of the lateral bonds. Since the adatoms are fixed in position with respect to the bulk lattice, they are unable to move in response to these forces. This results in tensile surface stress. In general, all clean metal surfaces should exhibit tensile stress regardless of their structure. While no theorem exists proving this to be the case, evidence to the contrary has yet to be found [23, 28]. It has recently been suggested [28], however, that certain magnetic metal surfaces may exhibit *compressive* stress.

The situation is more complicated in the case of semiconductor surfaces. Here, the sign of the surface stress depends on the structure of the surface. This is due to the effects of bonding topology and orbital re-hybridization in the surface region [23, 29]. The Si(111)–(1×1) surface serves as an ideal example of this behaviour [29, 30]. Here, each adatom possesses an unpaired valence electron arising from reduced coordination. This electron is localized in a "dangling bond" orbital at the top of the adatom. The presence of a dangling bond causes the adatom orbitals to re-hybridize from the tetrahedral sp^3 configuration of the bulk lattice to a quasi-planar sp^2 configuration. This causes the adatom to move downwards. Since the

Si (111)

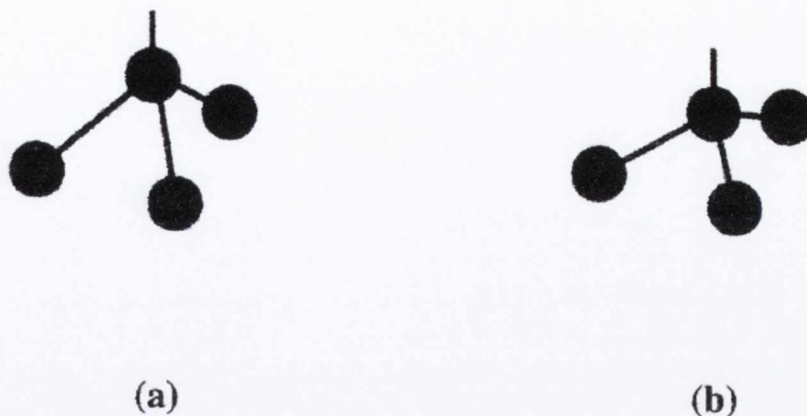


Fig. 2.2. Illustration of the transition between (a) sp^3 and (b) sp^2 structural configurations of adatoms on the Si(111)–(1 × 1) surface. In the sp^2 configuration, the adatom moves downwards. Since the Si–Si bond length does not change significantly during re-hybridization, the Si atoms in the underlying layer are pushed outwards. This results in compressive surface stress [23].

Si–Si bond length does not change significantly during re-hybridization, the Si atoms in the underlying layer are pushed outwards. This results in compressive surface stress. The transition from sp^3 to sp^2 configurations is illustrated in Fig. 2.2. Theoretical calculations predict a compressive surface stress of −0.54 eV per 1 × 1 unit cell and a bond angle of 114.0° [29]. These calculations also show that the sign and magnitude of the surface stress depend on the size and electronic configuration of the adatom. This implies that substitution of the adatom with atoms of varying size and electronic configuration will also change the surface stress.

In the case of substitution with Ga, a compressive surface stress of −4.45 eV per (1 × 1) unit cell and a bond angle of 119.3° are predicted [29]. The explanation for this behaviour lies in the fact that Ga has three valence electrons, each of which becomes involved in bond formation. The substituted Ga atom therefore has no dangling bond. This results in greater sp^2 hybridization relative to the unsubstituted surface, thus increasing the bond angle and decreasing the Ga–Si bond length. Since Ga is larger than Si, the new bond length is *less* than the ideal bond length, and the surface experiences compressive stress.

In the case of substitution with As, a tensile surface stress of 2.27 eV per (1 × 1) unit cell and a bond angle of 104.7° are predicted [29]. The explanation for this behaviour lies in the fact that As has five valence electrons, only three of which become involved in bond formation. The subsituted As atom therefore has a doubly-occupied dangling bond. Coulombic repulsion between the dangling bond and the As–Si bonds favours sp^3 hybridization, thus decreasing the bond angle and increasing the As–Si bond length. Since the new bond angle is *less* than that of a perfect tetrahedron and since the new bond length is *greater* than the ideal bond length, the surface experiences tensile stress.

In the case of substitution with B, a tensile surface stress of 4.87 eV per (1×1) unit cell and a bond angle of 119.0° are predicted [29]. The explanation for this behaviour lies in the fact that B is smaller than Si. B, like Ga, has three valence electrons, each of which becomes involved in bond formation. The substituted B atom therefore has no dangling bond. This results in greater sp^2 hybridization relative to the unsubstituted surface, thus increasing the bond angle and decreasing the B−Si bond length. Since B is *smaller* than Si, the new bond length is *greater* than the ideal bond length, and the surface experiences tensile stress.

Some semiconductor surfaces also exhibit anisotropic surface stress. In the case of the Si(100)–(2×1) surface, for example, theoretical calculations show that the surface is under tensile stress parallel to the Si−Si dimer bonds and compressive stress perpendicular to the dimer bonds [31–33]. This is crucial in explaining why externally applied strain alters the (2×1) and (1×2) domain populations following annealing [3–5]. In the presence of external strain, domains in which the dimer bonds are compressed are observed to grow at the expense of domains in which the dimer bonds are rotated by 90.0° [3]. This occurs due to relaxation of the energy associated with long-range strain fields extending into the bulk lattice [5]. These strain fields arise due to the anisotropic stress of the Si(100)–(2×1) surface [5, 31].

2.2.2 Adsorbate-induced surface stress

Chemisorption restores the coordination of surface atoms with respect to their bulk counterparts. In the case of metal surfaces, this should result in charge transfer from the lateral bonds and backbonds to the adsorbate-bound adatom. These bonds should become longer and weaker as a result, thus reducing the tensile stress intrinsic to the surface. In reality, however, the situation is more complicated. The chemical nature of both the adsorbate and the surface must be considered.

If the adsorbate is electronegative, charge transferred from the lateral bonds and backbonds is distributed between the adsorbate and the adatom to which it is bound. This reduces the intrinsic tensile stress as before, but can also induce compressive stress depending on the level of charge localized on the adsorbate. The level of charge localized on the adsorbate depends on how electronegative or electropositive it is with respect to the surface. Coloumbic repulsion between neighbouring adsorbates enhances the effect slightly. Some examples of systems characterized by this behaviour include O/Ni(100), C/Ni(100) and S/Ni(100) [34, 35]. The presence of weak compressive stress is not expected to alter the structure of a surface significantly, but strong compressive stress can induce soft phonon [36] and, in some cases, complete reconstructions [23]. Some examples of systems characterized by soft phonon reconstructions include C/Ni(100) [23, 37] and N/Ni(100) [23, 38].

If the adsorbate is electropositive, charge from the adsorbate-adatom bond is transferred to the lateral bonds and backbonds. These bonds become shorter and stronger as a result, thus leading to an increase in tensile stress. Coloumbic repulsion between neighbouring adsorbates retards the effect slightly. An example of the trends of electronegative and

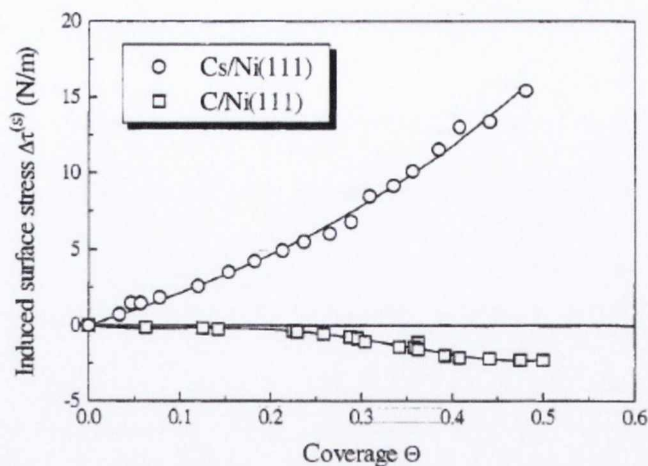


Fig. 2.3. Change in surface stress $\Delta\tau^{(s)}$ measured as a function of increasing coverage Φ for the adsorption of C (squares) and Cs (circles) on the Ni(111) surface. Coverages are scaled relative to the number of Ni atoms. [23, 39].

electropositive adsorbates is shown in Fig. 2.3. Fig. 2.3 shows a graph of the change in surface stress measured as a function of increasing coverage for the adsorption of C and Cs on the Ni(111) surface [23, 39]. The Pauling electronegativities of C, Cs and Ni are 2.55, 0.79 and 1.91 respectively. C is therefore electronegative with respect to Ni, while Cs is strongly electropositive. The adsorption of C relieves the intrinsic tensile stress [40] of the Ni(111) surface and induces compressive stress, consistent with the behaviour expected from an electronegative adsorbate. The adsorption of Cs, on the other hand, dramatically increases the tensile stress, consistent with the behaviour expected from an electropositive adsorbate.

While this reasoning is valid in many cases, it is not valid in all cases. This is exemplified by the anomalous case of H adsorption on the Pt(111) surface. The Pauling electronegativities of H and Pt are 2.20 and 2.28 respectively. H is therefore electropositive with respect to Pt. Based on previous reasoning, the adsorption of H on Pt(111) should increase tensile stress. However, both theoretical [41, 42] and experimental [42] evidence reveals that the adsorption of H *relieves* the intrinsic tensile stress [40, 41, 43] of the Pt(111) surface and induces compressive stress. This is clearly inconsistent with the behaviour expected from an electropositive adsorbate. Another inconsistency arises from the fact that the unoccupied orbitals of the Pt(111) surface are antibonding [41]. Occupation of these orbitals should lengthen and weaken the lateral bonds, thus reducing the intrinsic tensile stress and potentially inducing compressive stress [41]. This argument is, however, inconsistent with many experimental observations [23]. In spite of these problems, the use of electronegativity as a tool for qualitatively explaining adsorbate-induced changes in surface stress remains valid in many cases. First-principles calculations offer a more robust tool set for qualitatively and often quantitatively explaining the same changes with greater accuracy [23].

The situation is less complicated in the case of semiconductor surfaces. Here, adsorbate-induced changes in surface stress can be explained in terms of the chemical nature of the adsorbate, the bonding topology of the surface, and atomic size mismatch between the adsorbate and the surface [29]. For example, in the case of an O_2 molecule adsorbing on the unreconstructed Si(111) and Si(100) surfaces, a simple valence force model based on these arguments can be used to estimate the sign *and* magnitude of the adsorbate-induced change in surface stress [12]. The values obtained are found to be in reasonable agreement with experimental data for the reconstructed Si(111)–(7×7) and Si(100)–(2×1) surfaces [12]. It should be noted, however, that this model does not take account of dangling bonds or surface reconstruction, both of which have a profound influence on surface chemistry. Nevertheless, it highlights how adsorbate-induced changes in surface stress on semiconductor surfaces can be explained successfully using arguments based solely on changes in local bonding. In kind with metal surfaces, however, first-principles calculations offer a more robust tool set for explaining the same changes with greater accuracy.

2.3 Measurement

Intrinsic surface stress cannot be measured reliably using existing techniques [23, 41]. While previous attempts to do so have achieved limited success [44–47], reliable, general-purpose measurements remain impossible. Changes in surface stress, on the other hand, can be measured quite easily. This is generally performed using cantilever bending techniques [23]. These techniques are based upon measuring changes in the radius of curvature of a thin cantilever sample bending as a result of adsorbate-induced changes in surface stress. The radius of curvature can be measured by monitoring the change in capacitance between the sample and an adjacent reference electrode [12] or by monitoring the bending-induced deflection of a laser beam using a position-sensitive detector (PSD) [21]. This can be related to the surface stress using the Stoney equation [48]. An illustration of the operating principle is shown in Fig. 2.4.

2.3.1 The Stoney Equation

The radius of curvature of a thin cantilever plate is related to the differential surface stress between its upper and lower surfaces by the expression [49]

$$\sigma = \frac{Et^2}{6(1-\nu)} \cdot \frac{1}{R} \tag{2.2}$$

where σ is the differential surface stress, E is the Young's modulus of the plate, t is the thickness of the plate, ν is Poisson's ratio, and R is the radius of curvature. Eq. (2.2) is known as the Stoney equation [48]. It is valid for plates of uniform thickness composed of isotropic, homogeneous materials [49]. The differential surface stress caused by deposition

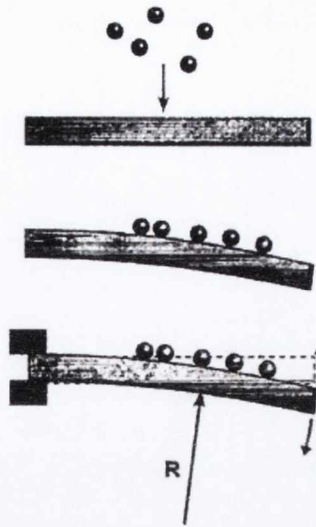


Fig. 2.4. Illustration of the operating principle of the cantilever bending technique [23]. The adsorption of material on the surface of a thin cantilever sample induces changes in surface stress that cause the sample to bend. This changes the radius of curvature R. The radius of curvature can be measured by monitoring the change in capacitance between the sample and an adjacent reference electrode [12] or by monitoring the bending-induced deflection of a laser beam using a position-sensitive detector (PSD) [21]. This can be related to the surface stress using the Stoney equation [48].

of a film of thickness h on one surface of the plate is given by the expression [50]

$$\sigma = \frac{Et^2}{6(1-\nu)h} \cdot \frac{1}{R}. \tag{2.3}$$

This expression is valid for the biaxial stresses caused by most films. Eq. (2.3) can be derived as follows. Consider an external force acting on the surface of an isotropic elastic body. Assume that the force propagates through the surface and into the body. If the external force is opposed by an internal force of equal magnitude, the body cannot undergo elastic deformation. This gives rise to stress. Fig. 2.5 shows an illustration of an internal force dP acting on an infinitesimal surface area dS. dN and dT are the components of dP acting in the planes normal and transverse to the surface respectively, while θ is the angle between dP and the plane normal to the surface. dN and dT can be expressed as

$$dN = dP\cos\theta, \qquad dT = dP\sin\theta.$$

The normal stress σ_n and transverse stress σ_t can be expressed as

$$\sigma_n = \frac{dN}{dS}, \qquad \sigma_t = \frac{dT}{dS}.$$

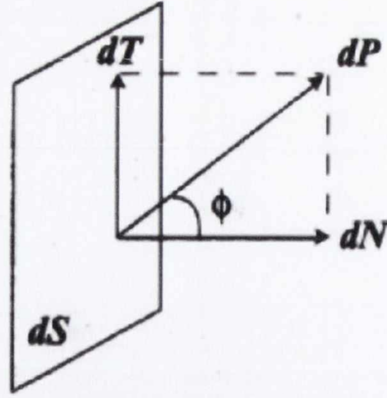


Fig. 2.5. Illustration of an internal force dP acting on an infinitesimal surface area dS. dN and dT are the components of dP acting in the planes normal and transverse to the surface respectively, while θ is the angle between dP and the plane normal to the surface [51].

Now consider an isotropic elastic bar of initial length L and initial diameter d. If the bar is stretched to a length L', the longitudinal strain ε and lateral strain ε' can be expressed as

$$\varepsilon = \frac{L' - L}{L}, \qquad \varepsilon' = \frac{d' - d}{d}.$$

where d' is the diameter of the bar at length L'. Note that strain is a dimensionless quantity. Since the bar is elastic, the diameter decreases with stretching and increases with compression. If the bar is stretched, ε will be positive while ε' will be negative. If the bar is compressed, ε will be negative while ε' will be positive. The ratio of ε and ε' is referred to as Poisson's ratio, v. Now consider two parallel planes intersecting the bar a distance L^* apart. If the bar is displaced along one plane by a distance l, the shear strain γ is given by

$$\gamma = \frac{l}{L^*} = \tan\alpha \tag{2.4}$$

where α is the shear angle between the two planes. If the strain is small, Hooke's law states that stress and the strain will be proportional. The longitudinal and shear strains then become

$$\varepsilon = \frac{\sigma_n}{E}, \qquad \gamma = \frac{\sigma_t}{G} \tag{2.5}$$

where E is Young's modulus and G is the shear modulus [52]. In any isotropic material, there are three orthogonal planes where no shear stresses act. These are called *principal planes*. Normal stresses acting on these planes are called *principal stresses*. The corresponding *principal strains* are given by

$$\varepsilon_{xx} = \frac{\sigma_{xx} - v(\sigma_{yy} + \sigma_{zz})}{E}, \quad \varepsilon_{yy} = \frac{\sigma_{yy} - v(\sigma_{xx} + \sigma_{zz})}{E}, \quad \varepsilon_{zz} = \frac{\sigma_{zz} - v(\sigma_{xx} + \sigma_{yy})}{E}. \tag{2.6}$$

When a film is deposited on a substrate, the film stress is confined to the plane parallel to the interface between the film and the substrate [53]. No stress acts in the direction normal to

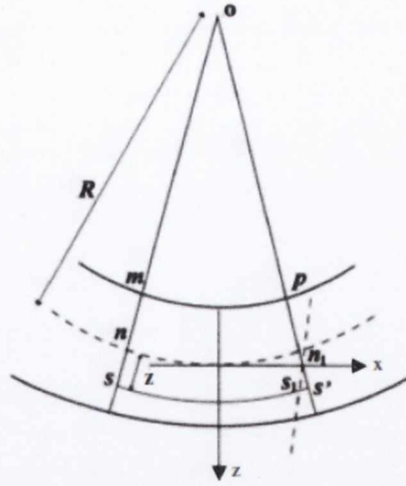


Fig. 2.6. Rectangular element bending under a film stress σ with a radius of curvature R. The dashed section through the centre of the element indicates the neutral axis [51].

this plane. By defining x and y axes in the plane parallel to the interface and a z axis in the plane normal to the interface, the film stresses can be expressed as

$$\sigma = \sigma_{xx} = \sigma_{yy}, \qquad \sigma_{zz} = \sigma_{zy} = \sigma_{xz} = \sigma_{xy} = 0. \tag{2.7}$$

The corresponding film strains can be expressed as

$$\varepsilon_x = \varepsilon_y = \frac{\sigma(1-\nu)}{E}, \qquad \varepsilon_z = \frac{-2\nu\sigma}{E} \tag{2.8}$$

Note that these expressions are valid only if the film stress is uniform. In practice, however, this is rarely true. Fig. 2.6 shows a rectangular element bending under a film stress σ with a radius of curvature R. The dashed section through the centre of the element indicates the *neutral axis*. No stresses act at the neutral axis. Since the triangles formed by n-o-n_1 and s_1-n_1-s' are similar, the film strain ε at a distance z from the neutral axis can be expressed as

$$\varepsilon = \frac{s's_1}{nn_1} = \frac{z}{R} \tag{2.9}$$

The film stress expressed as a function of z is given by

$$\sigma(z) = \varepsilon \cdot \frac{E}{(1-\nu)} = \frac{z}{R} \cdot \frac{E}{(1-\nu)}. \tag{2.10}$$

Two assumptions are now made. First, the film stress is constant if the substrate thickness t greatly exceeds the film thickness h. Second, the substrate stress varies as a function of z. With reference to Fig. 2.7, the balanced equations for the force evaluate to

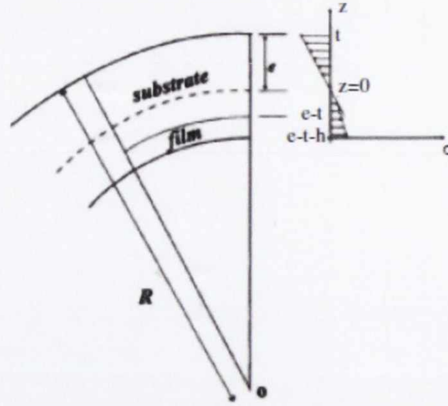


Fig. 2.7. Stress distribution for a film and substrate shown as function of distance z from the neutral axis [51].

$$0 = a\left(\int_{e-t-h}^{e-t} \sigma \, \mathrm{d}z\right) + a\left(\int_{e-t}^{e} \sigma(z) \, \mathrm{d}z\right) \tag{2.11}$$

$$0 = a(\sigma h) + a\left(\frac{Et}{2(1-v)R} \cdot (2e-t)\right) \tag{2.12}$$

where a is the width of the substrate. The balanced equations for the moment evaluate to

$$0 = a\left(\int_{e-t-h}^{e-t} \sigma z \, \mathrm{d}z\right) + a\left(\int_{e-t}^{e} \sigma(z) z \, \mathrm{d}z\right) \tag{2.13}$$

$$0 = a\left(\frac{\sigma h}{2} \cdot [2(e-t)-h]\right) + a\left(\frac{Et}{3(1-v)R} \cdot \left(3e^2 - 3et + t^2\right)\right). \tag{2.14}$$

Combining Eq. (2.11), Eq. (2.12), Eq. (2.13) and Eq. (2.14), the film stress σ evaluates to

$$\sigma = \frac{Et^2}{6(1-v)h} \cdot \frac{1}{R}. \tag{2.15}$$

This is the Stoney equation. The total change in film stress can be expressed as

$$\Delta\sigma = \sigma h = \frac{Et^2}{6(1-v)} \cdot \frac{1}{R}. \tag{2.16}$$

By measuring the change in the radius of curvature of a thin cantilever sample bending under adsorbate-induced surface stress, Eq. (2.16) can be used to calculate the change in surface stress causing the sample to bend.

References

[1] J. W. Gibbs, *The Scientific Papers of J. Willard Gibbs*, vol. 1 (Longmans, Green and Co., 1906).

[2] H. J. Gossmann, J. C. Bean, L. C. Feldman, E. G. McRae, and I. K. Robinson, Phys. Rev. Lett. **55**, 1106?1109 (1985).

[3] F. K. Men, W. E. Packard, and M. B. Webb, Phys. Rev. Lett. **61**, 2469 (1988).

[4] M. B. Webb, F. K. Men, B. S. Swartzentruber, and M. G. Lagally, J. Vac. Sci. Technol. A **8**, 2658 (1990).

[5] M. B. Webb, F. K. Men, B. S. Swartzentruber, R. Kariotis, and M. G. Lagally, Surf. Sci. **242**, 23 (1991).

[6] A. J. Schell-Sorokin and R. M. Tromp, Phys. Rev. Lett. **64**, 1039 (1990).

[7] R. M. Tromp, A. W. D. van der Gon, and M. C. Reuter, Phys. Rev. Lett. **68**, 2313 (1992).

[8] R. M. Tromp, Phys. Rev. B **47**, 7125 (1993).

[9] R. Koch, J. Phys.: Condens. Matter **6**, 9519 (1994).

[10] D. Sander, Rep. Prog. Phys. **62**, 809 (1999).

[11] D. Sekiba, Y. Yoshimoto, K. Nakatsuji, Y. Takagi, T. Iimori, S. Doi, and F. Komori, Phys. Rev. B **75**, 115404 (2007).

[12] D. Sander and H. Ibach, Phys. Rev. B **43**, 4263 (1991).

[13] J. J. Boland, Surf. Sci. **261**, 17 (1992).

[14] J. J. Boland and G. N. Parsons, Science **256**, 1304 (1992).

[15] M. Mavrikakis, B. Hammer, and J. K. Norskov, Phys. Rev. Lett. **81**, 2819 (1998).

[16] C. F. Herrmann, D. Chen, and J. J. Boland, Phys. Rev. Lett. **89**, 096102 (2002).

[17] M. Yata, Y. Uesugi-Saitow, M. Kitajima, A. Kubo, and V. E. Korsukov, Phys. Rev. Lett. **91**, 206103 (2003).

[18] N. T. Kinahan, D. E. Meehan, T. Narushima, S. Sachert, K. Miki, and J. J. Boland, Phys. Rev. Lett. **104**, 146101 (2010).

[19] M. Yata, Phys. Rev. B **81**, 205402 (2010).

[20] B. Gokce, D. Aspnes, and K. Gundogdu, Appl. Phys. Lett. **98**, 121912 (2011).

[21] D. Sander, A. Enders, and J. Kirschner, Rev. Sci. Instrum. **66**, 4734 (1995).

[22] T. Narushima, N. T. Kinahan, and J. J. Boland, Rev. Sci. Instrum. **78**, 053903 (2007).

[23] H. Ibach, Surf. Sci. Rep. **29**, 195 (1997).

[24] P. R. Watson, M. A. Van Hove, and K. Hermann, *NIST Surface Structure Database: Version 2.0* (National Institute of Standards and Technology, Gaithersburg, MD, 1996).

[25] P. T. Sprunger, K. Pohl, H. L. Davis, and E. W. Plummer, Surf. Sci. **297**, L48 (1993).

[26] C. Stampfl, M. Scheffler, and H. Over, Phys. Rev. B **49**, 4959 (1994).

[27] W. Berndt, D. Weick, C. Stampfl, A. M. Bradshaw, and M. Scheffler, Surf. Sci. **330**, 182 (1995).

[28] M. P. J. Punkkinen, S. K. Kwon, J. Kollár, B. Johansson, and L. Vitos, Phys. Rev. Lett. **106**, 057202 (2011).

[29] R. Meade and D. Vanderbilt, Phys. Rev. Lett. **63**, 1404 (1989).

[30] D. Vanderbilt, Phys. Rev. Lett. **59**, 1456 (1987).

[31] O. L. Alerhand, D. Vanderbilt, R. D. Meade, and J. D. Joannopoulos, Phys. Rev. Lett. **61**, 1973 (1988).

[32] M. C. Payne, N. Roberts, R. J. Needs, M. Needels, and J. D. Joannopoulos, Surf. Sci. **211-212**, 1 (1989).

[33] A. García and J. E. Northrup, Phys. Rev. B. **48**, 17350 (1993).

[34] D. Sander, U. Linke, and H. Ibach, Surf. Sci. **272**, 318 (1992).

[35] H. Ibach, J. Vac. Sci. Technol. A **12**, 2240 (1994).

[36] S. E. Trullinger and S. L. Cunningham, Phys. Rev. B **8**, 2622 (1973).

[37] T. S. Rahman and H. Ibach, Phys. Rev. Lett. **54**, 1933 (1985).

[38] W. Daum, S. Lehwald, and H. Ibach, Surf. Sci. **178**, 528 (1986).

[39] A. Grossmann, Ph.D. thesis, RWTH Aachen D80 (1996).

[40] P. Gumbsch and M. S. Daw, Phys. Rev. B **44**, 3934 (1991).

[41] P. J. Feibelman, Phys. Rev. B **56**, 2175 (1997).

[42] Z. Tian, D. Sander, N. N. Negulyaev, V. S. Stepanyuk, and J. Kirschner, Phys. Rev. B **81**, 113407 (2010).

[43] R. J. Needs and M. Mansfield, J. Phys.: Condens. Matter **1**, 7555 (1989).

[44] C. W. Mays, J. S. Vermaak, and D. Kuhlmann-Wilsdorf, Surf. Sci. **12**, 134 (1968).

[45] H. J. Wasserman and J. S. Vermaak, Surf. Sci. **22**, 164 (1970).

[46] H. J. Wasserman and J. S. Vermaak, Surf. Sci. **32**, 168 (1972).

[47] R. E. Martinez, W. M. Augustyniak, and J. A. Golovchenko, Phys. Rev. Lett. **64**, 1035 (1990).

[48] G. G. Stoney, Proc. R. Soc. Lond. A **82**, 172 (1909).

[49] J. E. Sader, J. Appl. Phys. **89**, 2911 (2001).

[50] D. Sander, Z. Tian, and J. Kirschner, Sensors **8**, 4466 (2008).

[51] N. T. Kinahan, Ph.D. thesis, University of Dublin, Trinity College (2009).

[52] K. R. Symon, *Mechanics* (Addison Wesley, 1971), 3rd ed.

[53] R. W. Hoffman, in *Physics of Nonmetallic Thin Films*, edited by C. H. S. Dupay and A. Cachard (Plenum Press, New York, 1976), pp. 273–353.

CHAPTER 3

The Si(111)–(7 × 7) Surface

Silicon crystallizes in a diamond cubic crystal structure. Each atom in the crystal lattice has sp^3-hybridized orbitals and is covalently bonded to four neighbouring atoms. Each bond is 2.35 Å long and has a bond strength of 2.35 eV [1]. When the crystal is cleaved to create a new surface, each surface atom is left with one or more unpaired electrons as a result of reduced coordination. These electrons are localized in so-called *dangling bond* orbitals. The number and spatial orientation of these orbitals depends on the crystallographic plane within which the crystal is cleaved. Dangling bonds increase the free energy of the surface. The surface minimizes this increase in free energy by reducing the number of dangling bonds. This is achieved through a process called *surface reconstruction*. Surface reconstruction involves reorganizing the atomic structure and bonding configuration of the surface in order to maximize bond formation and quench dangling bond electronic states. The formation of new bonds can occur between neighbouring surface atoms or between surface atoms and atoms supplied by diffusion from nearby sources such as steps [2]. Each new bond reduces the free energy of the surface by approximately 1 eV per atom [3], but at the expense of increasing surface stress [4]. This is minimized by relaxing atoms at and beneath the surface. Each atomic relaxation reduces the free energy of the surface by approximately 0.01 eV per atom [4]. Surface reconstruction can be either *activated* or *passive*. Activated surface reconstruction requires energy to occur, whereas passive surface reconstruction does not [2]. The surface obtained following reconstruction depends on the crystallographic plane within which the crystal is cleaved and the technique used to prepare the surface following cleavage [2]. It always corresponds to the lowest energy structure kinetically accessible under the conditions of preparation [2]. The lowest energy structure obtained following cleavage of a silicon crystal in the (111) plane is the Si(111)–(2 × 1) surface [4]. When annealed above 600 °C under ultra-high vacuum (UHV) conditions, it reconstructs irreversibly into the Si(111)–(7 × 7) surface [2]. The Si(111)–(7 × 7) surface is a complex, intrinsically tensile structure [5, 6] characterized by a large unit cell containing 12 adatoms, 6 "rest" atoms, 4 corner holes and 19 dangling bonds [2]. It is arguably one of the most controversial and widely studied structures in the history of surface science. This chapter presents an overview of the Si(111)–(7 × 7) surface, comprising a description of its physical structure and a discussion of its electronic, mechanical and chemical properties.

3.1 Physical structure

Since the first reported observation of the Si(111)–(7×7) surface with low-energy electron diffraction (LEED) [7], numerous models [8–19] have been proposed to explain its structure. The most widely accepted of these is the *dimer–adatom–stacking fault (DAS)* model [18, 19]. This model describes the (7×7) unit cell as a multilayer structure consisting of an unreconstructed (1×1) base layer, a dimer layer, a stacking-faulted rest atom layer, and an adatom layer. Fig. 3.1a shows an illustration of the (7×7) unit cell as described by the DAS model, while Fig. 3.1b shows an unoccupied state scanning tunneling microscope (STM) image of a clean Si(111)–(7×7) surface. STM is discussed in Chapter 1. A layer-by-layer illustration of the unit cell is shown in Fig. 3.2. The unreconstructed (1×1) base layer is shown in Fig. 3.2a. Each atom in this layer has a single dangling bond saturated by an atom in the dimer layer above.

The dimer layer is shown in Fig. 3.2b. Each atom in this layer saturates the dangling bond of an atom in the base layer below. The absence of an atom at each corner of the layer gives rise to so-called corner holes. These are holes in the unit cell structure extending from the top of the (1×1) base layer to the top of the adatom layer. They are present at each corner of the unit cell. Dimers are formed between atoms at the edges of the layer and between atoms at the edges of the short diagonal crossing its centre. These dimers divide the unit cell into two distinct triangular subunits. Each atom forming a dimer has two dangling bonds, while all other atoms have three dangling bonds. These dangling bonds are saturated by atoms in the rest atom layer above.

The rest atom layer is shown in Fig. 3.2c. Each atom in this layer saturates the dangling bonds of 3 atoms in the dimer layer below. In order to saturate all of these dangling bonds, the rest atoms in one triangular subunit must occupy positions out of registry with the positions of the rest atoms in the opposing subunit [16, 17]. This is called a stacking fault. The presence of a stacking fault gives rise to a difference in electronic structure between the faulted and unfaulted halves of the unit cell [20]. This manifests as a contrast difference between the adatoms in the two halves when their occupied electronic states are imaged with an STM [20] as shown in Fig. 3.3b. The unterminated rest atom layer has a total of 42 dangling bonds [1]. It is not possible to saturate each of these using tri-coordinated silicon atoms. In order to best overcome this problem, each of the 12 atoms in the adatom layer above saturates the dangling bonds of 3 atoms in the rest atom layer below. This reduces the number of dangling bonds in the rest atom layer to 6.

The adatom layer is shown in Fig. 3.2d. Each atom in this layer is classified according to its position within the unit cell. Adatoms adjacent to corner holes are called corner adatoms, while adatoms in the region enclosed by the corner adatoms are called centre adatoms. These are further classified as faulted and unfaulted according to which half of the unit cell they reside in. Following reconstruction, the (7×7) unit cell has a total of 19 dangling bonds [1]. 12 of these are localized on the adatoms, 6 on the rest atoms, and 1 is delocalized among the 4 corner holes [1]. In comparison, the (1×1) unit cell has a total of 49 dangling bonds [1].

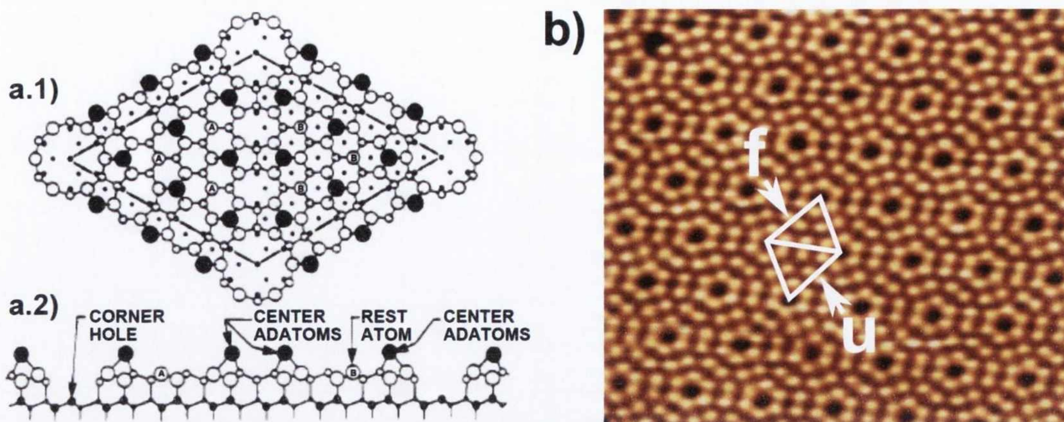


Fig. 3.1. Illustration of the Si(111)–(7 × 7) unit cell as described by the DAS model, showing: (a.1) Top view: filled circles represent adatoms, while open circles marked 'A' and 'B' represent rest atoms in the faulted and unfaulted halves of the unit cell respectively. Atoms increasing in distance away from the surface are denoted by circles of decreasing size. (a.2) Side view (cut along the diagonal of the unit cell): Atoms in the plane of the cut are denoted by larger circles than those behind the plane [18, 19]. (b) Clean Si(111)–(7 × 7) surface. Faulted and unfaulted halves of the (7 × 7) unit cell are labeled 'f' and 'u' respectively.

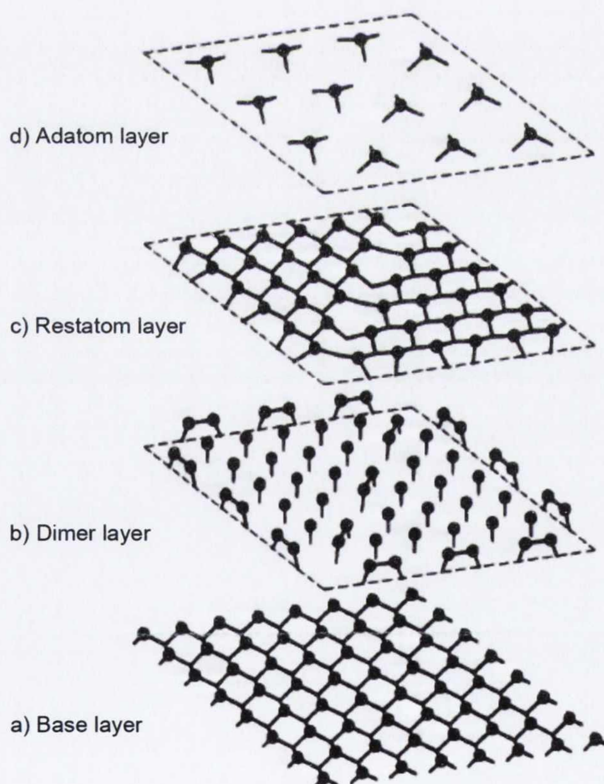


Fig. 3.2. Illustration of the layer-by-layer construction of the Si(111)–(7 × 7 unit cell as described by the DAS model showing (a) the unreconstructed (1 × 1) base layer; (b) the dimer layer; (c) the rest atom layer; and (d) the adatom layer. The dashed lines indicate the boundaries of the (7 × 7) unit cell [1].

3.2 Electronic properties

The Si(111)–(7 × 7) unit cell has a total of 19 dangling bonds [1]. 12 of these are localized on the adatoms, 6 on the rest atoms, and 1 is delocalized among the 4 corner holes [1]. Each adatom dangling bond is partially depleted due to charge transfer to the rest atoms in the underlying layer [21, 22]. Of these, centre adatom dangling bonds experience the greatest level of depletion [21, 22]. This is because centre adatoms have 2 rest atom neighbours, whereas corner adatoms have only 1. The presence of a stacking fault contributes additional charge to the adatom dangling bonds in the faulted half of the unit cell [20]. This manifests as a contrast difference between the adatoms in the two halves when their occupied electronic states are imaged with an STM [20].

Fig. 3.3a shows an unoccupied state STM image of the (7 × 7) unit cell acquired at a sample voltage of 2.0 V. Both adatoms and corner holes are clearly visible. Corner adatoms are labeled 'Co', while centre adatoms are labeled 'Ce'. Faulted and unfaulted halves of the unit cell are labeled 'F' and 'U' respectively. Fig. 3.3b shows a dI/dV map of the same area highlighting the adatom dangling bond state 0.35 eV below the Fermi energy. Corner adatoms appear brighter than centre adatoms, consistent with greater depletion of the centre adatom dangling bonds. Adatoms in the faulted half of the unit cell appear brighter than adatoms in the unfaulted half, consistent with additional charge contribution from the stacking fault.

Fig. 3.3c shows a dI/dV map of the same area highlighting the rest atom dangling bond state 0.8 eV below the Fermi energy. Both rest atoms and corner holes appear bright. Bright corner holes are associated with the delocalized corner hole dangling bond state. Fig. 3.3d shows a dI/dV map of the same area highlighting the backbond state 1.7 eV below the Fermi energy. Adatoms appear dark, while the adjoining backbonds appear bright. Bright corner holes associated with the delocalized corner hole dangling bond state are visible once again. It is clear from each of these images that the (7 × 7) unit cell possesses several electronically inequivalent sites. This diversity is strongly reflected in the chemistry of the surface.

3.3 Mechanical properties

The unreconstructed Si(111)–(1 × 1) surface is an intrinsically compressive structure [5, 26, 27]. This is explained by the presence of a singly-occupied dangling bond on each adatom [27]. The reduced occupation of this dangling bond favours re-hybridization of the adatom orbitals from the tetrahedral sp^3 configuration of the bulk lattice to a quasi-planar sp^2 configuration. This causes the adatom to move downwards. Since the Si–Si bond length does not change significantly during re-hybridization, the Si atoms in the underlying layer are pushed outwards. This results in compressive surface stress. In contrast, the Si(111)–(7 × 7) surface is an intrinsically tensile structure [5, 6, 27]. This is explained by assessing the individual stress contributions made by each unit cell component.

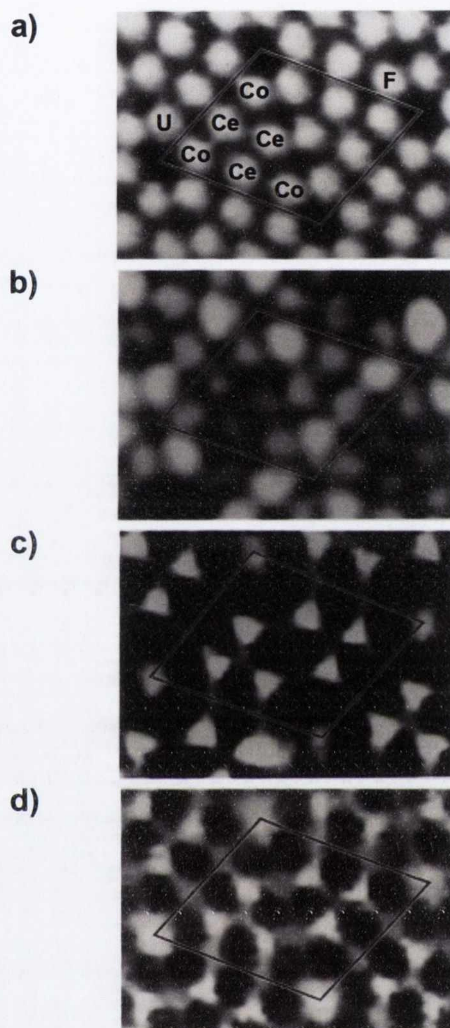
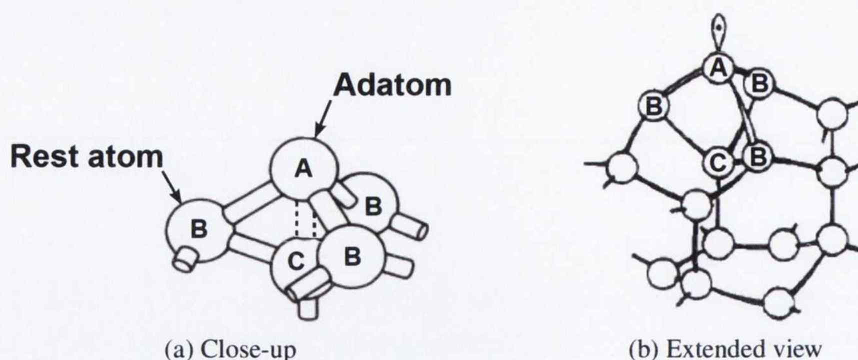


Fig. 3.3. STM and current-imaging tunneling spectroscopy (CITS) images of the Si(111)–(7 × 7) unit cell [23], showing the surface topography and spatial location of different surface electronic states: (a) topographic image acquired at a sample voltage of 2.0 V, with corner and centre adatoms labeled 'Co' and 'Ce' respectively, while faulted and unfaulted halves of the unit cell are labeled 'F' and 'U' respectively; (b) d*I*/d*V* image showing the adatom dangling bond state 0.35 eV below the Fermi energy; (c) d*I*/d*V* image showing the rest atom dangling bond state 0.8 eV below the Fermi energy; and (d) d*I*/d*V* image showing the backbond state 1.7 eV below the Fermi energy. Note that the electronic features observed in (b), (c) and (d) are consistent with the results of photoemission spectroscopy (PES) measurements [24, 25].

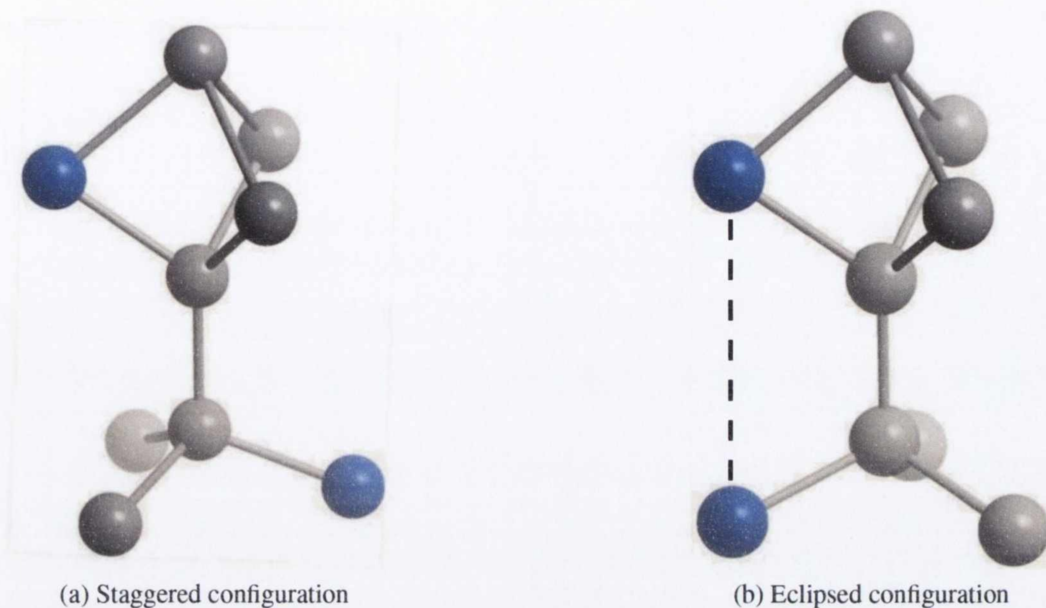


(a) Close-up (b) Extended view

Fig. 3.4. Illustration of the Si(111)–(7 × 7) unit cell adatom site, showing (a) a close-up of the adatom bonding structure [22]; and (b) an extended view of the same structure highlighting the four-member rings attached to the adatom [28]. The adatom is labeled 'A', the rest atoms are labeled 'B' and the dimer layer atom lying below the adatom is labeled 'C'. The dashed line between in (a) indicates a proposed weak bonding interaction between the adatom and the dimer layer atom [22].

First, adatoms make a tensile contribution [5, 27]. This arises due to a repulsive interaction between each adatom and the dimer layer atom lying directly below it. Their relative positions are shown in Fig. 3.4a. The driving force behing the interaction is Coulombic repulsion or the energy reduction associated with relieving the unnatural bond angles between the rest atoms [27]. Repulsion between the two atoms increases the length of the backbonds. Since the new bond length is *greater* than the ideal bond length, this results in tensile stress [27]. The rest atoms also move radially inwards, pulling the adatoms closer together [5, 5]. Additional tensile stress may result from the participation of each adatom in three four-member rings. The four- and five-member rings structures of the (7 × 7) unit cell are shown in Fig. 3.4b. Each four-member ring introduces backbond strain and forces the adatom to move closer to the dimer layer atom [22]. This increases Coloumbic repulsion between the two atoms, thus lengthening the backbonds. A proposed weak bonding interaction between the two atoms [22] would likely exacerbate the effect further by removing charge from the backbonds.

Second, the stacking fault makes a tensile contribution [5, 27]. This arises due to Coulomic repulsion between bonds in the two lower layers of the unit cell. Referring to Fig 3.5a, the bonds adjoining each of the blue-coloured atoms in the unfaulted half of the unit cell are staggered [27]. In the faulted half, these bonds are eclipsed [27], as shown in Fig 3.5b. This causes Coulombic repulsion, forcing the adatom to move upwards with bond angles of 112° and resulting in tensile stress [27]. Third, the dimer walls make a tensile contribution [5, 27]. Corner holes also make a tensile contribution [5, 27]. When the individual stresses of the unit cell components are considered together, it appears reasonable that the Si(111)–(7 × 7) surface should be intrinsically tensile. This is also in agreement with experimental measurements [6]. Calculated surface stresses for the (1 × 1) unit cell and the (2 × 2) islands modeling the halves of the (7 × 7) unit cell are shown in Table 3.1.



(a) Staggered configuration (b) Eclipsed configuration

Fig. 3.5. Illustration of the staggered and eclipsed bonding configurations found in the unfaulted and faulted halves of the unit cell respectively. The dashed line between the blue-coloured atoms in (b) indicates repulsive Coloumbic interaction.

Table 3.1. Calculated surface stresses σ_{ij} for Si(111) surfaces.

Structure	σ_{ij} from Ref. 5* [eV / (1 × 1) unit cell]	σ_{ij} from Ref. 27 [eV / (1 × 1) unit cell]
1×1	−0.5	−0.54
1×1 (faulted)	0.0	0.11
2×2 (adatom-covered)	1.8	1.66
2×2 (adatom-covered + faulted)	2.2	1.89
2×2 (corner hole)	–	1.18

* Additional calculations performed in Ref. 5 estimate the per–dimer and per–corner hole contributions to the (7×7) unit cell stress to be 5.1 eV and 4.2 eV respectively.

3.4 Chemical properties

In Sections 3.2 and 3.3, it was established that the Si(111)–(7 × 7) unit cell possesses both electronically *and* mechanically inequivalent sites. Consequently, the chemistry of the Si(111)–(7 × 7) surface is driven not only by the saturation of dangling bonds, but also by the relief of bond strain. The atomic hydrogen reaction, for example, is driven by the relaxation of strained surface bonds [29–31]. The NH_3 reaction, on the other hand, is driven by the saturation of dangling bonds [21, 22]. Both reactions highlight the wide variety of chemically active sites available for reaction on the Si(111)–(7 × 7) surface. Atomic hydrogen, for example, reacts preferentially with the strained adatom backbonds to form adatom trihydride species with one unstrained backbond [29–31]. NH_3, on the other hand, reacts preferentially with the rest atoms, followed by the adatoms, to form $Si-NH_2$ and $Si-H$ species [21, 22]. This can be explained by the greater occupation of the rest atom dangling bonds relative to the adatom dangling bonds.

NH_3 exhibits a further preference for reacting with centre adatoms over corner adatoms [21, 22], although the reasons for this are less clear. It has been shown that the reaction of a rest atom results in reverse charge transfer from the rest atom to the adatom dangling bonds [21, 22]. This increases the occupation of the adatom dangling bonds to similar levels [21, 22]. In spite of this, the preferential reaction of centre adatoms has been found to persist [22]. This suggests that the origin of the observed site selectivity is not electronic. One proposed explanation is that reaction at one site introduces more strain into the dimer bonds than reaction at the other site [22]. Inspection of the DAS model [18, 19] indicates that reaction with a corner adatom would strain *two* dimer bonds, whereas reaction with a centre adatom would strain only *one* dimer bond [22]. In this case, the reaction would be favoured at the centre adatom due to the smaller dimer strain involved.

Another example of site-selective reactivity is observed in the reaction of the surface with O_2 [32]. Here, O_2 molecules are found to preferentially react with adatoms in the faulted half of the unit cell. Furthermore, the insertion of oxygen atoms in the adatom backbonds is found to relieve the intrinsic tensile stress of the surface [32]. This relief is found to be greatest in the faulted half of the unit cell [32], consistent with the known larger tensile stress of the faulted half [5, 27]. Numerous other examples of site-selective reactivity on the Si(111)–(7 × 7) surface can be found throughout the literature [1, 33, 34].

It is possible to draw some general conclusions from the previous observations. First, the factors that drive surface reconstruction, namely the minimization of surface energy, the maximization of bond formation and the minimization of surface stress, also drive surface reactions. The products of these reactions are the balanced outcome of competition between each of these factors. Second, surface reactions are driven not only by the saturation of dangling bonds, but also by the relief of bond strain. Lastly, adsorbates often exhibit a preference for reacting at specific surface sites. For a more detailed discussion of chemical reactions on the Si(111)–(7 × 7) surface, several excellent resources are recommended [1, 33, 34].

References

[1] H. N. Waltenburg and J. T. Yates, Chem. Rev. **95**, 1589 (1995).

[2] J. A. Kubby and J. J. Boland, Surf. Sci. Rep. **26**, 61 (1996).

[3] D. J. Chadi, Ultramicroscopy **31**, 1 (1989).

[4] C. B. Duke, Appl. Surf. Sci. **65-66**, 543 (1993).

[5] D. Vanderbilt, Phys. Rev. Lett. **59**, 1456 (1987).

[6] R. E. Martinez, W. M. Augustyniak, and J. A. Golovchenko, Phys. Rev. Lett. **64**, 1035 (1990).

[7] R. E. Schlier and H. E. Farnsworth, J. Chem. Phys. **30**, 917 (1959).

[8] J. D. Levine and S. H. McFarlane, Phys. Rev. B **16**, 5415 (1977).

[9] L. C. Snyder, Z. Wasserman, and J. W. Moskowitz, J. Vac. Sci. Technol. **16**, 1266 (1979).

[10] D. J. Chadi, R. S. Bauer, R. H. Williams, G. V. Hansson, R. Z. Bachrach, J. C. Mikkelsen, Jr., F. Houzay, G. M. Guichar, R. Pinchaux, and Y. Pétroff, Phys. Rev. Lett. **44**, 799 (1980).

[11] J. C. Phillips, Phys. Rev. Lett. **45**, 905 (1980).

[12] M. J. Cardillo, Phys. Rev. B **23**, 4279 (1981).

[13] J. Pollmann, Phys. Rev. Lett. **49**, 1649 (1982).

[14] G. Binnig, H. Rohrer, C. Gerber, , and E. Weibel, Phys. Rev. Lett. **50**, 120 (1983).

[15] F. J. Himpsel, Phys. Rev. B **27**, 7782 (1983).

[16] E. G. McRae, Phys. Rev. B **28**, 2305 (1983).

[17] P. A. Bennett, L. C. Feldman, Y. Kuk, E. G. McRae, and J. E. Rowe, Phys. Rev. B **28**, 3656 (1983).

[18] K. Takayanagi, Y. Tanishiro, S. Takahashi, and M. Takahashi, Surf. Sci. **164**, 367 (1985).

[19] K. Takayanagi, Y. Tanishiro, M. Takahashi, and S. Takahashi, J. Vac. Sci. Technol. A **3**, 1502 (1985).

[20] R. M. Tromp, R. J. Hamers, and J. E. Demuth, Phys. Rev. B **34**, 1388 (1986).

[21] R. Wolkow and P. Avouris, Phys. Rev. Lett. **60**, 1049 (1988).

[22] P. Avouris and R. Wolkow, Phys. Rev. B **39**, 5091 (1989).

[23] R. J. Hamers, R. M. Tromp, and J. E. Demuth, Phys. Rev. Lett. **56**, 1972 (1986).

[24] T. Fauster and F. J. Himpsel, J. Vac. Sci. Technol. A **1**, 1111 (1983).

[25] F. J. Himpsel and T. Fauster, J. Vac. Sci. Technol. A **2**, 815 (1984).

[26] R. D. Meade and D. Vanderbilt, Phys. Rev. Lett. **63**, 1404 (1989).

[27] R. D. Meade and D. Vanderbilt, Phys. Rev. B **40**, 3905 (1989).

[28] I. Ohdomari, Surf. Sci. **227**, L125L129 (1990).

[29] J. J. Boland, J. Phys. Chem. **95**, 1521 (1991).

[30] J. J. Boland, Surf. Sci. **244**, 1 (1991).

[31] J. J. Boland, J. Vac. Sci. Technol. B **9**, 764 (1991).

[32] N. T. Kinahan, D. E. Meehan, T. Narushima, S. Sachert, K. Miki, and J. J. Boland,

Phys. Rev. Lett. **104**, 146101 (2010).

[33] J. T. Yates, J. Phys.: Condens. Matter **3**, S143 (1991).

[34] R. J. Hamers and Y. Wang, Chem. Rev. **96**, 1261 (1996).

CHAPTER 4

Development of a Combined Scanning Tunneling Microscopy and Surface Stress Measurement System

In Chapter 2, it was established that an adsorbate reacting with a surface could induce changes in surface stress. In the case of semiconductor surfaces, for example, these changes could be explained in terms of the chemical nature of the adsorbate, the bonding topology of the surface, and atomic size mismatch between the surface and the adsorbate. It was further shown that changes in surface stress could be measured. This enables the observation of mechanical forces acting at a surface during a chemical reaction. Since these forces arise due to changes in chemical bonding, their observation provides useful information about the underlying reaction mechanism. When considered alone, however, the utility of this information is limited. Surface stress is a *macroscopic* quantity, while its origins lie in structural and electronic changes occurring on the *atomic* scale. In order to maximize its value as a probe of chemical reactivity, it must be directly relatable to these changes. To this end, a combined scanning tunneling microscopy (STM) and surface stress measurement (SSM) system has been developed [1]. This system allows adsorbate-induced changes in surface stress to be measured and related to the structural and electronic changes causing them *in situ* and with atomic resolution. It consists of four components. These include (*i*) a high-resolution surface stress measurement system based on the capacitive cantilever bending technique [2]; (*ii*) a non-contact sample heating system capable of locally heating large, free-standing cantilever samples; (*iii*) a sample pushing system capable of inducing tensile and compressive stresses in the sample surface; and (*iv*) an ultra-high vacuum (UHV)-compatible STM. The first three components are integrated into a custom-designed sample holder. This, in turn, integrates into the STM. A front-view image of the assembled system is shown in Fig. 4.1. This chapter presents a detailed description of the design, construction and performance of the combined STM/SSM system. Note that this chapter contains material based in part on material previously published in Ref. 1.

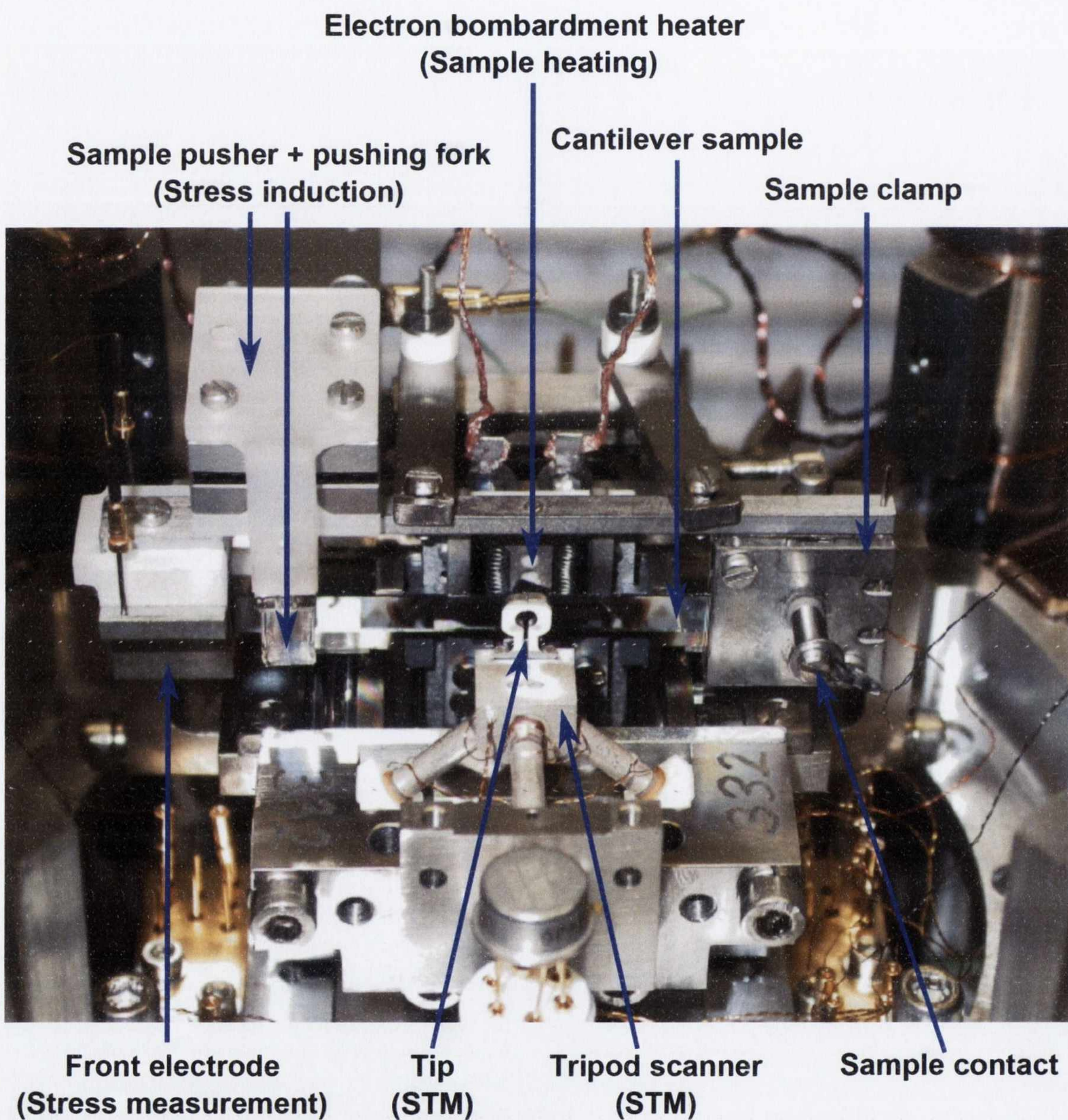


Fig. 4.1. Front view of the assembled RT-STM/SSM system.

4.1 Design

The RT-STM/SSM system consists of four components. These include (*i*) a surface stress measurement system; (*ii*) a non-contact sample heating system; (*iii*) a sample pushing system; and (*iv*) a UHV-compatible STM in the form of an Omicron STM-1 [3]. The first three components are integrated into a custom-designed sample holder. This, in turn, integrates into the STM. This section discusses the design of the system in terms of the sample holder, the components it incorporates, and some of the factors that must be considered in order to produce a functional system.

4.1.1 Sample holder

The sample holder is responsible for integrating the surface stress measurement, sample heating and sample pushing systems into a single platform that can be easily installed into the STM. Fig 4.2 shows an annotated 3-dimensional model of the sample holder viewed from the front-left side. A corresponding real-world image is shown in Fig. 4.3. Materials are colour coded, with green representing Super Invar, light blue representing quartz, dark blue representing tantalum, grey representing molybdenum, and light purple representing silicon.

The body of the sample holder is composed of Super Invar. Super Invar is a strong, UHV-compatible nickel-steel alloy with a low coefficient of thermal expansion across a wide temperature range ($\sim 0.4 \times 10^{-6}$ K^{-1} at room temperature). The sample clamp is a composite structure consisting of Super Invar and fused quartz layers arranged in an Invar–quartz–sample–quartz–Invar configuration. Fused quartz is an electrically-insulating ceramic. Like Super Invar, it has a low coefficient of thermal expansion across a wide temperature range ($\sim 0.5 \times 10^{-6}$ K^{-1} at room temperature). Since both materials have similar coefficients of thermal expansion, their combination minimizes unwanted movement of the sample due to thermal expansion of the sample clamp.

Fig 4.2 shows a rectangular silicon sample clamped at the right-hand side of the sample holder. The sample shown is 60 mm long $\times$ 10 mm wide $\times$ 0.28 mm thick, although samples 5 mm wide are more commonly used. This is due to the decreased outgassing and lower energy requirement associated with heating smaller volumes of sample material. The area of the sample enclosed by the sample clamp is 10 mm long $\times$ 10 mm wide. This produces a 50 mm-long cantilever sample.

The free end of the sample lies between two flat tantalum electrodes. Each electrode is 10 mm long $\times$ 10 mm wide $\times$ 1 mm thick and oriented parallel to the sample surface. The distance between the sample clamp and the centre of each electrode is 45 mm, while the distance between the sample and each electrode is $\sim$ 0.85 mm. By measuring the change in capacitance between the sample and one of the electrodes, changes in the radius of curvature of the sample can be detected. The remaining electrode is used as a reference electrode. This forms the basis of the surface stress measurement system described in Section 4.1.2.

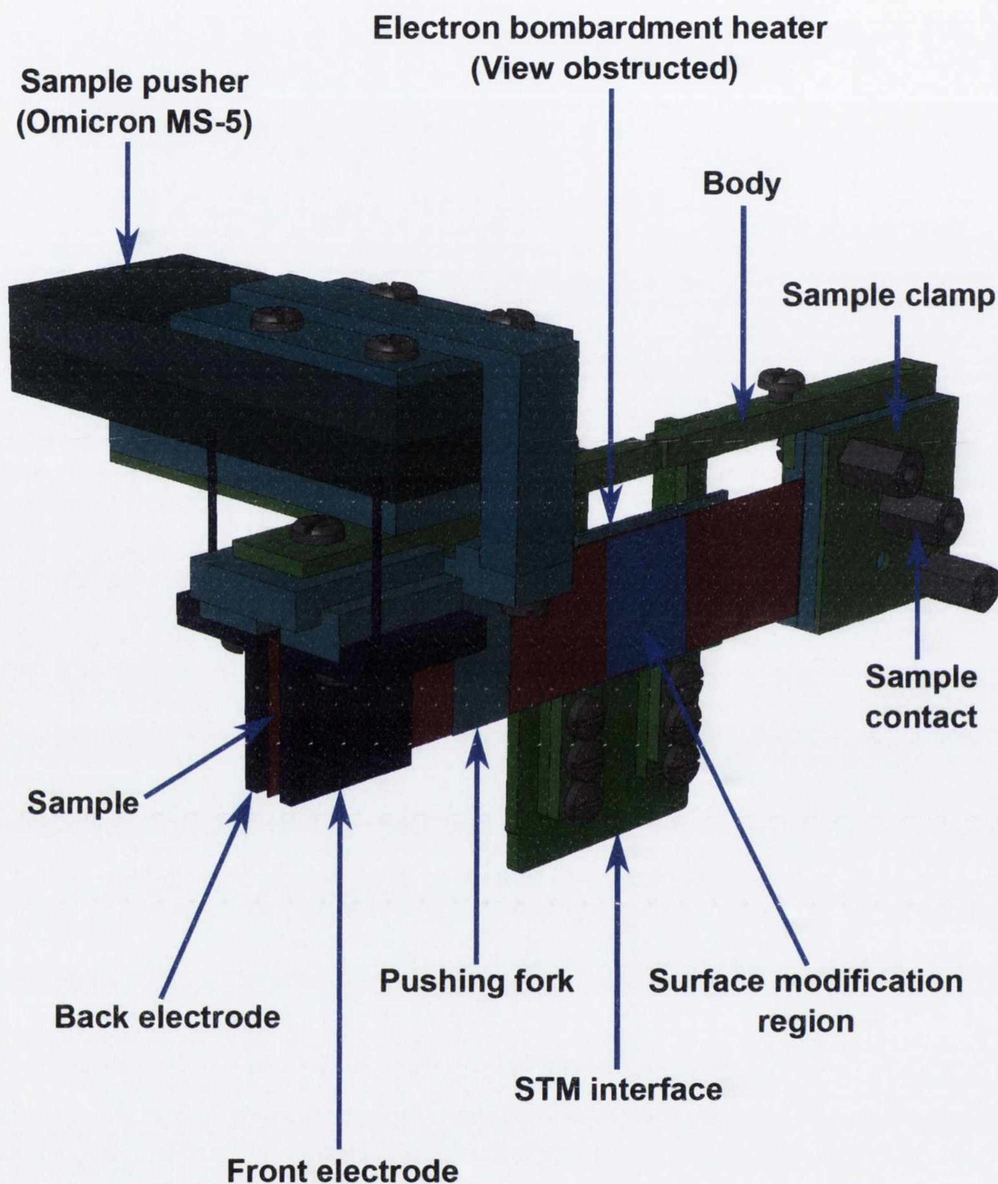


Fig. 4.2. 3-dimensional model of the RT-STM/SSM sample holder. Materials are colour coded, with green representing Super Invar, light blue representing quartz, dark blue representing tantalum, grey representing molybdenum, and light purple representing silicon.

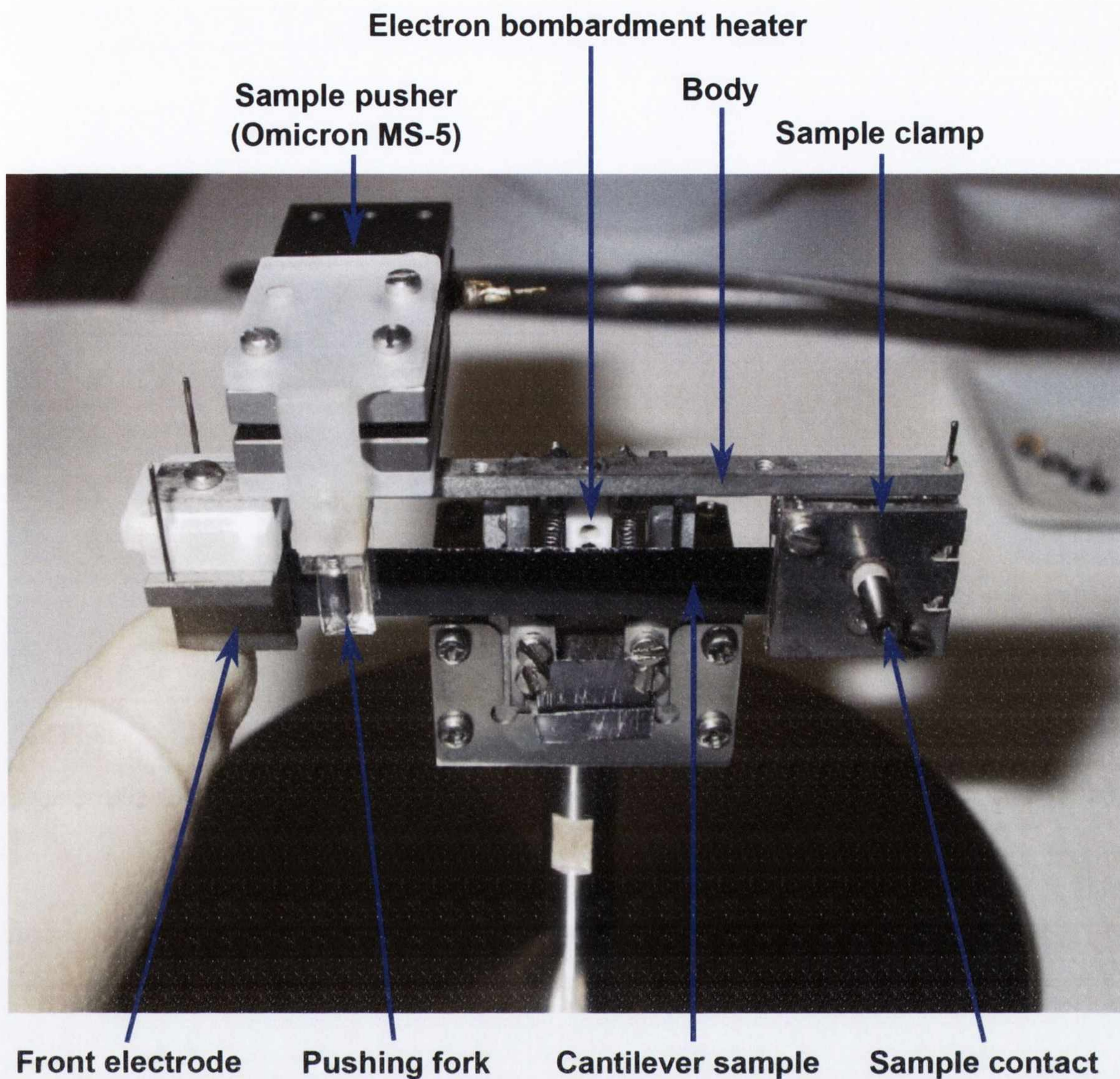


Fig. 4.3. Front view of the RT-STM/SSM sample holder.

An electron bombardment heater is centred 4 mm behind the sample at a point 15 mm away from the sample clamp. This consists of a T-shaped ceramic frame supporting two thoriated tungsten filaments, each of which is connected in series. When current is passed through the filaments, their temperature increases. If the temperature becomes high enough, thermionic emission [4–7] occurs. If a sufficiently large positive voltage is applied to the sample, the emitted electrons will accelerate towards and collide with it. Upon impact, the kinetic energy of the electrons is converted to heat energy and transferred to the sample. This forms the basis of the sample heating system described in Section 4.1.3.

A sample pusher is centred above the sample at a point 35 mm away from the sample clamp. This consists of a two-pronged quartz fork attached to an Omicron MS-5 piezoelectric motor [8]. The sample passes between the prongs of the fork at the 35 mm point. By moving the fork with the piezoelectric motor, the sample can be pushed backward and forward in the direction perpendicular to the sample surface. This forms the basis of the sample pushing system described in Section 4.1.4.

4.1.2 Surface stress measurement

Surface stress is measured using the capacitive cantilever bending technique [2]. This technique is based upon measuring changes in the radius of curvature of a thin cantilever sample bending as a result of adsorbate-induced changes in surface stress. The radius of curvature is measured by monitoring the change in capacitance between the sample and an adjacent electrode. This can be related to the total change in surface stress $\Delta\sigma$ using the modified Stoney equation derived in Section 2.3.1, which is given by

$$\Delta\sigma = \frac{Et^2}{6(1-v)R} \tag{4.1}$$

where E is the Young's modulus of the sample, t is the thickness of the sample, v is Poisson's ratio, and R is the radius of curvature.

In the context of the RT-STM/SSM, the radius of curvature is measured by monitoring the change in capacitance between a *low-resistivity sample* and one of the electrodes shown in Fig. 4.2. This is performed using an Andeen-Hagerling AH2550A capacitance bridge [9]. The remaining electrode is used as a *reference electrode*. This allows spurious or parasitic contributions to the measured capacitance to be detected. The effects of parasitic capacitance and the importance of using low-resistivity samples are discussed in Section 4.2.1.

Fig. 4.4 shows an illustration of a cantilever sample with its free end lying between two electrodes similar to those shown in Fig. 4.2. Here, L_0 is the distance between the sample clamp and the reaction area, L is the length of the reaction area, L' is the distance between the reaction area and the centre of the two electrodes, θ is the angle of curvature, R is the radius of curvature, and d_0 is the initial distance between the sample and the upper electrode. The initial capacitance $C(d_0)$ between the sample and the upper electrode is given by

$$C(d_0) = \frac{\varepsilon_0 A}{d_0} \tag{4.2}$$

where ε_0 is the permittivity of vacuum, and A is the area of the plates forming the capacitor. The change in capacitance $\Delta C(\Delta d)$ caused by a free-end deflection Δd is given by

$$\Delta C(\Delta d) = C(d_0 + \Delta d) - C(d_0) \approx \frac{\varepsilon_0 A}{d_0^2}\Delta d \tag{4.3}$$

Note that this expression is valid only for values of $\Delta d \ll d_0$ [1, 10]. Referring to Fig. 4.4, the free-end deflection Δd is given by

$$\Delta d = l\sin\frac{\theta}{2} + L'\sin\theta \approx \left(\frac{L^2}{2} + LL'\right)\frac{1}{R} \tag{4.4}$$

where l is the arc of a circle of radius R and L is its chord. Note that this expression is valid only if $l \approx L$, which is true when $\theta \ll \frac{\pi}{2}$. Combining Eq. (4.1), Eq. (4.2), Eq. (4.3) and

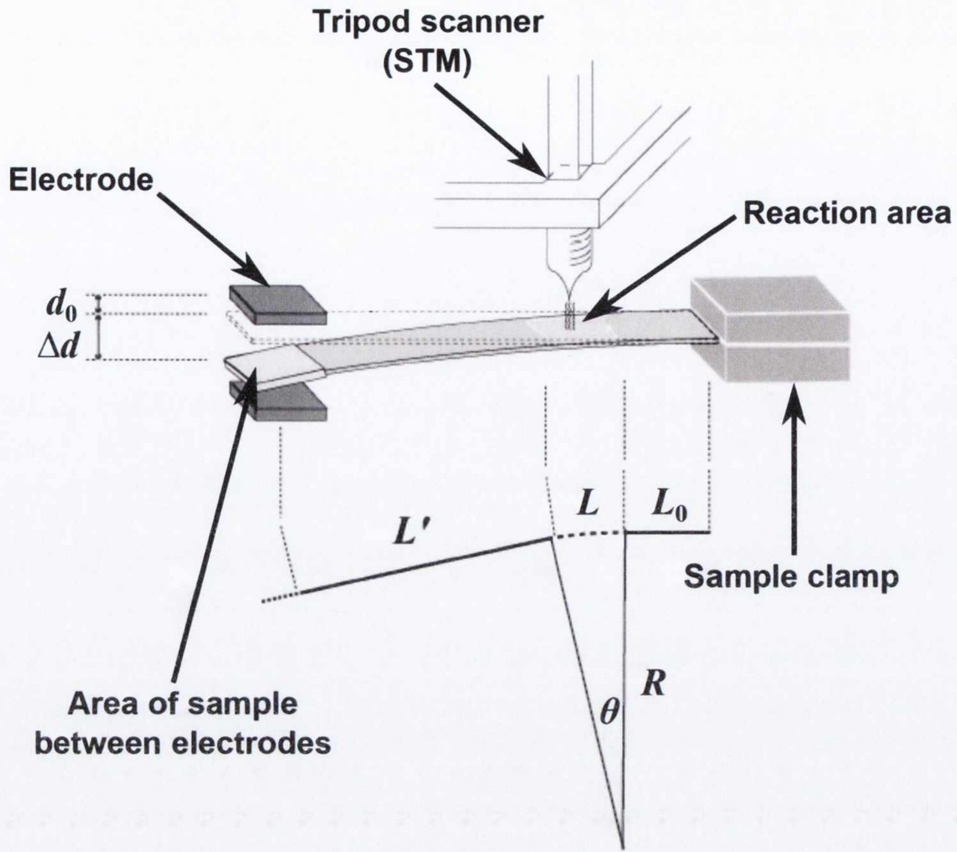


Fig. 4.4. Illustration of a cantilever sample with its free end lying between two flat electrodes. An STM tip is shown situated above the reaction area of the surface in a configuration similar to that of the RT-STM/SSM. L_0 is the distance between the sample clamp and the reaction area, L is the length of the reaction area, L' is the distance between the reaction area and the centre of the two electrodes, θ is the angle of curvature, R is the radius of curvature, d_0 is the initial distance between the sample and the upper electrode, and Δd is the change in distance between the sample and the upper electrode following an adsorbate-induced change in surface stress [1].

Eq. (4.4), the total change in surface stress $\Delta\sigma$ evaluates to

$$\Delta\sigma = \frac{Et^2}{6(1-\nu)} \cdot \left(\frac{1}{\frac{L^2}{2} + LL'} \right) \cdot \left(\frac{\varepsilon_0 A}{[C(d_0)]^2} \right) \cdot \Delta C(\Delta d). \tag{4.5}$$

Assuming $t = 0.525 \times 10^{-3}$ m, $[E/(1-\nu)] = 2.29 \times 10^{11}$ N m^{-1} [11], $L = 1.0 \times 10^{-2}$ m, $L' = 2.5 \times 10^{-2}$ m, $\varepsilon_0 = 8.854 \times 10^{-12}$ F m^{-1}, $A = 10^{-4}$ m^2, $C(d_0) = 3.0$ pF and a capacitance resolution of 0.5-0.8 aF [12], the theoretical deflection resolution and surface stress resolution calculated using the equations above are 0.71 Å and 2.48×10^{-3} N m^{-1} respectively. Experimentally, these values are reduced to 1.87 Å and 6.55×10^{-3} N m^{-1} due to the presence of capacitance measurement noise limiting the capacitance resolution to 1.9 aF. Note that the value chosen for $C(d_0)$ is an average value accounting for variations in the initial capacitance measured depending on the clamping configuration of the cantilever sample.

4.1.3 Sample heater

Sample heating is accomplished by localized, non-contact electron bombardment of the back side of the sample. This is performed using the electron bombardment heater shown in Fig. 4.5. The electron bombardment heater consists of a T-shaped ceramic frame supporting two thoriated tungsten filaments, each of which is connected in series. This structure is mounted 4 mm behind the sample at a point 15 mm away from the sample clamp. The frame is composed of Macor [13]. Macor is a UHV-compatible, electrically-insulating machinable glass ceramic with a low coefficient of thermal expansion across a wide temperature range ($\sim 3.4 \times 10^{-8}$ K^{-1} at room temperature). The use of Macor minimizes unwanted movement of the filaments due to thermal expansion of the frame. It also provides electrical isolation from the sample holder body.

Filaments are produced by coiling 0.15 mm-diameter 1%-thoriated tungsten wire around the grooves of a threaded rod. The filament length is determined by the thread pitch and the number of windings, while the coil diameter is determined by the thread diameter. Each filament is typically 8 mm long and has a coil diameter of 2 mm. Following production, the filaments are secured to the frame using a combination of M1 threaded nuts and rods. These are composed of molybdenum. Molybdenum is used due to the extreme temperatures produced while the filaments are operating. Once secured to the frame, the distance between the filaments is 6 mm, while the distance between the filaments and the sample is 2 mm. The reasons for using thoriated tungsten filaments are discussed later.

When current is passed through the filaments, their temperature increases. If the temperature becomes high enough, electrons near the filament surface gain sufficient energy to overcome the work function and escape into the vacuum. This process is called thermionic emission [5–7]. If a sufficiently large positive voltage is applied to the sample, the emitted electrons will accelerate towards and collide with it. This process is called electron bombardment. Upon impact, the kinetic energy of the electrons is converted to heat energy and transferred to the sample. In the context of the heater described above, typical values for the current and voltage are 1.2-1.9 A and 1000 V respectively. The current density J emitted by a metal filament undergoing thermionic emission is given by [7]

$$J = AT^2 \exp\left(\frac{-\phi}{kT}\right) \tag{4.6}$$

where T is the filament temperature, ϕ is the work function of the metal, k is the Boltzmann constant, and A is a constant given by the expression [7]

$$A = \frac{4\pi mk^2 e}{h^3} = 1.20173 \text{ A m}^{-2} \text{ K}^{-2} \tag{4.7}$$

where m is the mass of the electron, e is the elementary charge, and h is Planck's constant. Eq. (4.6) shows that the current density can be maximized by increasing the temperature or decreasing the work function. The work function can be decreased in two ways. First, an

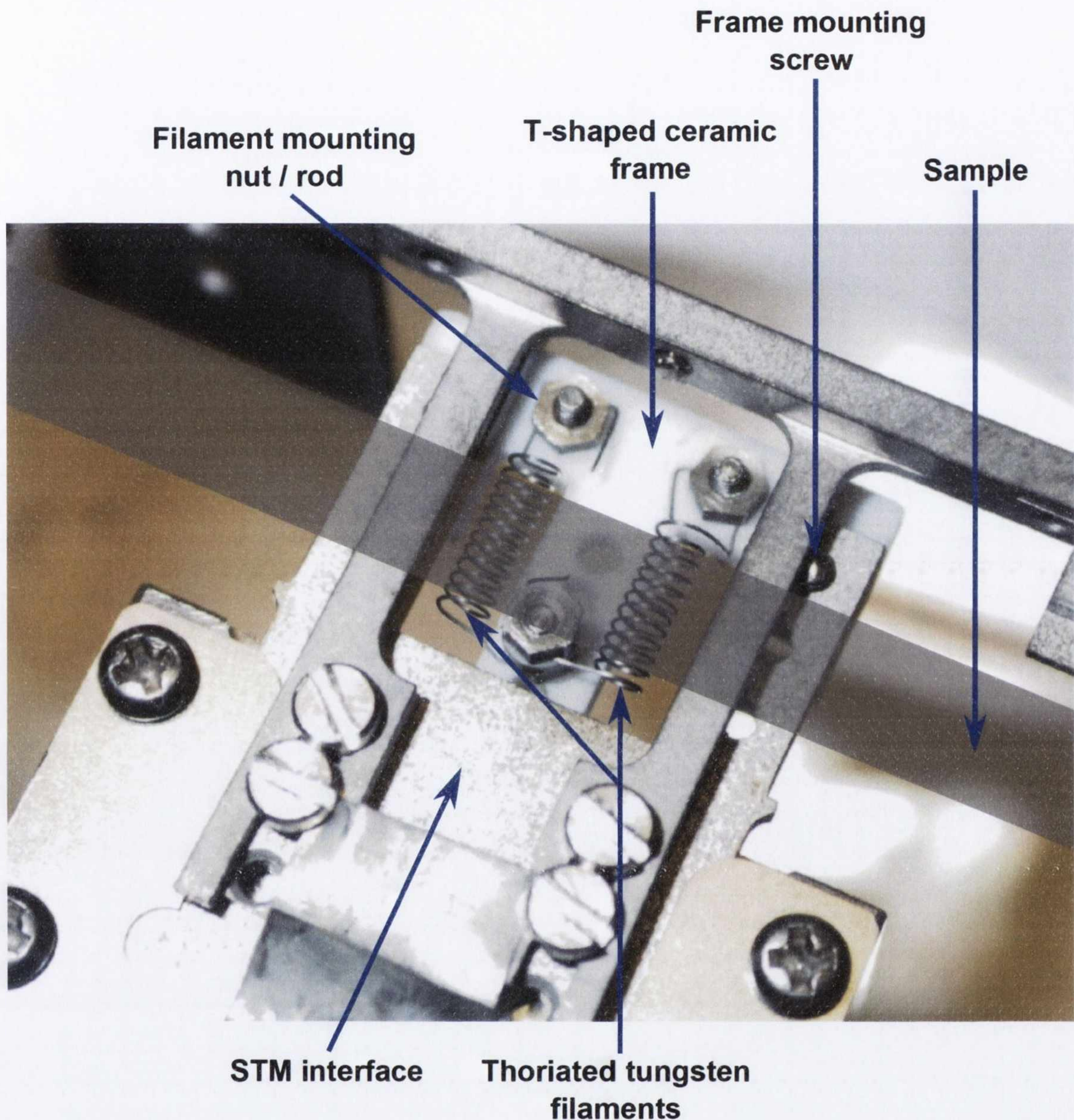


Fig. 4.5. Front-right view of the electron bombardment heater. A T-shaped ceramic frame is shown supporting two thoriated tungsten filaments, each of which is connected in series. It is secured to the sample holder body by two molybdenum screws. The frame is composed of Macor in order to minimize unwanted movement of the filaments due to thermal expansion. Filaments are secured to the frame by a combination of M1 threaded nuts and rods. These are composed of molybdenum due to the extreme temperatures produced while the filaments are operating. The position of the sample is also shown.

electric field can be applied. In the presence of an electric field F, Eq. (4.6) evaluates to [7]

$$J = AT^2 \exp\left(\frac{-(\phi - \Delta\phi)}{kT}\right) \tag{4.8}$$

where the change in work function $\Delta\phi$ is given by the expression

$$\Delta\phi = \sqrt{\frac{e^3 F}{4\pi\varepsilon_0}}. \tag{4.9}$$

Here, ε_0 is the permittivity of vacuum. Second, the filament can be coated with a metal possessing a lower work function than that of the filament. This explains the use of thoriated tungsten filaments. Tungsten has a work function of 4.54 eV [14], whereas thoriated tungsten has a work function of 2.96 eV [14]. A thoriated tungsten filament can therefore produce the same emission current as a tungsten filament operating at a much higher temperature. In practical terms, the use of thoriated tungsten filaments allows the current density required to heat the sample to be achieved at lower filament temperatures. This minimizes outgassing from the filaments and from nearby objects heated by their operation.

Thoriated tungsten filaments must be *activated* prior to use. Activation is a two-step process resulting in the growth of a thorium metal layer on the filament surface. Prior to activation, the majority of thorium present in the filament exists in the form of thorium oxide [4]. The first step of activation is the conversion of thorium oxide to thorium metal. This is achieved by operating the filament at a temperature of 2800 K for approximately 3 minutes [4]. Note that this step also results in the complete evaporation of thorium metal from the filament surface. The second step grows a new thorium metal layer on the filament surface by diffusion of thorium metal from the bulk of the filament. This is achieved by operating the filament at a temperature between 2000-2100 K [4] for approximately 60 minutes. Note that diffusion is insignificant below 1900 K, while evaporation occurs above 2200 K [4].

Activated filaments remain active indefinitely unless exposed to oxidizing or ionized gases [4]. Oxidizing gases react with thorium metal to form thorium oxide, while the presence of ionized gases may result in positive ion bombardment of the filament if the filament voltage is sufficiently high. These effects can be mitigated by adding small amounts of carbon to the filament [4–6]. This process is referred to as carburization. It should be noted, however, that carburization causes filaments to become brittle [15] and reduces their useful lifetime [4]. For these reasons, the filaments used in the electron bombardment heater are not carburized. For a more detailed discussion of thermionic emission, several excellent resources are recommended [4–7].

Given a Si(111) cantilever sample 50 mm long $\times$ 5 mm wide $\times$ 0.28 mm thick, the electron bombardment heater is capable of producing a clean region 7 ± 1 mm long $\times$ 5 mm wide. This region is centred between the filaments, distributed evenly either side of the centre point and fully reproducible across sample changes and heating cycles.

4.1.4 Sample pusher

Sample pushing is accomplished using the sample pusher shown in Fig. 4.3. This device consists of a two-pronged quartz fork attached to an Omicron MS-5 piezoelectric motor [8] centred above the sample at a point 35 mm away from the sample clamp. The sample passes between the prongs of the fork at the 35 mm point. By moving the fork with the piezoelectric motor, the sample can be pushed backward and forward in the direction perpendicular to the sample surface. This changes the radius of curvature of the sample and induces uniaxial stress. Referring to the front surface of the sample shown in in Fig. 4.3, tensile stress can be induced by pushing the sample toward the back electrode, while compressive stress can be induced by pushing it toward the front electrode. The sample pusher has a maximum travel range of 5 mm, a minimum step size of 40 nm, and can exert a maximum force of 0.5 N. In practice, however, travel is restricted to 0.85 mm in either direction because of the front and back electrodes shown in Fig 4.2. The stress can be calculated as follows. Consider a rectangular cantilever of length l, width w and thickness t subjected to a point load F at the free end. The free-end deflection z is given by [16]

$$z = F\left[\frac{4l^3}{Ewt^3}\right]. \tag{4.10}$$

Rearranging Eq. (4.10), the loading force F is given by

$$F = z\left[\frac{Ewt^3}{4l^3}\right]. \tag{4.11}$$

The bending moment M_x evaluated at a distance x from the clamp is given by [16]

$$M_x = F\left(l - x\right). \tag{4.12}$$

The uniaxial stress σ_x corresponding to M_x is given by [16]

$$\sigma_x = \frac{M_x c}{I} \tag{4.13}$$

where c is the distance from the neutral axis to the point of interest, and I is the area moment of inertia. The area moment of inertia for a rectangular cantilever is given by [16]

$$I = \frac{wt^3}{12}. \tag{4.14}$$

In order to evaluate the stress at the surface of the cantilever, the value of c is taken as $\frac{t}{2}$. Substituting Eq. (4.12) for M_x, Eq. (4.13) evaluates to [16]

$$\sigma_x = F\left[\frac{6\left(l - x\right)}{wt^2}\right]. \tag{4.15}$$

Substituting Eq. (4.11) for F, Eq. (4.15) evaluates to

$$\sigma_x = z\left[\frac{3Et}{2l^3}(l-x)\right]. \tag{4.16}$$

This expression describes the relationship between the free-end deflection z and the uniaxial stress σ_x at the surface of the cantilever evaluated at a distance x from the clamp. Assuming the uniaxial strain ε_x corresponding to σ_x is small, Hooke's law states that the two will be proportional. The relationship between σ_x and ε_x can then be described by

$$\sigma_x = E\varepsilon_x \tag{4.17}$$

where E is the Young's modulus of the cantilever. Rearranging Eq. (4.17), ε_x is given by

$$\varepsilon_x = \frac{\sigma_x}{E}. \tag{4.18}$$

Substituting Eq. (4.16) for σ_x, Eq. (4.18) evaluates to

$$\varepsilon_x = z\left[\frac{3t}{2l^3}(l-x)\right]. \tag{4.19}$$

This expression describes the relationship between the free-end deflection z and the uniaxial strain ε_x at the surface of the cantilever evaluated at a distance x from the clamp. If the load is applied at any other point along the cantilever, the length l is taken as the distance between the clamp and the point at which the load is applied. Consider, for example, a typical Si(111) cantilever sample 50 mm long $\times$ 5 mm wide $\times$ 0.28 mm thick. The Young's modulus and Poisson's ratio of Si(111) have been estimated as 168.9 GPa and 0.262 respectively [17]. If a load applied at a point 35 mm away from the sample clamp results in a 0.85 mm deflection towards the back electrode at the same point, the tensile stresses at the 12.5 mm and 17.5 mm points representing the boundaries of the clean area accessible to the STM tip are 31.64 MPa and 24.61 MPa respectively. The corresponding strains are 0.01873% and 0.01457% respectively. Note that stress and strain increase linearly with decreasing distance from the clamp.

4.2 Operational considerations

The operational success of the RT-STM/SSM depends on several factors beyond those discussed in Section 4.1. The most important of these are the effects of parasitic capacitance and sample oscillation. This section discusses each of these factors.

4.2.1 Parasitic capacitance

In Section 4.1.2, it was indicated that the radius of curvature of a cantilever sample could be measured by monitoring the change in capacitance between a *low-resistivity sample* and one of the electrodes shown in Fig. 4.2. It was further indicated that the remaining electrode could be used as a *reference electrode* in order to detect spurious or parasitic contributions to the measured capacitance. Low-resistivity samples are used because high-resistivity samples cause a spurious, resistance-related change in capacitance detectable only using a reference electrode. The magnitude of this change is substantially greater than that of the change in capacitance brought about by a change in the radius of curvature of the sample.

Before discussing this effect, it is necessary to discuss capacitors in greater detail. An ideal capacitor is characterized by a capacitance fully decoupled from inductive and resistive components; it does not lose charge unless discharged. A real capacitor is characterized by a capacitance coupled to inductive and resistive components present in the capacitor circuit; it loses charge through these components. The process of losing charge is called *loss*. Loss due to inductance becomes significant only at high frequencies. Since the AH2550A capacitance bridge [9] used to measure capacitance utilizes a 1 kHz test signal, loss due to inductance can be disregarded. Consequently, this section only discusses loss due to resistance.

The resistive components present in a capacitor circuit consist of those connected in series with the capacitor, such as the capacitor cables and capacitor plates, and those connected in parallel with the capacitor, such as the capacitor dielectric. These can be modeled using *equivalent circuits*. Equivalent circuits for a capacitor coupled to a series resistance, a parallel resistance, and both series and parallel resistances are shown in Fig. 4.6, Fig. 4.7 and Fig. 4.8 respectively.

The AH2550A reports capacitance and loss values in terms of the parallel resistance model shown in Fig. 4.7, with loss expressed in terms of either parallel conductance G_p, parallel resistance R_p or dissipation factor D. These values can also be expressed in terms of the series resistance model shown in Fig. 4.6. Given a parallel capacitance C_p and a parallel conductance G_p, the parallel resistance R_p is given by [12]

$$R_\mathrm{p} = \frac{1}{G_\mathrm{p}} \tag{4.20}$$

while the dissipation factor D is given by [12]

$$D = \frac{1}{2\pi f C_\mathrm{p} R_\mathrm{p}} \tag{4.21}$$

where f is the frequency of the applied test signal. The series resistance R_s is given by [12]

$$R_\mathrm{s} = R_\mathrm{p}\left[\frac{D^2}{1+D^2}\right]. \tag{4.22}$$

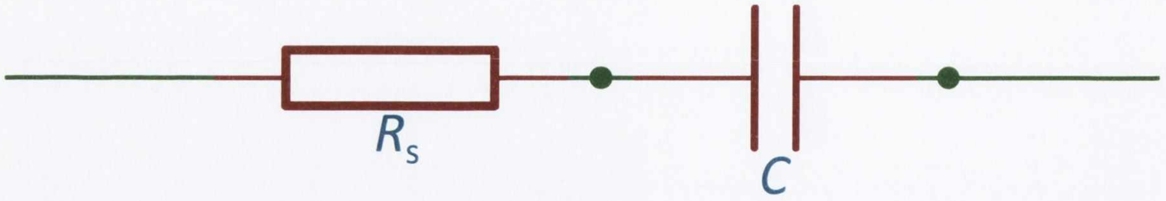


Fig. 4.6. Equivalent circuit for a capacitor of capacitance C coupled to a series resistance R_s.

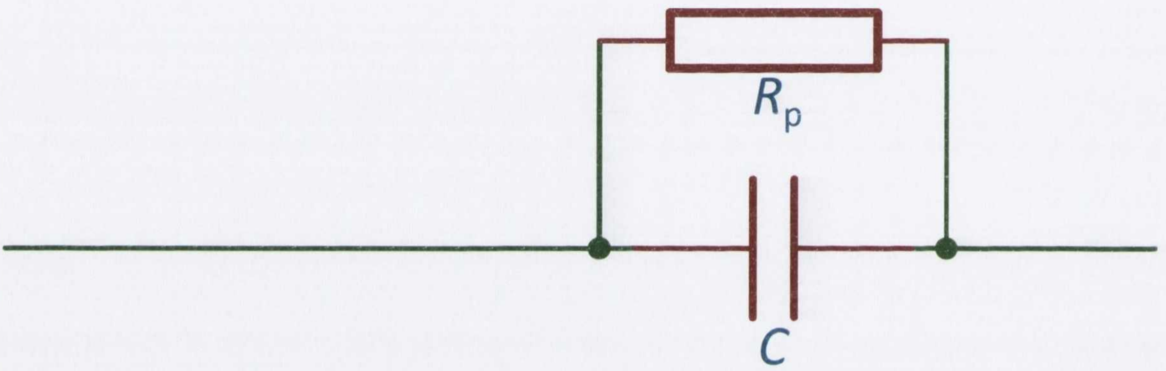


Fig. 4.7. Equivalent circuit for a capacitor of capacitance C coupled to a parallel resistance R_p.

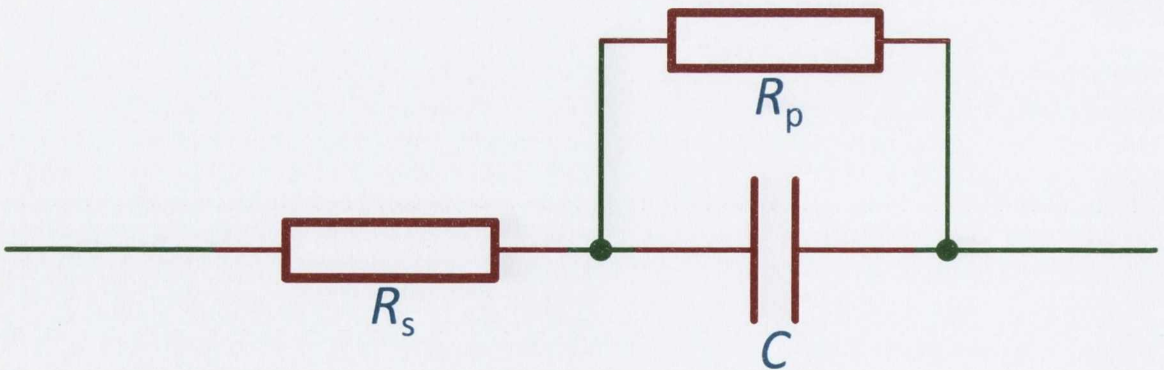


Fig. 4.8. Equivalent circuit for a capacitor of capacitance C coupled to a series resistance R_s and a parallel resistance R_p.

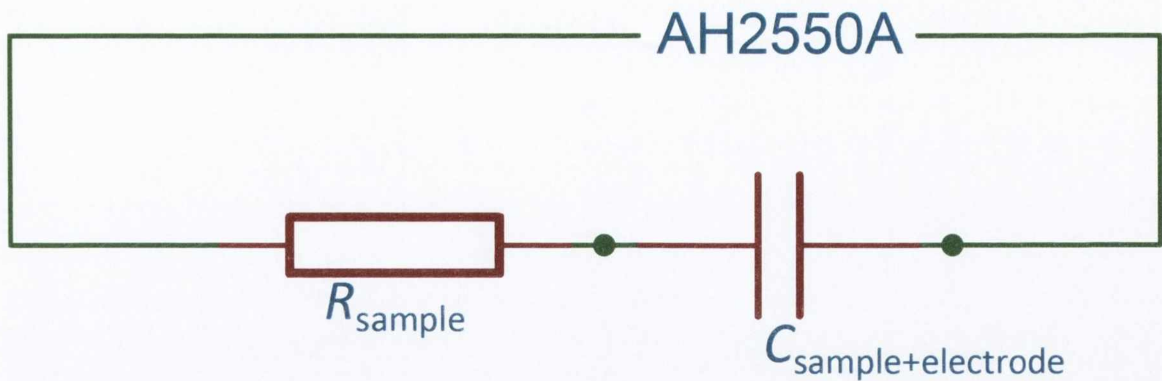


Fig. 4.9. Equivalent circuit for a parallel-plate capacitor of capacitance $C_{\text{sample + electrode}}$ formed between the free end of a cantilever sample of resistance R_{sample} and an adjacent tantalum electrode. The circuit is driven by an Andeen-Hagerling AH2550A [9] capacitance bridge. Note that the sample acts as a resistor connected in series with the capacitor.

A change in loss is indicative of an electronic change in the capacitor circuit. Thus, if a change in capacitance is accompanied by a change in loss, it must be partially or fully electronic in origin. Conversely, if a change in capacitance is not accompanied by a change in loss, it must be physical in origin. In the context of the surface stress measurement described in Section 4.1.2, this information is useful in determining whether a change in capacitance is the result of a change in the radius of curvature of the sample, or whether it is the result of an electronic change in the capacitor circuit.

As an example, consider the following investigation of the effects of light exposure on low-resistivity and high-resistivity silicon samples. Experiments were performed at room temperature in a light-free ultra-high vacuum chamber containing an Omicron STM-1 [3] equipped with the sample holder shown in Fig. 4.2. The base pressure of the chamber was 5×10^{-11} Torr and was monitored using a Televac 7FC [18] cold cathode gauge to eliminate sample heating effects. Low-resistivity samples were 60 mm long $\times$ 5 mm wide $\times$ 0.280 mm thick, while high-resistivity samples were 60 mm long $\times$ 10 mm wide $\times$ 0.525 mm thick thick. Both sets of samples were prepared from single-side-polished, n-type/phosphorus-doped Si(111) wafers with resistivities of 0.008-0.012 Ω cm and 800-1200 Ω cm respectively. Light was provided by a variable-intensity light source located behind a glass viewport situated $\sim$ 300 mm away from the sample. Capacitance measurements were performed using an Andeen-Hagerling AH2550A [9] capacitance bridge in the configuration shown in Fig. 4.9. Note that variations in the initial capacitance arise due to differences in sample dimensions, differences in sample clamping configurations, static capacitance contributions from nearby conductors, and the tendency of clamped samples to lie closer to the back electrode. Also note that the sample acts as a resistor connected in series with the capacitor; this will become important later.

Fig. 4.10 shows the change in capacitance and loss measured between a *high-resistivity sample* and the *front electrode* upon exposure of the front of the sample to light. When the light is turned on, the capacitance is observed to increase, while the loss is observed to

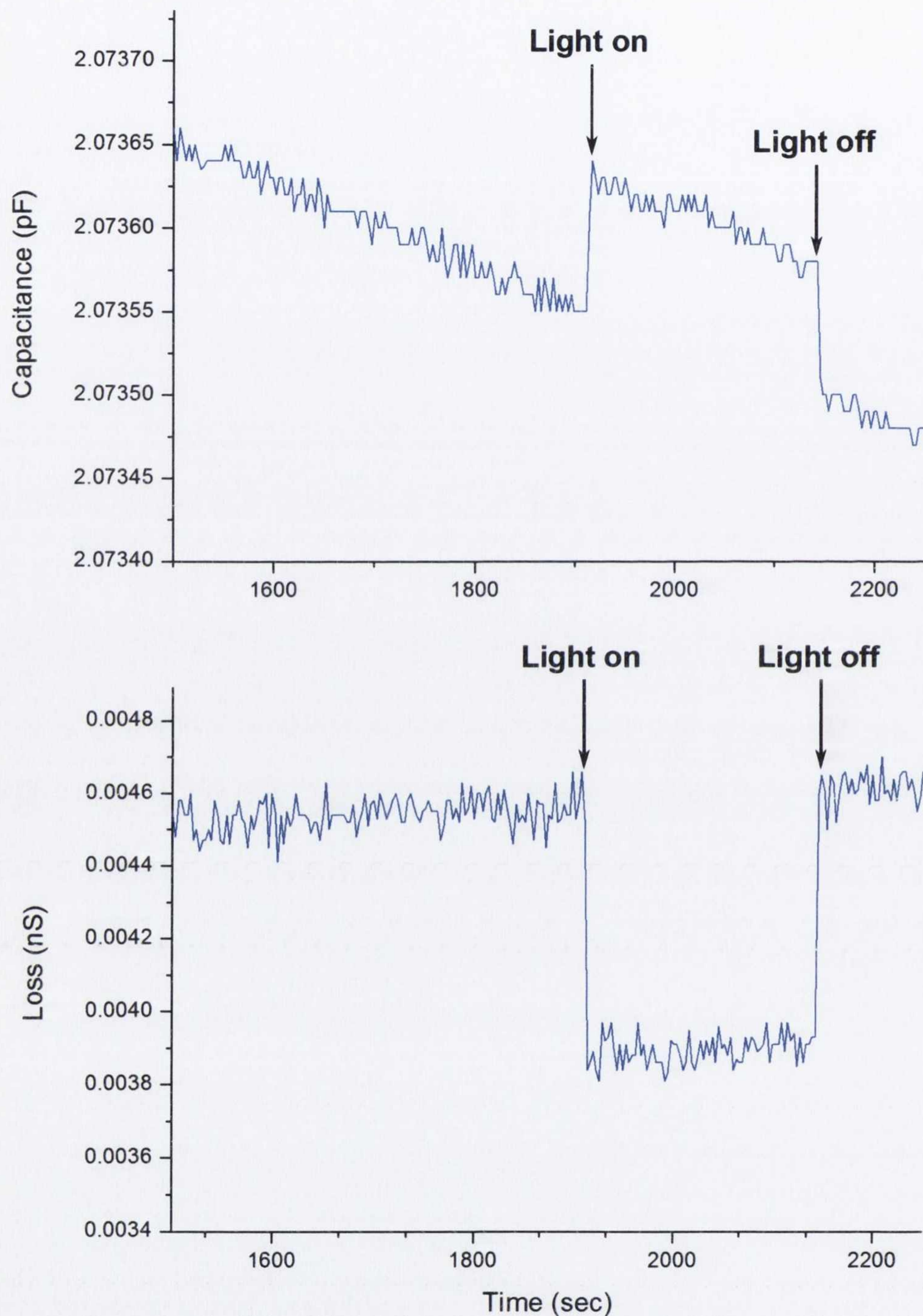


Fig. 4.10. Change in capacitance and loss measured between a *high-resistivity sample* and the *front electrode* upon exposure of the front of the sample to light. Loss is expressed in terms of parallel conductance. Note that the capacitance drifts downward due to post-flash cooling of the cantilever sample.

decrease. These changes correspond to a decrease in the series resistance of the sample. This is consistent with light-induced excitation of electrons from the valence band to the conduction band. When the light is turned off, both capacitance and loss return to baseline levels. Note that the transition between 'on' and 'off' states is instantaneous. This further suggests that the change in capacitance is electronic in origin.

Fig. 4.11 shows the results of the same measurement performed using the *back electrode*. When the light is turned on, the capacitance is observed to increase, while the loss is observed to decrease. When the light is turned off, both capacitance and loss return to baseline levels. Again, the transition between 'on' and 'off' states is instantaneous. This behaviour is identical to that observed in the previous measurement. Since a change in capacitance due to a change in the radius of curvature of the sample should have different signs when measured against front and back electrodes, the observed change in capacitance is almost certainly electronic in origin. In this case, it is ascribed to a decrease in the series resistance of the sample brought about by light-induced excitation of electrons from the valence band to the conduction band.

Fig. 4.12 shows the change in capacitance and loss measured between a *low-resistivity sample* and the *front electrode* upon exposure of the front of the sample to light. When the light is turned on, the capacitance is observed to decrease, while the loss is also observed to decrease. Since the change in loss is negligible, the change in capacitance is likely physical in origin. When the light is turned off, both capacitance and loss return to baseline levels. Note, however, that the transition between 'on' and 'off' states is slow rather than instantaneous. This further suggests that the change in capacitance is physical in origin.

Fig. 4.13 shows the results of the same measurement performed using the *back electrode*. When the light is turned on, the capacitance is observed to increase, while the loss remains unchanged. Since the loss remains unchanged, the change in capacitance is likely physical in origin. Note that the sign of the change in capacitance is opposite to that observed in the previous measurement. This behaviour is consistent with a change in the radius of curvature of the sample. When the light is turned off, the capacitance returns to a baseline level. Again, the transition between 'on' and 'off' states is slow rather than instantaneous.

Based on the above observations, it is reasonable to conclude that the observed change in capacitance is physical in origin. In this case, it is ascribed to a change in the radius of curvature of the sample brought about by light-induced heating and cooling. When the light is turned on, the front of the sample increases in temperature. This causes a temperature gradient to form between the front and back of the sample. Since the front of the sample is warmer than the back of the sample, it expands faster. This causes the sample to bend towards the back electrode. When the light is turned off, the front of the sample begins to cool. This causes it to contract, thus causing the sample to returns to its original position. The conclusions of the previous experiments can be summarized in two general rules:

- If a change in capacitance is *electronic* in origin, it will be accompanied by a change in loss, and it will have the same sign in front and back electrode measurements.

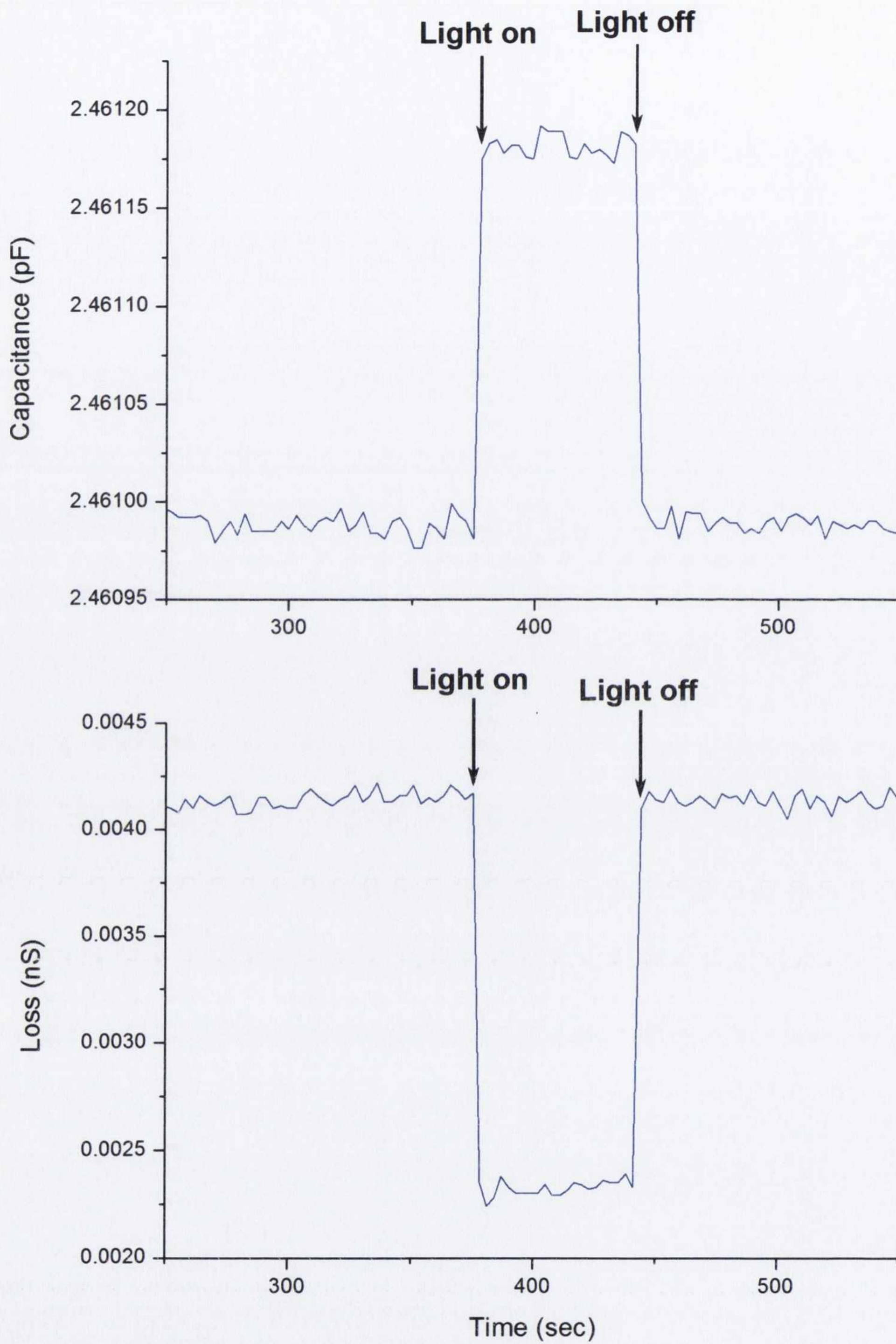


Fig. 4.11. Change in capacitance and loss measured between a *high-resistivity sample* and the *back electrode* upon exposure of the front of the sample to light. Loss is expressed in terms of parallel conductance.

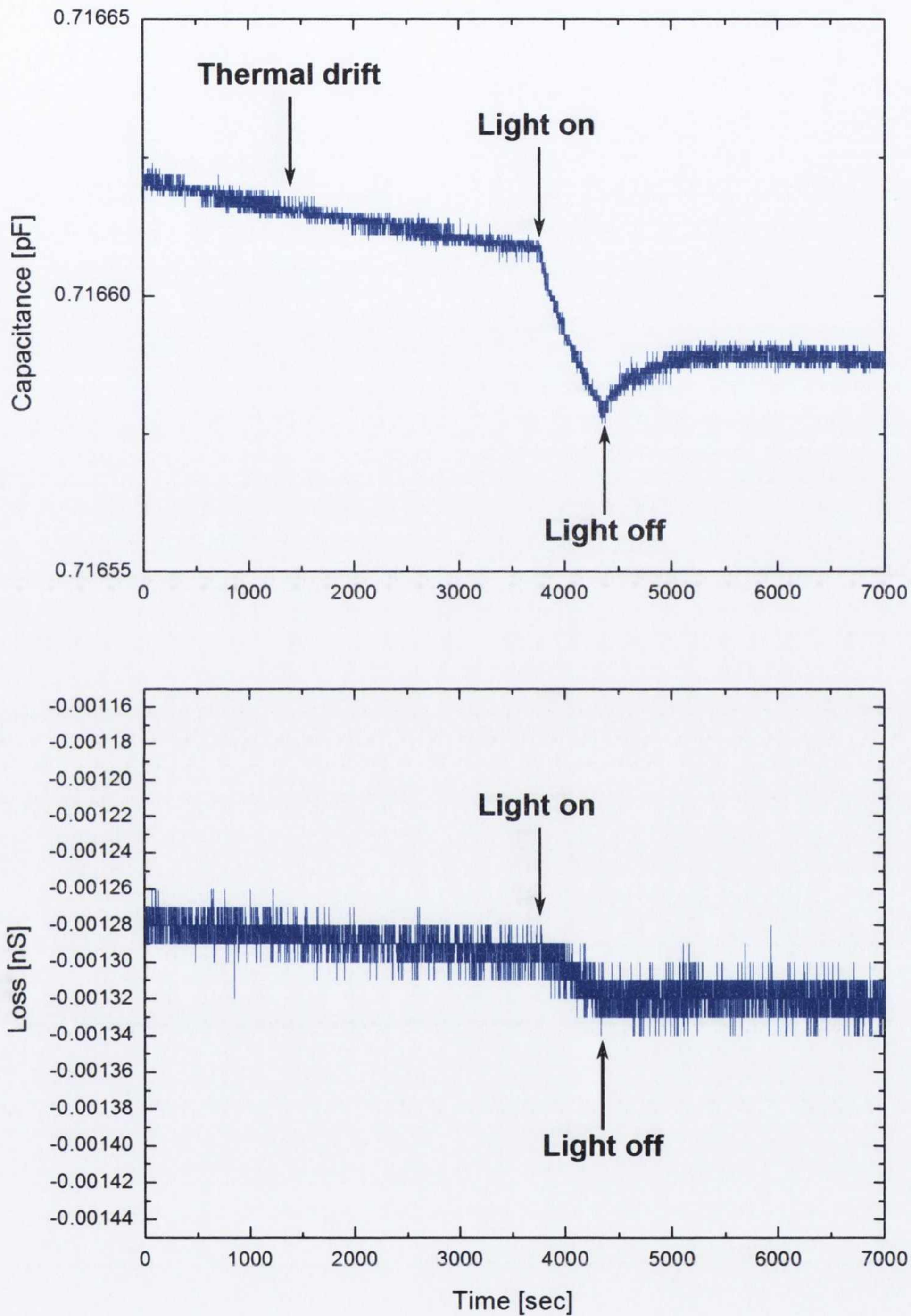


Fig. 4.12. Change in capacitance and loss measured between a *low-resistivity sample* and the *front electrode* upon exposure of the front of the sample to light. Loss is expressed in terms of parallel conductance.

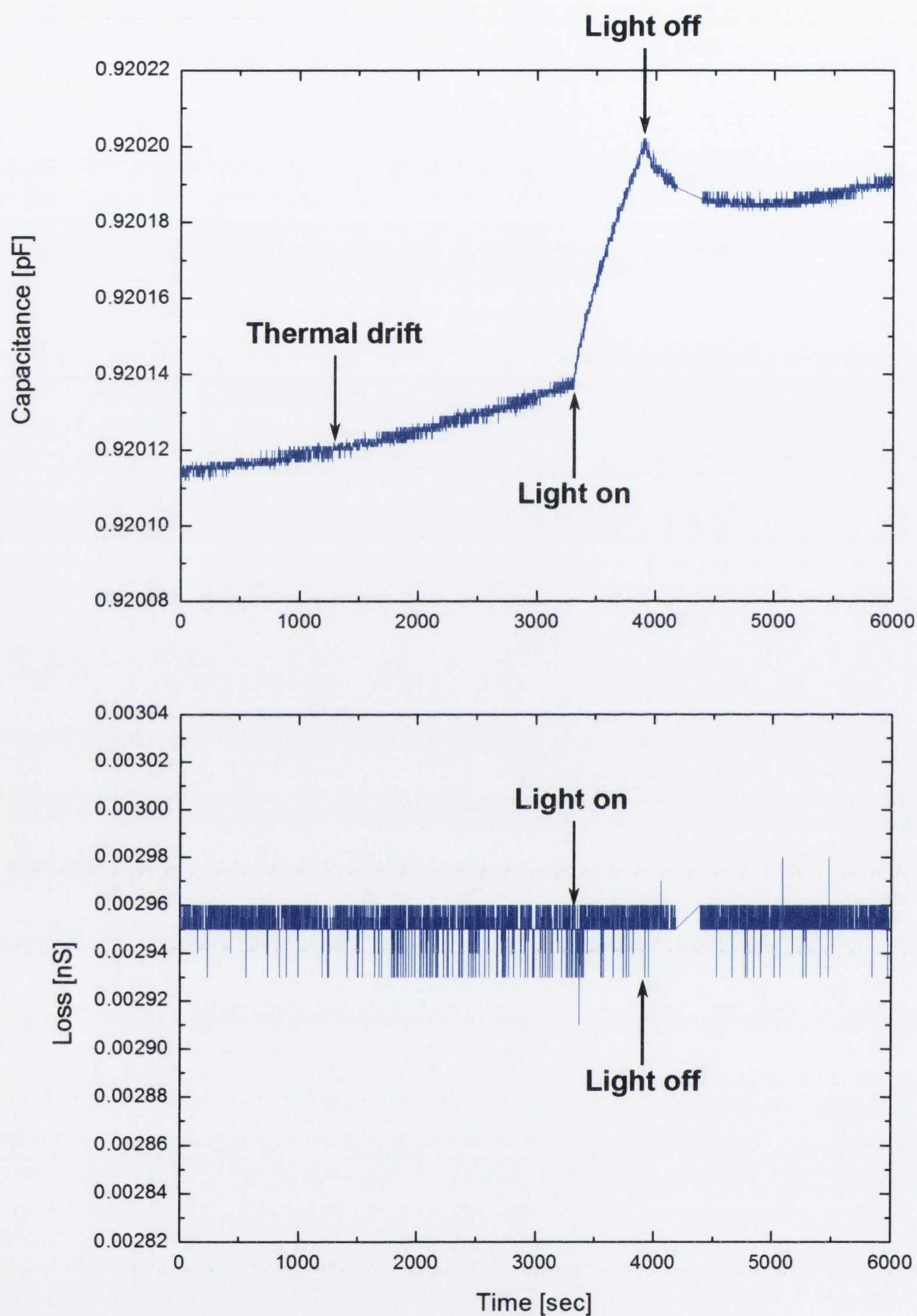


Fig. 4.13. Change in capacitance and loss measured between a *low-resistivity sample* and the *back electrode* upon exposure of the front of the sample to light. Loss is expressed in terms of parallel conductance.

- If a change in capacitance is *physical* in origin, it will not be accompanied by a change in loss, and it will have different signs in front and back electrode measurements.

Note that thermal effects will always be present regardless of the resistivity of the sample used. These rules can be applied to adsorbate-induced surface stress measurements. As an example, consider the following investigation of the reaction of O_2 with the Si(111)–(7×7) surface described in Chapter 3. Experiments were performed at room temperature in a light-free ultra-high vacuum chamber containing an Omicron STM-1 [3] equipped with the sample holder shown in Fig. 4.2. The base pressure of the chamber was 5×10^{-11} Torr and was monitored using a Televac 7FC [18] cold cathode gauge to eliminate gas cracking and sample heating effects.

The sample used was 60 mm long $\times$ 10 mm wide $\times$ 0.525 mm thick and prepared from a single-side-polished, n-type/phosphorus-doped Si(111) wafer with a resistivity of 800-1200 Ω cm. A clean Si(111)–(7×7) surface approximately 10 mm long $\times$ 10 mm wide was prepared using a local bimetallic strip heating method [19] that served as the forerunner to the electron bombardment heater described in Section 4.1.3. The quality of the surface was confirmed using STM. Capacitance measurements were performed using an Andeen-Hagerling AH2550A [9] capacitance bridge in the configuration shown in Fig. 4.9. Note once again that the sample acts as a resistor connected in series with the capacitor; this will become important later.

Fig. 4.14 shows the change in capacitance and loss measured between the sample and the *front electrode* upon exposure to O_2 gas at a pressure of 7.5×10^{-9} Torr. When the gas is introduced, the capacitance is observed to decrease, while the loss is observed to increase. These changes correspond to an increase in the series resistance of the sample. Using values of 1.7564 pF and 0.202 nS for the capacitance and loss before exposure and values of 1.7529 pF and 0.265 nS for after exposure, the increase in series resistance is estimated to be 0.525 MΩ. The initial serial resistance can be calculated using the expression

$$R = \frac{\rho l}{A} \tag{4.23}$$

where R, ρ, l and A are the resistance, resistivity, length and cross-sectional area of the sample respectively. Using values of 8-12 Ω m, 60×10^{-3} m and 5.25×10^{-6} m^2 for ρ, l and A respectively, the initial series resistance is estimated to be 0.091-0.137 MΩ. The increase in series resistance is therefore $\sim$ 3.8-5.8 times larger than the initial series resistance.

Fig. 4.15 shows the results of the same measurement performed using the *back electrode*. When the gas is introduced, the capacitance is observed to decrease, while the loss is observed to increase. The shapes of the capacitance and loss curves are broadly similar to the shapes of those shown in Fig. 4.14, with differences between the curves explained by the larger baseline slope and longer exposure time observed in Fig. 4.15. Applying the rules presented earlier, since the change in capacitance is accompanied by a change in loss, and since it has the same sign in front and back electrode measurements, it

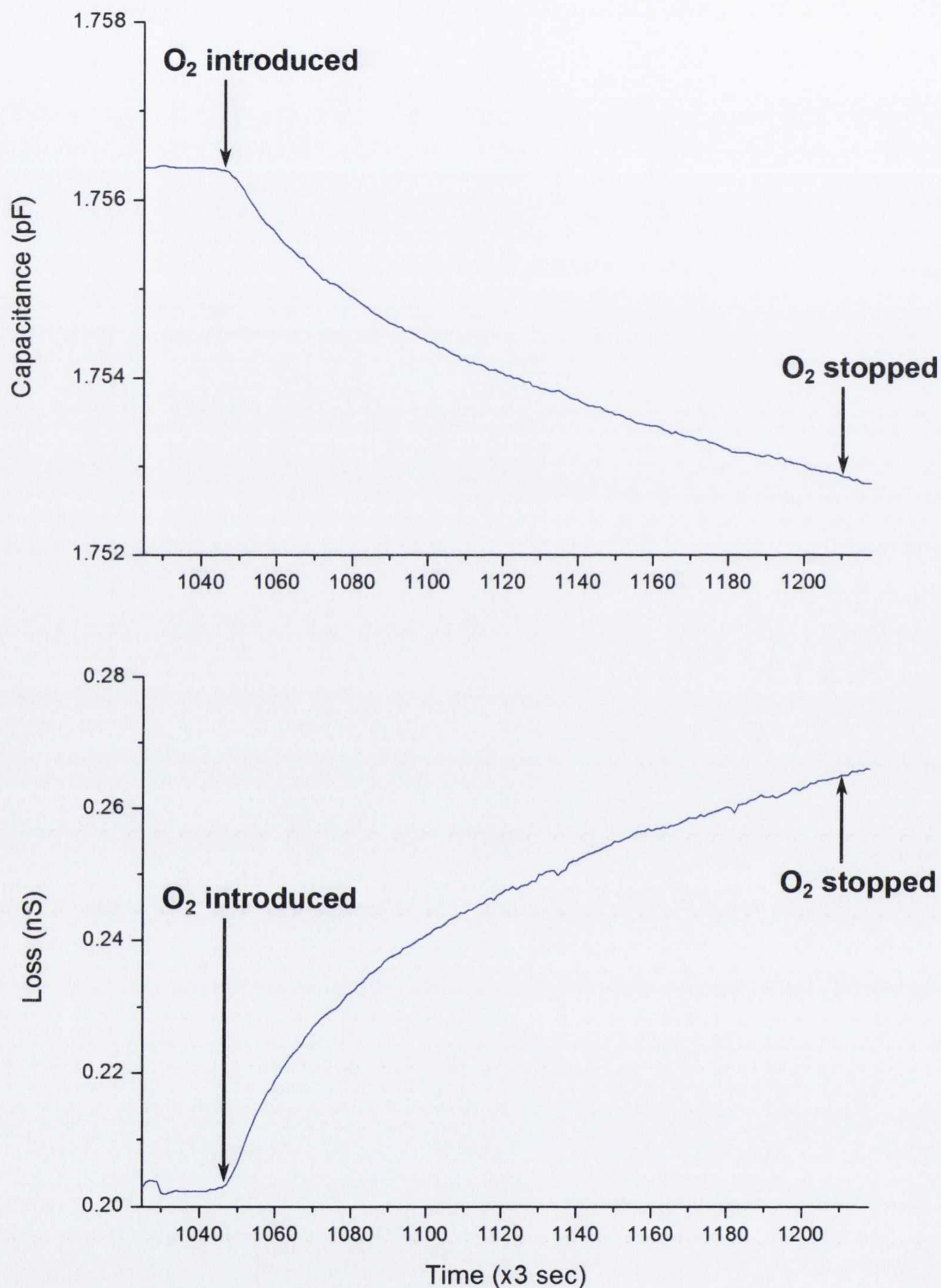


Fig. 4.14. Change in capacitance and loss measured between a sample with a resistivity of 800-1200 Ω cm and the *front electrode* upon exposure to O_2 gas at a pressure of 7.5×10^{-9} Torr. Loss is expressed in terms of parallel conductance.

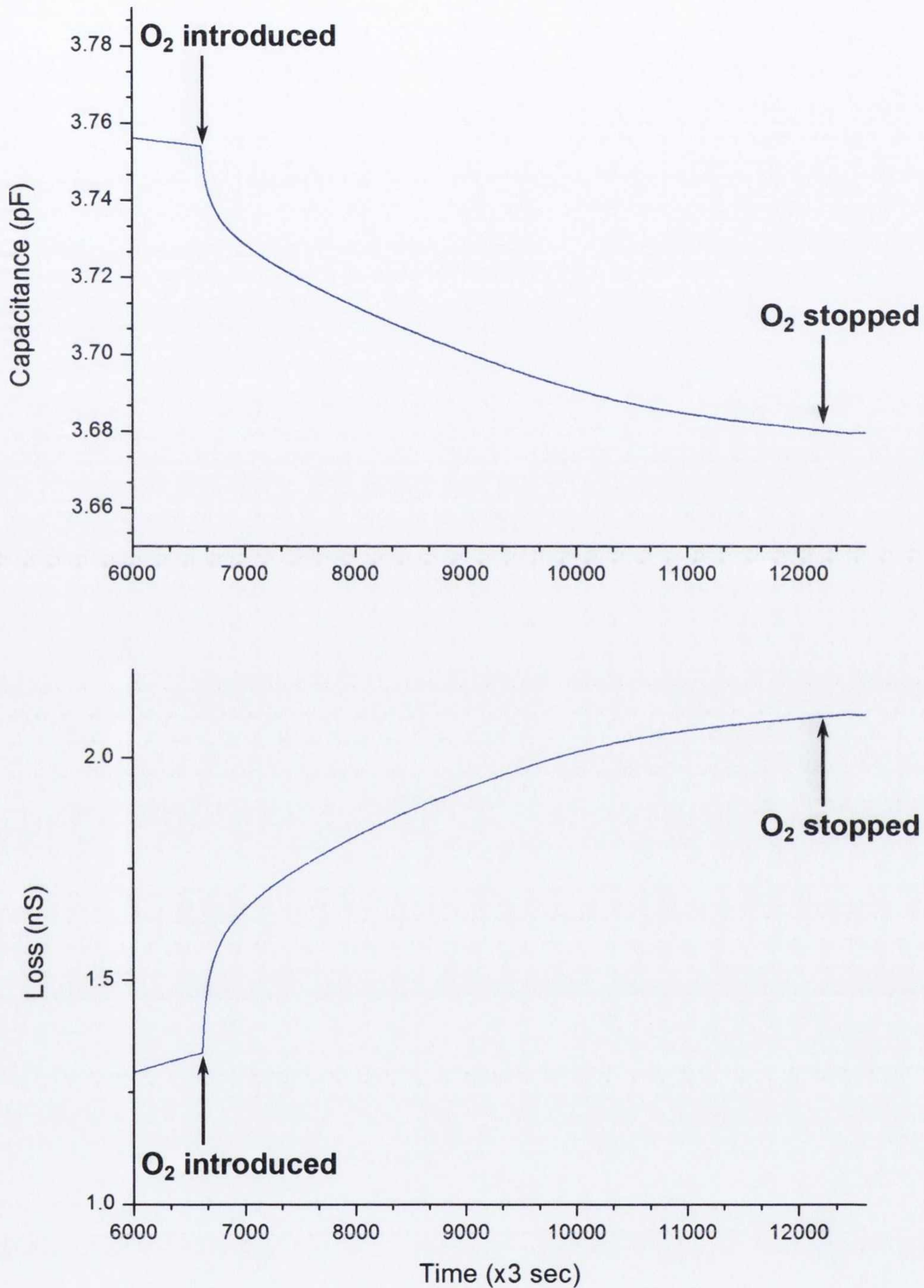


Fig. 4.15. Change in capacitance and loss measured between a sample with a resistivity of 800-1200 Ω cm and the *back electrode* upon exposure to O_2 gas at a pressure of 7.5×10^{-9} Torr. Loss is expressed in terms of parallel conductance.

must be electronic in origin. In this case, it is ascribed to an increase in the series resistance of the sample brought about by the reaction of adatom dangling bonds.

This can be understood in terms of a change in the surface contribution to the conductivity of the sample. The conductivity σ of the sample is given by the expression [20]

$$\sigma = \frac{1}{\rho} = \frac{l}{RA} = \int_0^d [\sigma_b + \sigma_s(z)] \, dz/d \tag{4.24}$$

where ρ is the resistivity of the sample, l is the length of the sample, R is the resistance of the sample, A is the cross-sectional area of the sample, d is the thickness of the sample, and σ_b and σ_s are the bulk and surface conductivities respectively. In the case of a low-resistivity sample, the surface contribution to the conductivity of the sample is substantially smaller than the bulk contribution. A change in surface conductivity would not be expected to significantly modify the conductivity of a low-resistivity sample. In the case of a high-resistivity sample, the surface contribution to the conductivity of the sample becomes significant. A change in surface conductivity would therefore be expected to significantly modify the conductivity of a high-resistivity sample.

It has previously been shown that the reaction of adsorbates with a surface can significantly alter its conductivity [21]. Since the reaction of O_2 with the Si(111)–(7 $\times$ 7) surface involves the eliminatation of adatom dangling bonds [22–26], and since adatom dangling bonds contribute to the surface conductivity [21], it is reasonable to conclude that the reaction would reduce the surface conductivity. Based on previous reasoning, this would significantly reduce the conductivity of a high-resistivity. Since the conductivity is inversely proportional to the series resistance, a corresponding increase in the series resistance would occur. This is consistent with the data presented earlier.

While high-resistivity samples are susceptible to parasitic electronic effects, low-resistivity samples are not. The following experiments were performed under the same conditions used in the high-resistivity experiments described earlier. In this case, however, the sample used was 60 mm long $\times$ 5 mm wide $\times$ 0.28 mm thick and prepared from a single-side-polished, n-type/phosphorus-doped Si(111) wafer with a resistivity of 0.008-0.012 Ω cm. In addition, the Si(111)–(7 $\times$ 7) surface was prepared using the electron bombardment heater described in Section 4.1.3. It was 7 mm long $\times$ 5 mm wide.

Fig. 4.16 shows the change in capacitance measured between the sample and the *back electrode* upon exposure to O_2 gas at pressures of 2×10^{-9} and 8×10^{-9} Torr respectively. When the gas is introduced, the capacitance is observed to increase. While the loss is not shown in this case, it was observed to remain unchanged. This is consistent with bending of the sample toward the back electrode.

Fig. 4.17 shows the results of the same measurement performed using the *front electrode* at an O_2 pressure of 2×10^{-9} Torr. When the gas is introduced, the capacitance is observed to decrease. While the loss is not shown in this case, it was observed to remain unchanged. This is consistent with bending of the sample away from the front electrode. Applying the rules presented earlier, since the change in capacitance is not accompanied by a change in

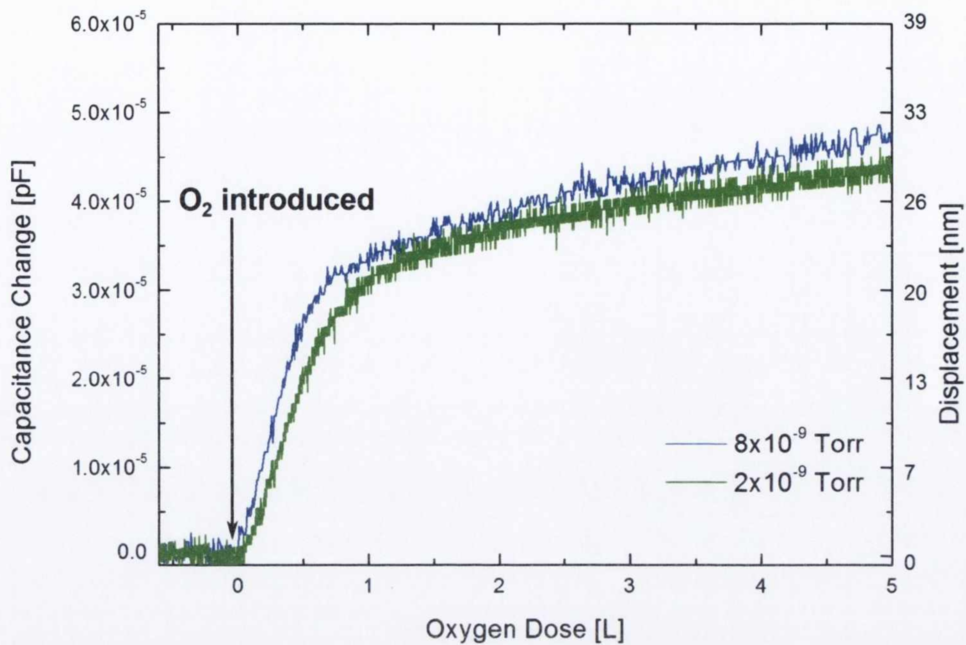


Fig. 4.16. Change in capacitance and sample displacement measured as a function of O_2 gas exposure between a sample with a resistivity of 0.008-0.012 Ω cm and the *back electrode*. Exposure was performed at pressures of 2×10^{-9} and 8×10^{-9} Torr respectively. Exposure is presented in units of Langmuir ($1\text{L} = 1.0 \times 10^{-6}$ Torr sec).

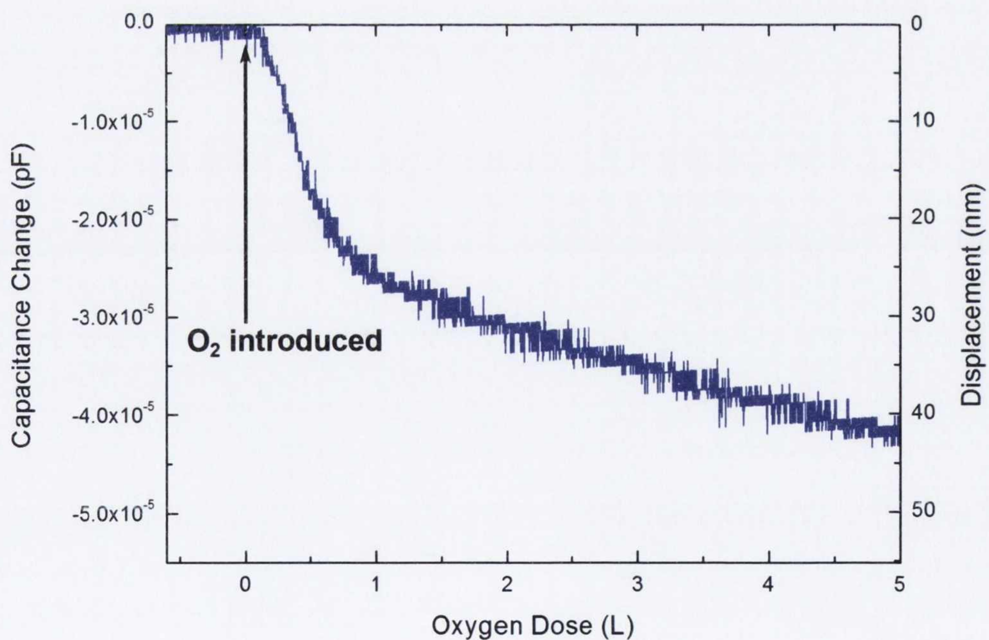


Fig. 4.17. Change in capacitance and sample displacement measured as a function of O_2 gas exposure between a sample with a resistivity of 0.008-0.012 Ω cm and the *front electrode*. Exposure was performed at a pressure of 2×10^{-9} Torr. Exposure is presented in units of Langmuir ($1\text{L} = 1.0 \times 10^{-6}$ Torr sec).

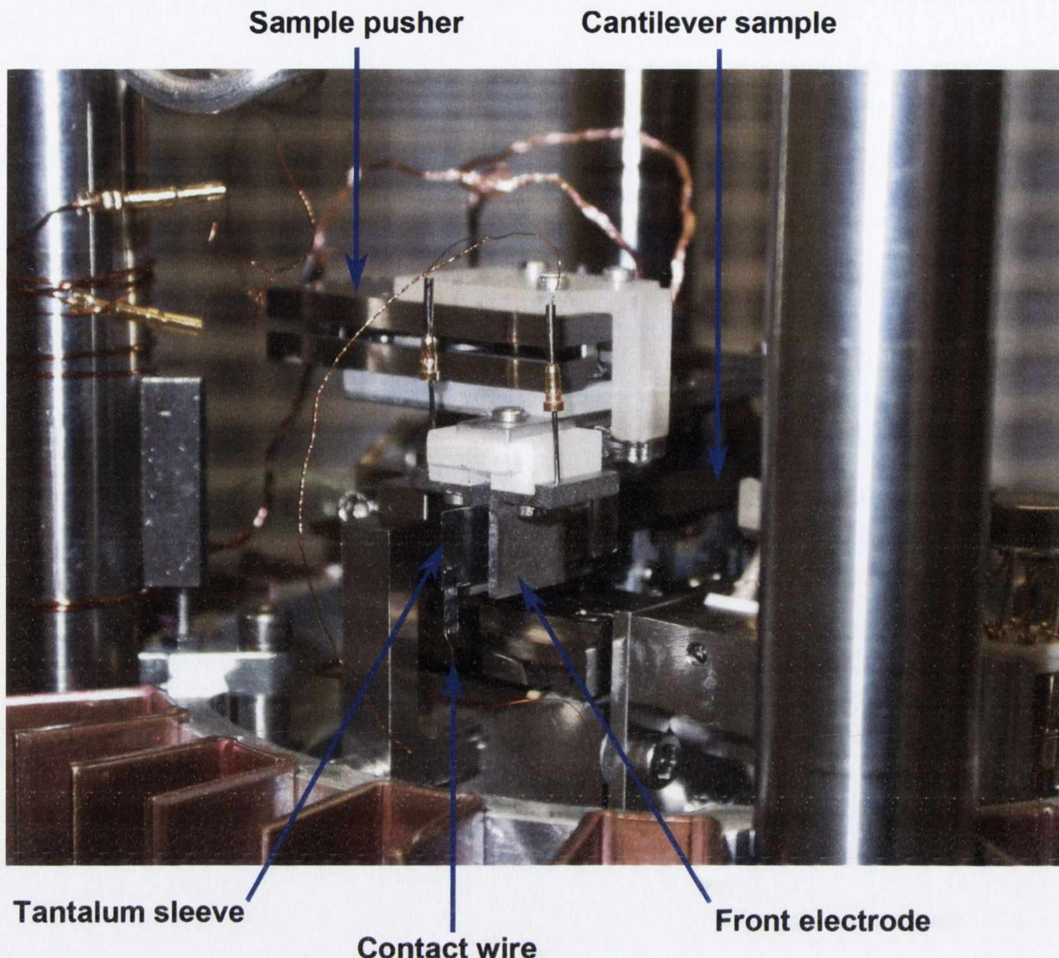


Fig. 4.18. Front view of the assembled RT-STM/SSM system showing the sample fitted with a typical tantalum sleeve bypass.

loss, and since it has different signs in front and back electrode measurements, it must be physical in origin. In this case, it is ascribed to a change in the radius of curvature of the sample brought about by expansion of the surface arising from a reaction-induced decrease in tensile surface stress. The reaction of O_2 with the Si(111)–(7×7) surface will be discussed in greater detail in Chapter 5.

In summary, low-resistivity samples *must* be used in order to achieve meaningful results from capacitive surface stress measurements. In situations where the use of low-resistivity samples is not possible, a metallic sleeve can be fitted to the free end of the sample and biased directly. This allows the resistance of the sample to be bypassed, thereby eliminating sample-related parasitic electronic effects. The sleeve is constructed by spot-welding the edges of two thin pieces of tantalum foil together, leaving a small hollow between the foils to allow it to be fitted over the free end of the sample. A fine, flexible wire is spot-welded to the sleeve to provide an electrical contact for performing capacitance measurements. A typical sleeve is shown fitted to a sample in Fig. 4.18.

4.2.2 Sample oscillation

The natural vibration of a cantilever sample must be suppressed before it can be imaged using STM. This is necessary based on the following analysis of the sample vibrational modes. The spectral density of the *peak amplitude* of a vibration is given by [27]

$$z_\mathrm{p}(f) = \sqrt{\frac{2k_\mathrm{B}T}{\pi k f_0 Q}} \cdot \left[\frac{1}{\left[1 - \left(\frac{f}{f_0}\right)^2\right]^2 + \left[\frac{f}{f_0 Q}\right]^2} \right]^{\frac{1}{2}} \tag{4.25}$$

where f is the vibrational frequency, k_B is Boltzmann's constant, T is the temperature, and k, f_0 and Q are the spring constant, fundamental frequency and quality factor of the sample respectively. The spring constant k is given by

$$k = \frac{Ewt^3}{4l^3} \tag{4.26}$$

where E, w, t and l are the Young's modulus, width, thickness and length of the sample respectively. The quality factor Q is given by

$$Q = \frac{f_0}{\Delta f} \tag{4.27}$$

where Δf is the bandwidth. Rearranging Eq. (4.27), Δf is given by

$$\Delta f = \frac{f_0}{Q}. \tag{4.28}$$

The peak amplitude of the vibration evaluated at f_0 is given by

$$z_\mathrm{p}(f_0) = \sqrt{\frac{2k_\mathrm{B}TQ}{\pi k f_0}}. \tag{4.29}$$

Assuming the peak amplitude spectrum is characterized by a Gaussian distribution centred at f_0, the root-mean-square amplitude of the vibration evaluated at f_0 approximates to

$$z_\mathrm{rms}(f_0) \approx z_\mathrm{p}(f_0) \cdot \Delta f \approx \sqrt{\frac{2k_\mathrm{B}TQ}{\pi k f_0}} \cdot \frac{f_0}{Q} \approx \sqrt{\frac{2k_\mathrm{B}T}{\pi k}}. \tag{4.30}$$

The peak-to-peak amplitude of the vibration evaluated at f_0 is given by

$$z_\mathrm{pp}(f_0) = 2\sqrt{2} \cdot z_\mathrm{rms}(f_0) \approx 4\sqrt{\frac{k_\mathrm{B}T}{\pi k}}. \tag{4.31}$$

Eq. (4.31) shows that the peak-to-peak amplitude of a vibration is proportional to the temperature T and the square root of the reciprocal of the spring constant k.

From Eq. (4.26), the square root of the reciprocal of k is given by

$$\sqrt{\frac{1}{k}} = \sqrt{\frac{4l^3}{Ewt^3}}. \tag{4.32}$$

Assuming T is constant, the relationship between the peak-to-peak amplitude and the spring constant can be described by the expression

$$z_{\mathrm{pp}}(f_0) \propto \sqrt{\frac{4l^3}{Ewt^3}}. \tag{4.33}$$

Eq. (4.33) implies that the peak-to-peak amplitude of a vibration can be controlled by choosing appropriate values for l, w and t. Note that the peak-to-peak amplitude is lower in the region 15 mm away from the sample clamp imaged by the STM. As an example, consider the Si(111)–(7 × 7) surface described in Chapter 3. The observable height of an adatom on the Si(111)–(7 × 7) surface is 50 pm [1]. In order to resolve these atoms, it is necessary to reduce the peak-to-peak amplitude of the sample vibration to less than 50 pm. Given a sample 50 mm long × 5 mm wide × 0.28 mm thick, and assuming the Young's modulus of Si(111) to be 168.9 GPa [17], the spring constant calculated using Eq. (4.26) is estimated to be 37.08 N m^{-1}. Substituting this value for k in Eq. (4.31) and assuming a temperature of 273.15 K, the peak-to-peak amplitude is estimated to be 22.76 pm. This is sufficient to resolve adatoms on the Si(111)–(7 × 7) surface.

A topographic STM image of the Si(111)–(7 × 7) surface acquired at a sample voltage of 2.0 V on the surface of a cantilever sample 50 mm long × 10 mm wide × 0.525 mm thick is shown in Fig. 4.19. Note that while the dimensions of a cantilever sample can be tailored to enable imaging with atomic resolution, there are practical limitations as to how large these dimensions can be made. This is particularly true with regard to sample heating, where increasing the width or the thickness of the sample greatly increases the input power required to heat it. For this reason, only samples 50 mm long × 5 mm wide × 0.28 mm thick are used. Note that while it is possible to image a sample of these dimensions with atomic resolution, it is usual to pin the sample before imaging using the sample pusher described in Section 4.1.4. This reduces the peak-to-peak amplitude of the sample vibration further, thus leading to an improvement in imaging resolution.

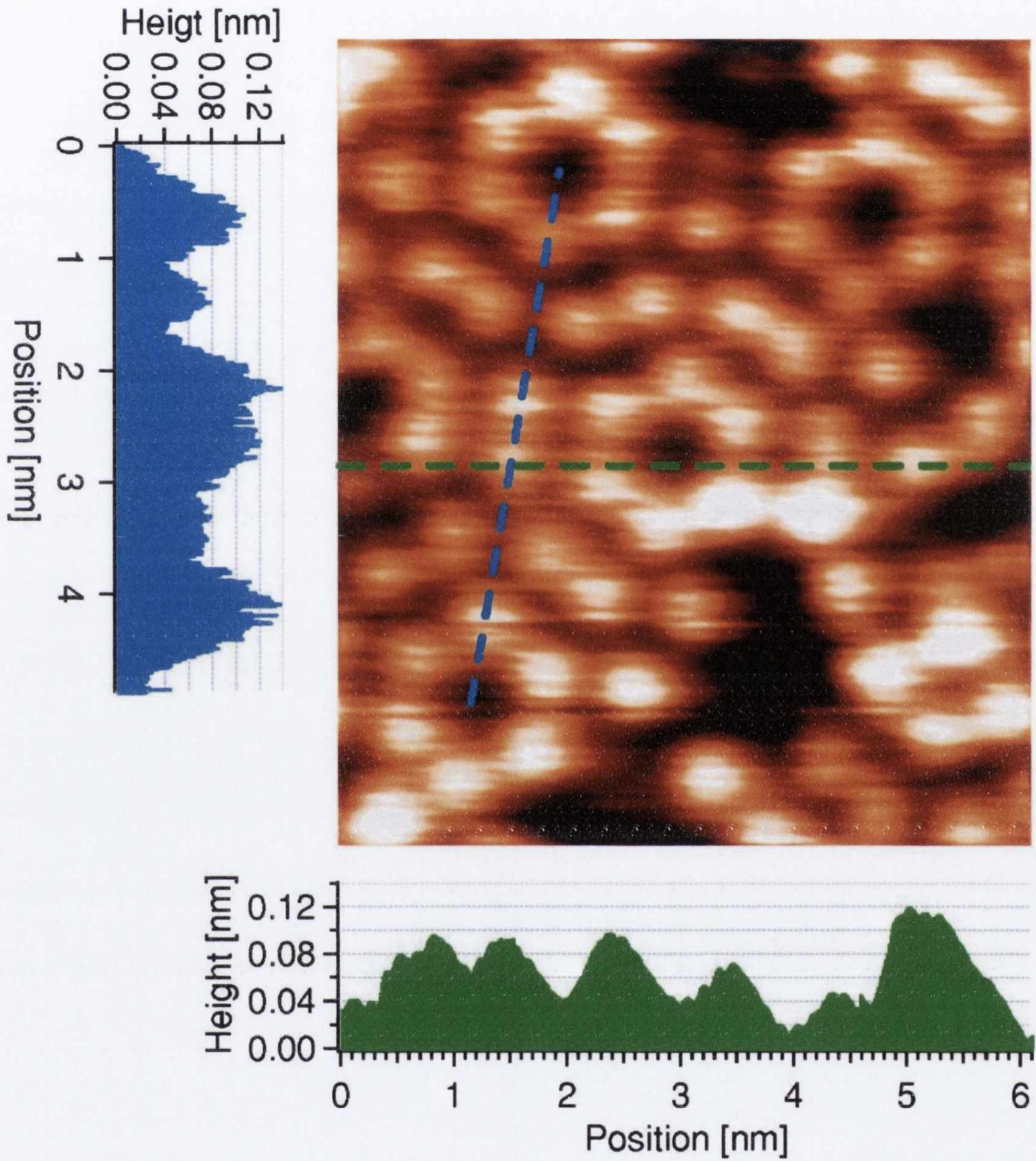


Fig. 4.19. Topographic STM image of the Si(111)–(7 × 7) surface acquired at a sample voltage of 2.0 V on the surface of a cantilever sample 50 mm long × 10 mm wide × 0.525 mm thick [1].

References

[1] T. Narushima, N. T. Kinahan, and J. J. Boland, Rev. Sci. Instrum. **78**, 053903 (2007).

[2] D. Sander and H. Ibach, Phys. Rev. B **43**, 4263 (1991).

[3] Omicron STM-1, Omicron NanoTechnology GmbH, Taunusstein, Germany.

[4] I. Langmuir, Phys. Rev. **22**, 357 (1923).

[5] S. Dushman, Rev. Mod. Phys. **2**, 381 (1930).

[6] S. Dushman, Rev. Mod. Phys. **3**, 190 (1931).

[7] C. Herring and M. H. Nichols, Rev. Mod. Phys. **21**, 185 (1949).

[8] Omicron MS-5, Omicron NanoTechnology GmbH, Taunusstein, Germany.

[9] Andeen-Hagerling AH2550A, Andeen-Hagerling, Inc., 31200 Bainbridge Road, Cleveland, Ohio.

[10] J. D. Wilcock and D. S. Campbell, Thin Solid Films **3**, 3 (1969).

[11] W. A. Brantley, J. Appl. Phys. **44**, 534 (1973).

[12] Andeen-Hagerling AH2550A Operation and Maintenance Manual, Andeen-Hagerling, Inc., 31200 Bainbridge Road, Cleveland, Ohio.

[13] Macor, Corning Inc., Corning, New York.

[14] J. T. Yates, *Experimental Innovations in Surface Science: A Guide to Practical Laboratory Methods and Instruments* (Springer, New York, 1997).

[15] P. Schneider and P. Mareš, Vacuum **12**, 255 (1962).

[16] S. P. Timoshenko and D. H. Young, *Elements of Strength of Materials* (D. Van Nostrand Company, 1968), 5th ed.

[17] J. J. Wortman and R. A. Evans, J. Appl. Phys. **36**, 153 (1965).

[18] Televac 7FC, Televac, 2400 Philmont Avenue, Huntingdon Valley, Philadelphia 19006-0662.

[19] T. Narushima, N. T. Kinahan, and J. J. Boland, Rev. Sci. Instrum. **76**, 095113 (2005).

[20] C. Jiang, S. Hasegawa, and S. Ino, Phys. Rev. B **54**, 10389 (1996).

[21] S. Hasegawa, C. Jiang, Y. Nakajima, T. Nagao, and X. Tong, Surf. Rev. Lett. **5**, 803 (1998).

[22] J. P. Pelz and R. H. Koch, Phys. Rev. B **42**, 3761 (1990).

[23] P. Avouris, I.-W. Lyo, and F. Bozso, J. Vac. Sci. Technol. B **9**, 424 (1991).

[24] F. Bozso and P. Avouris, Phys. Rev. B **44**, 9129 (1991).

[25] R. Martel, P. Avouris, and I.-W. Lyo, Science **272**, 385 (1996).

[26] G. Dujardin, A. Mayne, G. Comtet, L. Hellner, M. Jamet, E. Le Goff, and P. Millet, Phys. Rev. Lett. **76**, 3782 (1996).

[27] H. I. Rasool, P. R. Wilkinson, A. Z. Stieg, and J. K. Gimzewski, Rev. Sci. Instrum. **81**, 023703 (2010).

CHAPTER 5

The Oxidation of the Si(111)–(7 × 7) Surface at Room Temperature

The growth of silicon dioxide is a critical step in the production of virtually all silicon-based electronic devices. As the size of these devices approaches the atomic scale, the use of ultra-thin oxide films will become integral to sustaining device performance and reliability. In order to grow these films, a deeper understanding of the chemistry underlying the oxidation process is required. Substantial progress has been made in this regard, most notably in relation to the oxidation of Si(100) and Si(111) surfaces [1–8]. In spite of this, however, many questions remain unanswered, particularly with respect to the oxidation of the Si(111)–(7 × 7) surface at room temperature. This reaction is a consecutive, two-step process involving the reaction of an oxygen molecule with an unreacted adatom 'A' to form an intermediate state 'B' that subsequently reacts with an additional oxygen molecule to form a fully-reacted state 'D' [9, 10]. The mechanism underlying this reaction has proven controversial. Structures associated with bright and dark features observed in scanning tunneling microscopy (STM) images are disputed [9, 11–15], while the presence of long-lived metastable species remains a long-standing point of contention [11–13, 16–18]. The role of rest atoms is also uncertain [12, 17–20]. While numerous attempts have been made to address these issues, few [10, 21] have attempted to do so in the context of surface stress. In the preceeding chapter, a system combining STM and high-resolution surface stress measurement was described. This chapter presents the results of experiments involving the use of this system to investigate the oxidation of the Si(111)–(7 × 7) surface at room temperature. The site-specific evolution of surface stress is measured and associated with relief of the intrinsic tensile stress of the (7 × 7) unit cell. It will be shown that stress relief is greatest in the faulted half of the unit cell, consistent with the known larger tensile stress of that half [22, 23]. It will be further shown that metastable species proposed to exist on the surface at room temperature do not have significant stress signatures. A widely accepted reaction model [9] will also be extended to account for the known preference of oxygen for reacting with the faulted half of the unit cell [14]. Finally, the validity of three proposed reaction mechanisms will be assessed in terms of the above findings. Note that this chapter contains material based in part on material previously published in Ref. 10.

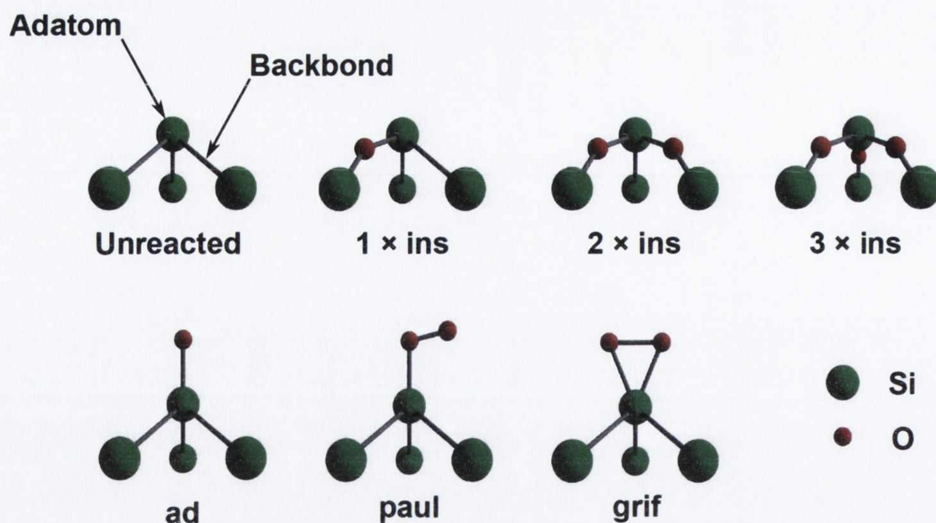


Fig. 5.1. Proposed products arising from the reaction of molecular oxygen with the Si(111)–(7 × 7) surface. Large (green) spheres represent silicon atoms, while small (red) spheres represent oxygen atoms. Structures are labeled '$n\times$ins', 'ad', 'paul' and 'grif', with 'ins' standing for inserted, 'ad' for adsorbing, 'paul' for Pauling, and 'grif' for Griffith. '$n\times$ins' structures may be combined with 'ad', 'paul' and 'grif' structures to form '$n\times$ins-ad', '$n\times$ins-paul' and '$n\times$ins-grif' structures respectively [10].

5.1 Background

The initial stages of the oxidation of the Si(111)–(7 × 7) surface have proven controversial, with conflicting evidence supporting both fully-dissociative and precursor-facilitated adsorption mechanisms [9, 11, 12, 19, 24–29]. The structures of the products formed following adsorption have proven equally controversial, with conflicting evidence supporting the presence of several different structures at different temperatures [12–15, 18, 19, 29–37]. Some of these structures are illustrated in Fig. 5.1. Large (green) spheres represent silicon atoms, while small (red) spheres represent oxygen atoms. Structures are labeled '$n\times$ins', 'ad', 'paul' and 'grif', with 'ins' standing for inserted, 'ad' for adsorbing, 'paul' for Pauling, and 'grif' for Griffith. '$n\times$ins' structures may be combined with 'ad', 'paul' and 'grif' structures to form '$n\times$ins-ad', '$n\times$ins-paul' and '$n\times$ins-grif' structures respectively.

The 'ins-ad' structure is the initial product formed following dissociative adsorption of an oxygen molecule at temperatures below 200 K [14, 15, 29, 37]. At temperatures above 200 K, the 'ins-ad' structure transforms spontaneously into the '2$\times$ins' structure [14, 15, 29, 37]. The reaction of an additional oxygen molecule with the '2$\times$ins' structure is proposed to result in the formation of the '3$\times$ins-ad' structure [14, 15, 29, 34, 35] suggested to be the final product at room temperature. An illustration of the the overall reaction mechanism is shown in Fig 5.2. Fig. 5.3 shows unoccupied state STM images of the Si(111)–(7 × 7) surface following exposure to 0.1 L of oxygen at 30 K and 300 K respectively. 'ins-ad' structures appear dark in the 30 K image, while '2 × ins' structures appear bright [14, 37], consistent with theoretical predictions [11, 38].

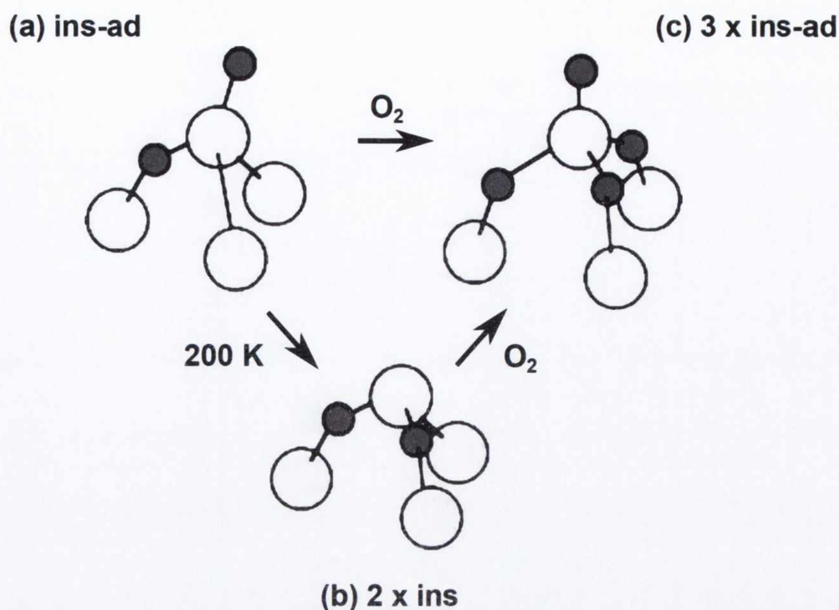


Fig. 5.2. Proposed mechanism for the reaction of molecular oxygen with the Si(111)–(7 × 7) surface at room temperature. Large unshaded circles represent silicon atoms, while small shaded circles represent oxygen atoms. The first step involves the reaction of an oxygen molecule with an unreacted adatom to form the 'ins-ad' structure. The second step involves the spontaneous transformation of the 'ins-ad' structure into the '2×ins' structure. The third step involves the reaction of an additional oxygen molecule with the '2×ins' structure to form the '3×ins'-ad structure suggested to be the final product at room temperature [15].

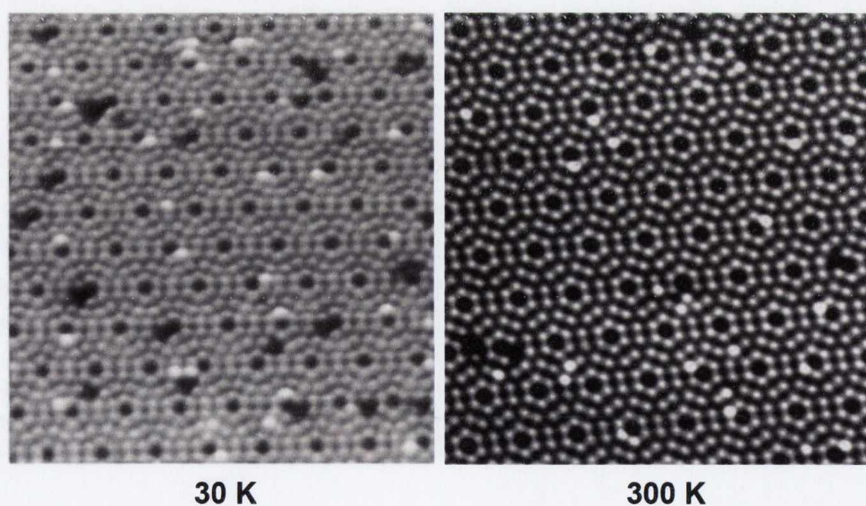


Fig. 5.3. Unoccupied state STM images of the Si(111)–(7 × 7) surface following exposure to 0.1 L of oxygen at 30 K and 300 K respectively. Each image was acquired over a 400 nm^2 area. The 30 K image was acquired at a sample voltage of 1.5 V and a tunneling current of 0.2 nA, while the 300 K image was acquired at a sample voltage of 1.5 V and a tunneling current of 1.0 nA. 1 L is defined as 10^{-6} Torr seconds [14].

Bright and dark sites are also observed in the 300 K image. Here, bright sites have been attributed to 'paul', 'para' and 'grif' structures based on the results of STM, photoemission spectroscopy (PES) and theoretical calculations [9, 34, 34]. They have also been attributed to 'n-ins' structures based on the results of alternative studies utilizing the same techniques [11, 13–15]. An oxygen bridge structure formed between a rest atom and an adatom has also been proposed based on the results of theoretical calculations [12]. Dark sites, meanwhile, have been attributed to '$n\times$ins-ad' and '$n\times$ins-paul' structures containing at least two inserted atoms based on the results of STM, PES and isotope-labeled vibrational spectroscopy [13–15]. Assignment of these structures is further complicated by the proposed presence of long-lived metastable species [11–13, 16–18]. These will be discussed later.

The coverage of bright and dark sites can be described using a simple kinetic model [9]. This model describes the reaction as a consecutive, two-step process [39] involving the reaction of an oxygen molecule with an unreacted adatom 'A' to form a bright intermediate state 'B' with a rate constant k_1, followed by subsequent reaction of the intermediate state with an additional oxygen molecule to form a dark, fully-reacted state 'D' with a rate constant k_2 [9]. Assuming both k_1 and k_2 are time- and pressure-independent, the coverages of bright sites, $\Phi_{\mathrm{B}}(d)$, and dark sites, $\Phi_{\mathrm{D}}(d)$, as a function of exposure d are given by [10]

$$\Phi_{\mathrm{B}}(d) = \frac{k_1\left[e^{-k_1 d} - e^{-k_2 d}\right]}{k_2 - k_1} \tag{5.1}$$

$$\Phi_{\mathrm{D}}(d) = 1 - \left[\frac{k_2 e^{-k_1 d} - k_1 e^{-k_2 d}}{k_2 - k_1}\right]. \tag{5.2}$$

Note that this model does not account for the known preference of oxygen for reacting with the faulted half of the unit cell [14, 32]. This will be discussed later.

5.2 Experimental

Experiments were performed at room temperature in a light-free ultra-high vacuum chamber containing an Omicron STM-1 [40] equipped with the sample holder shown in Fig. 4.2. The base pressure of the chamber was 5×10^{-11} Torr and was monitored using a Televac 7FC [41] cold cathode gauge to eliminate gas cracking and sample heating effects. The sample used was 60 mm long $\times$ 5 mm wide $\times$ 0.280 mm thick and prepared from a single-side-polished, n-type/phosphorus-doped Si(111) wafer with a resistivity of 0.008-0.012 Ω cm. A clean Si(111)–(7 $\times$ 7) surface approximately 7 mm long $\times$ 5 mm wide was prepared using the electron bombardment heater described in Section 4.1.3. The quality of the surface was confirmed using STM. Capacitance measurements were performed using an Andeen-Hagerling AH2550A [42] capacitance bridge in the configuration shown in Fig. 4.9. A reference measurement was performed to rule out the presence of parasitic electronic effects discussed in Section 4.2.1.

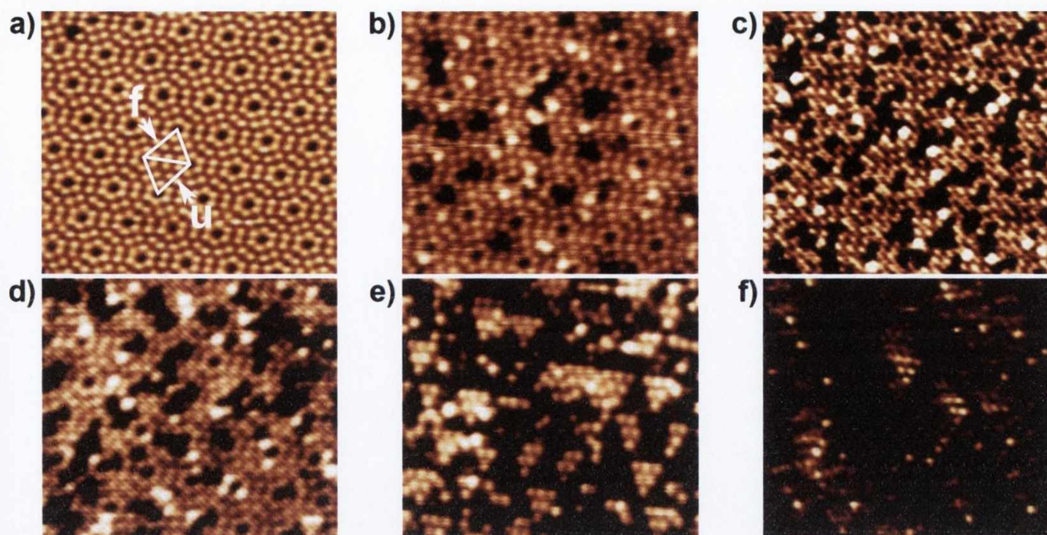


Fig. 5.4. Unoccupied state STM images of the Si(111)–(7 × 7) surface acquired before and after exposure to oxygen gas, showing: (a) clean surface; (b) 0.4 L exposure; (c) 0.9 L exposure; (d) 1.3 L exposure; (e) 7 L exposure; (f) 53 L exposure. Images were acquired at a sample voltage of 1.3 V and a tunneling current of 0.06 nA. Faulted and unfaulted halves of the (7 × 7) unit cell are labeled 'f' and 'u' respectively. 1 L is defined as 10^{-6} Torr seconds [14].

5.3 Results & Discussion

This section begins with the analysis of STM data acquired following oxygen exposures up to 53 L, where 1 L is defined as 10^{-6} Torr seconds. Fig. 5.4 shows unoccupied state STM images of the Si(111)–(7 × 7) surface following oxygen exposures of 1.3 L, 7 L and 53 L respectively. Faulted and unfaulted halves of the (7 × 7) unit cell are labeled 'f' and 'u' respectively. Both bright and dark sites are visible, with the latter found to dominate at higher coverages, consistent with the results of previous studies [9, 14]. A greater number of dark sites are present in the faulted half of the unit cell, consistent with the known preference of oxygen for reacting with that half [14, 32]. A small number of unfaulted-half subunits are visible following exposure to 53 L of oxygen, indicating preservation of the underlying 7 × 7 reconstruction [43–45].

Fig. 5.5a, Fig. 5.5b and Fig. 5.5c show the results of a statistical analysis of bright and dark site populations performed using data obtained from counting bright and dark sites in each STM image. Note that a minimum of 500 unit cells were counted per image. Fig. 5.5a shows the experimentally-determined coverage of bright sites (red dots) and dark sites (black squares) plotted as a function of oxygen exposure. The number of dark sites is found to increase with increasing exposure, while the number of bright sites is found to increase initially and decrease quickly thereafter. This is in good agreement with the kinetic model [9] described in Section 5.1.

Fig. 5.5b shows the experimentally-determined coverage of dark sites in the faulted (black squares) and unfaulted (red dots) halves of the unit cell plotted as a function of

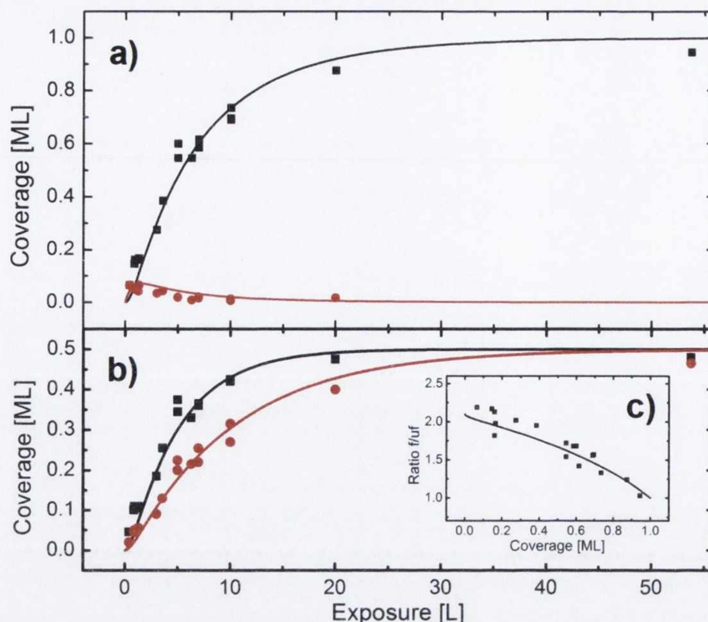


Fig. 5.5. Statistical analysis of bright and dark site populations, showing: (a) Experimentally-determined coverage of bright sites (red dots) and dark sites (black squares) plotted as a function of oxygen exposure. Note that one monolayer of bright (dark) sites corresponds to all adatoms appearing bright (dark). Solid lines represent bright and dark site coverages predicted by the extended kinetic model described in the text. (b) Experimentally-determined coverage of dark sites in faulted (black squares) and unfaulted (red dots) halves of the unit cell plotted as a function of oxygen exposure. Solid lines represent the dark site coverage predicted by the extended kinetic model described in the text. (c) Ratio of faulted-half dark site coverage to unfaulted-half dark site coverage plotted as a function of dark site coverage [10].

oxygen exposure. Bright sites are omitted for clarity. By assuming different rate constants for the faulted and unfaulted halves, the kinetic model can be extended to account for the the known preference of oxygen for reacting with the faulted half. The predicted coverages of dark sites in the faulted (black line) and unfaulted (red line) halves shown in Fig. 5.5b were calculated using Eq. 5.2 with rate constants $k_{1(\mathrm{f})} = 0.21\ \mathrm{L}^{-1}$, $k_{2(\mathrm{f})} = 1.7\ \mathrm{L}^{-1}$, $k_{1(\mathrm{uf})} = 0.1\ \mathrm{L}^{-1}$ and $k_{2(\mathrm{uf})} = 1.7\ \mathrm{L}^{-1}$. Note that the value of k_2 was taken from Ref. 9, while the ratio of $k_{1(\mathrm{f})}$ to $k_{1(\mathrm{uf})}$ was determined experimentally using data presented in Fig. 5.5c. These values are in good agreement with theoretical calculations [46] predicting that k_2 should be greater than k_1.

The next stage of the analysis focuses on the evolution of surface stress observed during oxygen exposure. Fig. 5.6b shows the change in capacitance measured between the sample and the *back electrode* plotted as a function of oxygen exposure. Red, blue and green curves correspond to the change in capacitance measured at pressures of 2×10^{-8} Torr, 6×10^{-8} Torr and 3×10^{-7} Torr respectively. Note that blue and green curves were obtained with a dosing tube directed at the sample surface. Sample displacement was calculated from the change in capacitance using the equations described in Section 4.1.2, while the change in surface stress was calculated using Eq. (4.5) with values of $2.29 \times 10^{11}\ \mathrm{N\,m}^{-2}$ for

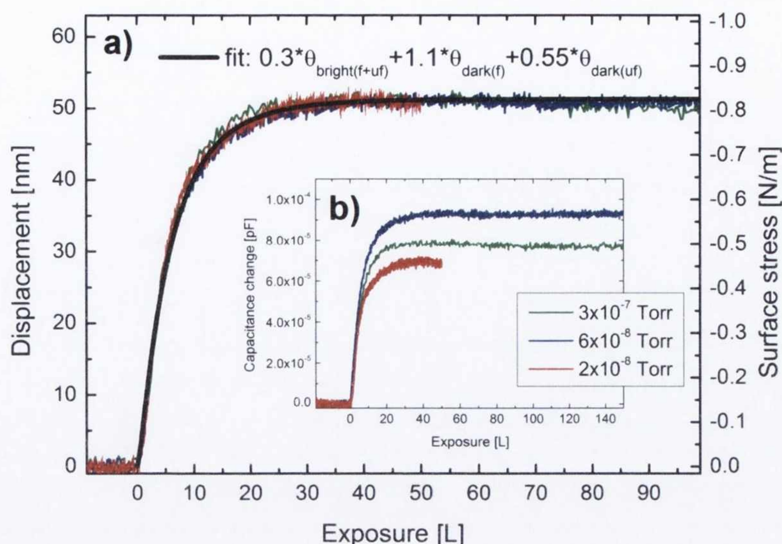


Fig. 5.6. Evolution of surface stress observed during oxygen exposure, showing: (a) Change in surface stress plotted as a function of oxygen exposure for pressures of 2×10^{-8} Torr (red line), 6×10^{-8} (blue line) Torr and 3×10^{-7} Torr (green line) respectively. Note that graphs have been rescaled using data shown inset. The solid black line represents the change in surface stress calculated using the kinetic stress model described in the text. (b) Change in capacitance measured between the sample and the *back electrode* plotted as a function of oxygen exposure for the same pressures as above. The red curve was obtained with near-equal pressures at the polished and unpolished sides of the sample, while the blue and green curves were obtained with a pressure ~ 60 times higher at the polished side [10].

$E/(1-\nu)$ [47], 0.3 ± 0.02 mm for t, 7 ± 1 mm for L, and 27 ± 1 mm for l. Saturation coverage of the surface yielded a compressive change in surface stress of -0.82 ± 0.25 N m^{-1} corresponding to an energy change of -0.66 eV per (1×1) unit cell. Note that the shape of each curve shown in Fig. 5.6b changes *only* as a function of oxygen coverage. Each curve is time- and pressure-independent, with differences in magnitude arising only due to variations in the size of the prepared (7×7) surface and the size of stress-accumulating terraces on the surface.

In spite of these variations, each curve can be arbirarily rescaled to produce the curve shown in Fig. 5.6a. This leads to two conclusions. First, the stress contribution of the unpolished side of the sample is negligible with respect to that of the polished side. This is supported by the fact that the shape of each curve remains the same irrespective of whether a dosing tube is used or not, and by the fact that there is only a small decrease in the saturation value associated with the red curve. Second, changes in surface stress occur *only* as a function of oxygen coverage. This implies that changes in surface stress are time- and pressure-independent. Consequently, metastable species proposed to exist at room temperature cannot have significant stress signatures.

By assuming different stress contributions for bright and dark sites in the faulted and unfaulted halves of the unit cell, and by assuming that these are additive, the extended kinetic model can be used to relate the change in surface stress to the site-specific oxygen coverage. The calculated change in surface stress best fitting the experimental data is given by

$$\Delta\sigma = 0.3\Phi_{B(f+uf)} + 1.1\Phi_{D(f)} + 0.55\Phi_{D(uf)}. \quad (5.3)$$

The fitted curve is shown in Fig. 5.6a. This leads to two conclusions. First, dark sites appear to contribute more to the change in surface stress than bright sites. This would be consistent with the larger number of oxygen atoms inserted in the backbonds of dark sites relative to bright sites. Note, however, that since the number of bright sites is small compared to the number of dark sites, they contribute less to the *total* change in surface stress, but this does not imply that an individual bright site contributes less to the change in surface stress than an individual dark site.

Second, dark sites in the faulted half of the unit cell contribute more to the change in surface stress than dark sites in the unfaulted half. This can be understood by considering the fact that the Si(111)–(7 × 7) unit cell is an intrinsically tensile structure [22, 23], with the faulted half known to be under greater tensile stress than the unfaulted half [22, 23]. It is reasonable to conclude that the insertion of an oxygen atom into a backbond would partially relieve this stress. This view is consistent with the fact that the calculated average tensile stress of the Si(111)–(7 × 7) surface is 1.9 eV per (1 × 1) unit cell [22], while the change in surface stress measured at saturation is -0.66 eV per (1 × 1) unit cell. This is also consistent with preservation of the underlying 7 × 7 reconstruction indicated by the STM data.

Since the insertion of an oxygen atom into a backbond partially relieves tensile stress, it is also reasonable to conclude that stress relief would be greatest in the faulted half of the unit cell since it is under the greatest tensile stress. Note that the change in surface stress measured above differs from that measured in an earlier study [21], which found a a compressive change in surface stress of -7.2 N m^{-1}. This corresponds to an energy change of -5.8 eV per (1 × 1) unit cell. Based on the calculated average tensile stress of the Si(111)–(7 × 7) cited above, this would result in a final *compressive* stress of 3.9 eV per (1 × 1). This is inconsistent with preservation of the underlying 7 × 7 reconstruction.

The final stage of the analysis focuses on application of the previous findings to assess the validity of three reaction mechanisms proposed in the literature. These are illustrated in Fig. 5.7, Fig. 5.8 and Fig. 5.9 respectively. Mechanism A involves the reaction of an oxygen molecule with a bright '2×ins' structure to form a dark '3×ins-ad' structure. No intermediate is formed. Since the number of oxygen atoms inserted in the '3×ins-ad' structure is greater than the number of oxygen atoms inserted in the '2×ins' structure, a change in surface stress is expected. This mechanism is fully consistent with the previous findings. Mechanism B involves the reaction of an oxygen molecule with a bright '2×ins' structure to form a metastable '2×ins-paul' intermediate, which subsequently transforms into a dark '2×ins-ad' structure. This transformation is time-dependent. Since the number

of inserted oxygen atoms remains constant across product states, no change in surface stress is expected. This could be consistent with the findings above, but only if bright and dark sites contribute equally to the change in surface stress. Mechanism C involves the reaction of an oxygen molecule with a bright 'ins' structure to form a metastable 'ins-paul' intermediate, which subsequently transforms into a dark '2×ins-ad' structure. Once again, this transformation is time-dependent. Since the number of oxygen atoms inserted in the '2×ins-ad' structure is greater than the number of oxygen atoms inserted in the 'ins' structure, a time-dependent change in surface stress is expected. This is inconsistent with the findings above.

5.4 Conclusions

The oxidation of the Si(111)–(7×7) surface has been studied using a combined STM and surface stress measurement system. A widely accepted kinetic reaction model [9] describing the coverage of bright and dark sites observed in unoccupied state STM images has been extended to account for the known preference of oxygen for reacting with the faulted half of the (7×7) unit cell [14, 32]. The site-specific evolution of surface stress has been measured and quantitatively associated with relief of the intrinsic tensile stress of the (7×7) unit cell. It was shown that stress relief is greatest in the faulted half of the unit cell, consistent with the known larger tensile stress of that half [22, 23]. It was also shown that the underlying 7×7 reconstruction is preserved even at saturation oxygen coverage. Furthermore, it was shown that that metastable species proposed to exist on the surface at room temperature do not have significant stress signatures. Using these findings, the validity of three reaction mechanisms proposed in the literature was assessed. The mechanism proposed in Ref. 14 and Ref. 15 was found to be in good agreement with the experimental data.

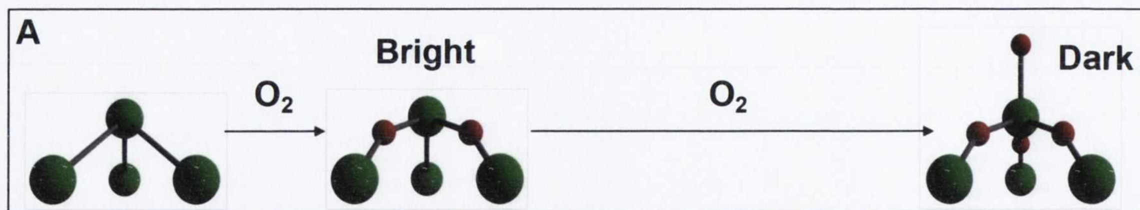


Fig. 5.7. Reaction mechanism proposed in Ref. 14 and Ref. 15. An unreacted adatom site reacts with an oxygen molecule to form a bright '2×ins' structure that subsequently reacts with an additional oxygen molecule to form a dark '3×ins-ad' structure. No metastable intermediate is formed.

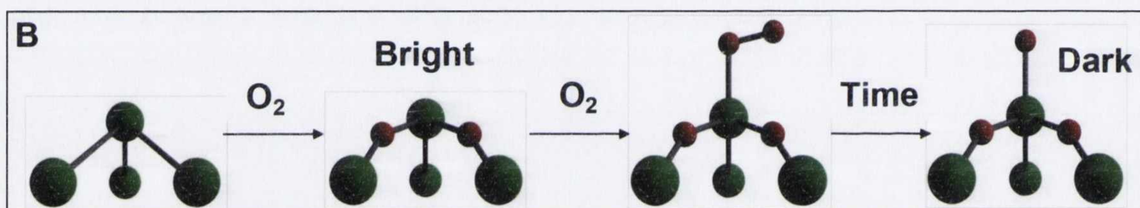


Fig. 5.8. Reaction mechanism proposed in Ref. 13. An unreacted adatom site reacts with an oxygen molecule to form a bright '2×ins' structure that subsequently reacts with an additional oxygen molecule to form a metastable '2×ins-paul' intermediate. This intermediate subsequently undergoes a time-dependent transformation into a dark '2×ins-ad' structure.

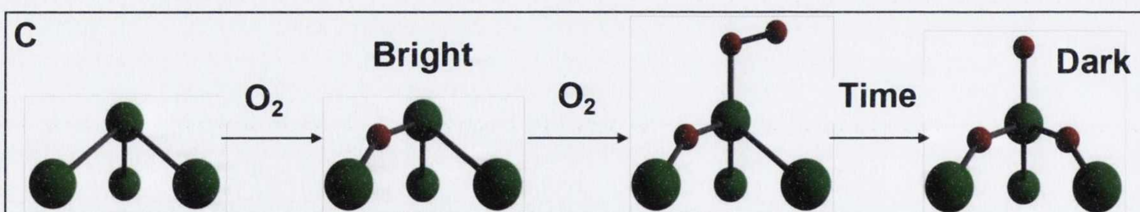


Fig. 5.9. Reaction mechanism proposed in Ref. 13 An unreacted adatom site reacts with an oxygen molecule to form a bright 'ins' structure that subsequently reacts with an additional oxygen molecule to form a metastable 'ins-paul' intermediate. This intermediate subsequently undergoes a time-dependent transformation into a dark '2×ins-ad' structure.

References

[1] E. A. Irene, Crit. Rev. Solid State Mater. Sci. **14**, 175 (1988).

[2] T. Engel, Surf. Sci. Rep. **18**, 93 (1993).

[3] H. N. Waltenburg and J. T. Yates, Chem. Rev. **95**, 1589 (1995).

[4] R. J. Hamers and Y. Wang, Chem. Rev. **96**, 1261 (1996).

[5] A. Stesmans and B. Nouwen, Phys. Rev. B **61**, 16068 (2000).

[6] B. Nouwen and A. Stesmans, Mat. Sci. Eng. A **288**, 239 (2000).

[7] A. Stesmans and V. V. Afanas'ev, Appl. Surf. Sci. **168**, 324 (2000).

[8] W. Futako, T. Umeda, M. Nishizawa, T. Yasuda, J. Isoya, and S. Yamasaki, J. Non-Cryst. Solids **299-302**, 575 (2002).

[9] G. Dujardin, A. Mayne, G. Comtet, L. Hellner, M. Jamet, E. Le Goff, and P. Millet, Phys. Rev. Lett. **76**, 3782 (1996).

[10] N. T. Kinahan, D. E. Meehan, T. Narushima, S. Sachert, K. Miki, and J. J. Boland, Phys. Rev. Lett. **104**, 146101 (2010).

[11] S.-H. Lee and M.-H. Kang, Phys. Rev. Lett. **82**, 968 (1999).

[12] T. Hoshino and Y. Nishioka, Phys. Rev. B **61**, 4705 (2000).

[13] K. Sakamoto, H. M. Zhang, and R. I. G. Uhrberg, Phys. Rev. B **68**, 075302 (2003).

[14] A. J. Mayne, F. Rose, G. Comtet, L. Hellner, and G. Dujardin, Surf. Sci. **528**, 132 (2003).

[15] H. Okuyama, Y. Ohtsuka, and T. Aruga, J. Chem. Phys. **122**, 234709 (2005).

[16] F. Matsui, H. W. Yeom, K. Amemiya, K. Tono, and T. Ohta, Phys. Rev. Lett. **85**, 630 (2000).

[17] K. Sakamoto, S. Thorbjörn Jemander, G. V. Hansson, and R. I. G. Uhrberg, Phys. Rev. B **65**, 155305 (2002).

[18] K. Sakamoto, H. M. Zhang, and R. I. G. Uhrberg, Phys. Rev. B **70**, 035301 (2004).

[19] R. Martel, P. Avouris, and I.-W. Lyo, Science **272**, 385 (1996).

[20] I.-S. Hwang, R.-L. Lo, and T. T. Tsong, Phys. Rev. Lett. **78**, 4797 (1997).

[21] D. Sander and H. Ibach, Phys. Rev. B **43**, 4263 (1991).

[22] D. Vanderbilt, Phys. Rev. Lett. **59**, 1456 (1987).

[23] R. D. Meade and D. Vanderbilt, Phys. Rev. B **40**, 3905 (1989).

[24] H. Ibach, H. D. Bruchmann, and H. Wagner, Appl. Phys. A **29**, 113 (1982).

[25] K. Edamoto, Y. Kubota, H. Kobayashi, M. Onchi, and M. Nishijima, J. Chem. Phys. **83**, 428 (1985).

[26] U. Höfer, P. Morgen, W. Wurth, and E. Umbach, Phys. Rev. Lett. **55**, 2979 (1985).

[27] U. Höfer, P. Morgen, and W. Wurth, Phys. Rev. B **40**, 1130 (1989).

[28] K. Sakamoto, S. Suto, and W. Uchida, Surf. Sci. **357-358**, 514 (1996).

[29] G. Comtet, L. Hellner, G. Dujardin, and K. Bobrov, Phys. Rev. B **65**, 035315 (2001).

[30] P. Morgen, U. Hfer, W. Wurth, and E. Umbach, Phys. Rev. B **39**, 3720 (1989).

[31] I.-W. Lyo, P. Avouris, B. Schubert, and R. Hoffmann, J. Phys. Chem. **94**, 4400 (1990).

[32] J. P. Pelz and R. H. Koch, Phys. Rev. B **42**, 3761 (1990).

[33] P. Avouris, I.-W. Lyo, and F. Bozso, J. Vac. Sci. Technol. B **9**, 424 (1991).

[34] B. Schubert, P. Avouris, and R. Hoffmann, J. Chem. Phys. **98**, 7593 (1993).

[35] B. Schubert, P. Avouris, and R. Hoffmann, J. Chem. Phys. **98**, 7606 (1993).

[36] S.-H. Lee and M.-H. Kang, Phys. Rev. B **61**, 8250 (2000).

[37] H. Okuyama, T. Aruga, and M. Nishijima, Phys. Rev. Lett. **91**, 256102 (2003).

[38] H. Kageshima, Y. Ono, M. Tabe, and T. Ohno, Jpn. J. Appl. Phys. **33**, 4070 (1994).

[39] P. Atkins and J. de Paula, *Physical Chemistry* (Oxford University Press, Great Clarendon St., Oxford, 2010), 9th ed.

[40] Omicron STM-1, Omicron NanoTechnology GmbH, Taunusstein, Germany.

[41] Televac 7FC, Televac, 2400 Philmont Avenue, Huntingdon Valley, Philadelphia 19006-0662.

[42] Andeen-Hagerling AH2550A, Andeen-Hagerling, Inc., 31200 Bainbridge Road, Cleveland, Ohio.

[43] F. M. Leibsle, A. Samsavar, and T. C. Chiang, Phys. Rev. B **38**, 5780 (1988).

[44] J. M. Gibson, Surf. Sci. Lett. **239**, L531 (1990).

[45] H. Tokumoto, K. Miki, H. Murakami, H. Bando, M. Ono, and K. Kajimura, J. Vac. Sci. Technol. A **8**, 255 (1990).

[46] P. Sonnet, L. Stauffer, and C. Minot, Surf. Rev. Lett. **6**, 1031 (1999).

[47] W. A. Brantley, J. Appl. Phys. **44**, 534 (1973).

Conclusions

- A novel combined scanning tunneling microscopy (STM) and surface stress measurement system was developed. This system uniquely allows absorbate-induced changes in surface stress to be measured and related to the structural and electronic changes causing them *in situ* and with atomic resolution.
- The combined system was used to investigate the oxidation of the Si(111)–(7×7) surface at room temperature. A widely accepted kinetic reaction model [1] describing the coverage of bright and dark sites observed in unoccupied state STM images has been extended to account for the known preference of oxygen for reacting with the faulted half of the (7×7) unit cell [2, 3].
- The site-specific evolution of surface stress observed during the oxidation of the Si(111)–(7×7) surface at room temperature was measured and quantitatively associated with relief of the intrinsic tensile stress of the (7×7) unit cell.
- It was shown that oxidation-induced relief of the intrinsic tensile stress of the Si(111)–(7×7) surface is greatest in the faulted half of the unit cell, consistent with the known larger tensile stress of that half [4, 5].
- It was shown that the Si(111)–(7×7) reconstruction is preserved even at saturation oxygen coverage.
- It was shown that metastable species proposed to exist during the oxidation of the Si(111)–(7×7) surface at room temperature do not have significant stress signatures. Using this information, the validity of three reaction mechanisms proposed in the literature was assessed. The mechanism proposed in Ref. 3 and Ref. 6 was found to be in good agreement with the experimental data.
- Following the successful investigation of the oxidation of the Si(111)–(7×7) surface at room temperature, the system will be used to investigate the oxidation of the technologically-important Si(100)–(2×1) surface at room temperature. It is expected that the results of this investigation will be of practical use in the fabrication of next-generation silicon-based devices.

References

[1] G. Dujardin, A. Mayne, G. Comtet, L. Hellner, M. Jamet, E. Le Goff, and P. Millet, Phys. Rev. Lett. **76**, 3782 (1996).

[2] J. P. Pelz and R. H. Koch, Phys. Rev. B **42**, 3761 (1990).

[3] A. J. Mayne, F. Rose, G. Comtet, L. Hellner, and G. Dujardin, Surf. Sci. **528**, 132 (2003).

[4] D. Vanderbilt, Phys. Rev. Lett. **59**, 1456 (1987).

[5] R. D. Meade and D. Vanderbilt, Phys. Rev. B **40**, 3905 (1989).

[6] H. Okuyama, Y. Ohtsuka, and T. Aruga, J. Chem. Phys. **122**, 234709 (2005).