Athermal Stress Release of Silicon Surface and the Microscopic Structure

Tetsuya NARUSHIMA

A dissertation submitted to the Doctoral Program
in Physics, the University of Tsukuba
in partial fulfillment of requirements
for degree of Doctor of Philosophy (Science)

October, 2001

Contents

0.1 General introduction 1
0.2 Organization of this thesis 2

1 Backgrounds **3**
1.1 Semiconductor surfaces: reconstructed surface structures of Si(100) 3
1.1.1 Introduction to Silicon 3
1.1.2 Clean surfaces 7
1.2 Surface and interface stress 24
1.2.1 Introduction to surface and interface stress 24
1.2.2 The charge distribution and surface stress on clean surfaces 28
1.2.3 Adosobate-induced stresses - general discussion 30

2 Principle of surface stress measurements and the experimental setup **37**
2.1 Experimental methods 37
2.1.1 Detection of surface stress 38
2.1.2 Cantilever bending methods 39
2.1.3 Stoney's formula 42
2.1.4 Temperature estimation of microcantilever with 'bimetallic effect' 47
2.2 Our experimental setup 48
2.2.1 Optical micro-cantilever bending technique 50

3 Electron-stimulated surface stress relaxation of Si **53**
3.1 Abstract of this chapter 53
3.2 Introduction 53
3.3 Experimental 55
3.3.1 Samples of Si(100) microcantilever and its cleaning 55
3.3.2 Methods for irradiations of ions and electrons 55

3.3.3 Measurements of stresses during ion and the following electron irradiation . . . 56
3.4 Results and discussions . . . 63
3.4.1 Ion bombardment induced surface stresses . . . 63
3.4.2 Electron irradiation onto the bombarded surfaces . . . 66
3.4.3 Non-thermal mechanism of surface stress relaxation . . . 75
3.4.4 Energy density estimation of the surface stress relaxation based on IED mechanism . . . 76
3.4.5 Thermal gradient effect due to further electron irradiation . . . 80
3.4.6 Electron irradiation in oxidation atmosphere . . . 85
3.5 Further work . . . 86

4 Microscopic structure observed with a Scanning Tunneling Microscopy 90
4.1 Abstract of this chapter . . . 90
4.2 Introduction . . . 90
4.3 Experimental . . . 92
4.3.1 Scanning Tunneling Microscopy observation of Plasma-modified surface . . . 92
4.3.2 Pretreatment and features of clean Si(100) surfaces . . . 94
4.3.3 Electron and ion irradiation methods with DC plasma 94
4.4 Atomic level identification of surface restoration by electron irradiation . . . 100
4.4.1 Modified surfaces of Si(100) by ion bombardment . . . 100
4.4.2 Modification of the disordered surface by low energy electron irradiation . . . 104
4.5 A possibility of ordered oxidation . . . 110

5 Conclusion 115

A Scanning Tunneling Microscopy 124
A.1 Basic principles . . . 124
A.1.1 Vacuum tunneling . . . 124
A.1.2 Scanning tunneling microscopy . . . 125
A.2 Theory of STM imaging . . . 128
A.2.1 Beyond Topography . . . 128
A.2.2 Tunneling Hamiltonian approach . . . 129
A.2.3 Modeling the tip . . . 130
A.3 Metal surfaces: STM as surface topography . . . 131
A.3.1 Calculation of the LDOS . . . 131

A.3.2 Atom-superposition modeling . . . 132
A.4 Semiconducting surfaces: role of surface electronic structure . 133
A.4.1 Voltage dependence of images . . . 133

B Basic plasma equations and equilibrium 141
B.1 Field equations, current . . . 142
B.2 The conservation equations . . . 145
B.2.1 Boltzmann's equation . . . 145
B.2.2 Macroscopic quantities . . . 147
B.2.3 Particle conservation . . . 147
B.2.4 Momentum conservation . . . 148
B.2.5 Energy conservation . . . 151
B.2.6 Summary . . . 151
B.3 Equilibrium properties . . . 152
B.3.1 Boltzmann's relation . . . 154
B.3.2 Quasineutrality . . . 155

C Discharge and diagnostics of cold plasmas 157
C.1 Introduction to plasmas . . . 157
C.2 Definition of plasma . . . 158
C.3 Plasma parameters . . . 161
C.3.1 The degree of ionization . . . 163
C.3.2 Plasma temperature . . . 164
C.3.3 Debye length . . . 167
C.3.4 Plasma frequency . . . 169
C.4 Conditions for plasma existence . . . 170
C.5 Plasma DC sheath calculations . . . 171
C.5.1 Basic concept and equations . . . 171
C.5.2 The Bohm sheath criterion . . . 173
C.5.3 The high voltage sheath . . . 178
C.6 Diagnostics of plasma: electrostatic probe diagnostics . . . 180
C.6.1 Planer probe with collisionless sheath . . . 183
C.7 Cold plasma generation . . . 184
C.7.1 DC grow discharges . . . 185
C.7.2 Radio frequency discharges . . . 191

Abstract

A new mechanism of recrystallization is presented in this thesis by using irradiation of very low energy electron on the disordered surfaces of Si. The recrystallization was observed by means of two experimental methods from both the macroscopic point of view and also from the microscopic one: surface stress measurement and Scanning Tunneling Microscopy. It was found that the recrystallization is dominated not by thermal activation mechanism, but by non-thermal mechanism induced by irradiated electrons.

It is well known that thermal annealing at high temperatures can recrystallize a surface disordered layer and release the surface stress of the disordered layer. However, is it possible that the disorder-induced stress in the disordered layer can be relaxed at low temperature and the surface disordered layer can be recrystallized? If the recrystallization of the disordered layer without thermal annealing at high temperature is possible, it would give important informations to understand the physical and chemical mechanisms of ordering of surface atoms, *i.e.*, the recrystallization of surface.

Defects in Si have different energy levels in the band gap according to the charge states. Carrier trapping at a defect site may create a localized energy release. Hence a possible way to relax disorder-induced stress and change a disordered structure is for the defect to capture an electron and change its own charge state using low energy electron irradiation.

This hypothesis is tested with two experimental methods: surface stress measurement and Scanning Tunneling Microscopy (STM). The surface stress is a quantity which affords a macroscopic description of surfaces and is extremely sensitive to phenomena induced at surfaces or interfaces. In contrast, the STM observation provides methods to measure accurately the structure and electronic properties of single surface atoms. Therefore, we have used both macroscopic and microscopic observation methods to reveal influences of atomic scale phenomena on the macroscopic surface stress. Following results were obtained from these two experimental methods.

The surface stress changes during ion-bombardment and the following electron irradiation on Si(100) was measured by means of an optical micro-cantilever technique. An atomically thin disordered layer was introduced by Ar ion bombardment. We have found that the compressive stress in the Si surface due to disorder induced by ion bombardment was completely relaxed by electron irradiation at low energy. The criterion for complete relaxation is found not to be the total energy of the irradiated electrons, but the number

of irradiated electrons. Hence, we have observed the non-thermal relaxation of surface stress in Si induced by electron irradiation at room temperature.

STM observation under similar electron irradiation after ion-bombardment, performed for cross-check of restoration of the disordered surface, demonstrated re-ordering of atomic level, *i.e.*, recrystallization. Its fine structure of reordered surface was, however, different from the typical structure by thermal annealing. In general, the structure generated by thermal annealing has 2x1 symmetry due to Si dimers. The fine structure of reordered surface, in contrast, was slightly closer to positions of a 'bulk terminated' 1x1 structure. It suggests that non-thermal genesis of a new surface structure was done by very low energy electron irradiation.

In conclusion, we found that the Si surface recrystallizes athermally under very low energy electron irradiation at room temperature without thermal annealing at high temperatures. In demonstrating that surfaces can have the capacity of self-healing, these results suggest new opportunities for atomic scale surface engineering.

0.1 General introduction

An understanding of the physical and chemical origins of ordering of surface atoms, *i.e.*, recrystallization of the surface, is helpful for advances on solid state surface science and also technological applications of atomic scale surface engineering, electronic device manufacturing, surface analysis. In Chapters 3 and 4, we present the new recrystallization method at room temperature for the Si surface under very low energy electron irradiation. In general, thermal annealing at high temperatures can recrystallize a surface disordered layer and would release the stress due to the disordered layer. Two complementary experimental techniques, a macroscopic surface stress measurement and microscopic STM observations, are used to investigate the response of the disordered surface to low-energy electron irradiation, and thereby connect the large and small length scales.

The surface stress and surface energy are quantities which afford a macroscopic description of surfaces and is extremely sensitive on phenomena induced at surfaces or interfaces. In the past decades, nevertheless, the science of solids surfaces has developed largely with the emphasis only to gain insight into microscopic structure of surfaces on an atomic scale. With an ever-increasing knowledge about the crystallographic structure, the electronic, magnetic and dynamical properties of surfaces, and with the ability to engineer surface and interface systems with particular properties, experimental and theoretical studies on macroscopic aspects of surfaces fell out of fashion. It is only relatively recent that the value of a thermodynamic, or coarse-grained, description is again applied. The particular value of a coarse-grained description rests with the fact that one is able to draw conclusions about qualitative and even some quantitative aspects of phenomena without explicit knowledge of all the atomistic process involved. Tutorial examples of this kind are the recent experimental studies on the fluctuations of monoatomic steps[1], the Brownian motion of islands[2] and vacancy islands, and the decay of small islands in the process of Ostwald ripening[3]. In these processes, scaling laws for the time dependencies can be derived without explicit knowledge of all atomistic processes. It becomes evident in these studies that the establishment of the relation between microscopic observations and macroscopic quantities is by no means trivial. For example, the individual hops of atoms on surfaces can be observed and quantitatively studied using the field emission microscope or STM. It is, however, a long way from such observations to the understanding of the complexity of all processes involved in the smoothening of a rough surface through mass transport via surface self-diffusion. It is of similar difficulty to link the atomic structure of the various crystallographic faces of a crystalline solid with the specific free energy of

these faces, which determine the equilibrium shape of the crystal. Substantial progress has been achieved in recent years in first principle calculations of the surface free energy and the surface stress. There are several well-studied examples of the interplay between surface stress and reconstruction and of the effect of locally anisotropic surface stresses on the restructuring of surfaces on larger, mesoscopic length scales. While it is not yet possible to determine the absolute surface stress experimentally with acceptable accuracy, techniques have been developed for the measurement of the changes in the surface stress when the surface undergoes reconstruction. To date, while such studies of the surface stress are mainly focused on metal surfaces, very few of these studies are on semiconductor surface. In contrast, the STM observation provides, powerfully, methods to measure accurately the structure and electronic properties of single surface atoms. Then, we used both of macroscopic and microscopic observation methods to reveal influences of atomic scale phenomena on the macroscopic surface stress complementarily.

It is the purpose of this thesis to reveal possibilities and problems of recrystallization by very low energy irradiation at low temperature and its origin with measuring the surface stress changes and observing the surface structure of atomic scale.

0.2 Organization of this thesis

In this thesis, the relaxation of surface stress on a disordered Si(100) surface via an athermal surface recrystallization induced by very low energy electron irradiation at room temperature are presented. An optical cantilever bending techniques and Scanning Tunneling Microscopy were performed for measuring the surface stress change and observing the surface structure. Ion-bombardment and electron irradiation at low energy were performed using a cold plasma, while a negative or positive dc bias was applied to the Si samples to attract ions or electrons in the plasma. In Chapter 1, from now on, the fundamentals of Si surfaces, and the surface stress will be introduced in Chapter 2. In Chapter 3, we will discuss the nonthermal relaxation of the compressive stress in the Si surface due to disorder induced by ion bombardment and, in Chapter 4, mention probabilities and problems of recrystallization at room temperature by very low energy electron irradiation. In Chapter 5, we will summarize these results and discuss future applications. In the appendixes, we will treat optional knowledge of STM and plasma etc.

Chapter 1

Backgrounds

We will proceed studies using two powerful experimental tools of surface stress measurements and STM observations. In this chapter, in order to understand these outputs, we will obtain the fundamental knowledges on the Si surface structures in section 1.1 and on the surface stresses induced at the surfaces and interfaces in section 1.2.

1.1 Semiconductor surfaces: reconstructed surface structures of Si(100)

1.1.1 Introduction to Silicon

The semiconductor industry, along with a large segment of the condensed matter physics community, devotes the lion's share of its resources to element 14 in the periodic table, silicon. Silicon lies just below carbon in Group IV, and, together with germanium, is an intrinsic semiconductor with an indirect bandgap energy of 1.17eV. Silicon principally forms tetrahedral sp3 bonds, taking the diamond crystal structure displayed by carbon. The stability of silicon oxide against attack by common processing regents such as water and methanol, coupled with the abundance of this element in the earth's crust, has made it pervasive in advanced technology – from semiconductor microelectronics ot power rectifiers and solar cells. The vast effort devoted for more than a quarter century to silicon technology has resulted in its continued dominance over other materials, notably the III-Vs, in all areas of solid state electronics save optical devices.

The first semiconductor surface imaged with the scanning tunneling microscopy (STM) was the 7×7 reconstruction of Si(111).[4] One of the properties of the elemental semiconductors is that their clean surfaces typically

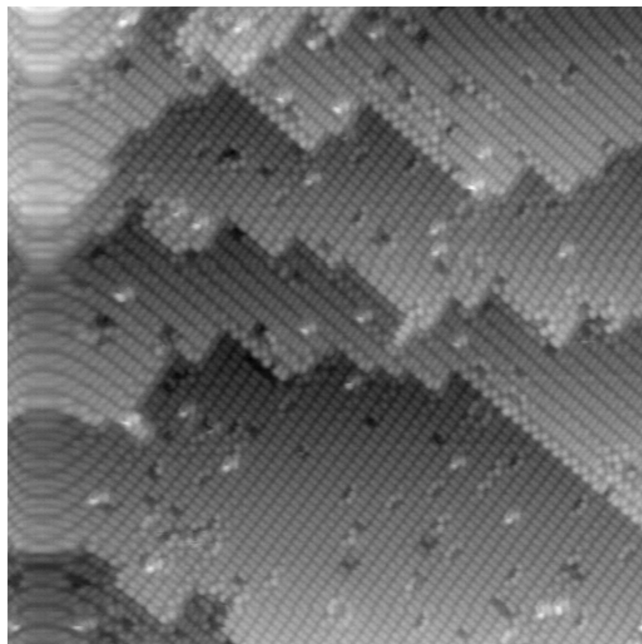


Figure 1.1: STM imaging example of reconstructed Si(100) surface. V_s=-1.5V, I=0.10nA, 40nm width

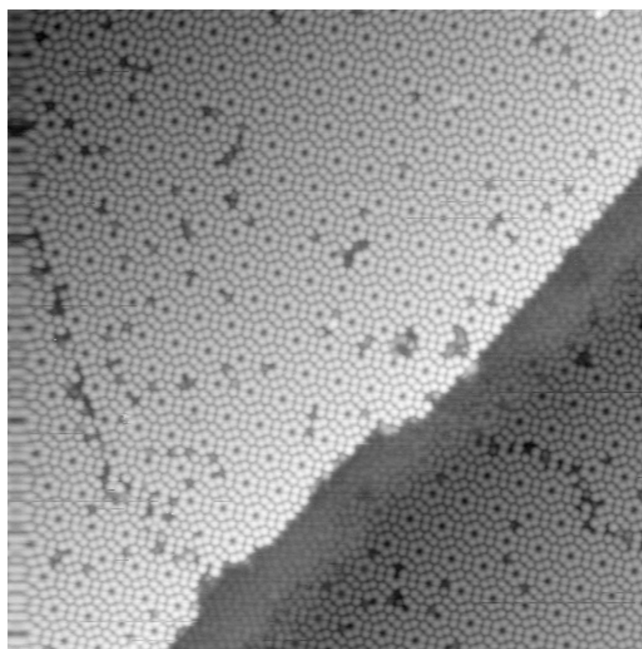


Figure 1.2: STM imaging Example of reconstructed Si(111) surface. V_s=+2.0V, I=0.10nA, 50nm width.

undergo a process termed "reconstruction," whereby the fundamental periodicity of the structure taken by the surface atoms is different from that of the underlying bulk material. This process is due to the covalent nature of their bonds; a simple bulk termination at the surface leaves a large a number of unsatisfied (dangling) bonds that result in a large free energy. In order to mitigate the energy associated with these dangling bonds, the surface atoms rearrange themselves to diminish the dangling bond density, reducing the free energy, generally at some cost in increasing the component of the free energy derived from "surface stress".[5] The terminology "$m \times n$" refers to the two-dimensional Miller-indices needed to describe the surface unit cell (unit mesh) in terms of bulk lattice vectors. The surface superlattice, since it is larger than the bulk lattice, will produce fractional order beams observed in diffraction studies of these reconstructions. The experimental and theoretical basis for surface reconstructions is a field that has occupied many surface scientists for the last 30 years. Since the initial observations, using low energy electron diffraction (LEED), of the 7×7 on Si(111) surfaces in 1959 by Schlier and Fransworth,[6] a large number of models have been tendered in explanation of the unusual symmetry displayed by this surface. The inability of the various studies to settle on particular structures reflects the difficulty of solving the inverse scattering problem in three dimensions; it is not an easy task to determine the scattering potential given the diffraction intensities. This problem is exacerbated in LEED with the large contribution of multiple scattering to the diffracted beams, making detailed analysis of LEED I-V data formidable for even the simplest structures. While the tunneling images of the 7×7 shown by Binnig *et al.* did not unequivocally determine the detailed structure of this phase, they did reject a number of models and cleared the way for the dimer-adatom-stacking fault (DAS) model proposed by Takayanagi *et al.*[7] Indeed, it can be argued that this single observation by Rohrer and Binnig "made" the new field of scanning tunneling microscopy, bringing it forcibly to attention of mainstream surface scientists world-wide. Even more significant is the effect the STM images have had on the outlook of surface studies, turning them from a collective, band structure-like perspective ot a more atomistic viewpoint. As has been stated many times, "A picture is worth at least a thousand words."

The pioneering studies by Zurich group were followed by further investigations with STM from other laboratories, many of which were carried out on silicon surfaces under ultra-high vacuum (UHV) conditions. The first images of the fundamental defect at a surface, the atomic step, on Si(111)-7×7 surface were shown by Becker *et al.*[8] The strong correlation between the position of the step riser with respect to the 7×7 unit mesh evident in the tunneling images suggested that a close connection exisisted between these

features. These workers further demonstrated that the partial stacking fault inherent in the 7×7 DAS structure may be directly imaged by measuring the spatial variation in the differential conductivity (dI/dV) at a bias condition near an unoccupied stacking fault state,[9] demonstrating that at least some of electronic features of this reconstruction were spatially distinct. The 7×7 is not the only stable phase that may be found on Si(111). A reconstruction of 2×1 symmetry exists on the cleaved surface; the 7×7, by contrast, is found on surfaces prepared by *in-situ* annealing to temperatures in excess of 875°C. The first tunneling images of Si(111)-2×1 reconstruction were shown by Feenstra *et al.*,[10] where the surface features demonstrated consistency with the π-bonded chain model proposed by Pandey.[29] In this study, the initial STM measurement of a tunnel junction I-V characteristic disclosed a surface state energy gap of $\sim$ 0.5eV. Tromp *et al.*[12] then showed high-resolution tunneling images of the Si(100)-2×1 surface, confirming that the dimerization of neighboring surface atoms clearly accounted for the principal features of this crystal face. This work is fueled, rather than quenched, the controversy over the detailed nature of Si-Si dimers on this surface, for the tunneling images showed both dimers that were symmetric and dimers that were asymmetric. Further exploration of spatial configuration of electron states on the Si(111)-7×7 was carried out by Hamers et al.,[13] using a variation on interrupted feedback method in Ref.[10], where selected parts of junction I-V characteristic were displayed in surface plan view. These current images clearly suggested that some of the surface state were associated with adatom features, while others were more characteristic of the partial stacking fault, in accordance with earlier work.[9] Plainly, the interest in characterizing all aspects of silicon surface features had driven the STM from an instrument with which only interesting pictures could be taken to a powerful tool for the determination of both surface geometry and surface electronic features.

Since this promising start, a large range of surface physics experiments have been performed on silicon surfaces. Metals such as Ag, Au, Al, Ga, B, K, Li, As, Sb, Sn, Pd, Cu, Ni and In have been deposited and their surface properties examined. Gases such as H, O, Cl, NH_3 and NO have been reacted with clean silicon surfaces under a variety of conditions. Transitions involving both surface phases and step phases have been explored under a range of temperatures, and the effects of varying the surface stress has been examined on Si(100). Experiments examining the atomic details of homo- and hetero-epitaxy and surface diffusion on silicon are in the literature, as well as measurements on the relative formation energies of fundamental surface defects such as steps and kinks. The remainder of this section devotes itself to these briefly mentioned results.

1.1.2 Clean surfaces

A fundamental understanding of the nature of the clean surfaces of silicon is necessary before examining more complex issues such as adsobate interactions, diffusion and epitaxy. Each of the low-index faces of silicon will be examined in turn, starting with the Si(100).

Si(100).

The Si(100) surface displays a reconstruction that is a relatively simple modification of the bulk terminated structure, in which each surface atom has two dangling bonds and is bonded to two subsurface atoms. A small displacement, involving no bond breaking, allows surface atoms to pair-up to form dimers, thereby attaining a stabilized structure in which the dangling bonds are reduced from two per atom to one. In their pioneering LEED studies, Schlier and Fransworth detected half-integral beams which they understood could not arise from surface atoms in bulk-like configuration.[6] They proposed that the observed 2×1 surface mesh was consistent with a structure created when adjacent rows of surface atoms moved together in a bonding interaction. A diagram of the Si(100) surface is shown in Fig.1.3.

Surface atom pairs, or *dimers*, were not widely accepted for many years because studies subsequent to those of Schlier and Fransworth revealed higher order diffraction spots, of an intensity and sharpness that was highly dependent on sample treatment.[14, 15, 16] It became apparent that a simple symmetric dimer structure could not account for all observations. Various vacancy[15, 17] and conjugated chain models[18] were proposed. When electronic structure calculations by Appelbaum *et al.* on the dimer and vacancy models,[19] and by Kerter *et al.* on the conjugated chain model,[20] became available for comparison with the photoemission data of Rowe,[21] the surface dimer model was found to give the best agreement. The chain model was supported by early LEED analyses that only considered displacements in one or two surface layers.[22, 23] However, after Appelbaum and Hamann showed that substantial atomic displacements were expected in the first five layers at least,[24] more elaborate LEED analyses were performed and the dimer model was then found to provide the best fit.[25]

Two decades after the first LEED study, dimers were generally recognized to be the principal feature of the reconstructed Si(001) surface, yet some dissatisfaction with the dimer model lingered because of two key inconsistencies. These were: (1) the dimer model could only account for integral and half integral diffraction, yet 1/4 order beams were sometimes observed, and (2) the half-filled band derived from the remaining dangling bonds im-

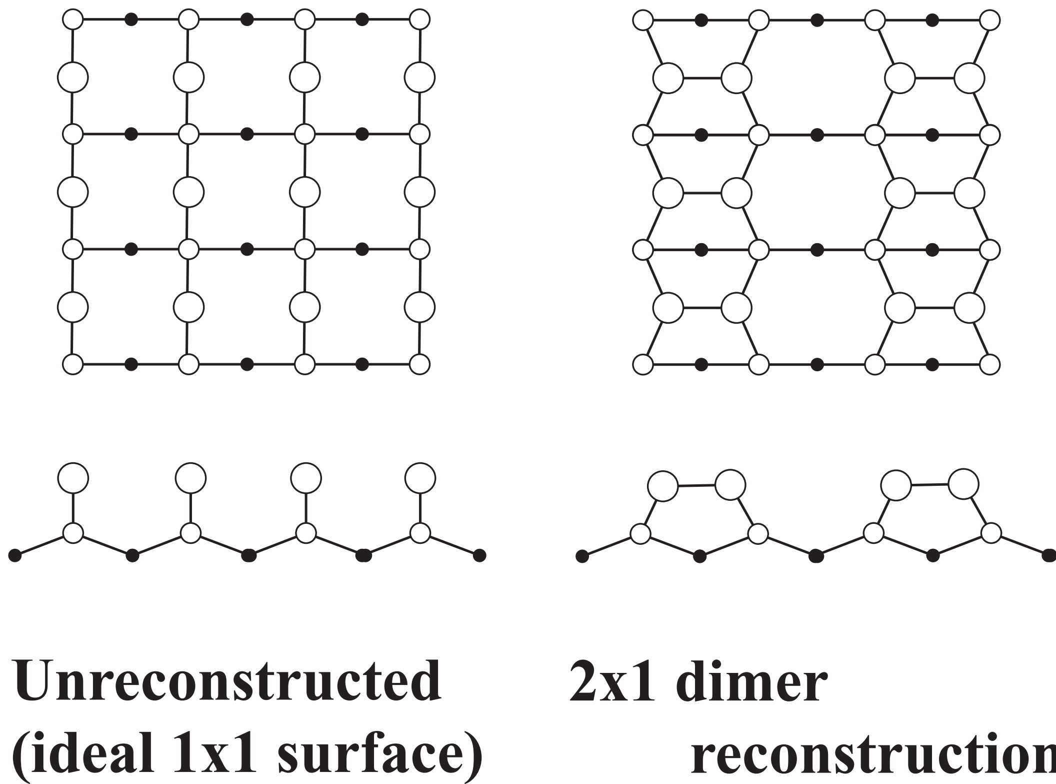


Figure 1.3: Schematic diagram the Si(100) reconstructed surface.

plies a metallic surface, but, angle-resolved photoemission measurements by Himpsel and Eastman indicated that the surface was not metallic.[26]

Because of these difficulties the idea of asymmetric dimers was pursued. It was recognized that symmetric dimers should be subject to a Peierls distortion which would result in a stabilized asymmetric structure. The buckling of dimers could make the degenerate dangling bonds of a symmetric dimer to split, causing a gap to open in the band associated with those states, in correspondence with the experimental observation that the surface is non-metallic. In addition, if arranged in a 4×2 mesh, buckled dimers could account for the observation of 1/4 order diffraction beams. Chadi explored these possibilities with an empirical tight-biding total energy calculation of the Si(001) reconstruction.[27] He found that the symmetric dimer was indeed unstable with respect to the buckled structure. Further, it was found that the geometric buckling is coupled to an electronic asymmetry, which results in a relatively full dangling bond on the up atom, and depleted occupation of the dangling bond on the low atom. The calculated band structure showed a gap of 0.6eV. These results, though very compelling, did have a weak point: the intra-atomic Coulomb repulsion (U) was not explicitly accouted for but only estimated, and since larger value of U could lead to a preference for symmetric dimers, debate continued. Yin and Cohen performed more through self-consistent pseudo-potential calculations and also found in favor of the buckled dimer,[28] but Pandey , also using the pseudo-potential method, subsequently pointed out that, because of sensitivity to computational parameters, the uncertainly in the method did not allow either the symmetric or the asymmetric structure to be identified with certainty as more stable.[29] Chadi's results were followed by new experimental efforts. Medium and low energy ion scattering measurements were suggestive of buckled dimers.[30, 31, 32] Complex dynamical LEED analyses were performed bit did not result in compelling evidence of the buckled dimer model.[33, 34, 35, 36]

Room temperature imaging of Si(001)-2×1. The emergence of the STM led to a turning point in the long effort to understand the Si(001) surface. Tromp, Hamers and Demuth[12, 37, 38] reported occupied-state images that revealed the dimers as oblong protrusions with the expected periodicity, and with corrugations along and between dimer rows of approximately 0.5Å and 0.1Å respectively, as shown in Fig.1.4. These images, while leaving a number of issues unresolved, clearly established a number of important points. Foremost among these was verification of the dimer model, other models such as the conjugated chain and missing row simply did not

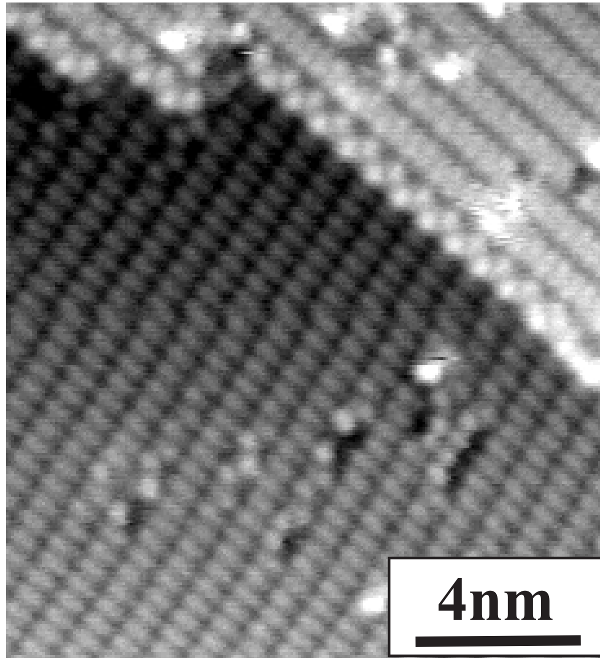


Figure 1.4: STM image of Si(100) showing dimers, both symmetric and buckled, and numerous defects.

match the observed topographic features. Interestingly, both symmetric and buckled dimers were observed, as were very numerous defects. The buckled dimers displayed local 2×2 and 4×2 symmetry. Further, it appeared that the buckling of dimers was associated with defects. These observations led to a plausible explanation for diffraction measurements that showed predominant 2×1 symmetry, and depending on sample preparation, variable degrees of 4×2 order. Higher-order diffraction should be associated with increased contamination.

Occupied state images of Si(100) fortuitously correspond to one's imagined view of the reconstructed surface. Unoccupied state images, however, are distinctly different than those of the filled states, and are more difficult to interpret because the topographic features do not simply confirm to the geometric structure of the underlying nuclei. Maxima are not observed over dimers, but between them. This can be seen when filled and empty state images of a step edge are juxtaposed as in Fig.1.5. Focusing on the upper terrace at the top side of the image, the occupied state image shows rows, of a regular width, up to the edge of the terrace. In the unoccupied state image, by contrast, the last row is clearly more narrow than the others. If

the unoccupied maxima are interpreted as dimers, we are forced to conclude that atoms in the last row on the terrace are unpaired. It is clear that this is not the case, for energetic reasons, and also since the resulting two dangling bonds per atom would be very distinct in both occupied and unoccupied state images. It can therefore be concluded that the unoccupied state electronic structure leads to maxima between the dimers, while occupied state images show maxima which confirm to the underlying dimers. In Fig.1.6, schematic STM contours for both occupied and unoccupied states are shown. The measured corrugation is typically 0.1Å and 0.2Å. Occasionally, with an uncommonly sharp tip[39, 40] a more shallow topographic minimum (shown as a dotted line) is observed between dimers.

STM images of the Si(001) surface led to universal acceptance of the dimer model while also explaining the confusing higher-order diffraction which had often been observed. The observation of both symmetric and buckled dimers, however, fueled rather than settled the controversy involving the dimer configuration. While it was recognized that the symmetric appearance of dimers might result from averaging by the STM as the dimer rapidly flipped between buckled configurations, this could not be confirmed.

Other investigations directed at the symmetric versus asymmetric dimer question were carried out. In one recent calculation spin effects were incorporated and it was concluded that symmetric dimers are most stable.[41] High-resolution photoemission data, on the other hand, indicated two in equivalent types of surface silicon atoms, which were interpreted to be the up and down atoms of buckled dimers.[42] Thorough pseudo-potential calculations by Pandey,[29] Payne *et al.,*[43] and Roberts and Needs[44] show that if there is an energy lowering associated with buckled dimers, it is subtle effect, perhaps 10 to 60 meV, and is too small to predict with certainly.

Dimers and STM imaging. The essential features of the dimer electronic structure are depicted in the simple energy level diagram shown in Fig.1.7. At the left extreme of the diagram are four degenerate levels representing the dangling bonds of two bulk-terminated surface atoms. Proceeding to the right, dimer bond formation is represented by the σ bonding level and the σ^* antibonding level. The remaining dangling bonds, one on each atom, remain noninteracting. This simple description of the electronic structure is consistent with the ball-and-stick model that it describes gross feature correctly. Two of the four available electrons fill the dimer bonding level and the remaining two electrons half-fill the dangling bonds. Since the STM is most sensitive to electron states nearest the Fermi level, dimer images are expected to be derived primarily from the dangling bonds. STM images

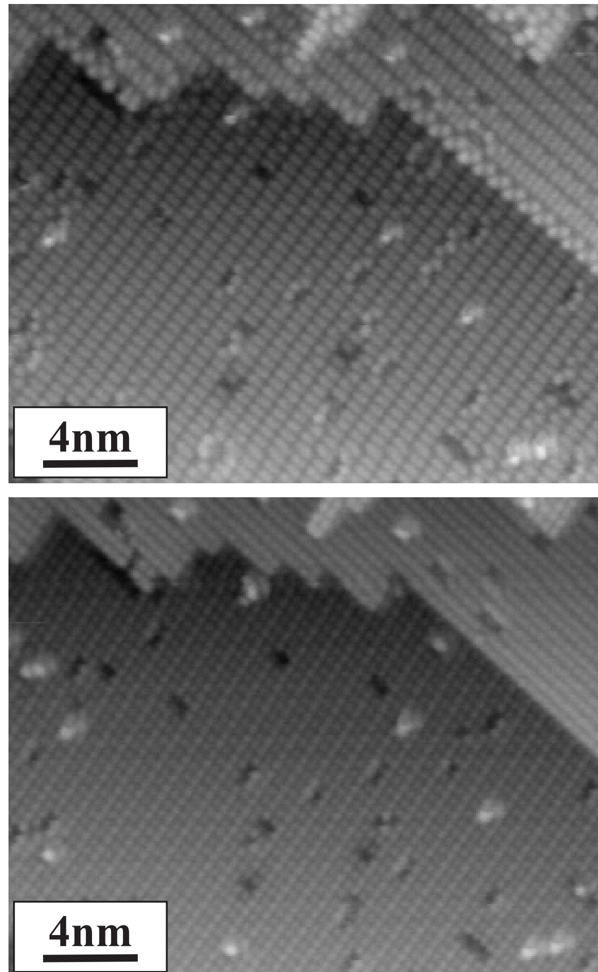


Figure 1.5: STM image of a step on the Si(100) surface. At the top is an occupied state image, and at the bottom is an unoccupied state image of the same area.

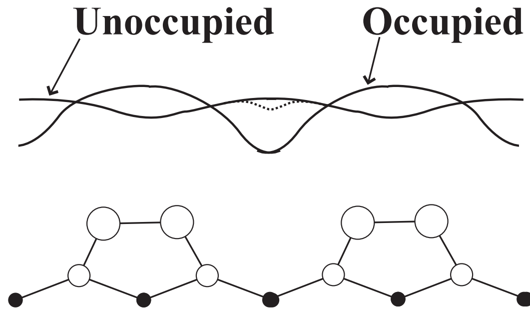


Figure 1.6: Schematic STM occupied and unoccupied state contours and their relation to the underlying dimers. The dotted position of the unoccupied state contour shows a secondary minimum seen occasionally with extraordinary tips.

of the Si(100) surface tend, however, to be described as due to the dimer bond,[38] or the dimer π bond.[39, 40] The term "dimer bond" is unfortunate as it suggests, quite incorrectly, that electrons are derived from the dimer σ bond. The σ bond is largely inaccessible to the STM tip both energetically and spatially; the electron density does not protrude into the vacuum but is concentrated between the atoms of the dimer pair. The use of dimer π-bond is a better choice as it refers to a state derived from the π-bonding interaction of the dangling bonds. In the original theoretical description of the (symmetric) dimer reconstruction Appelbaum, Baraff and Hamann[19] showed that the dangling bonds do interact to form a π-bonding, and a higher energy π-antibonding level, but that the interaction is very weak. The splitting due to the $\pi - \pi^*$ interaction is represented in Fig.1.7(c).

Since the true dimer configuration is asymmetric, not symmetric as assumed in the calculation by Appelbaum *et al.*, and since we associate distinctly different electronic structure with the buckled dimer [Fig.1.7(d)], we must re-examine the assignment of STM topographic features to the dimer π-bond. Theoretical calculations which contrast the electronic structure of the symmetric and asymmetric dimers show that the $\pi - \pi^*$ interaction induces a small splitting compared to that created by buckling.[45, 46, 47] Figure1.8 shows the spatial character of the π-bonding interaction associated with the symmetric dimer, and for comparison, the distinctly different buckled dimer. Buckling of the dimer leads to states which are like dangling bonds in spatial character, but which are energetically split such that the state at the dimer up atom becomes relatively full at the expense of the state localized at the down atom. Dimers which are forced by a nearby defect to buckle,

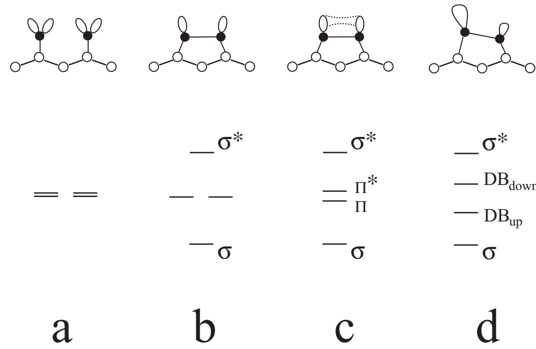


Figure 1.7: Energy level diagram representing (a) four degenerate dangling bonds (DB) from two bulk-like surface atoms, (b) dimer bond formation depicted by a σ bonding level and a σ^* antibonding level, two dangling bonds remain with one on each atom, (c) splitting of the DB levels as a result of $\pi - \pi^*$ interaction, (d) buckled dimer with sharply split DB states associated with the up and down dimer atoms.

or which naturally settle into a buckled configuration at low temperature, appear in occupied state images to be strongly tilted (the up atom is approximately 0.5Å higher than the down atom). This observation matches very well the picture of an asymmetric dimer with modified dangling bonds as described earlier. At room temperature the barrier to switching direction of tilt is readily overcome and the STM captures an average picture of the dimers fluctuating between buckled extremes, and an elongated, symmetric protrusion, centered over the dimer results. The symmetric-appearing dimer has routinely been described as due to the dimer π-bond. This is at least partially correct since the dimer will pass through a symmetric intermediate state while switching between extremes. The dangling bond state associated with the up and down ends of the buckled dimer also contribute. The weighting of these various contributions is unclear at this time.

Similar to the occupied state images, the states associated with the up atom and the down atom of the buckled dimer must figure into the room temperature unoccupied state image, however, the π^* description seems compelling. This state has a node at the center of the dimer bond, matching the observed minimum in STM images, and maximum density at the dimer ends where the topographic maximum is observed. It may be, however, that the unoccupied state image has contributions from the dimer antibonding level (which has maximum state density at the ends of the dimer) in addition to the dangling bonds. The calculations of Kruger *et al.*[46, 47] suggest that

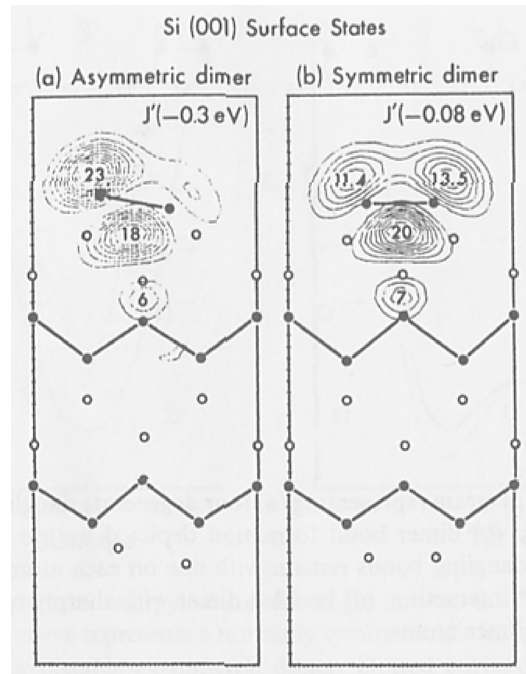


Figure 1.8: Charge density plots contrasting the character of occupied states near the Fermi level for the (a) asymmetric and the (b) symmetric dimers. From Ref.[45]

the dimer σ antibonding level is very near the bottom of the conduction band and is therefore enegetically accessible. While it does not have the spatial characteristics of a dangling bond, the dimer antibonding state does not decay with distance from the surface as rapidly as does a back bond.[47]

In addition to producing maxima which are not centered over the dimer units, unoccupied state images show another interesting feature. In contrast to occupied state images, unoccupied state images do not show buckling, even in dimers adjacent to buckle inducing defects. It has been suggested that the field imposed on a dimer during scanning with the STM could tend to flatten backled dimers.[48, 49] Kochanski *et al.* argued that when imaging unoccupied states, the electric field set up by the negative tip may repel the atom, causing the dimer to appear symmetric, while a positive tip would enhance backling. A dimer held flat would not show contributions characteristic of the strongly split buckled dimer. If this mechanism were acting, it might be expected that different degrees of asymmetry could be induced as a function of applied field, while no such dependence has as yet been observed. A simpler explanation for the absence of buckling in unoccupied state images arises from the symmetry in the electronic structure if the buckled dimer. In contrast to the occupied state situation where the up atom has the most pronounced state density, tending to amplify the buckling effect, when imaging unoccupied states the down atom has the greatest state density and tends to cancel the effect of the actual displacement of the nuclear centers. Figure1.9 contrasts charge density plots above and below the Fermi level.

Stepped surfaces. Steps on Si(001) display complex behaviour known to depend on a number of factors including the angle of miscut, annealing and growth conditions, contaminants and surface stress. Recent atomically resolved STM images, together with theoretical modeling, have led to considerable improvement in our understanding of the fundamental structures and mechanisms underlying these factors. In this paragraph we begin with a discussion of the basic step structure; we next consider kinks and conclude with a discussion of step interactions as effected by tilt angle, strain and annealing.

Step structure. The single step height in the (001) direction for the diamond lattice is $a/4 = 1.36$Å, where $a = 5.431$Å is the bulk silicon lattice constant. On traversing a single atomic step the dimer bond orientation rotates by 90°. Surfaces miscut from the (001) direction, known as "vicinal" surfaces, display a mean distance between single steps given by $a/4 \tan \alpha$ where α is the miscut angle. Surfaces cut toward the $[1\bar{1}0]$ or the [110]

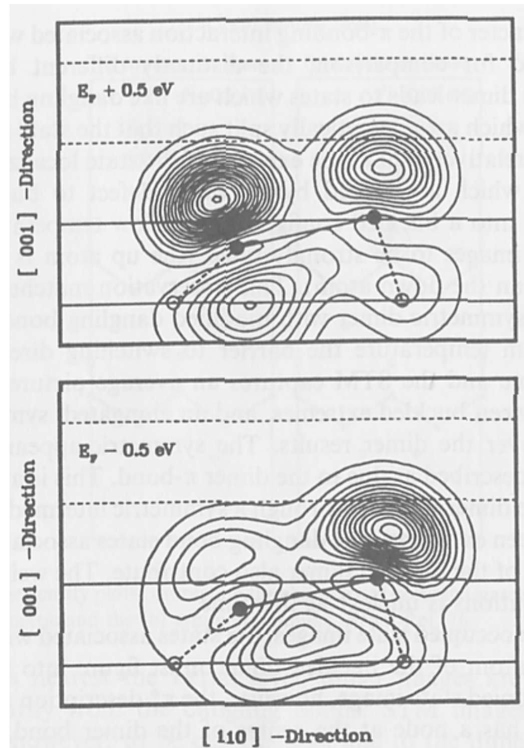


Figure 1.9: Contours of constant charge density above and below the Fermi level for the asymmetric dimer. (From Ref.[47])

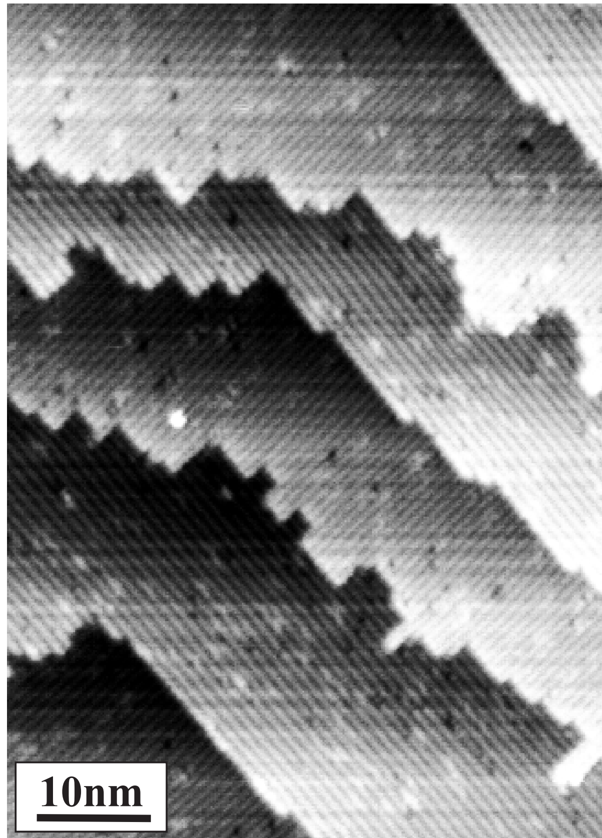


Figure 1.10: Tunneling image of a stepped Si(100) surface showing the type A (smooth) and type B (rough) step risers.

direction display two distinct types of single step, as seen in Fig.1.10. When dimer rows on the upper terrace run parallel to the step edge, the step is referred to as type A. The other possibility, referred to as a type B step, has rows on the upper terrace perpendicular to the step edge. When the surface is tilted toward a direction intermediate between $[1\bar{1}0]$ and [110], steps of mixed type A and type B character result. Type A steps are found to be very smooth, while type B steps show a marked tendency to form a kinks. A typical single stepped vicinal surface is shown in Fig.1.10. Surfaces with a miscut angle larger than a few degrees form type B double steps. Since the dimer bond orientation is the same on terraces separated by a double step, surfaces prepared to have exclusively double steps are referred to as single domain or primitive.

Chadi has proposed models for single and double steps of type A and B.[50] These structures, which are found to correspond well with STM im-

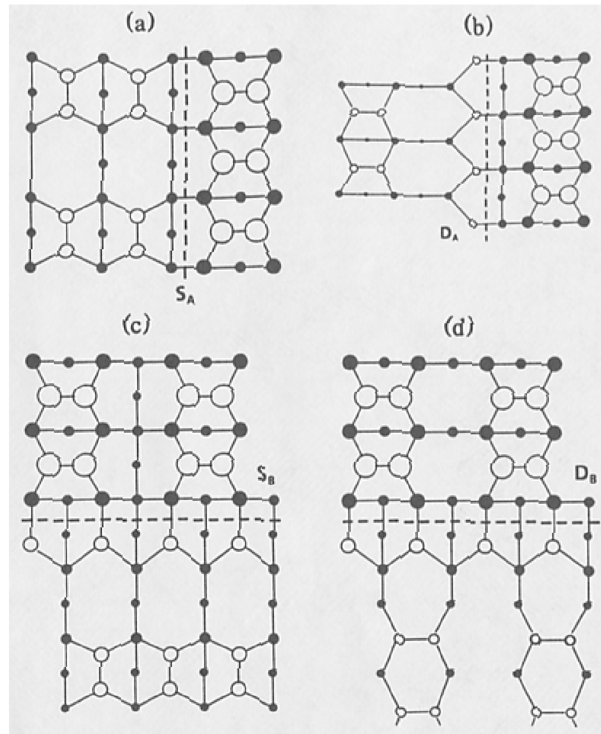


Figure 1.11: Top view of single and double steps of type A, labelled S_A and D_A, single and double steps of type B, labelled S_B and D_B. Dimer bonds are aligned in $[1\bar{1}0]$ direction. Dashed lines that run parallel to $[1\bar{1}0]$ or $[110]$ axes indicate the step positions. Open circles denote atoms with dangling bonds. Edge atoms (shaded circles) (b)-(d) are rebonded, i.e., they form dimerlike bonds with lower terrace atoms. Larger circles are used for upper terrace atoms. Only some sublayer atoms are shown for the sake of clarity. From Ref.[50].

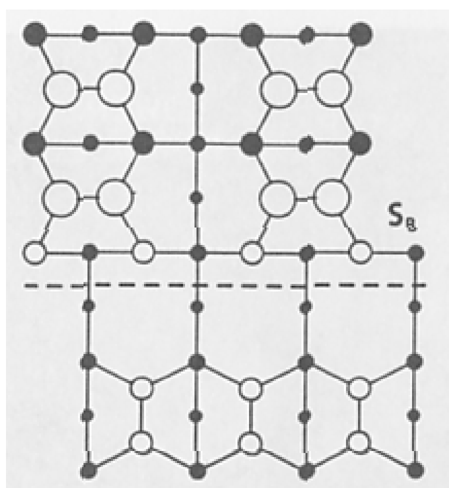


Figure 1.12: Top view of a simple “nonbonded” edge geometry for a type B single step. Open circles denote atoms with dangling bonds. This structure is characterized by having a dangling bond on each second-layer edge atom and is not as energetically favorable as the rebonded atomic configuration of Fig.1.11(c)

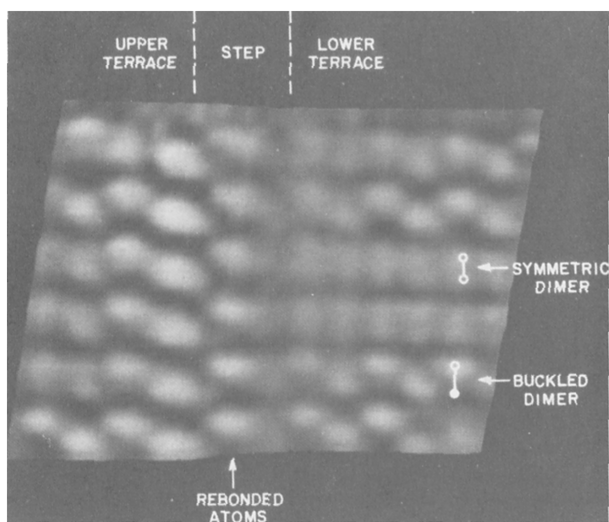


Figure 1.13: Occupied state STM image of a type B double step. The region covered is 46×50Å. Six dimer rows run laterally across the field of view while the step runs from bottom to top. Curvature keyed shading is used. The sample bias was -1.6V. From Ref.[52]

ages, are shown in Fig.1.11. Formation energies of type A and B single steps were calculated to be 0.01eV and 0.15eV respectively, while type A and B double step formation energies were calculated to be 0.54eV and 0.05eV respectively. Chadi found that type B step structures with "rebonded" atoms, as shown in Fig.1.11, were most stable. A type B single step without rebonded atoms was also considered and found to have a formation energy of 0.16eV. The corresponding structure is shown in Fig.1.12. Both kinds of type B single steps have been observed.[37] The double type A step which Chadi predicts to have a relatively high formation energy is not observed. Figure 1.13 shows an image of a type B double steps. An earlier model of the double step due to Aspens and Ihm[51] involed a π-bonded chain structure like that seen on Si(111)-(2×1) surface.[53] STM images do not support this model, however, since a periodic structure is observed at the step edge with twice the spacing expected for the π-bonded chain. While the spacing is also twice that expected for rebonded atoms, it is felt that Chadi's model correctly describes the double step. It appears that the rebonded atoms buckle, like the dimers on the (001) terrace, and only every other rebonded atom is visible.[52] The buckling of the rebonded dimers is caused by the asymmetric strain imparted from the asymmetric dimers on the upper terrace. Careful examination of both occupied and unoccupied state images reveals that all rebonded atoms predicted by Chadi are present.

Vicinality versus single double steps. For a number years it had been thought that vicinal Si(001) surfaces had only one equilibrium structure in which only double layer steps are present.[54] Calculated step energies substantiated this view by showing the double layer type **B** step to be energetically preferred over single layer steps.[50, 51] It has been observed, however, that small vicinal angles, on the order of 1° or less, lead to stable single stepped surfaces. Recently Alerhand *et al.* have provided a compelling explanation for these observations. They demonstrated that, at vicinal angles exceeding approximately 2°, the equilibrium surface is double stepped, but for small misorientations the single step is at equilibrium.[55] Further, Alerhand *et al.* established that, in addition to step energies, two other factors play a role in determining the relative stability of single and double steps. The first is the long range effect of the anisotropic stress associated with a dimerized terrace, and the second is the effect of thermal fluctuations (i.e., kinks) in step edge. At zero temperature the latter does not come into play. At angles of a few degrees or larger, the dominant factor is that double steps are energetically favored over single steps by approximately 0.1eV/atom. Since the dimerized surface has a compressive stress perpendicular to the dimer

bond and a tensile stress parallel to the dimer, crystal strain can be relieved by forming a single stepped surface since the dimers on adjacent terraces are then perpendicular. The strain per unit terrace width becomes more significant at small miscut angles and eventually dominates at a calculated angle of $0.05°$ at 0K. The entropic contribution of thermally induced kinks in step edges are found to significantly shift this transition angle to larger angles at higher temperatures well above room temperature but below typical anneal temperatures. The temperature at which mass transport stops during cooling of a crystal is not clear, if we take a reasonable value of 500K[55] we can read the experimentally relevant transition angle from the calculated phase diagram to be approximately $2°$. This value is in very good agreement with experimentally determined upper and lower bounds of approximately $3.5°$ and $1°$ respectively.[55]

Strain effects on Si(001) steps. The LEED investigations of Saloner *et al.*, established that steps on vicinal surfaces were evenly spaced, clearly suggesting a step-step repulsive interaction.[56] Menn *et al.* investigated the effect of applying a uni-axial stress to the Si(100) surface and provided striking evidence that the repulsion between steps was related to strain in the crystal.[57] A surface, which previous to the application of stress, showed an even distribution of 1×2 and 2×1 terraces, could, by bending the crystal, be transformed to a nearly primitive surface. Alerhand *et al.* successfully accounted for this behavior, and in so doing, showed that the bulk elastic strain is the primary cause of interaction between steps.[55] It is demonstrated that the dimerized surface has a compressive stress perpendicular to the dimer bond and a tensile stress along the dimer direction, and as a result, domains for which an applied compressive stress is directed along the dimer bond, grow at the expense of the other domain. Figure1.14 compares the experimental data of Menn *et al.* with the calculations of Alerhand *et al.* The experimental data clearly shows the coordinated shift in 1×2 and 2×1 LEED intensities as a function of applied strain. The light lines result from a model which assumes that the mean number of steps changes as a function of stress. The heavy lines, which fit the data very well, result when the number of steps is conserved.

Kinks and step-kink interactions. Step and kink energies have been the subject of many investigation.[59] Recently Swartzentruber *et al.* have analyzed equilibrium distributions of kinks and steps to determine kink step separations and lengths, and from these extract measures of step and kink energies.[60] They recorded STM images of thoroughly annealed Si(001) sur-

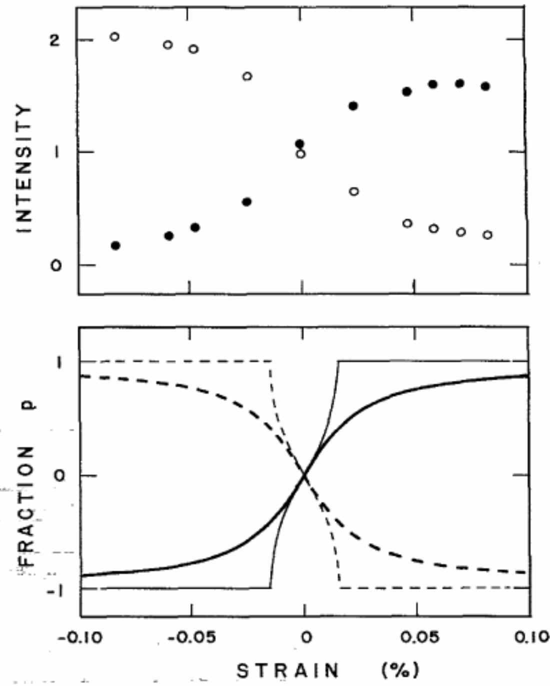


Figure 1.14: Top: Experimental intensity of (1×2) (open circles) and (2×1) domains on the Si(100) surface as a function of applied external strain. Bottom: Fit by theory, thick (thin) kines correspond to quasi (global) equilibrium. From Ref.[58].

faces miscut 0.3° toward [110]. Kink lengths and the separations between kinks were tabulated for type A and type B steps, and the probability of finding kinks separated by a given distance was plotted as a function of separation between kinks. The data were fit very well by a function that assumed the creation of each kink was a statistically independent event. This is somewhat surprising, since the long-range strain that leads to step-step repulsion[58] must be manifest in kink formation probabilities, otherwise steps would not tend to lie at the midpoint between adjacent steps. It is possible that the large separation between steps on the surfaces studied masked this effect by producing too small a gradient in the local strain field. In any case the assumption of independent kinks allows the number of kinks as a function of length to be written $N(n) \propto e^{-E(n)/kT}$, where $E(n)$ is the energy of a kink of length n. Plots of $E(n) = -\ln(N(n)/2N(0))$ versus n were fit according to $E(n) = n\epsilon_S + C$, where ϵ_S is the unit single step energy and C is described as an additional energy due to the corner structure at kinks. Values of $\epsilon_{S_A} = 0.028 \pm 0.002$eV/atom, $\epsilon_{S_B} = 0.09 \pm 0.01$eV/atom and $C = 0.08 \pm 0.02$eV were found.

A more detailed discussion, particularly of double steps and earlier developments which provided the background to work on steps described here, may be found in the review by Griffith and Kochanski.[59]

1.2 Surface and interface stress

1.2.1 Introduction to surface and interface stress

The concepts of a free energy of a surface or interface stress and of a surface stress were developed by Gibbs[61] as a part of his treatment of thermodynamics of surface and interface. Qualitatively speaking, the surface free energy is defined as reversible work per unit area to create surface. The surface stress is a reversible work per area to stretch a surface elastically. It is apparent from these definitions that the two quantities are of a different nature. The specific free energy of a surface, e.g., must be positive, since otherwise the solid would gain energy upon fragmentation and, therefore, would not be stable. The surface stress of the solid need not be positive (tensile), and we shall encounter many examples where it is not. For a liquid surface, the specific free energy and the surface stress are equal. As shown later, this equality is due to the fact that for a liquid surface the specific free energy does not change when surface is strained. It is presumably because of the equality of surface stress and specific energy for liquid surfaces, that the two quantities, while being in principle different, are frequently refered

to by the common name "surface tension". This terminology continues to be a source of confusion, particularly, in those scientific communities where the language does not provide two different words bearing a similar connotation, "tension" and "stress". For solid surfaces, surface stress and the surface free energy become different. The difference in their numerical values can be as large as a factor of 3. Because of this difference and the ambiguity of the term "surface tension", in this thesis we avoids the use of the latter term altogether.

Definition of surface stress

The surface stress can be defined in various ways. While the definitions are equivalent, they are of greater or lesser use in a particular framework of an experimental or theoretical investigation. Depending on the issue considered, one may prefer one or another definition. We begin with a definition which involves the change in the bulk stress tensor near a surface or an interface. We denote the coordinate perpendicular to the surface/interface as z and the coordinates in the surface as x and y. The tensor components which refer to the x and y axes are denoted as i and j, respectively.(Fig.1.15) The surface stress is then defined as the integral

$$\tau_{ij}^{(s)} = \int_{-\infty}^{\infty} (\tau_{ij}(z) - \tau_{ij}^{(b)}) dz \tag{1.1}$$

in which $\tau_{ij}^{(s)}$ denotes the component of the surface stress tensor and $\tau_{ij}(z)$ are the components of the bulk stress tensor as a function of z (which can be different from the bulk stress $\tau_{ij}^{(b)}$ in the vicinity of the surface). The dimension of the surface stress is force per unit length, not force per area as in the case of bulk stress. The sign if the surface stress is positive if the surface would like to contract under its own stress. This sign convention is consistent with the usual sign convention for bulk stress. The surface stress is called "tensile" when the surface stress is positive and "compressive" when the stress is negative. Eq. (1.1) denotes either a surface stress or an interface stress for an interface between two solids, two liquids or a liquid and a solid, provided that the bulk stresses are the same for either phase.

The definition above is particularly useful for a consideration of experimental means to determine the surface stress. Because of the (positive) surface stress, small crystallites are compressed in their volume and the compression can be calculated using the above definition and elastic theory. The measurement of the mean lattice constant of small particles as a function of their size, therefore, can be used to determine the surface stress.[62] A change in the surface stress on either one of the two surfaces of a thin crystal

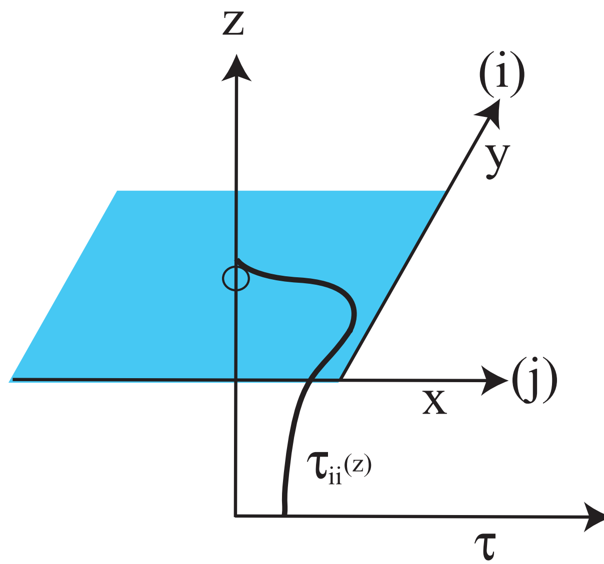


Figure 1.15: Illustration of the variation of the bulk stress $\tau_{ij}(z)$ near the surface (solid flat line) which defines the surface stress according to Eq.1.1. The indexes i and j denote the components of the stress tensor in the x and y direction, respectively.

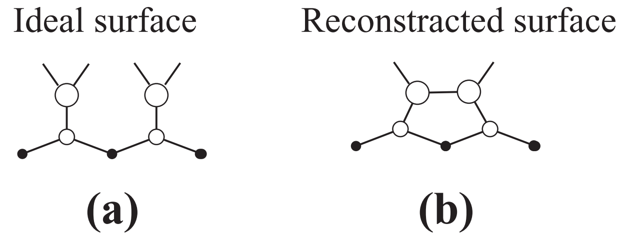


Figure 1.16: Surface atoms with dangling bonds on a Si(100) or Ge(100) surface (a). The surface reconstructs to form Si or Ge dimers at the surface (b). The bending involved in the dimer formation causes a tensile stress in the dimer direction.

plate leads to a bending of the plate. The bending can be calculated from the elastic bulk properties of the plate using the definition of Eq. (1.1). A measurement of the bending allows, therefore, the determination of the change of a surface stress, e.g., upon the adsorption or deposition of atoms.

For cases of interests in this work, the symmetry of the surface is such that the non-diagonal terms of stress tensor vanish when the coordinate system is that of the crystallographic axis. For the frequently considered cases of the stress on (100) and (111) surfaces of cubic metals, the surface stress is even isotropic. The fourfold symmetry of the (100) surface and the threefold symmetry of the(111) surface of cubic substrates may be broken however, when the surface is reconstructed, with the consequence that surface stress becomes locally anisotropic. The Si(100) surface with the pairing row reconstruction is mainly discussed in this paper. On this surface, the pairing of dangling bonds causes a locally anisotropic surface stress in either the [010] or the [001] direction. (Fig.1.16, and also see Fig.1.17) The overall symmetry of the crystal requires the existence of two domains of the pairing row reconstruction. We, therefore, encounter a difficulty in the definition of a surface stress. If the bulk stress in Eq. (1.1) is assumed to be genuine macroscopic stress, i.e., averaged over all domains of the reconstruction, the surface stress would be isotropic. As one aims for understanding of the macroscopic properties from the microscopic structure, such an averaged surface stress is not very useful. Whenever possible therefore, one defines a quasi-local surface stress which averages over the atoms in a single domain of reconstruction. Such a quasi-local surface stress is a meaningful operational concept if the linear extension of a domain involves more than 10-50 atoms, which is frequently

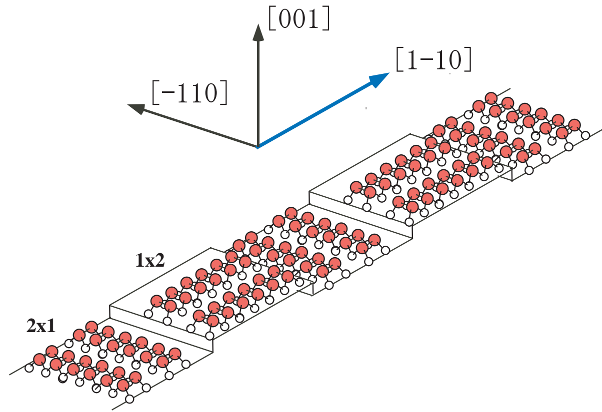


Figure 1.17: Stress domains on Si(100). The pairing row reconstruction with the formation of Si dimers introduces an anisotropy in the surface stress. The orientation of the dimer bonds and, therefore, the orientation of the stress tensor depends on the termination of the crystal. Stress domains of mutual orthogonal orientation are separated by monatomic steps of the "S_B"-type.

the case.

1.2.2 The charge distribution and surface stress on clean surfaces

When a surface is created, the electrons respond to the absence of atoms above the surface so that the charge distribution near the surface becomes different from that in the bulk. The surface stress which is the integral over the difference between the stress near the surface and the bulk stress (Eq. (1.1)) is a consequence of this redistribution of charge. The surface stress would be zero, if the surface charge density would remain the same as in the bulk. Quantitatively, the surface stress can be obtained from the Hellmann-Feynman forces[63, 64] by considering the following "Gedanken experiment": Assume a plane which intersects the surface at a right angle and remove all atoms and the electronic charge density on the one side of the plane without allowing the electronic charge density to relax in response to the missing atoms.(Fig.1.18) The surface stress is then the sum of the forces per unit length of intersection which keep the remaining atoms in place, minus these forces in the bulk. While this definition may sound artificial at first glance, it has the advantage that it immediately relates the surface stress to the difference between the electronic charge distribution near the surface and the bulk. The consideration of Hellmann-Feynman forces is, therefore, useful

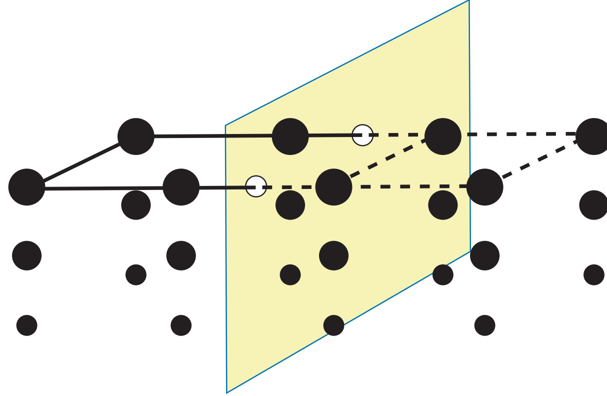


Figure 1.18: Illustration for the discussion of the surface stress in terms of the electronic charge density between the atoms near the surface. While the distribution of electronic charge is determined by quantum mechanics, the forces on the atoms are Coulomb forces arising from the charge distribution. The surface stress is the sum of all forces on the atoms per unit length minus the corresponding forces in the bulk when all atoms and the electronic charge density to the right of the shaded plane are removed.

for a qualitative discussion of the effect of adsobates and the role of the charge transfer on the surface stress. The definition also provides a prescription for calculating the surface stress as a sum of Hellmann-Feynman forces on the atoms. Technically, it is more straightforward to calculate the surface stress either using the stress theorem of Nielsen and Martin[65, 66] or from the dependence of the energy on the strain.[67]

In the following, we discuss qualitatively the consequences of the charge redistribution for the surface stress. We first consider materials where the electronic charge is accumulated between the atoms in the form of a bond charge but the bonding does not have the directionality of covalent bonding. This scenario is typical for transition metals and noble metals. The result of these considerations is that the surface stress should be positive (tensile) for a clean surface. A tensile stress is also typical for free electron metals.[68, 69] Because of the missing bonds, the bond charge moves from above the surface into the selvage of the solid.(Fig.1.19) Insofar as the charge is relocated to the area between the first and second layer atoms the backbonds of the surface atoms are strengthened and would like to become thereby shorter. Hence, such a redistribution of the electronic charge would lead to a contraction of the distance between the first and second layer. To a lesser or larger degree, such a contraction is found in many metal surfaces.[70] There are ex-

ceptions, however. These include the nearly free electron metals Al(100)[71], Al(111)[72], and Mg(0001).[73] For metals where one finds a contraction, the contraction is the larger, the more open structure of the surface. Electronic charge may also accumulate between the surface atoms. This causes a contraction of the *equilibrium* bond length of the surface atoms. Since the atoms cannot respond to the contraction of equilibrium length of the lateral bonds because their positions are required to remain in registry with crystal structure a tensile (positive) surface stress arises. A natural consequence of this qualitative picture is that the surface stress on a clean metal surface should be tensile. We do not know whether this is true for all metal surfaces since no general theorem to that effect has been proven. The first principles calculations carried out so far, however, reveal that the surface stress on clean metal surfaces is tensile.

While such considerations call for a tensile stress on clean metal surfaces regardless of the surface structure, the sign of the surface stress depends on the structure for the surfaces of semiconductors because of the effect of the bond directionality in combination with the rehybridization of orbitals near the surface. The clean, unreconstructed Si(111) surface is an instructive example in that regard.[74, 75] Because of the dangling bonds, the surface atoms rehybridize from the bulk sp^3 configuration towards a planer sp^2 configuration. Since the Si-Si equilibrium bond distance remains nearly the same for the sp^2 hybridization, the Si surface atom is drawn inwards and pushes the Si atoms in the second layer sideways.(Fig.1.20) This causes a compressive stress. The compressive stress on the clean surface is comparatively small (-0.63N/m [74]). The compressive stress becomes larger when the Si surface atom is substituted by Ga (-5.6N/m [75]) since Ga has only three valence electrons which cause an ever stronger trend toward sp^2 hybridization. Substitution by As reverses the sign of the stress to a tensile one (+3.95N/m [75]). The reason is that As has two electrons in a lone pair orbital and prefers to hybridize with bond angles less than those of perfect tetrahedron. The resulting surface stress in these cases does not only depend on the hybridization but also on the size of the atom by which the Si is substituted. If the Si surface atom is substituted by the small atom B, then the stress becomes tensile, despite the Sp^2 hybridization.[75] As discussed above (Fig.1.16), the tensile stress arises from the overlap of the half filled dangling bonds.

1.2.3 Adosobate-induced stresses - general discussion

We consider the case of metal surfaces first. The chemisorption of atoms on a surface re-establishes the bonds between the surface atoms and the

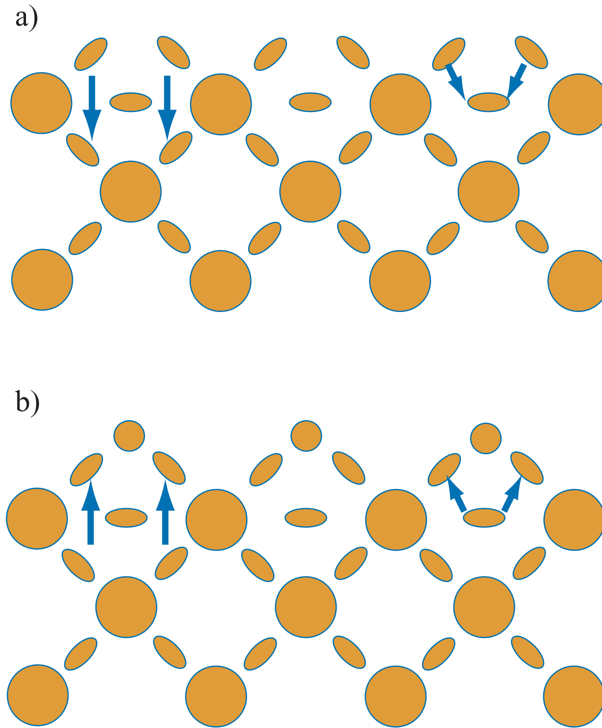


Figure 1.19: Illustration of a simple picture of the origin of the surface stress. Bond charges near the surface for (a) a clean surface and (b) a surface which is covered by an electronegativity adsobate. On a clean surface, the bond charge in the missing bonds is redistributed to strengthen the backbonds, on the one hand, and to reduce the *equilibrium* bond length between the surface atoms, on the other. The latter effect causes a tensile stress. The adsorption of electronegative atoms on the surface removes charge between the surface atoms and causes a compressive stress.

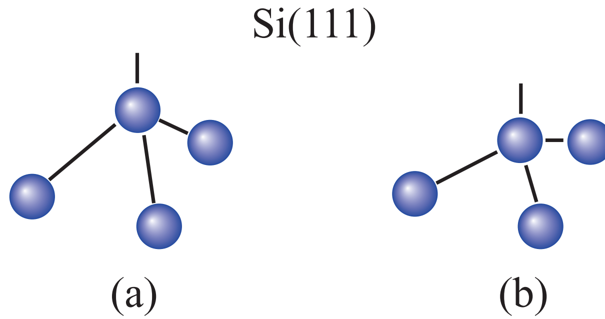


Figure 1.20: Surface atom configuration at the Si(111) surface. Because of the dangling bond of the Si(111) surface atom, the electronic configuration changes from sp^3 in the bulk towards sp^2. Compared to a perfect tetrahedron (a) the surface atom is pulled inward and the Si atoms in the second layer are pushed sideways (b). This causes a compressive stress on the clean Si(111) surface.

atomic entities above them. In the spite of our qualitative reasoning, this should cause a reduction of the tensile stress provided that bonding to the surface saturates the bonding possibilities of the adsobate atoms. One may expect particularly strong changes in surface stress with the deposition of atoms which form heteropolar bonding with the surface. If the adsobate is electronegative with respect to the substrate, charge should be removed from the bonds between the substrate surface atoms to reside on the adsobate atoms. According to our model, the removal of charge between the surface atoms should reduce the tensile stress and may even turn the surface stress into a compressive one. The effect is (though only by a small amount[76]) amplified by an additional repulsion between the charged adsobate atoms. For the electronegative adsobate atoms carbon, nitrogen and oxygen the induced stress was in fact experimentally found to be compressive and of such an amount that tensile stress of the clean surfaces turned into a compressive one.[76]

A negative (compressive) surface stress does not violate a stability criterion. If the compressive stress becomes very large, certain surface vibrational modes can become soft, creating structural transitions of the surface. Examples that negative surface stresses on adsobate covered surfaces and the particular structural instabilities which can arise from large compressive or from large tensile surface stresses are discussed in Ref.[76, 77].

In the case that the adsobate atoms are electron donors, the charge density between the surface bonds should be enhanced thereby increasing te

tensile surface stress. Because of the charge on the adsobate atoms themselves, this tensile stress is partly compensated by the repulsion between the adsobate atoms. The net effect should nevertheless be a tensile stress, since the distance between the charged adsobate atoms is larger than the distance between the enhanced bond charges in the surface and positively charged surface ion cores. An example of the trends with electropositive and electronegative adsobates [78] is shown in Fig.1.21. For carbon adsorption on the Ni(111) surface, the change in the surface stress towards the compressive side. The change is so large that the intrinsic surface stress of about 1.44N/m on the clean surface [79] turns into a compressive one. For Cs adsorption, on the other hand, the tensile stress increases dramatically. Interestingly, the increase persists well above a monolayer (ML) coverage which is at $\Theta = 0.26$. It is noted that the two experimental results concern cases of strong heteropolar bonding. Calculation of the adsobate-induced stress for H on Pt(111) by Feibelman [80] indicate that at least for an adsorption system which involves only small charge transfers the sign of the induced stress may not follow the rule presented above. From the work function change hydrogen appears to be an electron donor. The induced surface stress is nevertheless compressive. Feibelman has also pointed out an inconsistency in the above argument. On the Pt as well as on the Ni surface, the empty orbitals are antibonding. Charge donation into the surface should, therefore, result in a filling of antibonding orbitals. Feibelman argues that filling of antibonding orbitals should increase the equilibrium bond distance and should, therefore, cause a compressive stress which is in contrast with the majority of experimental observations. In summary, it is concluded that simple rule established by the (possibly questionable) reasoning above seems to be followed by most experiments carried out so far. Theory is also in a position to calculate adsobate-induced surface stresses which are in good agreement with the experimental data. The qualitative understanding of the adsobate-induced surface stresses on metal surface is still in its infancy.

The situation appears to better for semiconductors. There, the adsobate-induced surface stress can qualitatively understood in terms of the local covalent bonding. For the adsorption of oxygen, e.g., the sign and even the magnitude of the induced stress can be estimated from the bond angles and bond distance of the Si-O bond and the stiffness of these bond lengths and angles with regard to distortions.[81]

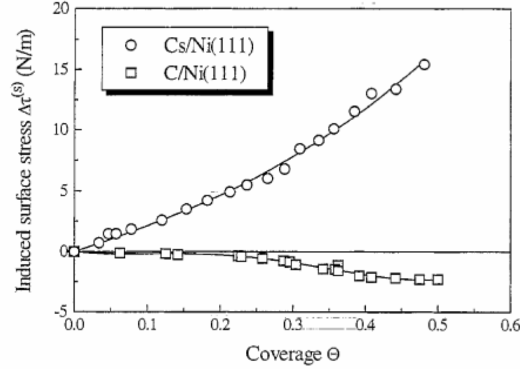


Figure 1.21: Change in the surface stress on a single crystal Ni(111) surface due to deposition of carbon (squares) and cesium (circles). The coverages are scaled relative to the number of Ni atoms. For cesium the first monolayer is completed at $\Theta = 0.26$. Beyond this coverage, the Cs atoms are deposited in the second layer. The continuing increase in the stress beyond monolayer coverage is not well understood presently. This effect is discussed in the context of heteroepitaxial growth. From Ref.[78].

Surface stress and the specific free energy of a surface

We consider the work δW involved in straining a thin crystal plate of thickness t by a strain $\delta\epsilon_{ij}$

$$\delta W = A \int_{-t/2-0}^{-t/2+0} dz \sum_{i,j} \tau_{ij}(z)\delta\epsilon_{ij}, \tag{1.2}$$

where A is the surface area of the slab and the coordinate z is perpendicular to the surface. The boundaries of the integral area taken such that the interfaces on either side where the stress might differ form the bulk are induced. The sign convention for the strains is that ϵ_{ij} is positive for an enlargement of the dimension. Hence for positive diagonal terms of the stress δW is positive and would increase the free energy of the slab by the same amount under the side condition of constant temperature and constant chemical potential of the constituents of the slab. The work on the slab and change in the free energy can be split into a bulk and surface contribution. Using the definition of the surface stress in Eq. (1.1) one obtains

$$\delta W = \delta W^{(s)} + \delta W^{(b)} = \delta F^{(s)} + \delta F^{(b)} = 2A \sum_{i,j} \tau_{ij}^{(s)} \delta\epsilon_{ij} + At \sum_{i,j} \tau_{ij}^{(b)} \delta\epsilon_{ij}. \tag{1.3}$$

The factor 2 in the first term is because of the two surfaces, one on each side of the slab. The specific free energy γ is introduced by

$$\delta F^{(s)} = \delta(\gamma A) = \gamma \delta A + A \delta\gamma. \tag{1.4}$$

After specifying Eq. (1.3) to a particular strain ϵ_{ij} one obtains a relation between the surface stress $\tau^{(s)}$ and the surface specific free energy γ:

$$\tau_{ij^{(s)}} = \gamma \delta_{ij} + \frac{\partial \gamma}{\partial \epsilon_{ij}}. \tag{1.5}$$

This is the venerable Shuttleworth equation. As noted before, for a liquid, surface stress and a surface specific free energy are equal. In other words, the second term in Eq.1.5 vanishes for a liquid. The reason is that for a liquid there is no resistance to plastic deformation. If the surface is expanded, atoms or molecules flow from the interior to the surface, so that the local environment of each surface atom is the same as before the expansion. Hence all area-specific quantities as the specific free energy do not change. For solid surfaces the absolute value of the second term in the Shuttleworth equation can exceed the specific free energy. The sign of the term can be positive or negative.

The definition of the surface stress with Eq.1.3 is frequently used in theoretical calculations of the surface stress. One calculation the change in the total energy of a slab of several monolayer thickness upon the expansion of the surface of the slab (Eq.1.3). The derivation of the energy with respect to the strain is a linear function of the thickness of the slab. The surface stress is obtained by extrapolation to zero thickness.

We conclude this section with a qualitative discussion of the thermodynamic meaning of the second term in the Shuttleworth equation (Eq.1.5). This term, the derivative of the specific free energy with respect to strain $\partial\gamma/\partial\epsilon$, can be either positive or negative, and the absolute value may outweigh the specific free energy by a factor of 2 or 3. The term is zero for a liquid surface. A liquid is also characterized by the fact that there is no resistance to a flow of atoms from the bulk to the surface. For solid surfaces one may, therefore, consider $\partial\gamma/\partial\epsilon = \tau^{(s)} - \gamma$, a thermodynamic driving force to move atoms from the bulk into the surface layer. When $\tau^{(s)} - \gamma > 0$ then the surface has the propensity to accumulate more atoms in the surface layer. When $\tau^{(s)} - \gamma < 0$ the surface prefers less atoms in the surface layer. For a solid, to have either more or less atoms in the surface layer is equivalent to a symmetry-breaking reconstruction, which places atoms in unfavorable bonding positions with respect to the next layer of the substrate. A reconstruction occurs, therefore, only if $|\tau^{(s)} - \gamma|$ is large enough to compensate

for this energy expense. The reconstruction of the Au(111) surface is an example where the large tensile stress causes a reconstruction to a surface layer of higher density and the quantity $\tau^{(s)} - \gamma$ is recovered as the driving force for the reconstruction in the microscopic modeling.

Chapter 2

Principle of surface stress measurements and the experimental setup

2.1 Experimental methods

In this section, we denote our attention to the experimental techniques by which the surface stress is measured and to the theory behind the experimental methodology. Unfortunately, up to now there is no practical method to determine the absolute surface stress. It has been proposed that the absolute surface stress could be determined by measuring the detailed curvature of a thin membrane of the material as it bends by gravitation through its own weight. The method does not appear, however, to be of practical usefulness because the resulting curvature is only significant for the surface stress if the membrane is a completely homogeneous single crystal plate. Furthermore, it seems difficult to achieve a proper mount of the sample.

An alternative route to measure the absolute surface stress could be to measure the elastic extension of the length of the thin wire under an applied force. In principle, the surface stress offsets the force obtained from the force vs. strain relation by a constant which is the surface stress multiplied by the perimeter around the wire. Alas, the effect is rather small and may, therefore, be extremely difficult to extract from the noise. No one seems to have attempted to use this method, although the extension of a wire upon a change in the surface stress was used as a measure of the change in the surface stress in an electrochemical environment. As mentioned before, the only attempt to determine the absolute surface stress was with the observation of the contraction of small gold, silver, and platinum particles under the influence of

the surface stress by Vermaak and coworkers[82]. The experimental results of 1.175, 1.415 and 2.574 N/m, respectively are, however, low by a about a factor of 2 compared with first principles calculations on unreconstructed (111) or (100) surfaces[83]. These calculations are considered to be fairly reliable. The discrepancy is, therefore, most likely to be attributed to the experiment. It is possible that the surfaces of the small particles could have been reconstructed and the reconstruction reduces the surface stress[77]. The change in the surface stress due to the reconstruction is, however, only about 0.3N/m[77] which still leaves a large gap between the experimental and the theoretical results. A more likely reason for at least part of the discrepancy between the experiment and theory is that the experimental measurement of the lattice constants gold particles were not performed under UHV conditions, but rather under relatively poor vacuum conditions in an electron microscope. Consequently, the surfaces of the gold particles were likely covered with a carbonaceous deposit or with a surface oxide. As we have leaned (Fig.1.21) that such deposits lower the tensile stress, the experimental values for the absolute surface stress may indeed be low. No one has attempted to repeat these early experiments with the improved instrumentation currently available. Finally, we note that the data were evaluated by assuming that the particles were elastic isotropic materials. Since the noble metals are known to be extremely anisotropic in their elastic properties, the evaluation procedure itself could have introduced an additional error.

2.1.1 Detection of surface stress

To obtain surface stress, we must evaluate strain. There are roughly two evaluation methods of

1. Measurement of strain from lattice distortion, and
2. Measurement of substrate deflection.

The former method is to measure of plane spacing or lattice constant of surface and determine the strain using deviation from normal value. Examples of this method are X-ray and electron beam diffraction. The X-ray diffraction need film of more than 10nm thickness and the electron beam diffraction is impossible in plasma which includes both positively and negatively charged particles. Hence, in this thesis, each method are inadequate to observe the surface stress of disorder-order processes at top surface layer ($\sim$1nm) in plasma.

Consequently, more appropriate method to evaluate the surface stress in this case is the latter one. For this substrate deflection method, the substrate

deflects according to its shape, the elastic constants and the stress if the surface modified by the plasma. Using detecting the deflection of the substrate, we can fundamentally calculate the surface stress. But if the deflection is complicated, the analysis becomes extremely difficult. It is important to select the simplest shape to analyse for easy calculations of the surface stress and higher precision. For the reasons, the shapes of rectangular and circle which are thin and simple are frequently used. Uniformity of the surface stress is compensable for measurably small.

The circle shaped substrate is to be uniformly deflected by the stress in the films. If the substrate is initially flat plane, we can consider it to be a part of the sphere. By measuring the curvature, the stress is determined. Some of the methods were using of light interference (e.g. Newton ring method, Michelson interferometer etc.). But these has difficulties in fixing the samples fundamentally and is not appropriate for *in-situ* measurements.

In case of the substrate which is thin and rectangular shape, monitoring the displacement of one free end enable to calculate the stress, while another is fixed. There are the measurements of optical and electrical method (using capacitance etc.). In plasma, the electrical methods is impossible to measure the displacement. Hence, we selected optical cantilever bending methods as one of the useful optical methods. Next, we review common 'cantilever bending methods' to understand history of the surface stress measurements using 'cantilever bending methods'.

2.1.2 Cantilever bending methods

While the experimental determination of the absolute surface stress remains a challenge, the experimental technique to measure changes in the surface stress is well established. In their pioneering work, Koch and Abermann[84] demonstrated that the bending of a cantilever can be measured with sufficient sensitivity that the change in the stress due to the deposition of a single monolayer on one side can be detected. The principle of the measurements is shown in Fig.2.1(a). In the work of Koch and Abermann the bending of a cantilever was measured as the change of the gap of the capacitor using the three terminal capacitance method (Fig.2.1(b)). The bending can also be measured with sufficient sensitivity by deflection of the beam of a laser using a position sensitive detector, provided that the sample is thin enough. A simple add-on system for the measurement of epitaxial stress in thin films using the laser beam deflection technique was described Sander *et al* [85]. The method was also employed earlier by Martinez *et al.* [86] and Schell-Sorokin *et al.* [87] on Si samples of about 50μm thickness. The preparation of large area single crystal metal sheets of such a small thickness is more

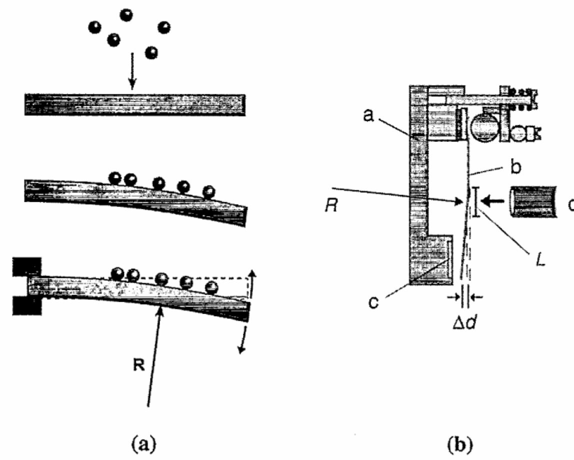


Figure 2.1: Illustration of the principle of the cantilever bending method(a). Upon deposition of material on one surface the stress is changed and the cantilever bends. Here, the case of a compressive stress where the surface wants to expand is drawn. The radius of curvature R is measured as the change of the gap of a capacitor by Δd (b). The two electrodes of the capacitor are formed by the sample and a capacitor electrode c. The capacitor electrode is surrounded by a guard electrode (not shown) in order to minimize the effects of stray capacitances. The sample b is clamped on one end in the sample holder a. From Ref.[84].

difficult. A further problem with metals is that the preparation of clean metal surfaces typically involves cycles of sputtering and annealing. During those cycles, thin metal stripes tend to bend which increases the effective stiffness of the sheet and thereby reduces the bending in an uncontrolled way. By employing the three terminal capacitance method one can detect extremely small changes of the gap which forms the capacitor (less than 0.1Å). This allows the use of thicker samples of 0.3-0.5mm. The ultimate sensitivity of the capacitance method is not limited by the sensitivity of the detection but the rather by vibrational noise and thermal drift. Exposure of the substrate material on one surface to adsobates is performed via a tube with an opening diameter d in Fig.2.1(b) placed close to the sample. If the crystal is cleaned only in a particular section of the sample which is comparable or smaller than the orifice of the tube (see Fig.2.1(b)), then the adsorption takes place only there and the sample bends only in this section. Cleaning the sample in a defined area by sputtering and annealing is the typical situation for metal surfaces. In case of semiconductors the cleaning is frequently facilitated by flashing the sample to a high temperature whereby the entire surface on both sides of the sample becomes clean. If the sticking coefficient is near unity, the arrangement as shown in Fig.2.1(b) still allows for a measurement of the change in the surface stress upon adsorption, since the exposure on the side facing the open tube is larger by about an order of magnitude. The relation between the change in the gap Δd and the radius of curvature R in the regime where the sample bends due to adsorption is:

$$\Delta d = (L^2/2 + Ll)/R \tag{2.1}$$

in which L is the linear dimension of the section which bends and l is the distance between the center of the exposed area and the center of the capacitor electrode. It is assumed that L is larger or equal to the width of the sample, and smaller than its length. The radius of curvature is related to the change in the capacitance ΔC by

$$\left(\frac{1}{R} - \frac{1}{R_0}\right) = \frac{\epsilon_0 A}{L^2/2 + Ll}\frac{\Delta C}{C_0^2} \tag{2.2}$$

in which ϵ_0 is the absolute dielectric permeability, A the area of the capacitor and R_0 and C_0 are the radius of curvature and the capacitance before the change in the surface stress, respectively. Without loss of generality R_0 can be assumed to be infinite.

We use optical micro-cantilever bending technique as one of cantilever bending methods. The detailed discussions will be appeared in subsection2.2.1. Here, first of all, we prepare the relations between strain and

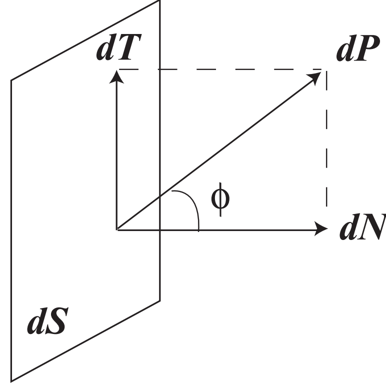


Figure 2.2: internal force acted on infinitesimal area: dP. dT: parallel component of dP, dN: normal component of dP.

surface stress and about thermal effects for the measurements using cantilever bending methods.

2.1.3 Stoney's formula

Surface stress occurs in film or layer, resulting from being bound by its substrate. Here, it is important to know the stress distribution. In general, theory of isotropic elastic body enable us to understand most of such problems. Strain, which is detected, can be converted into the stress. We derive Stoney's formula which describe relation between the strain and the stress.

If a solid body is acted by external force or thermal effects at interface, the action propagates into the solid body inside through the interface. Producing internal drag against the external action prevents the form of the body from being destructed. If the solid body is balanced, it keeps the balance during internal force acts each other at both side of a plane in the body.

As in Fig.2.2, if dP is internal force acted on infinitesimal area and ϕ is an angle between the plane and the normal axis, normal component of dP, dN, and parallel one, dT, are written below,

$$dN = dP\cos\phi,\ dT = dP\sin\phi\ . \tag{2.3}$$

In this case, stress along normal axis, σ_n, and parallel to the plane, σ_t are respectively written,

$$\sigma_n = dN/dS,\ \sigma_t = dP/dS. \tag{2.4}$$

Next, if a bar of length, L, and diameter, d, is extended and the length and the diameter become L', d' respectively while its length difference is L_1, longitudinal strain and transverse strain are defined as follows,

$$\epsilon = (L' - L)/L = L_1/L, \ \epsilon' = (d' - d)/d. \tag{2.5}$$

The ϵ and ϵ' are dimensionless quantities. The ϵ of compressive and tensile stress correspond to positive and negative, respectively. Since the diameter decreases if the bar is extended, the sign of the ϵ' is opposite of the ϵ. Poisson's ratio is defined as a following equation,

$$\nu = -\epsilon'/\epsilon. \tag{2.6}$$

If two parallel planes is set in distance, L, and slip by a length, l, with acting along the direction of tangent line of the plane, while the sheared angle is α, shear strain, γ, is defined as follows,

$$\gamma = l/L = \tan\alpha. \tag{2.7}$$

Being very small, the γ is approximately equal to the α.

If we restrict the discussion to small strains, the stress is directly proportional to the strain. This is called *Hook's* law.

$$\epsilon = \sigma_n/E, \ \gamma = \sigma_t/G, \tag{2.8}$$

where the proportional coefficient of E and G are $Young's$ modulus and Rigidity modulus, respectively.

In general, for a plane in the isotropic materials, we can always select orthogonal set of three planes which no shear stress acted on. These three plane is called *principal planes* and normal stress which acts each planes is *principal stress* of crystal. With the linearity relations between the stress and the strain, the strain is described with summation of strains which the stress components each correspond to. Hence, the principal strains in the isotropic materials is given by the principal stresses,

$$\epsilon_1 = [\sigma_1 - \nu(\sigma_2 + \sigma_3)]/E, \epsilon_2 = [\sigma_2 - \nu(\sigma_3 + \sigma_1)]/E, \epsilon_3 = [\sigma_3 - \nu(\sigma_1 + \sigma_2)]/E. \tag{2.9}$$

Afterwards, we assume that substrate and film are isotropic elastic body for convenience. On growing the film on the substrate, the stress in the film is parallel to the interface while being bound within the plane parallel to the interface between the substrate and the film. On the contrary, being bound free from the substrate along normal direction, no stress occurs in

this direction. Therefore, if the stress is while x, y-axes is set in the interface and the z-axis is set to be normal to the plane, the stresses are given by

$$\sigma = \sigma_{xx} = \sigma_{yy},\ \sigma_{zz} = \sigma_{zx} = \sigma_{zy} = \sigma_{xy} = 0 \tag{2.10}$$

in film. In this time, the strains are also given by

$$\epsilon_x = \epsilon_y = \sigma(1-\nu)/E,\ \epsilon_z = -2\nu\sigma/E. \tag{2.11}$$

We assumed the stress uniformly acted on the sectional plane of the sample so far. But, it is not uniform in general. As one example, let us take the rectangular shaped shell of the radius of curvature, R, which is stressed by the thin film of the film and substrate system. In Fig.2.3, strain ϵ at distance y from neutral plane, nn_1, which doesn't expand or shrink is derivated as follows, because triangle non_1 and s_1n_1s' are similar,

$$\epsilon = s's_1/nn_1 = y/R \tag{2.12}$$

where line nm is parallel to the n_1s_1. Hence, the stress as a function of thickness y is given by

$$\sigma(y) = \epsilon E/(1-\nu) = yE/R(1-\nu). \tag{2.13}$$

To derivate the Stoney's formula which describe the relation between the stress and the radius of curvature, As in Fig.2.4, we can assume that a stress in film is constant if $h \gg t$, where h and t are thickness of substrate and film respectively, while the stress in film is as a function of thickness y. Using Young's modulus E, Poisson's Ratio ν and width of substrate a perpendicular to the neutral plane and the thickness direction, the stress $\sigma(y)$ is as follows. From balanced equation of force,

$$a\int_{e-h-t}^{e-h} \sigma dy + a\int_{e-h}^{e} \sigma(y)dy = a\sigma t + \frac{aEh}{2R(1-\nu)}(2e-h) = 0 \tag{2.14}$$

and also from balanced equation of moment,

$$a\int_{e-h-t}^{e-h} \sigma y dy + a\int_{e-h}^{e} \sigma(y)y dy = \frac{a\sigma t}{2}\{2(e-h)-t\} + \frac{aEh}{3R(1-\nu)}(3e^2-3eh+h^2) = 0 \tag{2.15}$$

Here, because we assumed that $h \gg t$, the constant stress in the film is given by

$$\sigma = \frac{Eh^2}{6R(1-\nu)t} \tag{2.16}$$

where $e = h/3$.

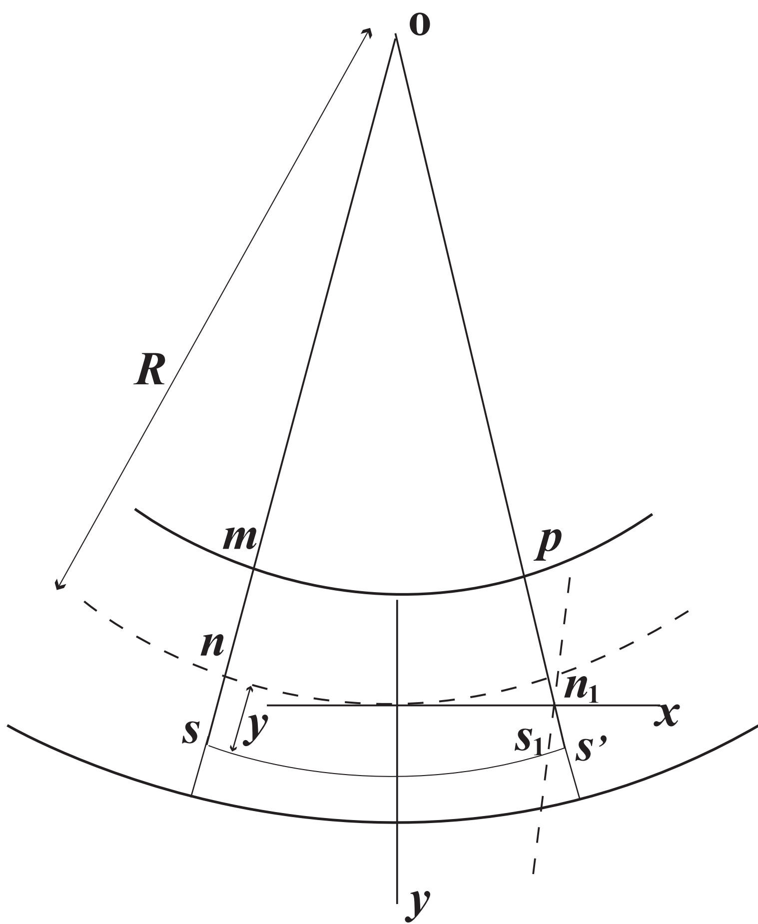


Figure 2.3: A shell of the radius of curvature, R.

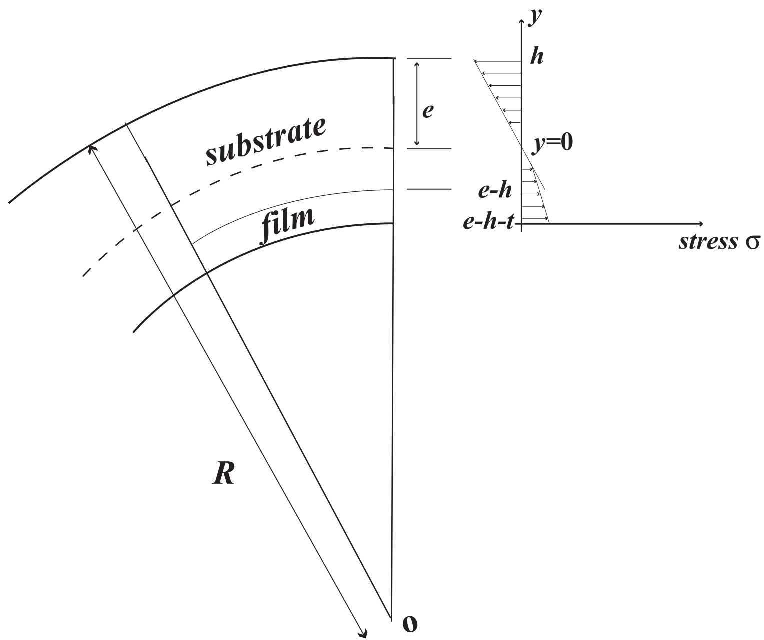


Figure 2.4: A stress distribution as a function of thickness of film and substrate system: $\sigma(y)$.

In case of cantilever which is the one end is free while the other is fixed, displacement of the free end, δ, is converted into the surface stress as follows using relation of $1/2R = \sigma/L^2$ where L is length of the substrate,

$$\sigma = \frac{Eh^2}{3L^2(1-\nu)t}\delta. \tag{2.17}$$

This equation is Stoney's formula[105].

In this thesis, we observe the total stress change $\Delta\sigma_t$ is given by the Stoney's formula,

$$\Delta\sigma_t = \frac{Eh^2}{3L^2(1-\nu)}\delta = 1.19 \times 10^6\delta[\mathrm{N/m}] \tag{2.18}$$

where L is the length of the cantilever (450μm), h the thickness(2.0μm), E Young's modulus and ν Poisson's ratio of Si(100) ($E/(1-\nu) = 1.805 \times 10^{11}\mathrm{N/m^2}$)[106].

2.1.4 Temperature estimation of microcantilever with 'bimetallic effect'

Expansion or shrinkage due to temperature change produces the stress, *i.e.*, thermal stress. The thermal stress is not intrinsic. In this subsection, we will discuss the thermal stress and derivate equation of the sample deflection δ.

Consider a rod of length L, and the initial temperature T_1 changes to T_2 by being heated. As the length change is $\alpha(T_2 - T_1)L$ according to temperature change, the strain ϵ is given by

$$\epsilon = \alpha(T_2 - T_1)L/L = \alpha(T_2 - T_1), \tag{2.19}$$

where α is a thermal expansion coefficient of line. Under bound by its internal drag in material, the stress appeared due to the temperature change:

$$\sigma = \epsilon E = \alpha(T_2 - T_1)E. \tag{2.20}$$

If the two rods having different thermal expansion coefficients was attached and bound by each other, they have same temperature change by heat conduction, *i.e.*, 'bimetallic effect'. In case of $\alpha_1 > \alpha_2$, the rod of 1 expands by $\alpha_1(T_1 - T_0)L$, while the rod of 2 shrinks by $\alpha_2(T_1 - T_0)L$. Consequently, the rod of 1 is shrinks by the compressive stress by $\epsilon_1 L = \sigma_1 L/E_1$. The actual expansion of the rod of 1 is

$$\alpha_1(T_1 - T_0)L - \sigma_1 L/E_1. \tag{2.21}$$

As in case of 1, the actual expansion of the rod of 2 added by the tensile stress of 1 is

$$\alpha_2(T_1 - T_0)L + \sigma_2 L/E_2. \tag{2.22}$$

Since the final length of 1 and 2 is same,

$$\alpha_1(T_1 - T_0)L - \sigma_1 L/E_1 = \alpha_2(T_1 - T_0)L + \sigma_2 L/E_2. \tag{2.23}$$

From balance of force,

$$\sigma_1 S_1 + \sigma_2 S_2 = 0 \tag{2.24}$$

where S_1 and S_2 is cross sections of 1 and 2. Combining Eq.2.23 and 2.24,

$$\sigma_1 = (\alpha_1 - \alpha_2)(T_1 - T_0)E_1/(1 + S_1 E_1/S_2 E_2). \tag{2.25}$$

Henceforth, we suppose that the rod of 1 is thin film (F) and the rod of 2 is substrate (S). In this case, the relation of $S_1 E_1 \ll S_2 E_2$ comes as $t \ll h$, and then

$$\sigma_1 = (\alpha_1 - \alpha_2)(T_1 - T_0)E_1. \tag{2.26}$$

And displacing the temperature difference $T_1 - T_0$ to ΔT,

$$\sigma_F = (\alpha_F - \alpha_S)\Delta T E_F/(1 - \nu_F), \tag{2.27}$$

where ν_F means that the stress acts parallel to the interface.

Further, using the Stoney's formula of Eq.2.17, we can obtain the sample deflection δ due to the thermal stress,

$$\delta = 3L^2 E_F(1 - \nu_S)t(\alpha_F - \alpha_S)\Delta T/E_S(1 - \nu_F)h^2. \tag{2.28}$$

In fact, in this thesis, the temperature change during the electron irradiation was measured using the bimetal effect (difference of thermal expansion) of the Si cantilever coated on its back side with Al as film. The sample dimensions were $L = 450\mu$m, $h = 2\mu$m and $t_{Al} = 30$nm. And the physical constants were $E_{Si(100)}/(1 - \nu_{Si(100)}) = 1.805 \times 10^{11}$[N/m], $E_{Al}/(1 - \nu_{Al}) = 1.073 \times 10^{11}$[N/m], $\alpha_{Si} = 3.3 \times 10^{-6}[^\circ\mathrm{C}^{-1}]$ and $\alpha_{Al} = 25 \times 10^{-6}[^\circ\mathrm{C}^{-1}]$. Hence, if the sample deflection δ[m] was induced due to the thermal stress, the temperature change is described in

$$\Delta T = 1.701 \times 10^7 \delta [^\circ\mathrm{C}]. \tag{2.29}$$

2.2 Our experimental setup

In this section, we denote the our experimental setup and needful knowledges for it.

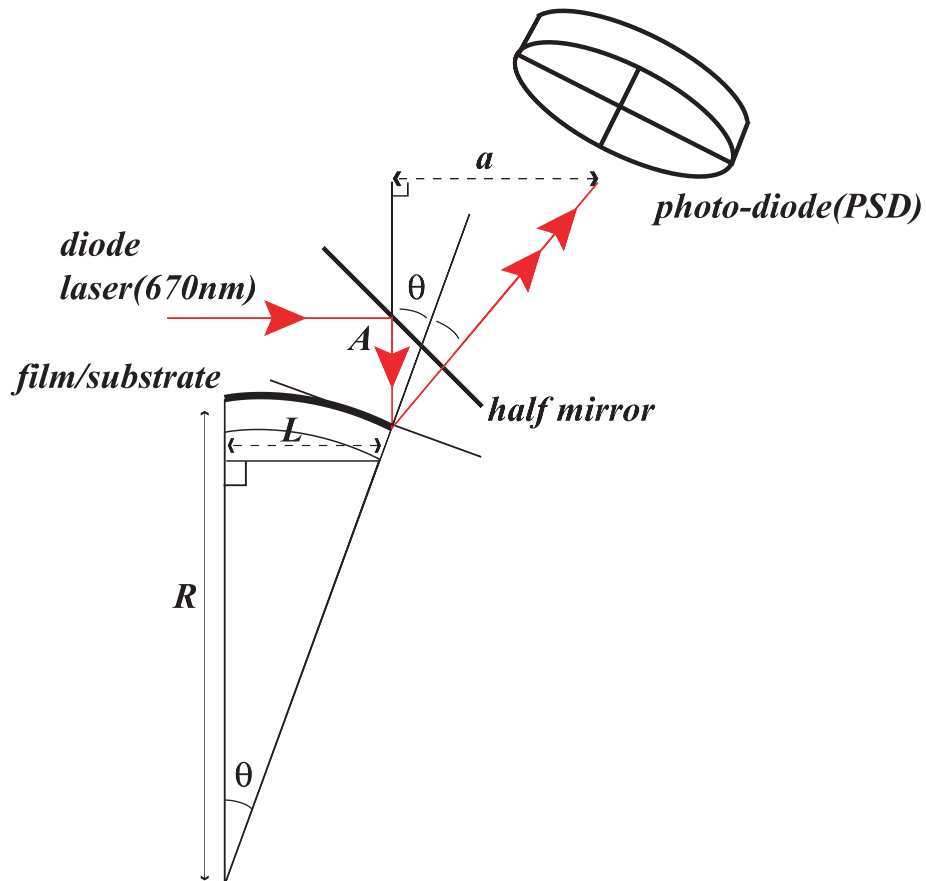


Figure 2.5: principle of Optical cantilever bending method. If reflector tilts by angle θ, direction of reflected beam changes by 2θ. By position sensitive detector (PSD), distance a is measured and the radius of the curvature R is obtained. From $\theta \ll 1$, $L \approx R\theta, a \approx 2A\theta$. Then, $R = 2AL/a$

2.2.1 Optical micro-cantilever bending technique

In this subsection, we discuss the optical microcantilever method. At first, we treat a principle for the optical cantilever bending technique. In Fig.2.5, the optical cantilever bending technique uses position change of reflection by the substrate. We used as samples Si microcantilevers with dimensions of $450 \times 50 \times 2.0\mu$m (Fig.2.6). This sample enables to measure surface stress with high accuracy, since it is a sensitive to the surface strain for being small and thin. One end of the cantilever sample was fixed, while the other was free. A laser light incident on the sample at the free end was reflected and detected by a position sensitive detector (*PSD*) etc., while the sample bent by angle θ and the reflective angle became to be 2θ. The radius of the curvature R of the substrate is obtained from the detected position change, and then we can determine the displacement δ at the free end.

We used a four splitted type photodiode as the *PSD*. The structure of the photodiode is similar to a diode of *pn* junction. Photon illumination to depletion layer of the of *pn* junction produces lots of electron-hole pairs according to degree of the illumination. Photon having a higher energy than the band gap induces diffusion of hole and electron to *p*-type and *n*-type, respectively. If recombination of hole and electron is negligible, current flows along to the inverse direction of diode proportional to the illumination.

We used the four splitted type photodiode as the two splitted type photodiode. First, we set reflection to center of the two photodiodes and then the detected potential difference (DIF) between two photodiodes is 0V, because same number of photons incident to each part of the photodiodes. If the substrate bends and position of the reflection incident to the *PSD* changes, the DIF becomes finite value. Using the DIF, the displacement of the substrate is given.

In our system of the optical microcantilever method, under total potential difference of two diodes to be 9.36V, the change in the DIF of 12.9mV corresponds to the displacement of 1nm. In other words, the ratio of (DIF/ADD) of 1.38×10^{-3} corresponds to 1nm.

UHV-optical micro-cantilever bending technique

Our system is adaptable even under UHV condition as in Fig.2.7. One end of the microcantilever sample was fixed on a mount near a quartz window in a UHV chamber, while the other was free. The distance between the backside of the sample and the quartz window was approximately 0.4 cm. The surface stress changes developed during the ion bombardment and the electron irradiation were monitored through the sample bending using the

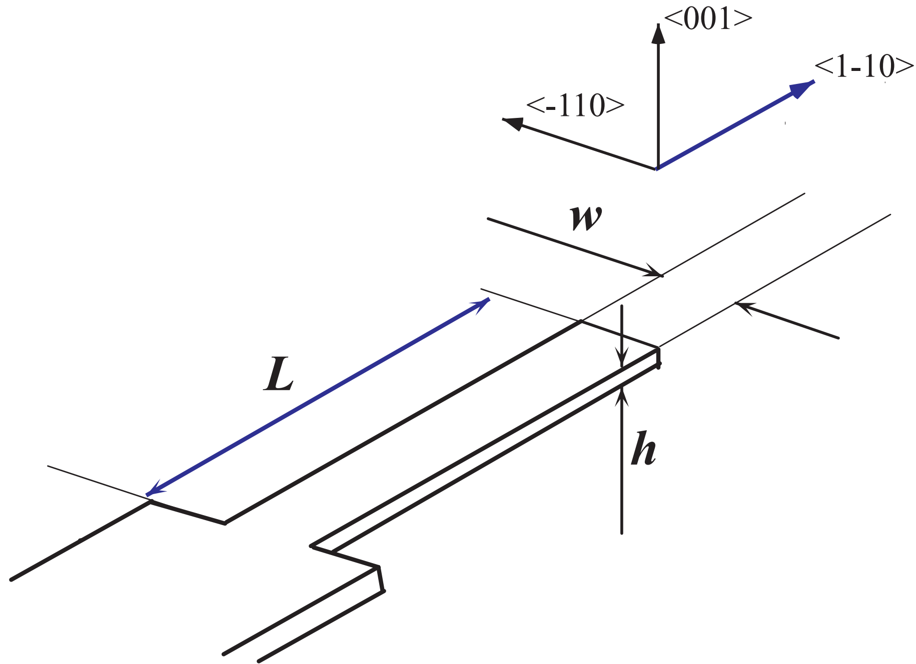


Figure 2.6: Our sample for the optical micro-cantilever bending technique. This is a very small and thin rectangular shaped sample with dimensions of $450 \times 50 \times 2.0\mu$m. It enable to measure surface stress with high accuracy, since it is a sensitive to the surface strain for being small and thin.

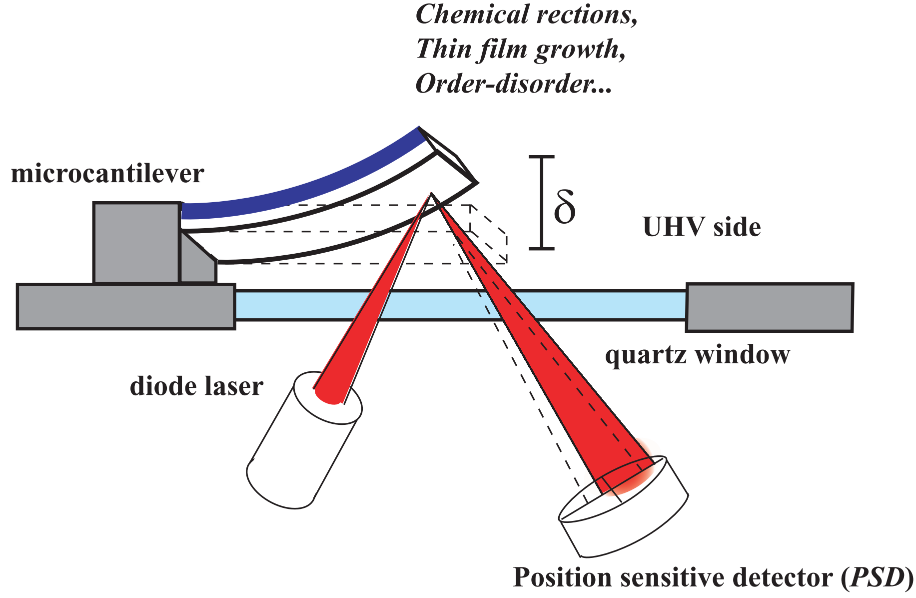


Figure 2.7: Schematic drawing of high sensitive optical microcantilever bending method under UHV condition.

optical microcantilever method. A laser light incident on the backside of the sample at the free end was reflected and detected by a position sensitive detector (*PSD*). A deflection of the free end of the lever, δ, was determined by monitoring the position of the reflected laser light on the *PSD*. This method allows us a non-contact, real-time detection of the deflection without any electric noise from the plasma.

In this system from the ratio of (DIF/ADD), the deflection δ of the substrate and the total stress $\Delta\sigma_t$ is given by

$$\delta = (\mathrm{DIF/ADD})/(1.38 \times 10^{+6})[\mathrm{m}] \tag{2.30}$$

$$\Delta\sigma_t = 0.8612(\mathrm{DIF/ADD})[\mathrm{N/m}]. \tag{2.31}$$

Furthermore, we calibrated our detection system with a piezoelectric element, and confirmed a resolution of higher than 0.1nm in δ.

Chapter 3

Electron-stimulated surface stress relaxation of Si

3.1 Abstract of this chapter

We have observed the nonthermal relaxation of surface stress in Si induced by electron irradiation at room temperature. An atomically thin disordered layer was introduced by Ar ion bombardment. The surface stress change during ion bombardment and the following electron irradiation of Si(100) was measured by means of an optical microcantilever technique. We have found that the compressive stress in the Si surface due to disorder induced by ion bombardment was completely relaxed by electron irradiation at low energy. The criterion for complete relaxation is found not to be total energy deposition, but number of irradiated electrons.

3.2 Introduction

Surface stress has been extensively investigated, and it is known that these stresses are generally caused by strains at surfaces (surface reconstruction, bonding topology or roughness) or at film/substrate interfaces (lattice mismatch, different thermal expansion)[91, 92]. Defects at surfaces should also affect surface stress through forming strains around them. However, most previous studies have focused on stress changes in ordered surfaces or films, and the effect of surface disorder has not been well understood. It is well known that defect disorder at surfaces and interfaces of semiconductors influences not only crystalline periodicity but also electrical and optical properties such as electric conductivity, dielectric strength and energy levels of mediate centers[93]. Recently, we have observed growth of compressive stress during

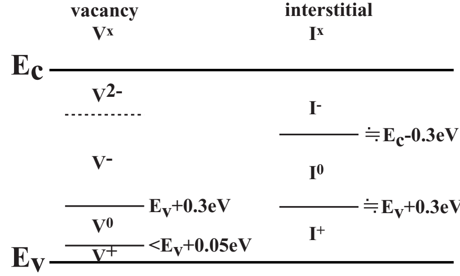


Figure 3.1: Energy levels of charge state of interstitial and vacancy in band gap.

ion-bombardment on Si(100) to introduce surface disorder[94]. This result indicates that the surface layer is expanded relative to the Si substrate lattice due to the defects. In general, thermal annealing at high temperatures can recrystalize a surface disordered layer and would release stress. An interesting question is, however, is it possible that the disorder-induced stress can be relaxed at low temperature? Defect in Si have different levels in the band gap depending on the charge states[95, 96]. Carrier trapping at a defect site may create a localized energy release. Hence a possible way to relax disorder-induced compressive stress is for the defect to capture an electron and change its own charge state with very low energy electron irradiation.

Low energy electron beams of 10-1000eV have long been used to investigate the chemical and structural character of surfaces[97, 98] (Fig.3.2). For ionic materials, electron stimulated desorption of constituents is known to occur[99], and this implies atomic-scale structural changes. For clean metal and semiconductor surfaces, however, it is generally believed that these energetic electrons do not cause structural changes[100]. To examine this conventional wisdom, we performed real-time measurements of surface stress changes of Si during electron irradiation. If motion of surface atoms is led by energetic electrons, the change of surface stress will have a finite value. In this chapter, we show, using a highly sensitive cantilever technique, that very low energy electron irradiation is quite effective in causing the recovery of defect structures into ordered Si surfaces through the inter-relationship between surface stress and defect disorder.

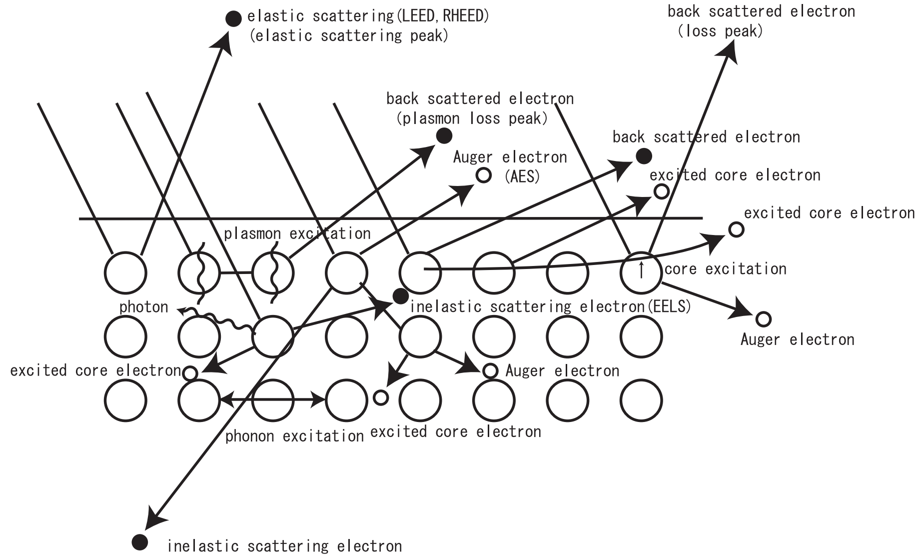


Figure 3.2: Schematics of electron excitations at surface frequently used in surface analysis.

3.3 Experimental

3.3.1 Samples of Si(100) microcantilever and its cleaning

We used as samples n-type Si(100) microcantilevers (resistivity $0.01-0.025\Omega$cm) with dimensions of $450 \times 50 \times 2.0\mu$m whose long side is parallel to the $\langle 011 \rangle$ axis(Fig.3.3). To obtain clean surface, the samples were treated by being dipped in 10% HF acid solution for 5 minutes, rinsed with deionized water for 5 minutes and annealed at 1000K for 30 minutes in ultra high vacuum (UHV) ($\sim 10^{-7}$Pa).

3.3.2 Methods for irradiations of ions and electrons

Ar ion-bombardment of the samples to introduce disorder into the surface was performed at room temperature for 0 to 3,000s using an argon plasma. The argon plasma was generated with the radio frequency (*rf*) discharge of Ar gas (1.3Pa, 99.9999%) at a frequency of 13.56MHz and a power of 500W in a UHV chamber of Fig.3.4. An adequately long distance of about 1.5m between the sample and a coil of *rf* discharge prevents the sample temperature

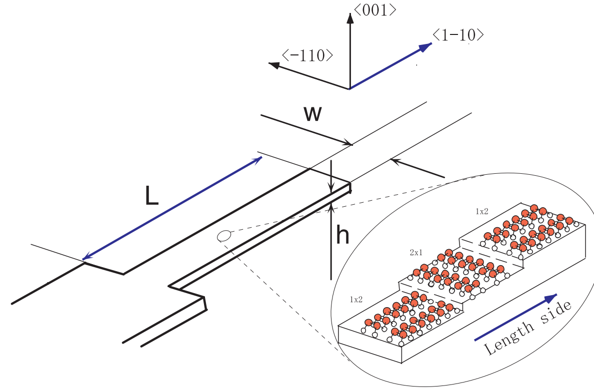


Figure 3.3: Orientations of Si(100) microcantilever: This sample with dimensions of $450 \times 50 \times 2.0\mu$m whose long side is parallel to the $\langle 011 \rangle$ axis.

from being high. From the Langmuir probe measurements[101](Fig.3.5), the electron temperature, the electron density and the space potential (V_p) were 1.4eV, 6×10^6cm^{-3} and +5V at the sample position, respectively. For Ar ion-bombardment, negative dc biases (V_b) of $-30, -60$ and -100V was applied to the Si sample to attract the ions(Fig.3.6). The average damage depth in the sample was obtained by TRIM calculation[104] to be about 1nm in these any case(Fig.3.7). No peak of LVV-Auger electron signal of Ar was observed on the surface after the ion-bombardment(Fig.3.8). Electron irradiation was performed after the bombardment at room temperature, using the argon plasma at positive biases from +8.75 to +45.0V. The electron penetration depth in this energy range is comparable to the damage depth[102](Fig.3.9 and 3.10). The plasma has a space potential and the incident energy of the irradiated species of ion and electron was given by $E_i = e|V_b - V_p|$.

3.3.3 Measurements of stresses during ion and the following electron irradiation

One end of the microcantilever sample was fixed on a mount near a quartz window in a UHV chamber, while the other was free as in Fig.3.11. The Debye length was 0.36 cm, and the sheath length was estimated to be 1.2cm from the Child law[101]. To minimize the back side of the sample from being bombarded with ions, the sample was set in the sheath; the distance between the sample and the quartz window was approximately 0.4 cm.

The surface stress changes developed during the ion bombardment and

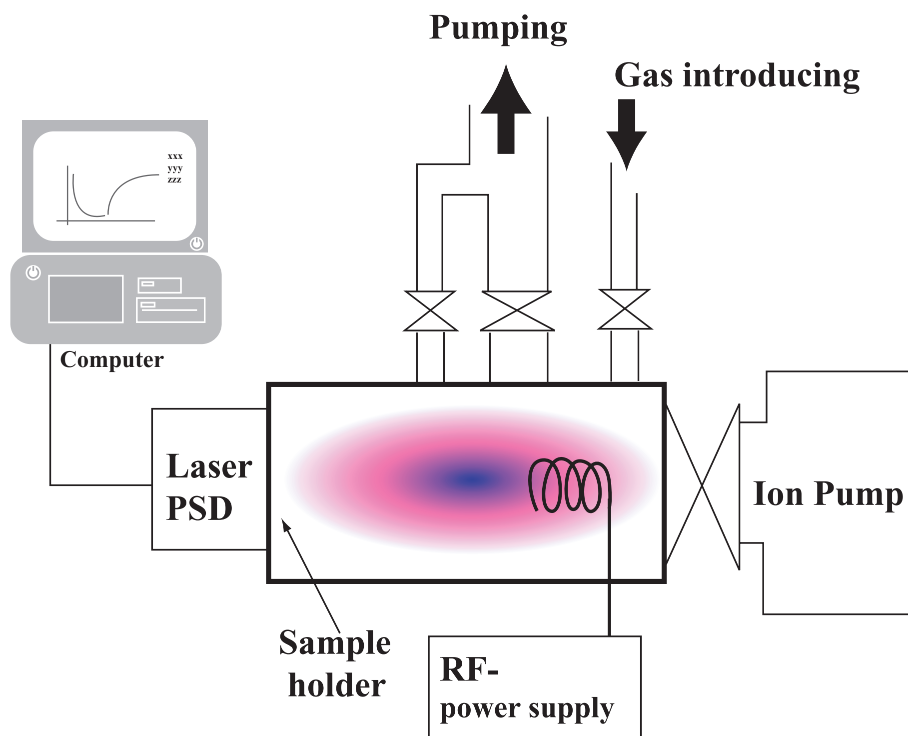


Figure 3.4: An UHV chamber system in this experiment. The UHV system achieves ultrahigh vacuum of $\sim 10^{-7}$Pa using Turbo molecular pump and Ion pump. The sample is set on a sample holder at leftside wall of the chamber. A distance between the sample and a coil of radio frequency (*rf*) discharge is adequately long of about 1.5m not to make the sample temperature high with the plasma.

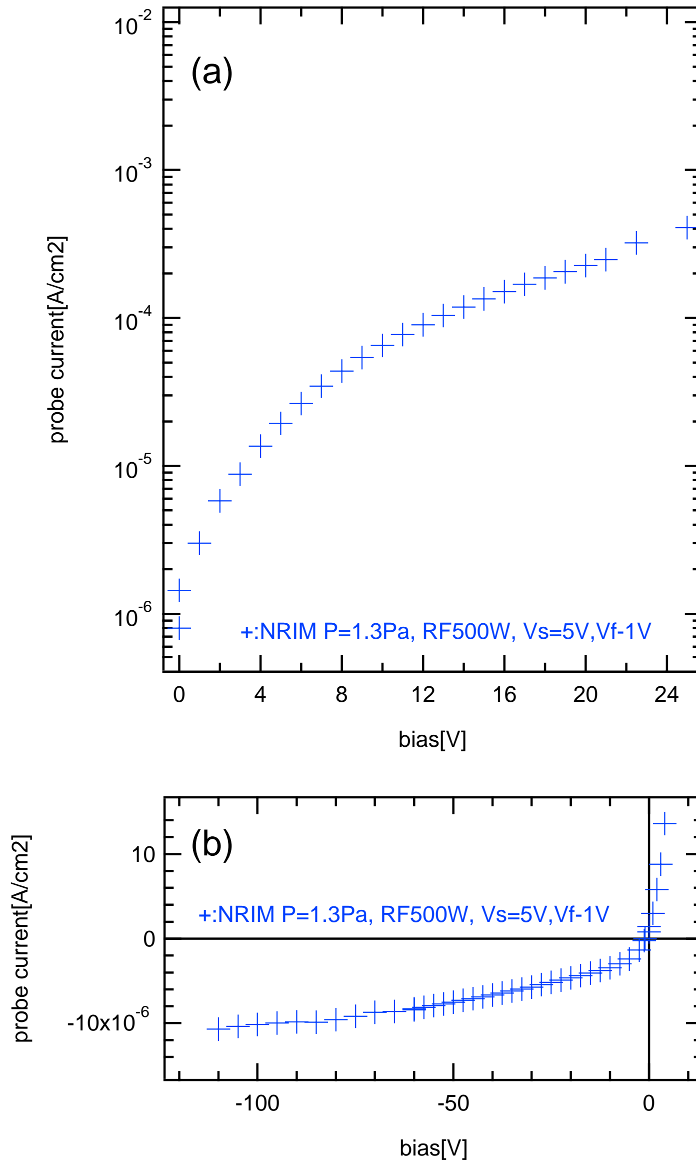
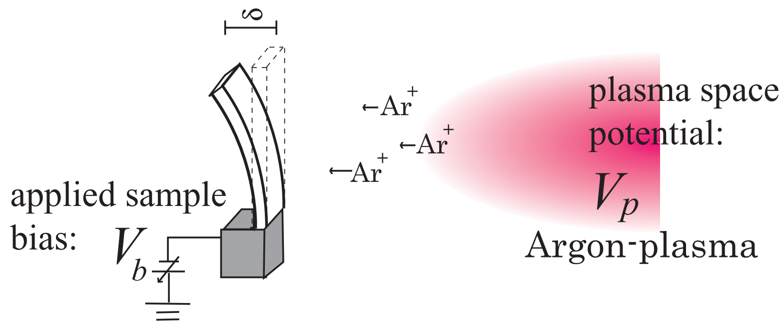


Figure 3.5: $I - V$ characteristic of sample holder as a single Langmuir probe: mainly (a) electron current region (positive bias); (b) ion current saturation region (negative bias). The sign '+' indicates $I - V$ characteristic of the RF discharge plasma. The electron temperature of this RF plasma, the electron density, the space potential (V_p) and the floating potential (V_f) were 1.4eV, $6 \times 10^6\text{cm}^{-3}$, +5V and -1V at the sample position, respectively.



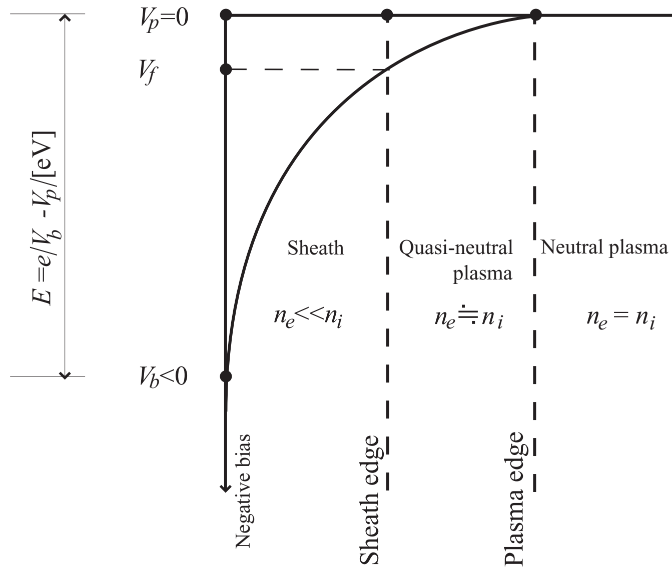


Figure 3.6: Ion irradiation onto Si(001) microcantilever and its spatial potential structure of Ar plasma. Detailed spatial structure is discussed in subsectionC.5.3.

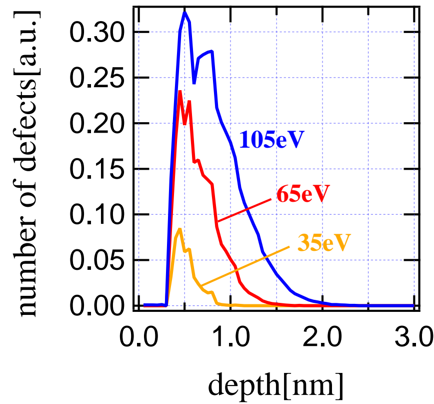


Figure 3.7: Defect distribution produced by ion bombardment. This figure shows calculated defect distributions which is obtained with TRIM code[104] as a function of depth with Ar ion energy at 35, 65 and 105eV. Average damage depths obtained from these distribution for 35, 65 and 105eV are 0.66, 0.87 and 1.1nm, respectively.

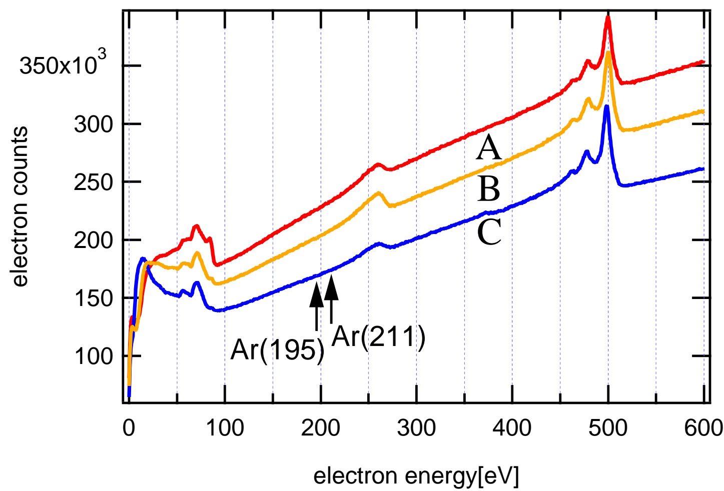


Figure 3.8: Examples of Auger spectra recorded in counting mode. These spectra were observed *ex-situ* on (A) clean Si(100) surface, (B) electron irradiated Si(100) surface for 1000s after ion-bombardment for 2000s and (C) ion bombarded Si(100) surface for 2000s. Two arrows indicate positions of LVV-Auger peak of Ar.

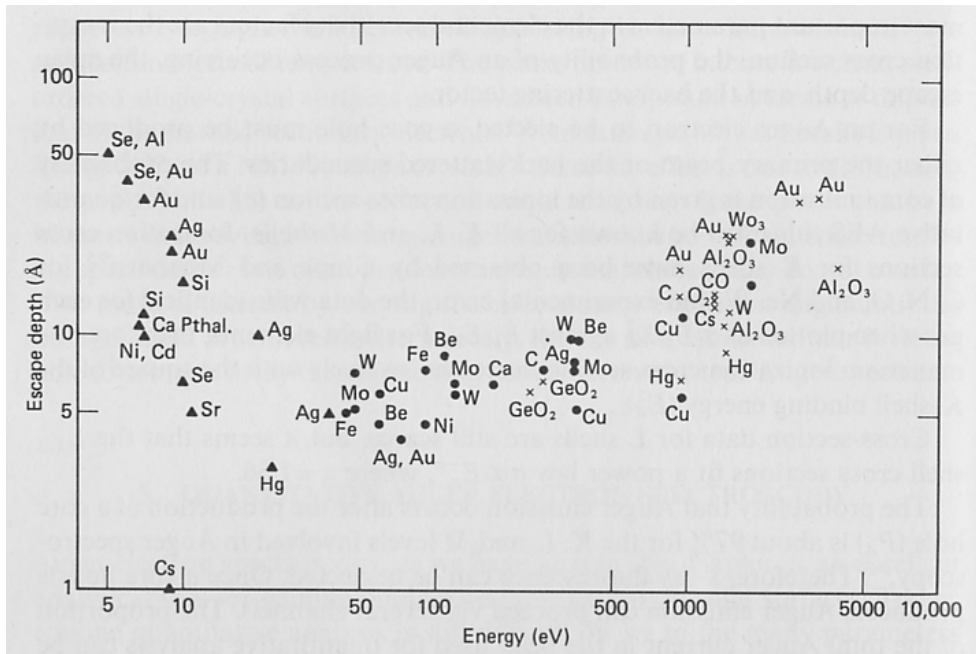


Figure 3.9: Escape depths for a number of materials as a function of electron energy. (From Ref.[103].)

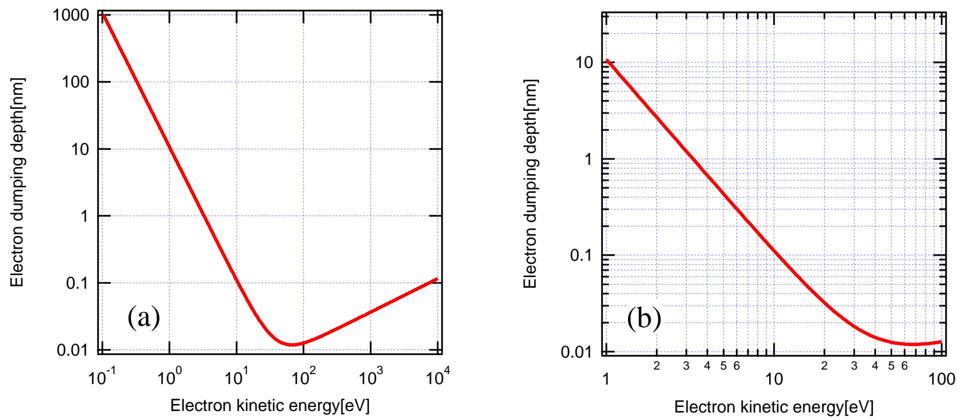


Figure 3.10: Universal curve of electron penetration depth obtained from lots of experimental results. (b) is the horizontal enlargement of (a). This curve is expressed with $\lambda_{dump} = 538aE^{-2} + 0.41a^{3/2}E^{1/2}$[nm]; $a^3 = (10^{21}A)/(\rho N_A)$[nm^3]. For Si, atomic weight $A = 28.09$, density $\rho = 2.33$[g/cm^3], Avogadro's number $N_A = 6.02 \times 10^{23}$. (From Ref.[102].)

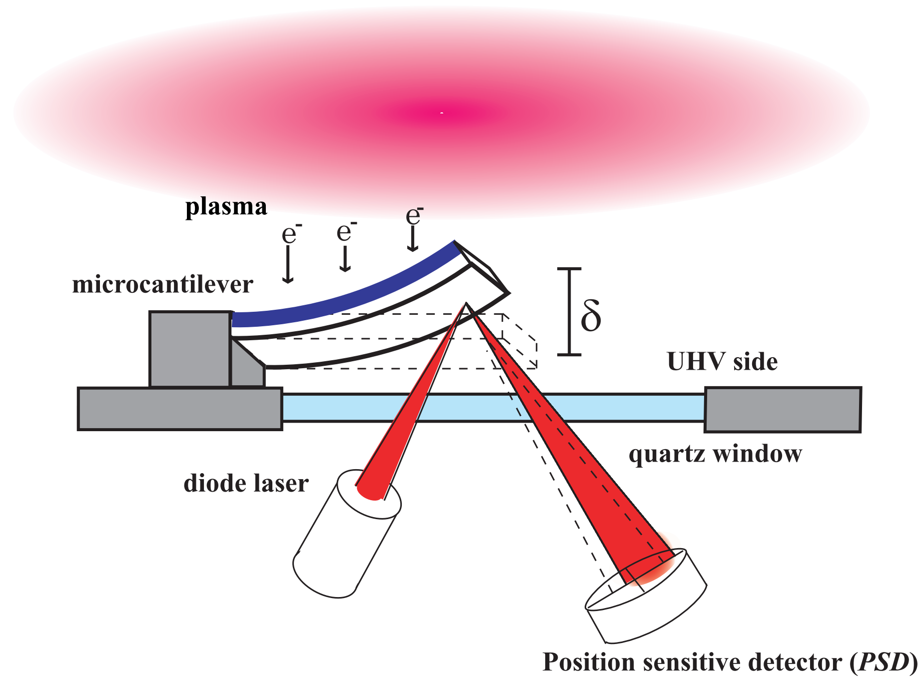


Figure 3.11: Redrawing of Fig.2.7. The high sensitive optical microcantilever bending system is set at the leftside wall of the chamber as in Fig.3.4. This system keep high sensitivity of the sample deflection by avoiding vibrations of pumps and considerable any other sources.

the electron irradiation were monitored through the sample bending using the optical microcantilever method. A laser light beam incident on the back side of the sample at the free end was reflected through the UHV-window and detected by a position sensitive detector (*PSD*). A deflection of the free end of the lever, δ, was determined by monitoring the position of the reflected laser light on the *PSD*. This method allows *in situ*, real time detection of the deflection. We calibrated our detection system with a piezoelectric element, and confirmed a resolution of higher than 0.1nm in δ of the cantilever sample. The high sensitivity of the sample deflection results from avoiding vibrations of pumps and considerable any other sources. The surface stress change $\Delta\sigma_t$ was obtained from the δ using Stoney's formula[105],

$$\Delta\sigma_t = \frac{Eh^2}{3L^2(1-\nu)}\delta[\mathrm{N/m}], \tag{3.1}$$

where L is the length of the cantilever (450μm), h the thickness(2.0μm), E Young's modulus and ν Poisson's ratio of Si(100) ($E/(1-\nu) = 1.805 \times 10^{11}\mathrm{N/m^2}$)[106]. Detailed detection method and treatment of the Stoney's formula are discussed in subsection 2.1.1 and 2.1.3, respectively.

3.4 Results and discussions

3.4.1 Ion bombardment induced surface stresses

Results

Little research has been performed on surface stress changes induced by ion bombardment. Here we discuss on the changes of stress during Ar ion bombardment. Figure 3.12(a) shows the time evolution of the total stress during Ar ion bombardment for different electron incident energies. The total stress became quickly negative indicating a development of compressive stress, and reached a saturation of -0.52N/m after bombardment for 2,000s for a bias of 105eV. The stress change was bigger as electron incident energy was getting higher. The compressive stress implies that the surface is expanded by ion bombardment.

Discussions

We attribute the compressive stress observed here to the weakening of Si-Si bonds due to the formation of surface defects. This is consistent with the longer bond length in amorphous Si (0.235-0.2375nm) compared with crystal (0.235nm).[108] The radial distribution function (RDF) obtained by

X-ray and electron diffraction is consistent with a 3-dimensional random network of tetrahedrally bonded atoms. The first and second neighbor distances (r_1, r_2) and coordination numbers (C_1, C_2) are almost identical in crystalline and amorphous silicon. The main features of the RDF do not depend on the method of amorphization, including the method by ion bombardment. In detail, the densities of the amorphous Silicon are 5-30% less than the densities of crystalline material depending on the method and conditions of amorphization.

The degree of lattice disorder can be evaluated quantitatively in terms of the number of defects[114]. The number of defects induced by ion bombardment N_s is given by

$$N_s = \alpha N_d \nu \xi t, \tag{3.2}$$

where N_d is the effective area density of Si atoms, ξ the ion flux, α the displacement cross section, ν the damage function and t the duration of ion-bombardment. The quantity ($\alpha N_d \nu$) is the number of defects per incident ion, and was calculated using the TRIM98-code[104]. Because the ion flux ξ can be measured for different electron incident energies[101], we use areal density of defects N_s as the index of degree of disorder. In other words, the N_s is appoximately proportional to the energy deposition by electron. Figure 3.12(b) shows the total stress induced by ion bombardment as a function of N_s. The change of total stress depended only on the number of defects and was independent of the bias, indicating that the compressive stress is determined by the defects induced by ion-bombardment.

There would be two main origins of the defective compressive stress during ion bombardment. One of the origins is decay of intrinsic stress due to surface reconstruction. Dimers are sequentially formed on the reconstructed Si(100) surface, and form 2×1-*dimer row structure* in wider area as in subsection 1.1.2. It is known that the anisotropy in local structure of the dimer bonding directly influences the surface stress. Consequently, the anisotropy on the surface stress is caused by the dimer formation, as the stress parallel to the dimer bonds is tensile, while that perpendicular is compressive. The difference was theoretically expected to be between 1.1 and 3.2N/m[109, 110], where the higher value refers to first principles calculations and is, therefore, presumably more reliable. Since the theoretical value is much bigger than the surface stress change during the ion bombardment, the dimer row structure of the reconstructed surface should be *partially* destroyed after the ion bombardment.

Here, we cannot explain the compressive stress based only on destruction of the surface reconstructed structure, because peak of surface disorder exists nearby 1nm depth as in Fig.3.7, which is 7 layers from the surface under this

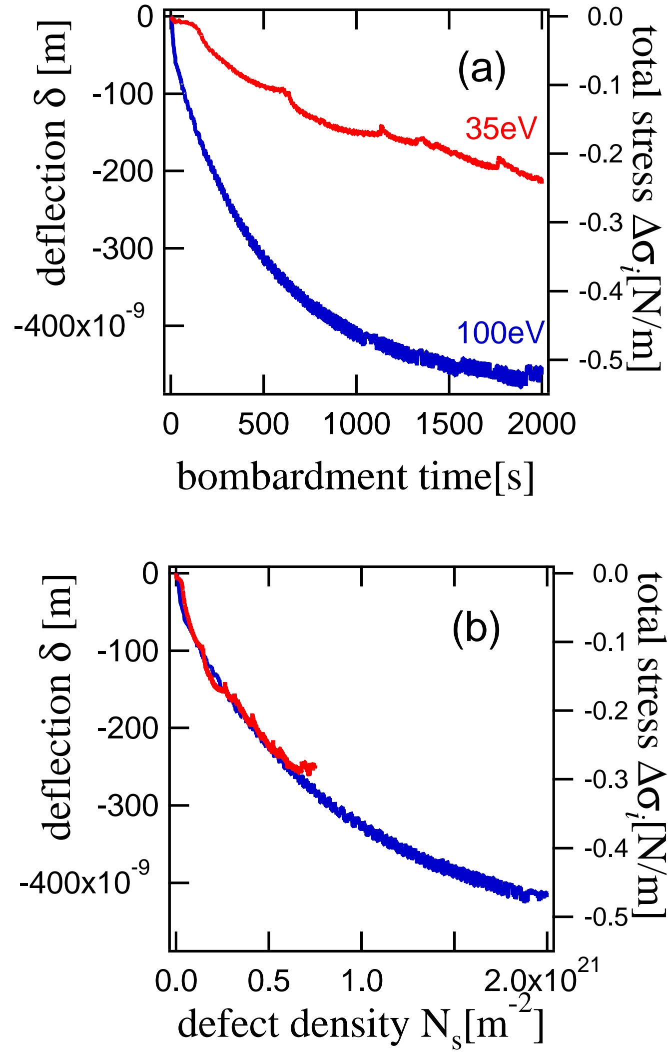


Figure 3.12: The total stress induced by ion bombardment (a) as a function of duration of bombardment by argon ions, and (b) as a function of areal density of defects, Ns

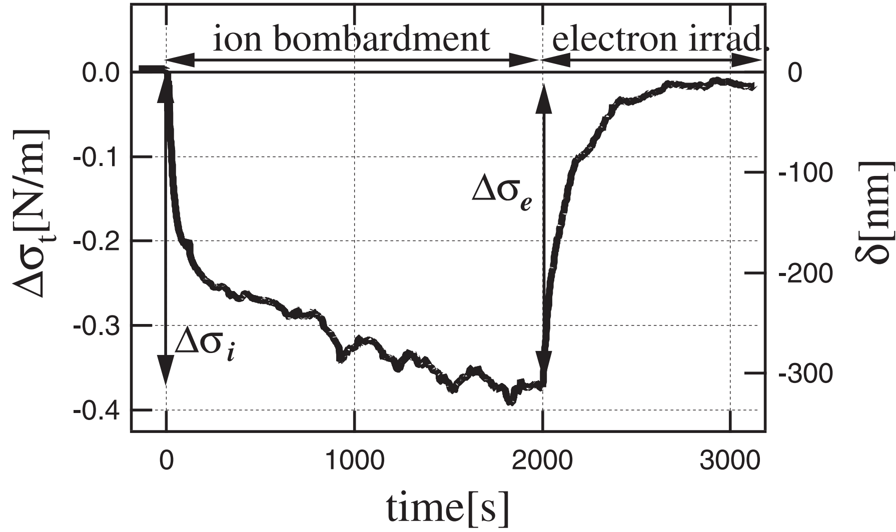


Figure 3.13: Time dependence of the surface stress change during ion-bombardment followed by electron irradiation. Ion-bombardment was performed at energy of 65eV for 2,000s and the subsequent electron irradiation was at 10eV.

bombardment condition. Hence, another origin is considered to be volume expansion of defective layer. The collision of an implanted argon ion breaks Si-Si bonds near the surface. Ions able to transfer more kinetic energy on impact than the displacement energy of Si ($E_d = 14$eV[111]) will thus create a lattice defect in the form of an interstitial silicon atom and vacancy pair[112]. The observed growth in surface compressive stress is due to the volume expansion of the defective layer of 1 nm depth and the evolution was explained as a result of the number of defects caused by the incident ions[94, 114]. The major contribution to the volume expansion will be the interstitials produced by the ion bombardment. Intuitively, vacancies are likely to cause shrinkage, or at least less expansion, because of volume loss at the sites although there may be anisotropy in the strain around vacancies, such as the Jahn-Teller effect[113].

3.4.2 Electron irradiation onto the bombarded surfaces

Results

Figure 3.13 shows the time dependence of the surface stress evolution during the ion-bombardment at 65eV for 2,000s and the subsequent electron irradi-

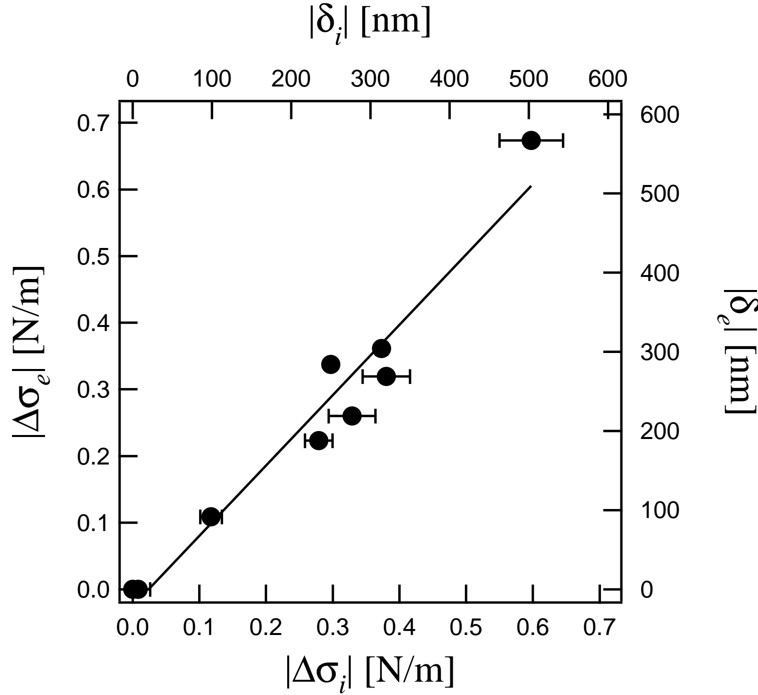


Figure 3.14: Stress relaxation under electron irradiation versus degree of disorder in the surface. Horizontal axis shows disorder-induced stress and deflection by ion-bombardment, $\Delta\sigma_i$ and δ_i, and vertical axis is stress and deflection at saturation induced by the following electron irradiation at fixed 40eV, $\Delta\sigma_e$ and δ_e.

ation at 10eV. Upon the ion-bombardment the value of the compressive (*i.e.* negative) stress rapidly increased and then gradually levelled off. This compressive stress was, on the contrary, found to decrease rapidly immediately after electron irradiation. The change of stress with time gradually decreased and stress was saturated. What is interesting here is that the compressive stress generated during ion-bombardment returned to zero under electron irradiation. The surface layer is expanded relative to the Si substrate lattice by the bombardment and the following electron irradiation acts on this layer to relax it, during which complete relaxation of the disorder-induced stress occurs.

Fig.3.13 is one example of surface disordering. Figure 3.14 plots values of surface stress release at saturation $\Delta\sigma_e$ under electron irradiation at a fixed energy for various disorder-induced stresses $\Delta\sigma_i$ (see Fig.3.13 for the symbols). On the non-bombarded surface: $\Delta\sigma_i = 0$, no significant change of

the surface stress was observed. Whether on less or more disordered surfaces, similar complete relaxation was observed. Linear curve fitting to these points yields the simple linear function $\Delta\sigma_e = +(1.06 \pm 0.09)\Delta\sigma_i + (-0.025 \pm 0.029)$ within errors. Complete relaxation was independent of the degree of disorder.

Discussions

Electron energy dependence of the stress relaxation Next, for relaxation of disorder-induced stress by electron irradiation, Fig.3.15(a) shows time evolution of surface stress during electron irradiation for different electron incident energies 3.75, 10 and 25eV on the surface disordered by fixed bombardment conditions at 65eV for 2,000s. The vertical axis was normalized as electron-irradiation-induced-stress at time t, $\Delta\sigma_e(t)$, divided by the disorder-induced stress, $\Delta\sigma_i(= -0.37[\mathrm{N/m}] = \mathrm{coast.})$: $|\Delta\sigma_e(t)/\Delta\sigma_i|$. Complete relaxation corresponds to the relaxation ratio $|\Delta\sigma_e(t)/\Delta\sigma_i| = 1$. It was found at any incident energy of electron. With higher energy, the time until complete relaxation becomes shorter.

The complete relaxation occurred if irradiated electron only had larger energy than 3.75±1.41eV. Fig.3.16 shows the saturated value of the relaxation ratio of the stress relaxation by the electron irradiation. The saturated values of lower energy than 2.5-3.75eV did not reach the complete relaxation. This results would results from including ions in the electron beams, *i.e.*, the electron temperature of the Ar plasma was to be 1.41 eV and thus the electron beams in this energy range of 2.5-3.75eV could include less than 7-17% ions having infinitesimal energy. But these ions has not so enough energy to break Si-Si bonds of a few eV for the bias condition of excessing the plasma space potential. Furthermore, from this consideration, this results indicates the relaxation has no critical threshold energy. For example concerning electron excitation, electron excitation process of Plasmon is 16.4-16.9eV and 11.6-12.0eV for bulk and surface of Si[113]. The value could not determine the threshold energy value.

Electron number dependence of the stress relaxation To investigate the dependence of number of irradiated electrons, Fig.3.15(b) plots the evolution of $|\Delta\sigma_e(t)/\Delta\sigma_i|$ as a function of the number of electrons. Here, the number of electrons was estimated from measurement of the electron current. Unexpectedly, the change in the relaxation ratio depended only on the number of electrons and was independent of the incident energy. This shows that the relaxation is independent of total energy deposition, because roughly the total energy deposition into the disordered layer should be the product of the number of irradiated electrons and the incident energy. For all incident ener-

gies the stress was completely relaxed for the number of irradiated electrons $n_{el} \sim 5 \times 10^{21}\mathrm{m}^{-2}$. This number of electrons corresponds to about 3 times as many as the estimated number of introduced defects ($\sim 1.7 \times 10^{21}\mathrm{m}^{-2}$) at 0 Kelvin[94].

Why is the temperature rise so low?

In our experimental condition, annealing by electron heating would not occur, because the temperature rise during electron irradiation was no more than about 10°C from room temperature as in Fig.3.17. The temperature change was measured experimentally by using the bimetal effect (difference of thermal expansion) of the cantilever coated on its back side with Al as is described in subsection 2.1.4.

Why is the temperature rise so low under the electron irradiation? To answer the question, we perform estimation of the temperature rise from energy balance quantitatively. Areal energy influx by electron irradiation, W, is expressed as

$$W = j \times \epsilon [\mathrm{W/cm^2}] \tag{3.3}$$

where j and ϵ are current and incident energy of irradiated electrons, respectively. Here, areal density of the micro cantilever sample which having thickness of 2.0μm is given by

$$m = \rho V = (2.33) \times (1 \times 1 \times 2 \times 10^{-4}) = 4.66 \times 10^{-4} [\mathrm{g/cm^2}], \tag{3.4}$$

where ρ and V are mass density and volume per area, respectively. From specific heat at constant pressure of Si, $c_p = 0.7[\mathrm{J/g^\circ C}]$, areal heat capacity is

$$mc_p = 3.262 \times 10^{-4} [\mathrm{J/^\circ C cm^2}]. \tag{3.5}$$

Therefore, if all the energy of electron influx contributes increase of temperature, rate of the temperature rise is

$$\frac{\Delta T}{\Delta t} = \frac{W}{mc_p} [^\circ\mathrm{C/s}]. \tag{3.6}$$

Since the areal energy influxes are $W = 3.75 \times 0.03 = 0.1125[\mathrm{mW/cm^2}]$ for 3.75eV, $10 \times 0.2 = 2.0[\mathrm{mW/cm^2}]$for 10eV, $25 \times 1.0 = 25[\mathrm{mW/cm^2}]$ for 25eV and $40 \times 50 = 2000[\mathrm{mW/cm^2}]$ for 40eV, respectively, from the current measurements in Fig.3.5, $\Delta T/\Delta t = 0.345[^\circ\mathrm{C/s}]$ for 3.75eV, $6.13[^\circ\mathrm{C/s}]$ for 10eV, $76.6[^\circ\mathrm{C/s}]$ for 25eV and $6,131[^\circ\mathrm{C/s}]$ for 40eV, respectively. These results without thermal emission mean the sample becoming very high temperature after longer irradiation against the experimental results with the bimetal effect, and then we add black body radiation term as the thermal emission.

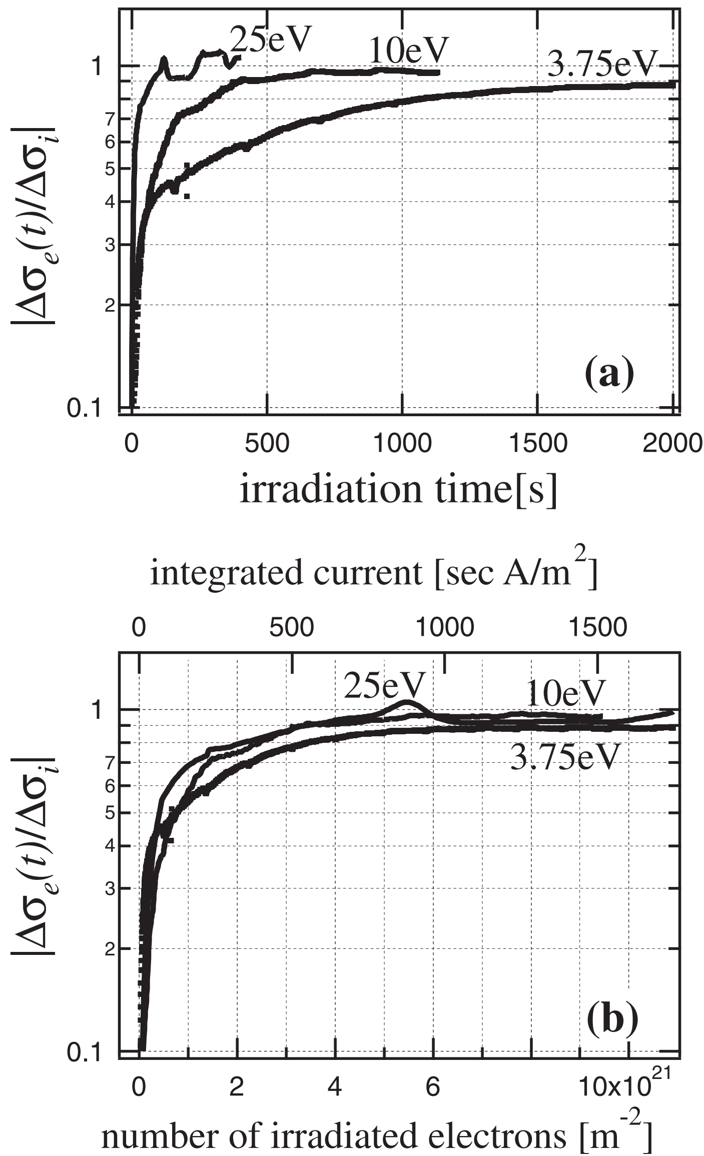


Figure 3.15: Relaxation ratio, $|\Delta\sigma_e(t)/\Delta\sigma_i|$ as a function of (a) the electron irradiation time and (b) the number density of irradiated electrons. All of the disordered surfaces are bombarded at 65eV for 2,000s and values of the stress, $\Delta\sigma_i$, are about -0.37N/m.

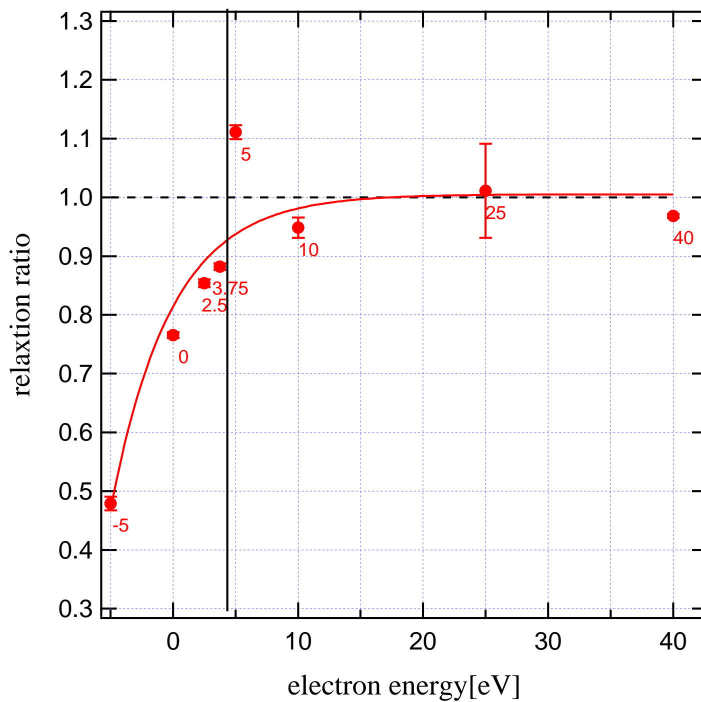


Figure 3.16: Electron energy dependence versus the relaxation ratio. The complete relaxation occurred if irradiated electron only had larger energy than 3.75±1.41eV. These would result from including ions in the electron beams, because the electron temperature of the Ar plasma was to be 1.41 eV. The electron beams in this energy range of 2.5-3.75eV could include less than 7-17% ions having infinitesimal energy.

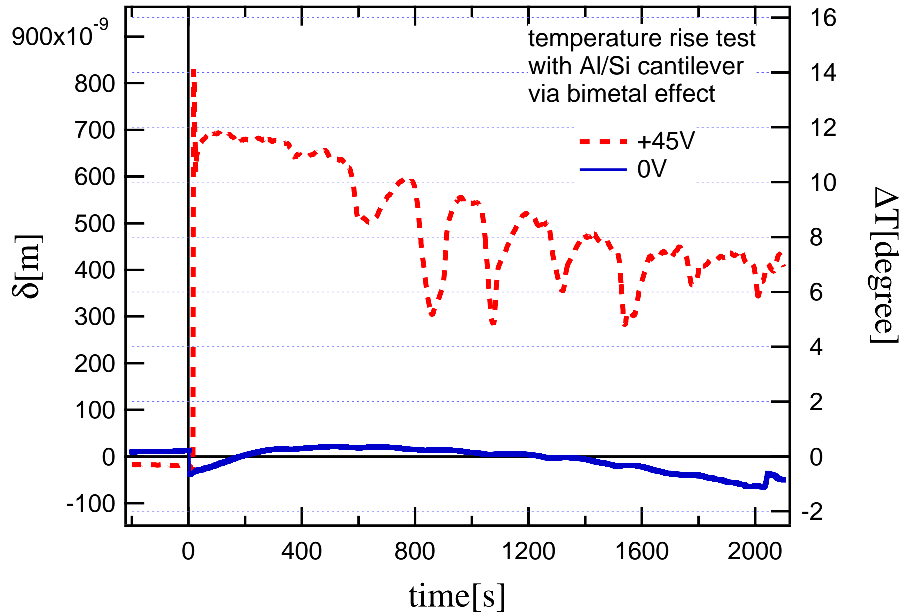


Figure 3.17: Estimation of temperature change during plasma irradiations using the bimetal effect (difference of thermal expansion) of the cantilever coated on its back side with Al. These results show the temperature rise of positive values for the appling dc biases of +45V and 0V to the samples. Plasma irradiation began at time 0 s. The temperature rises were estimated to be within about 10°C for +45V and 1.1°C for 0V from the room temperature. After the stopping the plasma irradiations, the value of the temperature change returned to zero. Hence, the temperature rises are very low during the plasma irradiations.

The matter emits radiation according to its own temperature, or performs the black body radiation. The amplitude depends Stefan-Boltzmann's law of σT^4 where $\sigma = 5.67 \times 10^{-12}$[W/°K^{4}cm^2] is Stefan-Boltzmann's constant. This system is considered to be steady state, because the typical diffusion times of $\tau = l^2/\kappa_{Si}$ are enough small of 44ns and 2.25ms for sample thickness direction (2μm) and length one (450μm), respectively, compared with the time interval of the surface stress measurements of 1 sec. κ_{Si} is thermal diffusion coefficient for Si. Hence, we can obtain the following equation from the energy balance for the sample temperature T and the total surface area $S \left(= [\{(450 \times 50) + (450 \times 2.0)\} \times 2 + (2.0 \times 50)] \times 10^{-4\times 2}[\mathrm{cm}^2]\right)$ during the electron irradiation,

$$\omega - (\sigma T^4 - \sigma(300°\mathrm{K})^4)S = 0[\mathrm{W}] \quad (3.7)$$

where is the energy gain on irradiated surface of $s = (450 \times 50) \times 10^{-4\times 2}[\mathrm{cm}^2]$, which is expressed as $\omega = sW$ for the micro cantilever sample of the dimensions of 450μm$\times$50μm$\times$2.0μm. According to Eq.3.7, temperature rises for 3.75, 10, 25 and 40eV are estimated to be 0.088, 1.56, 17.9 and 349°K from room temperature of 300°K, respectively. The value of 40eV is too high, compareed with the experimental results.

Next step, we can include thermal diffusion term to holder of the cantilever to explain the discrepancy in case of high electron energy. The cantilever is fixed to a silicon holder. We used cantilever part of *POINTPROBE Silicon-AFM-Sensor* of NANOSENSORS GmbH & Co. KG in Germany as samples. Fig.3.18 shows a sketch of the holder together with cantilever and tip. As we can see the holder being very big relative to the microcantilever, the holder is considered to be perfect heat sink at room temperature[116]. Adding the thermal diffusion term of $\eta = k_{th}\frac{\partial T}{\partial x}A$ to Eq.3.7,

$$\omega - (\sigma T^4 - \sigma(300°\mathrm{K})^4)S - \eta = 0[\mathrm{W}]. \quad (3.8)$$

The derivation term of $\frac{\partial T}{\partial x}$ is able to be displaced with $\frac{T-300°\mathrm{K}}{L}$, because we can use the sample temperature T instead of $T(x)$ for being in steady state. The symbols of L, A and k_{th} are the length, the cross section of cantilever and the thermal conductivity of Si. According to Eq.3.8 including the thermal diffusion term, the temperature rises for 3.75, 10, 25 and 40eV are estimated to be 0.001, 0.013, 0.167 and 13.4°K from room temperature of 300°K, respectively. These values including both the thermal radiation and the thermal diffusion are consistent with the measured temperature of Fig.3.17 by using the bimetal effect even at high electron energy. Hence, in this electron irradiation case, we concluded that temperature rise was very

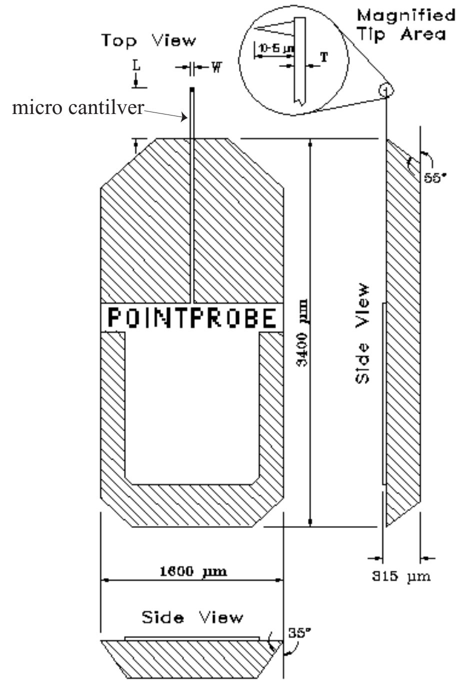


Figure 3.18: We used cantilever part of *POINTPROBE Silicon-AFM-Sensor* of NANOSENSORS GmbH & Co. KG in Germany as samples. This is a versatile silicon SPM cantilever for very high resolution imaging and fits to all well-known commercial SPMs. It consists of a single crystal silicon with integrated single crystal silicon tip. The tip is pointing into the $\langle 100 \rangle$ direction. The cantilever and the tip are supported by a single crystal silicon holder. The volume of the holder is very big and then we can consider it to be perfect heat sink for the cantilever.

low, or at highest about 10 °K from room temperature by both this estimation and the bimetal experiment. This means that the sample temperature mainly affected by the thermal emission of the black body radiation at lower than 25eV, and also it is suggested that the contribution of thermal diffusion for temperature change is very small, because the thermal diffusion depends on temperature gradient and typical diffusion time. On the other hand, even at high electron energy case of 40eV, we could estimate the temperature rise by including the thermal diffusion term.

In addition, considering these results of both the stress relaxation depending on the number of irradiated electrons and the very small temperature rise, the relaxation is evidently promoted not by thermal annealing effect, but by a non-thermal mechanism.

3.4.3 Non-thermal mechanism of surface stress relaxation

Since any mechanism of thermal annealing by electron beam, plasmon excitation, electron-hole pair creation, exciton creation or plasma model[113, 117] must depend on total energy deposition by the irradiated electrons, the non-thermal stress relaxation could not been explained by them.

A plausible mechanism is suggested for the stress relaxation later. The mobility of defects in bulk crystal Si is strongly enhanced by a charge state transition through successive potential change under electron irradiation[95]. This mechanism is known to be as Bourgoin-Corbett ionization enhanced diffusion (IED)[119]. The detailed diffusion process of the IED mechanism was also surveyed by a first principle calculation[120]. Such successive charge state transitions promote migration athermally. This reason is that any charge state has different ordered potential corrugation of crystal Si(Fig.3.19). In other words, it does not require the help of temperature to occur. The athermal migration may also occur under electron irradiation of the Si surface. After electrons are captured by defects, vacancies and interstitials migrate and recombine. In consequence of such stabilization, the atomically thin disordered layer should be restored into ordered Si and the disorder-induced stress should be released.

The IED mechanism, however, is not so enough to explain the restoration of the highly disordered surface where the surface stress evolution was saturated during the ion-bombardment, because the mechanism was structured while assuming a little deviation from the ordered structure, *i.e.*, dealing with migration of defects in highly ordered Si crystal. Our present understandings and theoretical predictions for highly disordered structures are very limited

and infant for having no mathematics to describe them, computational limitation etc. Nevertheless, we can expect qualitatively the possibility of the restoration by extending the IED mechanism. Speculatively thinking migration in disordered Si, it is possible to suppose that potential around defect locally change after electrons are captured by the defect, and then the defect searches new stable site and migrates. In addition, it is possible to suppose epitaxially growth from the interface between ordered region and disordered region for thin thickness of 1nm, *i.e.*, solid phase regrowth[118]. Hence, the restoration in the disordered case should occur by similar migration mechanism via the charge state transition.

3.4.4 Energy density estimation of the surface stress relaxation based on IED mechanism

Next, in this subsection, we will discuss detailed process of the stress relaxation, focusing on energy density. In general, the surface stress is defined as a reversible work per area to stretch a surface elastically. Hence, strained surface region worked to surroundings or was worked by surroundings, *i.e.*, whose energy changes. In our case, the energy of the ion-bombarded surface increased. For the ion bombardment condition of 65eV for 2000s, the disordered region of 1nm thickness has compressive surface stress of $\Delta\sigma_i = -0.37$N/m. This stress value is convertible into specific strain energy density, $\epsilon_d([\mathrm{N/m^2}] = [\mathrm{J/m^3}])$:

$$\epsilon_d = \frac{|\Delta\sigma_i|}{t_d} = 3.7 \times 10^8[\mathrm{J/m^3}] = 3.7 \times 10^{-19}[\mathrm{J/nm^3}] = 2.31[\mathrm{eV/nm^3}] \quad (3.9)$$

where thickness of this disordered layer t_d is estimated to be about 1nm with TRIM-calculation[104]. In fact, calculated values which is average depth of defect distribution are 0.7nm for 35eV, 0.9nm for 65eV and 1.1nm for 105eV shown in Fig.3.7.

The first principle study of migration of interstitial Si atom in bulk Si crystal based on IED mechanism[120] concluded that the successive charge state transition under electron irradiation leads the migration with capturing electron or hole. According to this framework, if the interstitial captured electron and transited another charge state at a stable site, it obtains additional energy, E_t, of about 1eV relatively to the state before the charge state transition(Fig.3.19 and Fig.3.20 from [120]). And the interstitial searches for the new stable site and migrate, while it is releasing the energy of E_t. In this time, we assume that the released energy corresponds the macroscopic strain energy release from the disordered region, because it is expected that

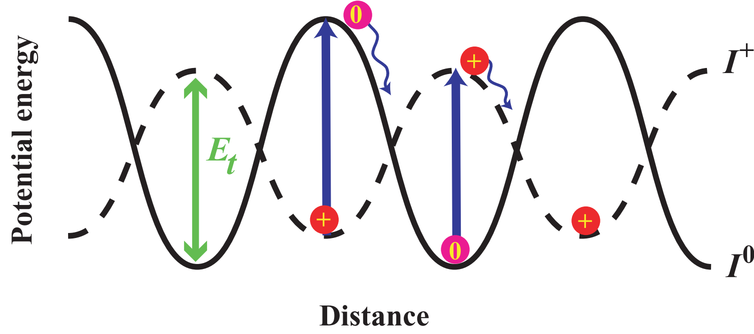


Figure 3.19: Schematic drawing of Bourgoin-Corbett type ionization enhanced diffusion (IED). Assuming a charge state transition under electron irradiation, a defect has various charge states such as $(+, 0, -$ or $2-)$. An interstitial I^+ at a stable site, for example, may capture one electron, $I^+ + e^- \rightarrow I^0$. The site is then no longer stable. The interstitial then seeks a new stable site by migration. Such successive charge state transitions under electron irradiation promote migration athermally. E_t indicates energy change of the defect by one charge state transition. From Fig.3.20, $E_t \approx 1$eV.

such migration collectively deform to stable structure, *i.e.*, more crystalline one.

From the obtainable energy by the charge state transition, number of the IED type migration, N_t, can be estimated,

$$N_t = \frac{\epsilon_d}{E_t} = 2.31[\text{times/nm}^3] \tag{3.10}$$

where are $E_t = 1.0$eV and 1eV $= 1.6102 \times 10^{-19}$J. This value is possible to convert into atomic scale value: $N_t = 0.37$[times/"unit cell"] or 0.046[times/"atom"].

Hence, these values indicate some other results:

1. The charge state transition of 2.31 times per nm^3, or 0.046 times per one atom makes disorder-induced surface stress relaxed.

2. Rate for electron capture by defects can be estimated from number of incident electron, $n_{el} \approx 5 \times 10^{21}$m^{-2} and the number of the expected charge state transition, N_t,

$$N_t/n_{el} = (2.31 \times 10^{18}[\text{m}^{-2}\text{nm}^{-1}])/(5 \times 10^{21}[\text{m}^{-2}]) \tag{3.11}$$
$$\approx 0.046\%(1/2,000). \tag{3.12}$$

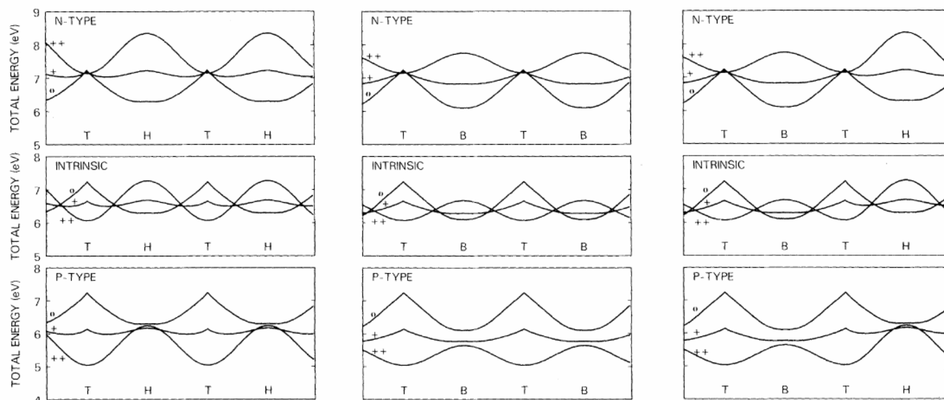


Figure 3.20: The total energy of a self-interstitial along the *TH*, *TB* and *TBTH* paths. Green's function calculations are done only at the high-symmetry sites. (From Ref.[120] and See Fig.3.21 for symbols.)

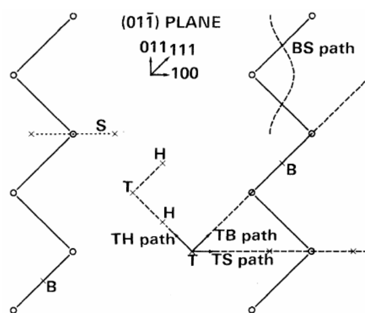


Figure 3.21: Schematic definition of various forms of the interstitial and migration paths in Si: T, tetrahedral; H, hexagonal; B, bond-centered; S, split interstitial. In the calculations, the nearest neighbors are allowed to relax. The *TBTH* path corresponds to motion in the [111] direction and this *TS* path corresponds to motion along the [100] direction. (From Ref.[120].)

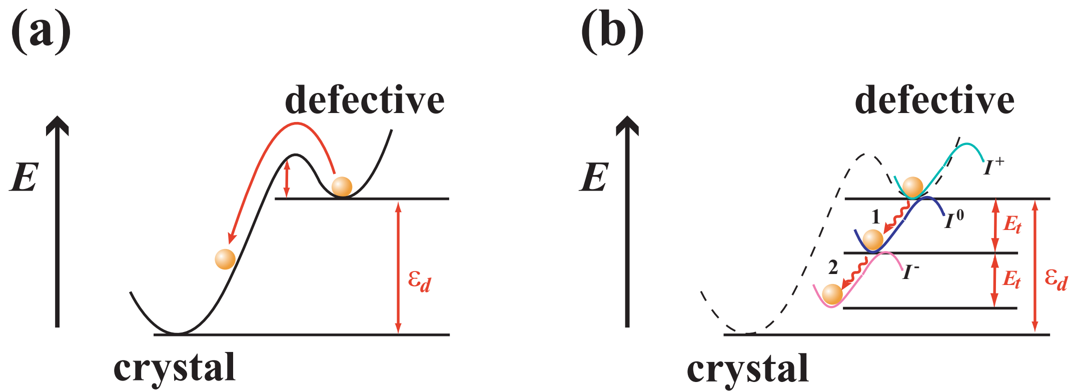


Figure 3.22: These are schematic drawing of recrystallization processes, which are two types of recrystallization methods due to (a)thermal activation path and (b)athermal migration path. Balls represent 'typical' interstitial atoms introduced by ion bombardment. For the sake of recrystallization, the interstitial atom in higher unstable energy state must move into lower stable 'crystal' energy state in the energy diagram. In case of (a), the interstitial atom moves into the 'crystal' state beyond activation barrier by obtaining additional thermal energy. On the other hand, in case of (b) which is one example of charge state transition that $I^+ + e^- \rightarrow I^0$ and then $I^0 + e^- \rightarrow I^-$ where symbols of I^+, I^0 and I^- indicate the interstitial and also their potentials around the interstitial having the charge state $+, 0$ and $-$ after Fig.3.19 and 3.20, the interstitial atom moves into 'crystal' state without passing beyond the activation barrier. By changing its own charge state and surrounding potential corrugation repeatedly, the interstitial atom migrate to 'crystal' state athermally.

3. We obtained the calculated number of defects at 0 Kelvin, n_{d0}, to be about $1.7\times10^{21}[\mathrm{m}^{-2}]$ in this bombardment condition as in Fig3.12. But number of atoms in Si crystalline in 1nm thickness, n_{Si}, is $5\times10^{19}[\mathrm{m}^{-2}]$, *i.e.*, $n_{Si} \ll n_{d0}$. This relationship between n_{Si} and n_{d0} must $n_{Si} > n_{d0}$, at least, according to requirement for reality. Then, n_{d0} was overestimated. And a part of the defects could be vanished for annealing effect at room temperature. A *real* number of defects is estimated to be $2.31\times10^{18}/q[\mathrm{m}^{-2}]$, if it is assumed that the charge state transitions of q turns make defect pair recombine for the stress relaxation. Hence, the *real* number of introduced defects at room temperature, n_d, is estimated to be $\frac{2.31\times10^{18}[\mathrm{m}^{-2}\mathrm{nm}^{-1}]/q}{1.7\times10^{21}[\mathrm{m}^{-2}\mathrm{nm}^{-1}]} = 1/(736.054q)$ times as many as n_{d0}. This consideration is consistent with common sense of 'production method of amorphous Si with ion-bombardment' requiring a number of introduced defects 100-1,000 times as many as an expected number of defects[108].

According to these results, we can conclude that the surface stress in ultrathin disordered layer including *real* defects of $2.31\times10^{18}/q[\mathrm{m}^{-2}]$ is completely and athermally relaxed by the charge state transitions of the defects under extremely low energy electron irradiation.

3.4.5 Thermal gradient effect due to further electron irradiation

We observed Fig.3.23 shows the surface stress change in case of electron irradiation of 40eV for 2000s after ion irradiation of 65 eV for 2000s. There are two noticeable points. One of the points is that the compressive surface stress appeared again after the complete relaxation of the disordered compressive surface stress during the electron irradiation of 45eV. This indicates that further electron irradiation after the complete relaxation makes the surface expanded.

Another is that the newly compressive surface stress after stopping electron irradiation vanished, *i.e.*, the surface stress re-relaxed completely. Hence, we assign the complete re-relaxation after the stopping electron irradiation to not universal dislocation of the structure, but transient modification. Similar behaviours of the two points were observed for any other electron irradiation conditions of 3.75-40eV.

This effect of further electron irradiation can be considered to be resulting from the temperature difference between front and back side of the sample, which the electron irradiation could induce. The thermal analysis will described below.

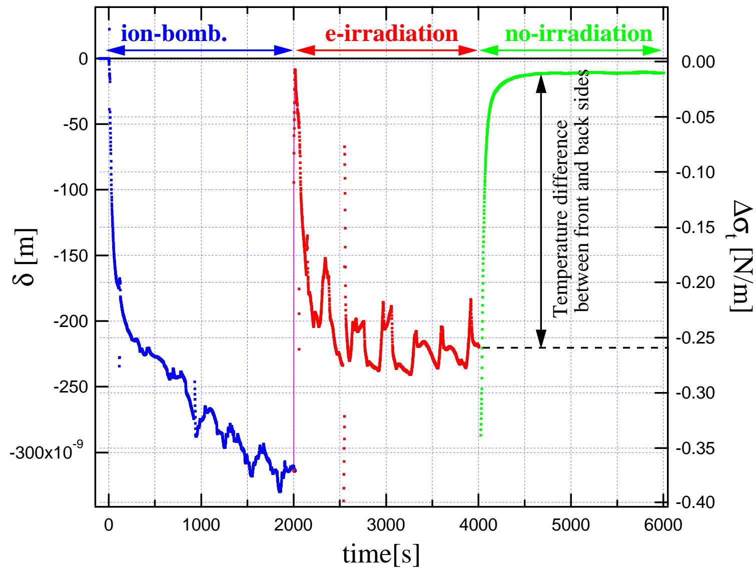


Figure 3.23: The surface stress changes before and during and after the electron irradiation at 45eV. The ion-bombardment before the electron irradiation was performed at 65eV for 2000s. During and after the electron irradiation, it is found that sample deflection of 2.09 and 2.14×10^{-7}m, respectively, due to temperature difference between the front and back side is induced.

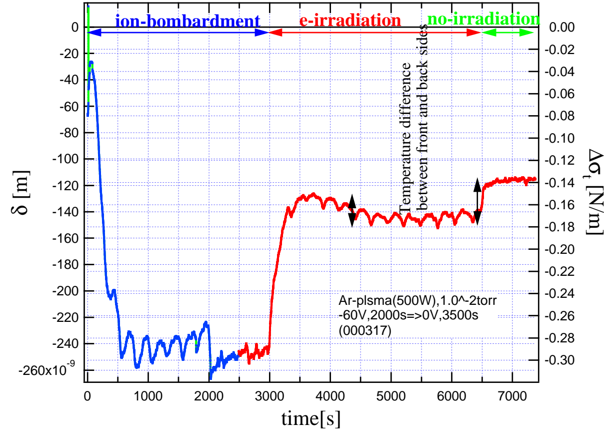


Figure 3.24: The surface stress changes before and during and after the electron irradiation at 0eV which is grounded. The ion-bombardment before the electron irradiation was performed at 65eV for 2000s. During and after the electron irradiation, it is found that sample deflection of 2.09 and 0.19×10^{-7}m, respectively, due to temperature difference is induced.

Thermal analysis of further electron irradiation

From page 99 in Ref.[115], we consider the solid bounded by two parallel planes of the region $0 < x < l$ which the ends are kept at constant temperatures v_1 and v_2 and whose initial temperature $f(x)$. In this case, we have the equations

$$\frac{\partial v}{\partial t} = \kappa \frac{\partial^2 v}{\partial x^2} \quad (0 < x < l) \tag{3.13}$$

where κ is thermal diffusion constant,

$$v = v_1, \text{ when } \ x = 0, \tag{3.14}$$

$$v = v_2, \text{ when } \ x = l, \tag{3.15}$$

and

$$v = f(x), \text{ when } \ t = 0. \tag{3.16}$$

Put

$$v = u + w \tag{3.17}$$

to reduce this to a case of steady temperature and a case where the ends are kept at zero temperature, where u and w satisfy the following equations:

$$\frac{\partial^2 u}{\partial x^2} = 0 \quad (0 < x < l), \tag{3.18}$$

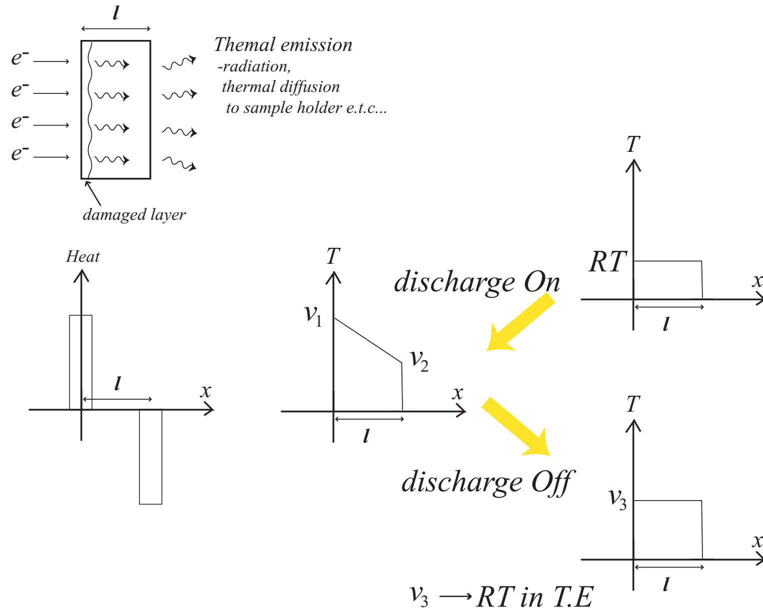


Figure 3.25: Schematic of heat balance of the sample and temperature distribution before and during and after the electron irradiation. The electron irradiation to the front side of the sample induces temperature difference between the front and back side.

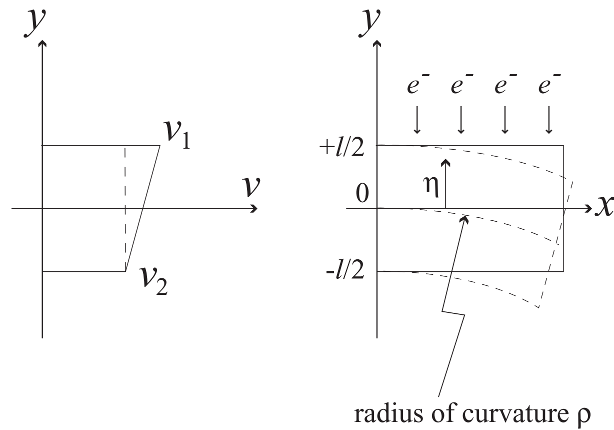


Figure 3.26: Schematic of sample deflection due to the temperature difference by electron irradiation.

$$u = u_1, \text{ when } \ x = 0, \tag{3.19}$$

$$u = u_2, \text{ when } \ x = l, \tag{3.20}$$

and

$$\frac{\partial w}{\partial t} = \kappa \frac{\partial^2 w}{\partial x^2} \quad (0 < x < l), \tag{3.21}$$

$$w = 0, \text{ when } \ x = 0 \text{and } x = l, \tag{3.22}$$

$$w = f(x) - u, \text{ when } \ t = 0. \tag{3.23}$$

In this case, we find the temperature distribution at time t,

$$v = v_1 + (v_2 - v_1)\frac{x}{l} + \frac{2}{\pi}\sum_{n=1}^{\infty} \frac{v_2 \cos n\pi - v_1}{n} \sin \frac{n\pi x}{l} e^{-\kappa n^2 \pi^2 t/l^2} \tag{3.24}$$

$$+ \frac{2}{l}\sum_{n=1}^{\infty} \sin \frac{n\pi x}{l} e^{-\kappa n^2 \pi^2 t/l^2} \int_0^l f(x') \sin \frac{n\pi x'}{l} dx'. \tag{3.25}$$

Here, we assume that the initial temperature distribution is $f(x) = v_0\delta(x)$to match this experimental condition and we can obtain the following equation ((fourth term in rhs) $\rightarrow 0$):

$$v = v_1 + (v_2 - v_1)\frac{x}{l} + \frac{2}{\pi}\sum_{n=1}^{\infty} \frac{v_2 \cos n\pi - v_1}{n} \sin \frac{n\pi x}{l} e^{-n^2 \pi^2 t/\tau}, \tag{3.26}$$

time constant to the stationary state is found to be

$$\tau \equiv \frac{l^2}{\kappa}. \tag{3.27}$$

In our microcantilever sample case, the time constant is appeared with the physical constants and dimensions($l = 2\mu\text{m}$, $\kappa_{Si} = 0.9\text{cm}^2/s$):

$$\tau_{canti} = 4.4 \times 10^{-8}\text{s} = 44\text{ns}. \tag{3.28}$$

The time is much far shorter than our experimental period ($\sim$ 1s, $t/\tau \gg 1$). Hence, the exponential part in the third term of Eq.3.26 is to be negligible. The realistic temperature distribution is

$$v = v_1 + (v_2 - v_1)\frac{x}{l}. \tag{3.29}$$

In this case of Fig.3.26, the deflection of the sample having the temperature gradient of $\Delta v = (v_1 - v_2)\frac{y}{l}$ was given as follows.

$$\Delta v = \frac{v_1 - v_2}{l}\eta \equiv \beta\eta \tag{3.30}$$

and, from the feature and the stress, the strain ϵ are given as

$$\begin{cases} \epsilon = \frac{\eta}{\rho} \\ \epsilon = \frac{\sigma}{E/(1-\nu)} + \alpha \Delta T \end{cases} \tag{3.31}$$

where α is thermal expansion coefficient and the temperature difference $\Delta T = v_1 - v_2$. And, the balance equations of force and moment are described in

$$N_x = b \int \sigma dA = 0, \tag{3.32}$$

$$M_x = b \int \sigma \eta dA = 0 \tag{3.33}$$

where b is width of substrate. Combining with these equations,

$$\frac{1}{3}\left(\frac{E}{1-\nu}\right)\left(\frac{1}{\rho} - \alpha\beta\right)\left[\eta^3\right]_{-l/2}^{+l/2}. \tag{3.34}$$

Hence,

$$\frac{1}{\rho} = \alpha\beta = 1.25\Delta T \tag{3.35}$$

where $l = 2.0\mu$m and $\alpha_{Si} = 2.5 \times 10^{-6}[^\circ\mathrm{C}^{-1}]$. Furthermore, using the relation between the radius of curvature ρ and the sample deflection δ of $1/2\rho = \delta/l^2$,

$$\delta = 1.266 \times 10^{-7}\Delta T[\mathrm{m}]. \tag{3.36}$$

From the relation in Eq.3.36, the sample deflection of $2.09 \sim 2.14 \times 10^{-7}$m in Fig.3.23 which could be due to resulting from the temperature difference between front and back side of the sample corresponds to the temperature difference of 1.65∼1.69°C, in the case after the electron irradiation at 40eV for the bombarded surface. On the non-bombarded surface, the electron irradiation at 45eV induced the sample deflection of 2.54×10^{-7}m and it corresponds to 2.00°C. On the electron irradiation at 0V after the ion-bombardment at 65eV for 2000s, the sample deflection of $0.17 \sim 0.19 \times 10^{-7}$m was induced and corresponding to $0.13 \sim 0.15$°C. Hence, according to these results and Fig.3.17, the temperature difference between front and back side of the sample is considered to be, empirically, the about 1/8 times as high as the total bulk temperature rise of the sample (about 10°C for electron irradiation at 45eV).

3.4.6 Electron irradiation in oxidation atmosphere

Electron irradiation in oxidation atmosphere has same effects with only electron irradiation as shown later. The electron irradiation in oxidation atmosphere was performed following ion bombardment at room temperature

using an oxygen plasma in an anodic oxidation condition during attracting electrons. The oxygen plasma was generated by rf discharge of oxygen gas (1.3Pa, 99.99%) at power 500 W, and a bias of +45V was applied to the sample. We observed the surface stress evolutions during applying positive bias to the samples after the Ar ion bombardment, which enable to oxide the Si surface during attracting lots of electrons. On the stress evolution after the ion bombardment, it was found for the ion-bombarded surfaces that two stages are developed in the stress evolution up to about 300 s with different slopes of stress change. Note that the absolute values of σ_i after the bombardment and σ_{e+oxid} of the first stage during the oxidation agree well with each other. The result suggests that the first stage of the oxidation abrupt increase corresponds to complete relaxation of the compressive stress induced by ion bombardment, and the second stage slow increase to the growth of the oxide free from the disorder-induced stress.

Fig.3.27 is one example of surface disordering about anodic oxidation. Figure 3.28 plots values of surface stress release at saturation $\Delta\sigma_e$ under the anodic oxidation and also electron irradiation of Fig.3.14 at a fixed energy for various disorder-induced stresses $\Delta\sigma_i$ (see Fig.3.13 and 3.27 for the symbols). Whether on less or more disordered surfaces, similar complete relaxation was observed. Linear curve fitting to these points yields the simple linear function $\delta_e = (1.06\pm0.07)\delta_i + (-18.024\pm20.4)$ nm within errors. Complete relaxation was independent of the degree of disorder, and also whether being oxygen species or not. Hence, these results indicate that the complete relaxation of the surface stress would be due to only electrons, *i.e.*, the oxygen species have no contributions to the stress relaxation. Furthermore, it implies that the oxidation under anodic condition could bring structure relaxation to some ordered structure, nevertheless proceeding the oxide formation.

3.5 Further work

Recently, non-thermal diffusion induced by photon illumination has been reported for Si surfaces. The study also reveals significant non-thermal effects that appear to be mediated by charged vacancy of the surface[121]. But no information was given on these stabilization or relaxation of the crystalline structure resulting from collective motion of atoms, nevertheless they suggested similar diffusion of nonthermal diffusion.

It has been reported that electron irradiation but at higher incident energy 90-2,000eV and a higher electron dose ($\sim$ 20 times as many as ours) causes surface defects on clean Si(100) and Si(111) surfaces with a STM[100]. Needless to say, they are different from ours about not only experimental

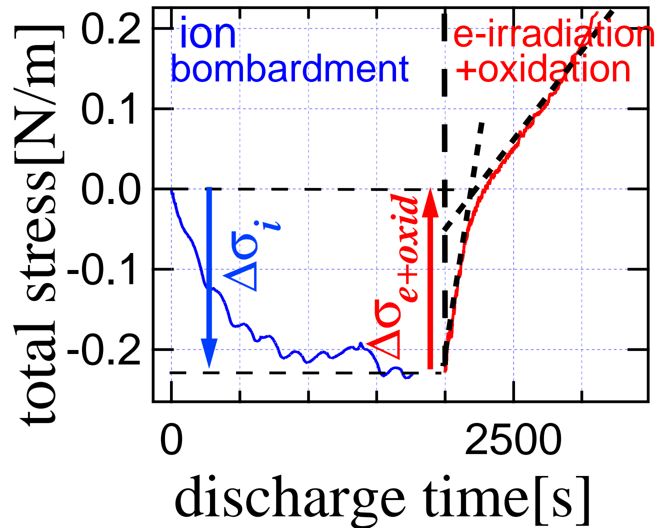


Figure 3.27: The surface stress evolutions during the ion-bombardment and the following anodic oxidation in an oxygen plasma. The electron irradiation in oxidation atmosphere was performed following ion bombardment at room temperature using an oxygen plasma in an anodic oxidation condition. On the stress evolution after the ion bombardment, it was found for the ion-bombarded surfaces that two stages are developed in the stress evolution up to about 300 s with different slopes of stress change. The absolute values of σ_i after the bombardment and σ_{e+oxid} of the first stage during the oxidation agree well with each other. Upon oxidizing the bombarded Si surfaces, the disorder-induced compressive stress became relaxed completely.

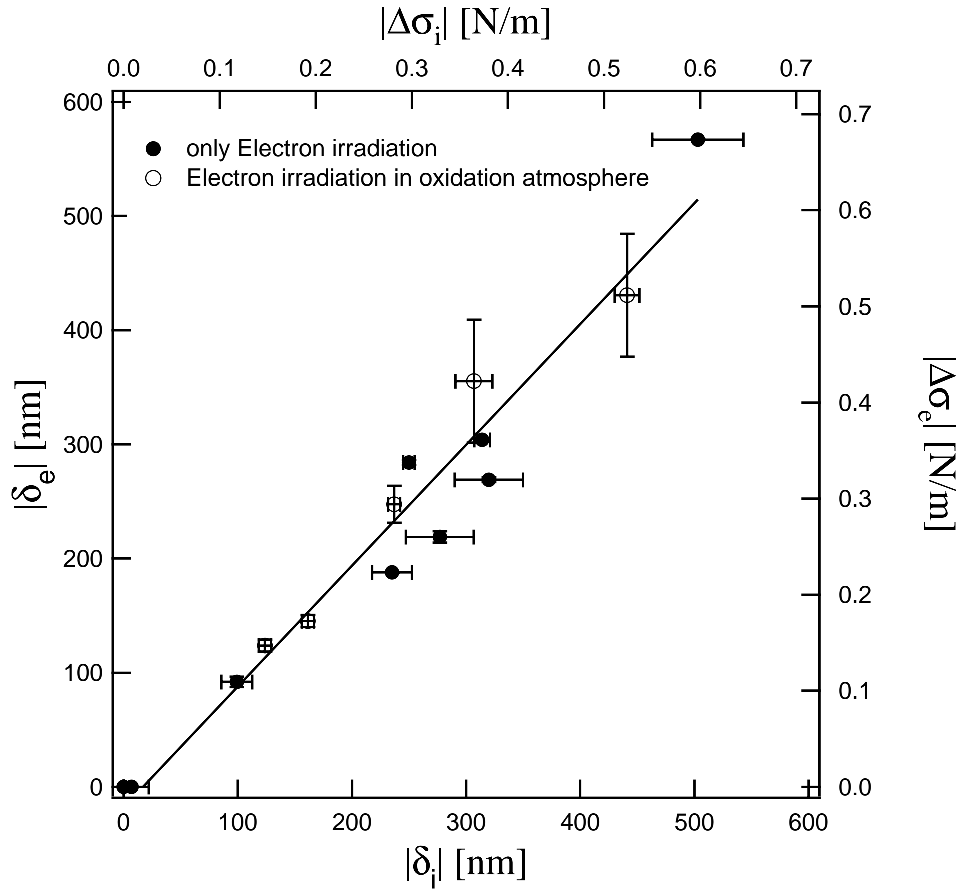


Figure 3.28: Stress relaxation under anodic oxidation (open circle) and electron irradiation (close circle) of Fig.3.14 versus degree of disorder in the surface. Horizontal axis shows disorder-induced stress and deflection by ion-bombardment, $\Delta\sigma_i$ and δ_i, and vertical axis is stress and deflection at saturation induced by the following the anodic oxidation and the electron irradiation at fixed electron energy of 40eV, $\Delta\sigma_e$ and δ_e.

methods of surface stress measurement and STM, but also initial condition of disordered surface by ion bombardment and clean Si surfaces. It is, however, interesting that the opposite behaviour was observed, *i.e.*, restoration versus destruction, depending on whether the electron incident energy is between 3.75 and 40.0eV or 90 and 2,000eV.

In next chapter, to give answers to the interesting problem, we will discuss detailed atomic scale surface structures with STM observation on the growth and the relaxation of the surface stress.

Chapter 4

Microscopic structure observed with a Scanning Tunneling Microscopy

4.1 Abstract of this chapter

A STM observation of electron irradiated Si(100) surface after ion-bombardment to be performed for cross-check of restoration of the disordered surface presents re-ordering of atomic level, *i.e.*, recrystallization. The fine structure of re-ordered surface was different from the structure by thermal annealing. In general, the structure generated by thermal annealing has 2×1 symmetry due to Si dimers. The fine structure of reordered surface, in contrast, were close to positions of a 'bulk terminated' 1×1 structure, but with a slight pairing, while dimers retain a 2×1 configuration. It suggests that non-thermal genesis of new surface structure was done by very low energy electron irradiation. In demonstrating that surfaces can have the capacity of self-healing, these results suggest new opportunities for atomic scale surface engineering.

4.2 Introduction

Electrons sometimes behaves against our expectations. In general, we believed that higher energetic particles can penetrate to materials inside deeper. The electron having extremely lower kinetic energy than 50eV is, however, penetrates to fabulous depth, because it cannot lose its own energy via any energy loss process. The critical energy depends on whether excite plasmons or not, whose values are 16.4-16.9eV and 11.6-12.0eV for bulk and surface of Si[122]. These are demonstrated better in Fig.4.1.

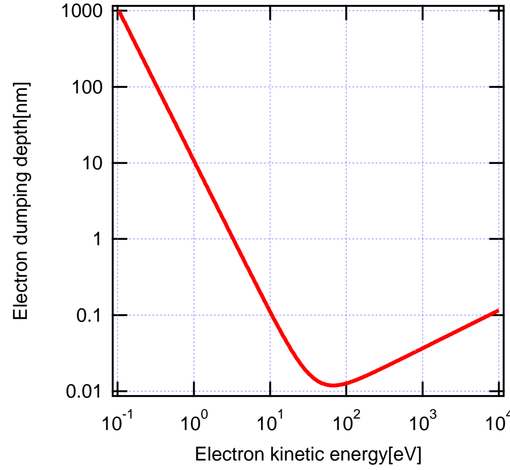


Figure 4.1: Electron penetration depth as a function of kinetic energy (Redrawing of Fig.3.10). This curve is expressed with $\lambda_{dump} = 538aE^{-2} + 0.41a^{3/2}E^{1/2}$[nm]; $a^3 = (10^{21}A)/(\rho N_A)$[nm^3]. Higher energetic electrons can penetrate to materials inside deeper, which has higher than about 50eV. On the contrary, less energetic electrons can also penetrate to materials. This threshold is given by whether excite plasmons or not.

The higher energy electron irradiation than 90eV induces surface defects of Si surfaces. Nakayama and Weaver [123] reported that structural modifications that involved atom desorption and displacement following 'mild' irradiation by electrons of 90-2000 eV with STM studies. For Si(111)-7×7 as in Fig.4.2, adatom layer vacancies increased monotonically with incident energy. For Si(100)-2×1 as in Fig.4.3, irradiation produced dimer vacancies, and ad-dimers as Si atoms transferred to the terrace. They argued that the modification processes are tied to the energy distribution of electron-hole and electron attachment states achieved by inelastic cascade scattering.

Their 'mild' electron irradiation condition is more violent than our irradiation condition; higher incident energy 90-2000eV and a higher electron dose $\sim$ 20 times as many as ours. In the chapter3, we showed, using a highly sensitive cantilever technique, that very low energy electron irradiation is quite effective in causing the recovery of defect structures into ordered Si surfaces through the inter-relationship between surface stress and defect disorder, macroscopically. Then, it is interesting that the opposite behaviour was observed, *i.e.*, restoration versus destruction, depending on whether the electron incident energy is between 3.75 and 40.0eV or 90 and 2000eV. Hence, in this chapter, we will investigate, with STM, contributions of the surface

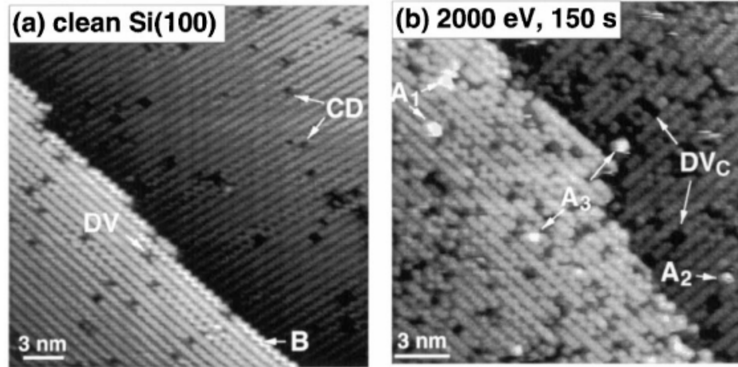


Figure 4.2: (a) Filled state image of clean Si(100)-2×1 showing dimer vacancies (DV) and c-type defects (CD), and a row of asymmetric buckled dimers at a step (B). (sample -2.0V, 0.2nA). (b) Image after 2000eV irradiation for 150 s showing dimer vacancy complexes (DV$_c$). Ad-dimers are labeled A_1 if their axis is parallel to the dimer row, A_2 if they lie perpendicular to the dimer row, and A_3 if they reside between dimer rows. From Ref. [123].

morphologies and structures on the surface stress of surface restoration by extremely low energy electron irradiation, while searching traces of differences of the 'restoration versus destruction'. If aims of the surface stress relaxation by very low energy electron irradiation to be athermal process are true, some traces of the athermallity on surface structure could be observable.

4.3 Experimental

4.3.1 Scanning Tunneling Microscopy observation of Plasma-modified surface

For the sake of the surface confirmation, we used a Scanning Tunneling Microscopy (STM). This is one of most powerful tools to understand the surface morphologies and structures, because it has a resolution of atomic scale and also allows us to observe the surface not in reciprocal space, but in *real space* according to their charge densities.

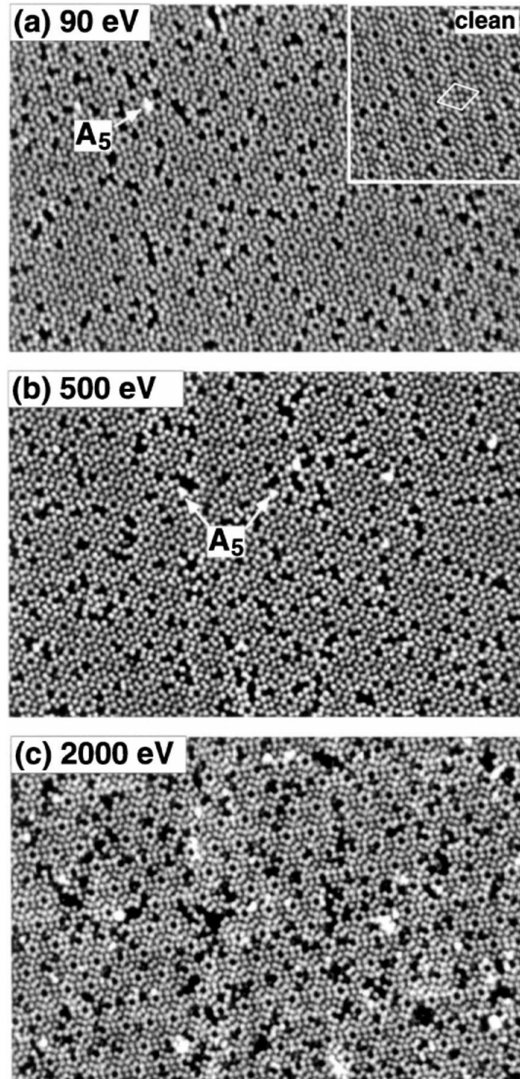


Figure 4.3: Images after irradiation of Si(111)-7×7 by electrons of primary energy of (a) 90eV, (b) 500eV, and (c) 2000eV. Vacancies in the adatom layer appear as dark features. The development of these vacancies is largely due to Si atom desorption though feature like A_5 represent Si atoms trapped at a rest atom dangling bond. Desorption is random when the vacancy concentration is low. All images are $33 \times 50\text{nm}^2$; sample bias 1.7V, tunneling current 0.4nA. From Ref. [123].

4.3.2 Pretreatment and features of clean Si(100) surfaces

STM can directly observe the change of the surface morphologies and structures during which the surface stress occurs, however to achieve this not disturbing the reconstructed surface of Si(100), there is an important point to keep in mind about surface contamination. As already confirmed by STM, C-type defects that may be related to contamination influence on surface structure as inducing buckled dimers running from the C-type defect, while A, B-type have no influence on the structure as in Fig.4.4.

Nishizawa *et al.*[124] will report that the origin of the C-type defect is residual water even under a UHV condition, therefore we have carefully baked the STM chamber, the Si sample and gas line. The Si(001) sample was cleaned by an H_2SO_4:H_2O_2 after removal of sacrificial oxidation surface and then it was prebaked in a UHV at around 600°C overnight. A clean surface was obtained by flashing the sample around 1050°C. The cleanliness of the initial surface was checked, and the absence of C-type defects was confirmed as in Fig.4.5.

4.3.3 Electron and ion irradiation methods with DC plasma

The samples for STM measurements were exposed by Ar-plasmas generated by DC glow discharges as shown in subsectionC.7.1. Distance between the sample and DC grids was set about 15cm and the sample surface was set so as not to look towards DC grids to avoid contamination of sputtered materials in Fig.4.6. And actual and typical generated plasmas are shown in Fig.4.7 and Fig.4.8. This discharge method enables to make similar conditions of the ion bombardment and electron irradiation in the case of the surface stress measurements in the chapter3. To compare the characteristics of DC glow plasma with those of the RF plasma during the surface stress measurements, the plasmas of the $I-V$ characteristics were also diagnosed by the Langmuir probe in Fig.4.9 as in subsection C.6. The DC plasma of Ar gas was discharged at pressure of 10Pa, current between grids of 20mA and grid voltage of 400V. The sample temperature during the plasma exposures was at room temperature (300K).

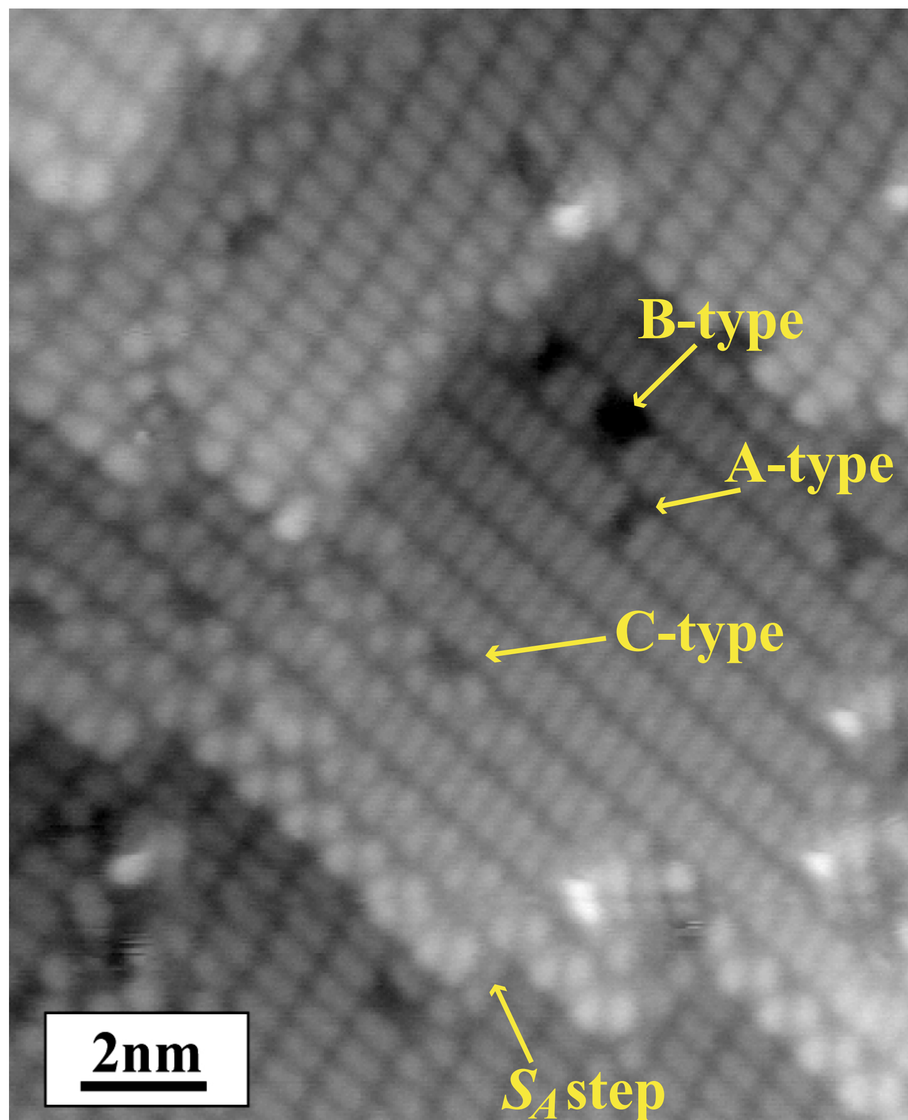


Figure 4.4: A occupied state STM image of A, B and C-type defects and their surrounding structures. The C-type defect induce buckled dimers running from the C-type defect, while A, B-type have no influence on the structure. And also, buckled dimers exist at S_A type step edge.

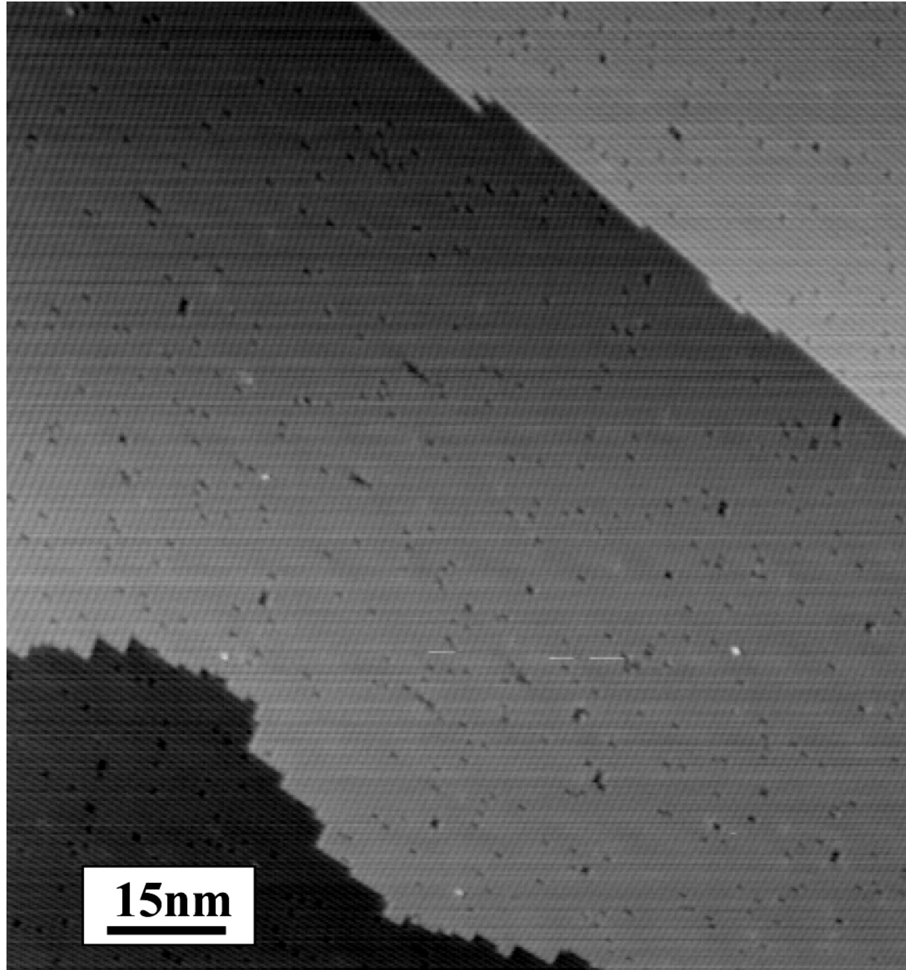


Figure 4.5: Careful cleaned Si(100) surface. Defects exits only a few percent, *i.e.*, it's negligibly small amount.

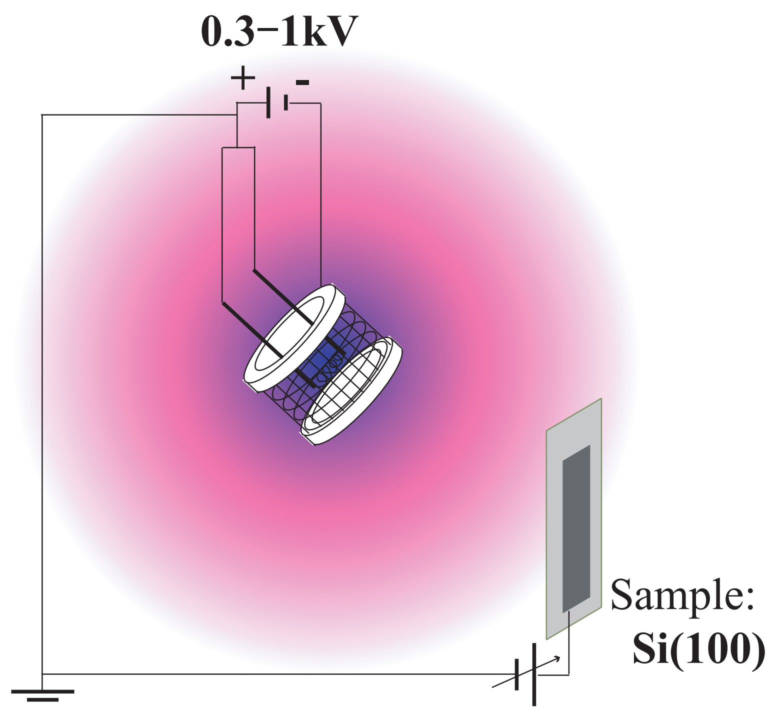


Figure 4.6: A schematic drawing of DC plasma source and sample. The configuration of this instrument is composed with a W-filament and a ring-shaped SUS mesh which diameter is 2cm. An electric potential difference to be about 400V between the ring-shaped mesh and the W-filament is given by a DC power supply and current through the grids flows at about 10-50 mA. While the current flows stably, the plasma is maintained.

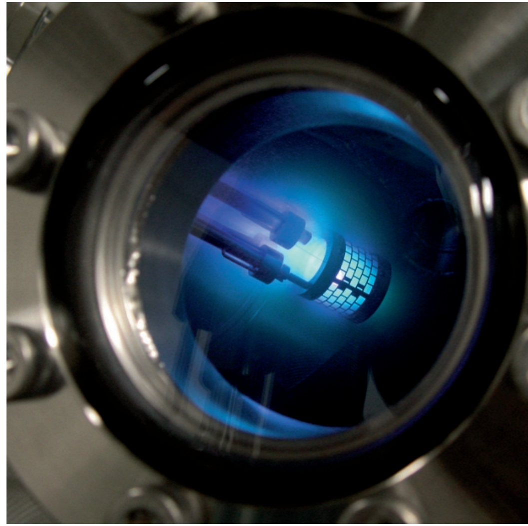


Figure 4.7: A DC discharge Plasma (cold cathod type).

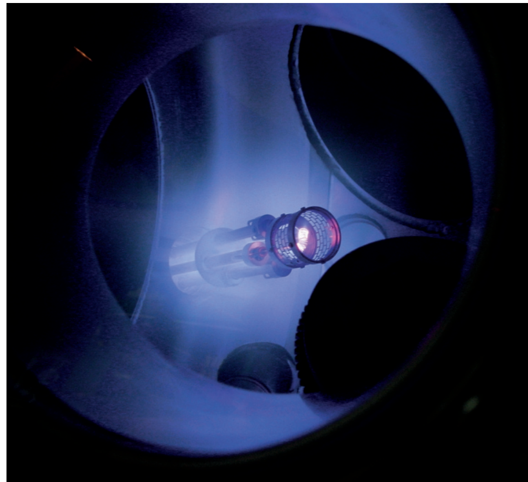


Figure 4.8: Similar DC discharged plasma of hot cathod type. This instrument differs, on lightening W-filament, from Fig.4.7. The lightening filament generates free electron to ionize lots of atoms and promote to maintain the plasma.

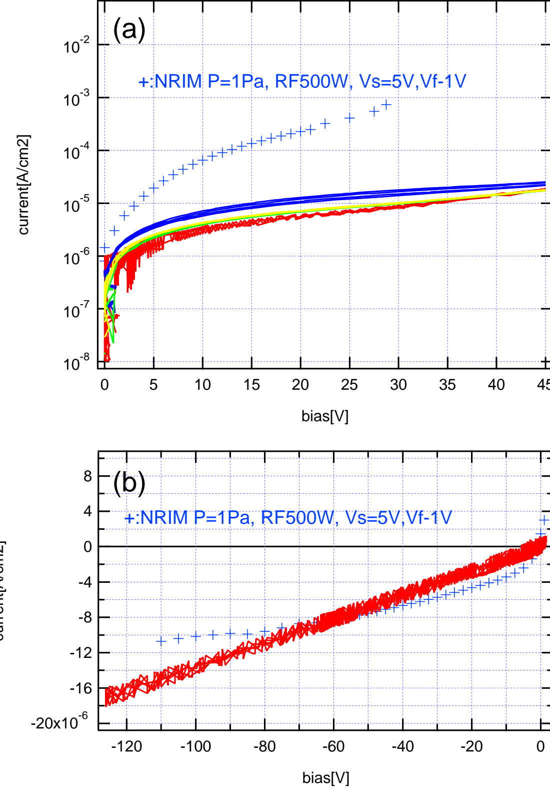


Figure 4.9: Comparison the characteristic of DC glow plasma with that of the RF plasma during the surface stress measurements. $I-V$ characteristic of sample holder as a single Langmuir probe: mainly (a) electron current region (positive bias); (b) ion current saturation region (negative bias). The sign '+' indicates $I-V$ characteristic of the RF discharge plasma and other lines indicate that of the DC discharge. The electron temperature of this DC plasma of Ar gas at pressure of 10Pa, current between grids of 20mA and grid voltage of 400V, the electron density, the space potential (V_p) and the floating potential (V_f) were 3eV, 6.5×10^5cm^{-3} +5V and -1V at the sample position, respectively.

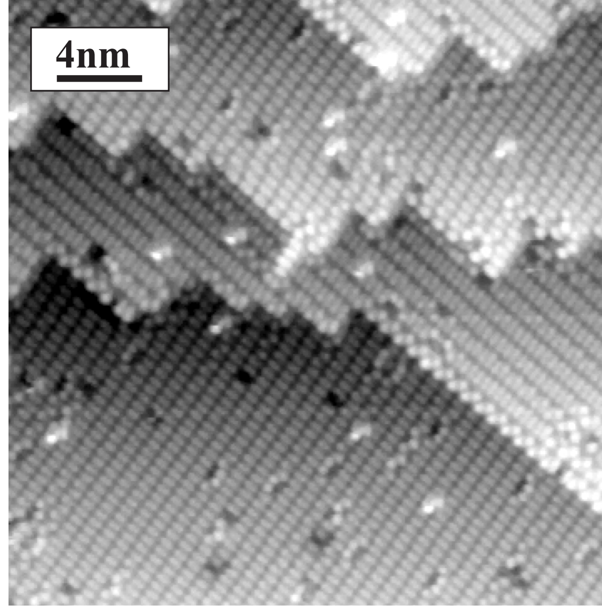


Figure 4.10: typical clean Si(100) surface: $V_s = -1.5\text{V}$, $I = 0.10\text{nA}$.

4.4 Atomic level identification of surface restoration by electron irradiation

Fig.4.10 shows typical Si(100)-2×1 surface prepared for this STM experiment.

4.4.1 Modified surfaces of Si(100) by ion bombardment

We performed ion-bombardment to these surfaces using DC-plasma during appling negative bias of -60V to the samples. Fig.4.12 and 4.14 show the ion-bombarded surfaces of Si(100). It seems that degree of disorder, or number of defects, of Fig.4.12 is less from its appearance, while total detected ion current of Fig.4.12 is as many as that of Fig.4.14.

Less disordered case of Si(100) surface

On less disordered surface case of Fig.4.12, at first, it has some different features from those of clean Si(100) surface that

1. 2×1 dimer rows are partially destroyed and they are cut and bent,

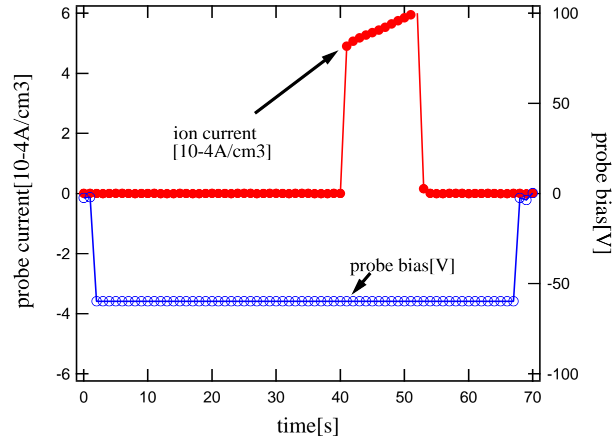


Figure 4.11: Ion current of Fig.4.12. The ion current was measured directly with real time by using sample holder as a probe. What the current was detected is equivalent to Ar ions atacking the Si(100) surface, *i.e.*, it means that ion bombardment is ensured.

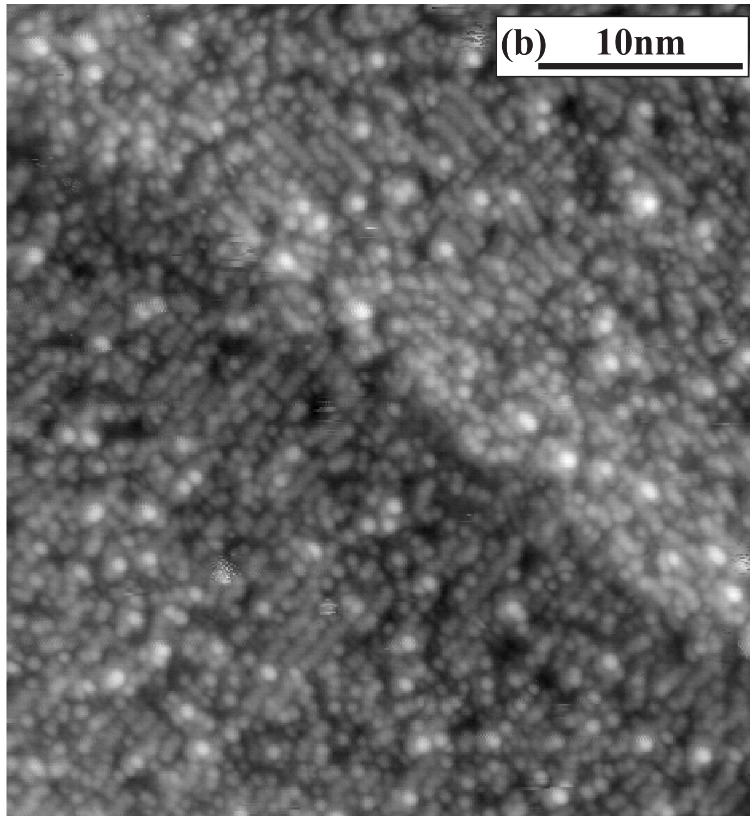


Figure 4.12: Ion bombarded Si(100) surface at 65eV for 12s.: $V_s = -1.5$V, $I = 0.10$nA. The integrated ion-bombarded current is 66.2×10^{-4}A·s/cm^2.

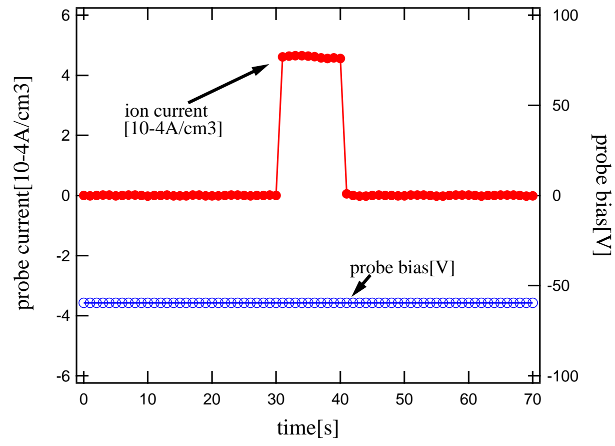


Figure 4.13: Ion current of Fig.4.14.

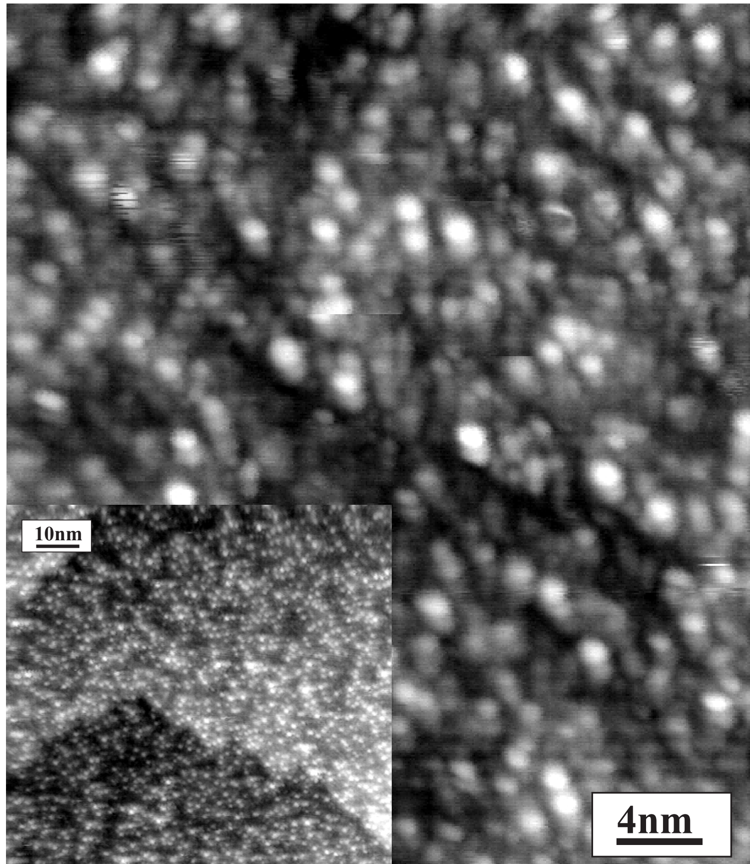


Figure 4.14: Ion bombarded Si(100) surface at 65eV for 10s.: $V_s = -1.5$V, $I = 0.10$nA. The integrated ion-bombarded current is 46.1×10^{-4}A·s/cm^2.

2. bright points are observed at intervals,
3. structure of monoatomic height step is roughly kept.

Feature 1 could result from energetic implanted Ar ions. Since the Ar ions have higher energy than displacement energy, Si-Si bonds on the surface should be cut and/or rebonded with another Si atom by the ions, inducing the surface defects. Substrate atoms could also be moved another sites by the Ar ions and dimer rows on the top surface are seen as bending.

Feature 2 could be due to surface dangling bonds. The highly energetic Ar ions having kinetic energy of 65eV would be able to cutting bonds of Si-Si. From the result, some of rest unoccupied bonds may leave alone, while the others might bond nearby similar unoccupied bonds to release energy. The bond which could not bond, i.e., dangling bond has electron, so that this site brights under this occupied state.

Feature 3 indicates that surface morphology and substrate structure is roughly kept. If structures of substrate crystal were absolutely broken by ion-bombardment, the monoatomic height step should not be conserved. Therefore, the ion-bombardment is able to do only short range transfer of Si atom from original site to another site, which does not exchange each other, while the surface is expected to be amorphous in this bombardment condition.

Harder disordered case of Si(100) surface

Next, in case of Fig.4.14, feature 1 of Fig.4.12 is not identified at all, while Feature 2 and 3 is identified. But due to more disorder, bright points form groups being sparsely. The monoatomic height step feature is also conserved. It also indicates that the ion-bombardment is able to do only short range transfer of Si atom.

Understanding from these STM images, it would be true that decay of intrinsic stress due to surface reconstruction is one of origins of the compressive surface stress by ion-bombardment, because the feature 1 is observed. Furthermore, it could not be false that another origin is considered to be volume expansion of defective layer, because the features of 1 and 3 are observed. Hence, our expected origins of disorder induced surface stress in discussions of subsection 3.4.1 were consistent with the detailed surface structures identified by our STM imaging of disordered surfaces.

4.4.2 Modification of the disordered surface by low energy electron irradiation

Let us discuss structures of electron irradiated surface of Si(100) after ion bombardment. Fig.4.16 shows occupied state image of electron irradiated Si(100) surface at 25eV for 11s after ion bombardment at 65eV for 7s (See also Fig.4.17 for guide to easily see features around the monoatomic step.). Surprisingly, upon the whole, disordered surface by ion bombardment restored to the 'highly' *ordered structure*. Some observed bright points would be due to being dangling bonds or Si adatoms, while they are not distinguished at the present time.

It is ensured that each ions and electrons strike the surface by measured ion and electron currents in Fig.4.15. Hence, the disordered surface by ion-bombardment were restored by very low energy electron irradiation. This surprising result of surface re-ordering is consistent with the complete surface stress relaxation by the electron irradiation in chapter3, *i.e.*, our expected the (athermal) recrystallization under electron irradiation in subsection3.4.3 must be true. Furthermore, it was understood that the restoration, *i.e.*, re-ordering phenomena influenced not only top surface structure and also substrate periodicity, because of the 'highly' *ordered structure*.

The electron irradiated surface structure resembles dimerized 2×1 surface generated (annealed) at high temperature well. This surface should be compared with clean Si(100) surface of Fig.4.10. At first glance at the 'highly' *ordered* surface, similar structure of dimer row is identified. If this restoring process results from athermal process, some traces of the athermallity on surface structure must be observable.

Obtained images by STM (Fig.4.10, 4.12, 4.14, 4.16) were converted into diffraction images in reciprocal space by FFT (fast Fourier transformation) corresponding to LEED (low energy electron diffraction) patterns in Fig.4.18; 'virtual' LEED. Fig.4.18a shows the LEED pattern of clean Si(100). The pattern indicated 1×1 and 2×1 spots due to dimer row well. In Fig.4.18c, both the 1×1 and 2×1 spots were disappered by ion-bombardment, while only the 2×1 was remained in Fig.4.18b. This remainder of the 2×1 suggests that Si dimerization was conserved in an intermediate state even during being bent or cut itself.

In Fig.4.18d, both the 1×1 and 2×1 spots were restored absolutely. Because of this restoration of the 'virtual' LEED pattern and the occupied image of STM in Fig.4.16, we can argue that the disordered surface structure by ion bombardment restores to the 'highly' *ordered* surface structure by low energy electron irradiation.

But detailed comparisons of the fine structure have not been made. As

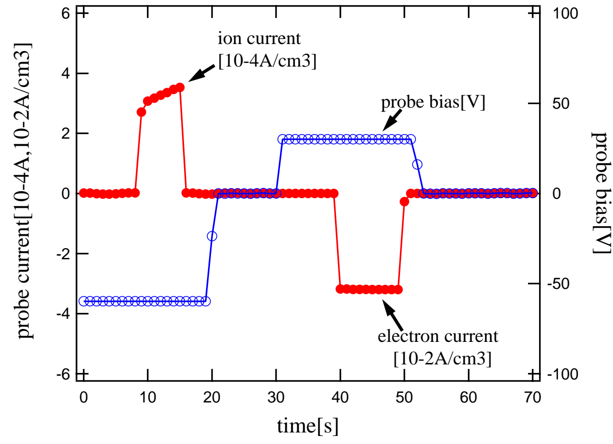


Figure 4.15: Ion and electron current of Fig.4.16.

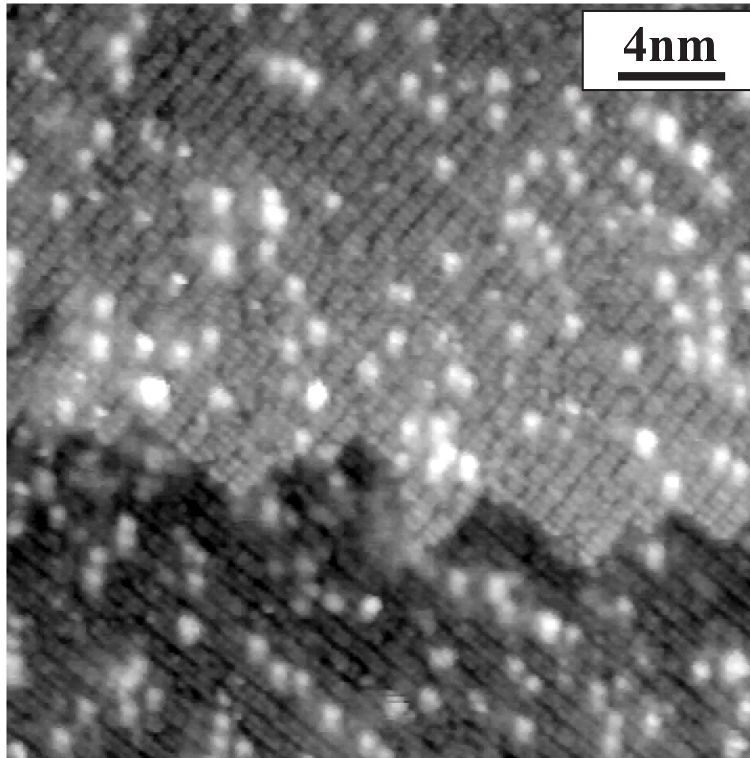


Figure 4.16: Electron irradiated Si(100) surface at 25eV for 11s. after ion bombardment at 65eV for 7s.: $V_s = -1.5$V, $I = 0.10$nA. The integrated ion-bombarded current and irradiated electron current are 22.6×10^{-4}A·s/cm^2 and -32.2×10^{-2}A·s/cm^2 respectively.

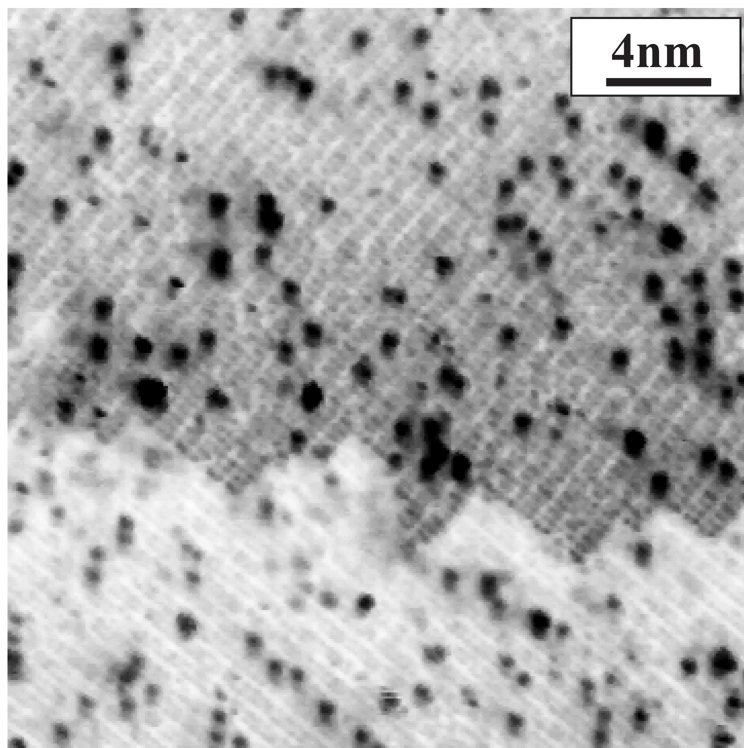


Figure 4.17: The reversed image of Fig.4.16 for guide to easily see features around the monoatomic step.

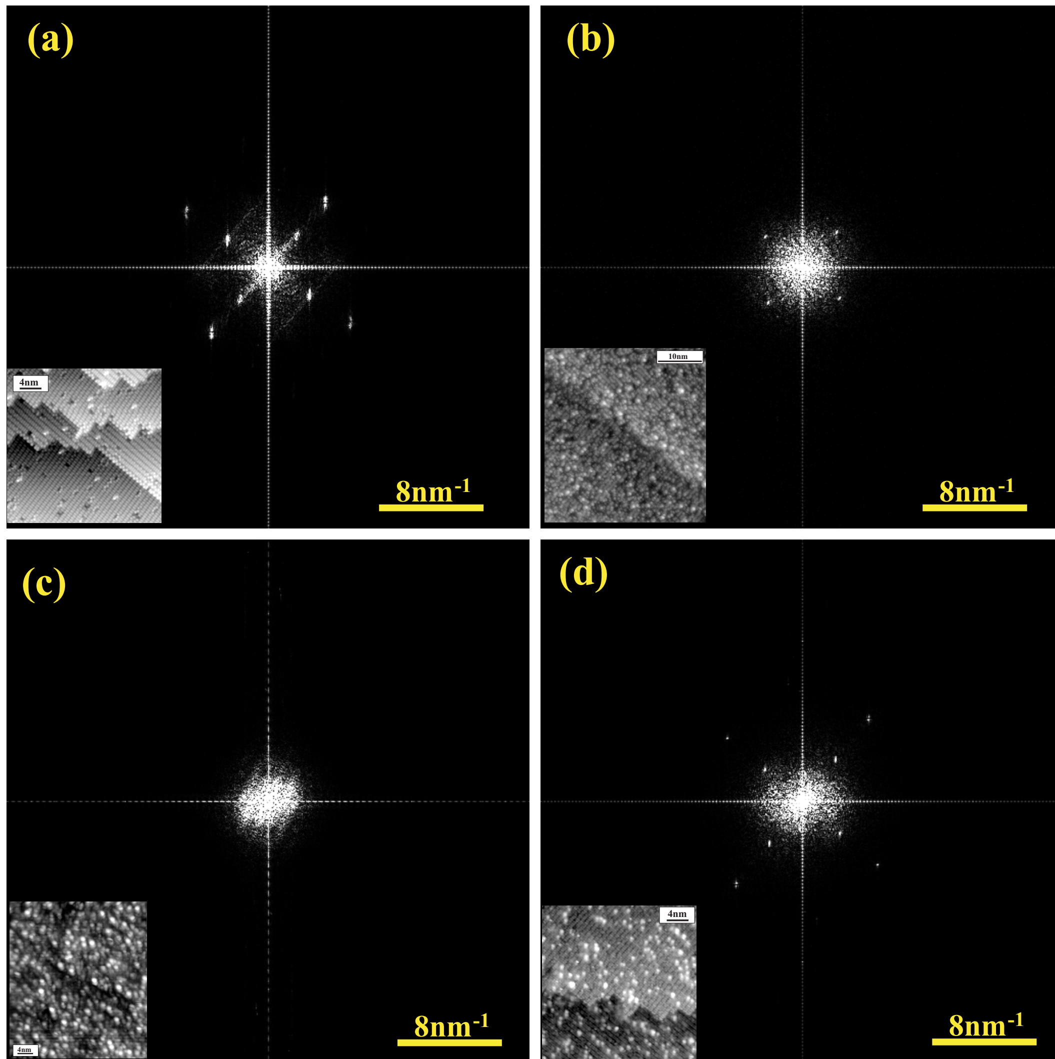


Figure 4.18: 'Virtual' LEED by FFT (fast Fourier transformation). Obtained images by STM were converted into diffraction images in reciprocal space by FFT corresponding to the LEED (low energy electron diffraction) patterns, which is shown in (a), (b), (c) and (d) for Fig.4.10, 4.12, 4.14 and 4.16, respectively.

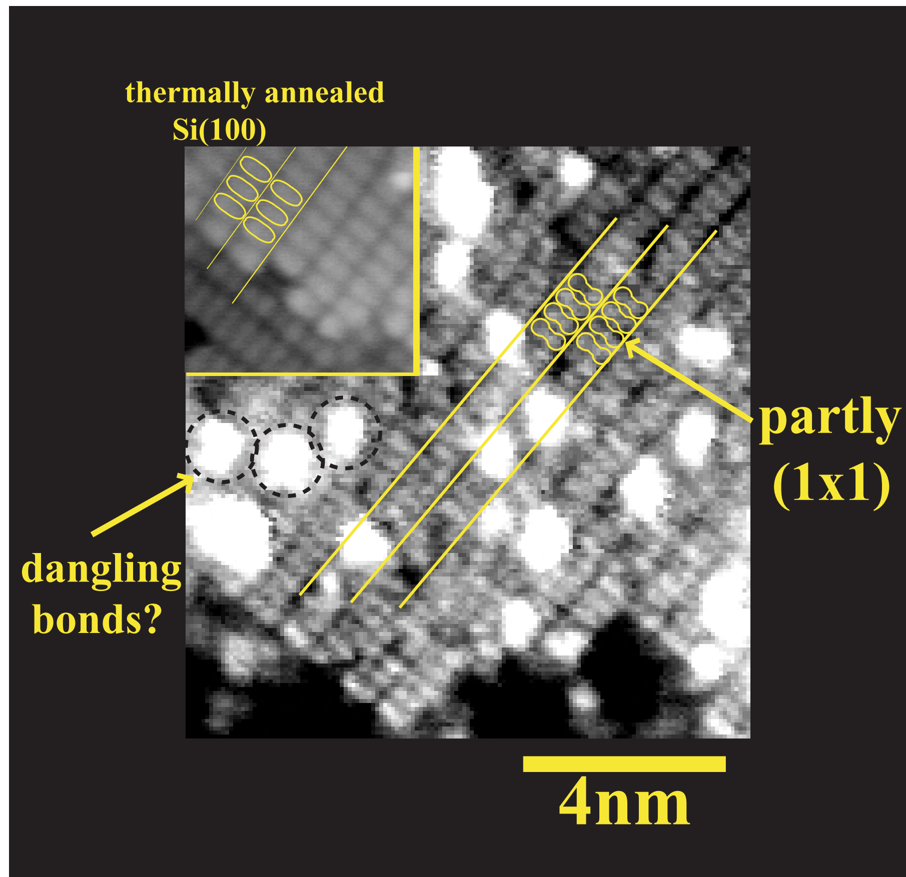


Figure 4.19: Expanded image of Fig.4.16 electron irradiated Si(100) surface after ion bombardment: $V_s = -1.5$V, $I = 0.10$nA. An inset is typical clean Si(100) surface thermally annealed at high temperature. The inset shows 'the 2×1 symmetry' due to the Si dimers. It is well known that, if the surface is dimerized, the dimer atoms cannot be resolved under these conditions as in the inset. On the contrary, the surface after electron irradiation shows that most of atoms are pairing, while our measurements show slight resolved. Hence, periodicity of the surface was partly translated into 1×1 symmetry from 2×1.

seen in Fig.4.19, there are a little difference between thermally annealed Si(100) surface and restored Si(100) surface by low energy electron irradiation. In general, Si dimers of the thermally annealed surfaces appear as elliptically shaped protrusions with minor axes aligned along the dimer row direction, i.e., perpendicular to the dimer bond direction. It is believed that dimerization involves recombining the four dangling sp^3 electrons of two surface atoms into occupied σ and π orbitals[125, 126] accompanied by contraction of the interatom spacing. All of the above images represent the occupied states since they were measured using -1.5 V sample bias with 0.1nA of tunneling current. It is well known that, if the surface is dimerized, the dimer atoms cannot be resolved under these conditions. In contrast, while most of atoms are pairing on the restored surface, our measurements show *slightly resolved*. Hence, periodicity of the surface was partly translated into 1×1 from 2×1.

A similar image of the slightly pairing 2×1 surface has been reported by Xu *et al*[127]. They performed STM observation of laser-irradiated Si(100) surfaces shows that the dimerized outer-most layer can be selectively removed by a pulsed Nd:YAG laser with a fluence below the melt threshold as in Fig.4.20. The atoms in the laser-uncovered second layer are close to positions of a 'bulk terminated' 1×1 structure, but with a slight pairing, while dimers retain a 2×1 configuration in the first layer.

They believe that the structure observed in the uncovered layer of 1×1 is a metastable state stabilized by either the presence of the unremoved top-layer Si atoms or the absence of vacancies in the uncovered layer. Based on the fact that the 2×1 dimerized structure is recovered by thermal annealing or by high fluence of laser irradiation, there is an *energy barrier* for transition from the laser-induced to the dimerized 2×1 structures. The barrier may depend on the fraction and distribution of vacancies produced by the laser irradiation. The atoms in the uncovered layer do not rebond after small fractions of surface dimers are removed by laser, in contrast to Pandey's prediction.[128] As pointed out by Roberts and Needs[129], rebonding of two atoms beneath a dimer vacancy is inhibited by increased lattice strain energy. For a small fraction of dimers removed by laser irradiation, the strain may be one of factors that prevent such rebonding. According to their interpretation, our restored Si(100) surface by low energy electron irradiation also could be similar metastable state, or athermal state not to be thermally annealed.

Furthermore, the traces of the athermal restoration were identified at steps and around C-type defects. It is known that on thermally annealed Si(100) surface buckled dimer features are observed at the S_A-steps and around the C-type defects as in Fig.4.4. In addition, our restored Si(100) surface has 'no buckled dimer features' around C-type defects and at the

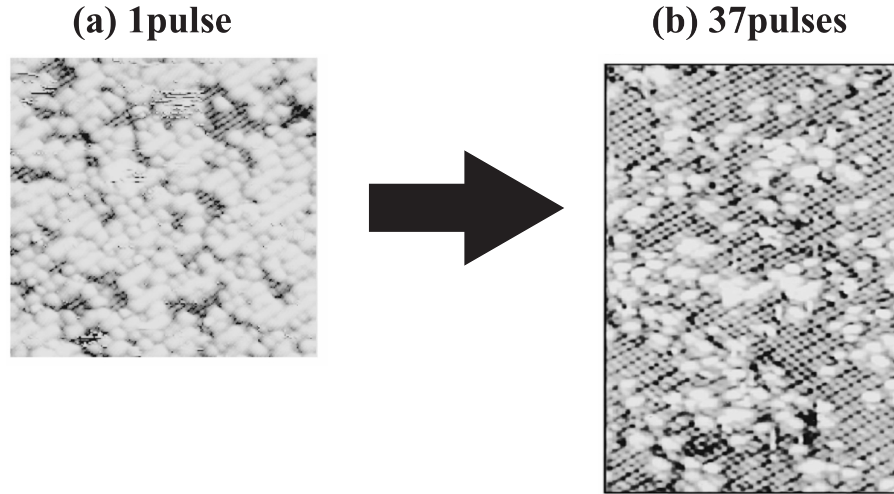


Figure 4.20: Surface dimers are removed by laser irradiation(Nd:YAG 532nm; 7ns). (a) Occupied-state image 16.1×19.2nm of a Si(100) surface that has been irradiated by one pulse of laser with a fluence of 150 mJ/cm^2. (b) Occupied state image of a Si(100) surface irradiated by multiple 37 pulses (12.5×27.0nm). The fluence is the same as used for (a). (From Ref.[127].)

S_A-type steps(Fig.4.21). In fig.4.22, 'recomposing of dimer bonds' occured in the dashed line circle, which is never seen on thermally annealed Si(100) surface. These two traces also support the restored Si(100) surface would be produced by very low energy electron irradiation, i.e., athermal process.

Hence, the low energy electron irradiation to the Si(100) surface investigated in this thesis indicated that it is so useful for not only the relaxation of the disorder-induced compressive surface stress during ion-bombardment but also the restoration of the destroyed structure. And its structure had differences of slightly pairing (1×1) structure and no bucking features from thermally annealed surface, derived from athermal process of electron irradiation.

4.5 A possibility of ordered oxidation

In general, it is believed that oxidation is disordering process, *i.e.* the periodicity of the surface is vanishing during an initial oxidation. In subsection3.4.6, we noticed a possibility from the results of the surface stress measurement that the oxidation under anodic condition could make the structure ordered, nevertheless proceeding the oxide formation. However, we have never been

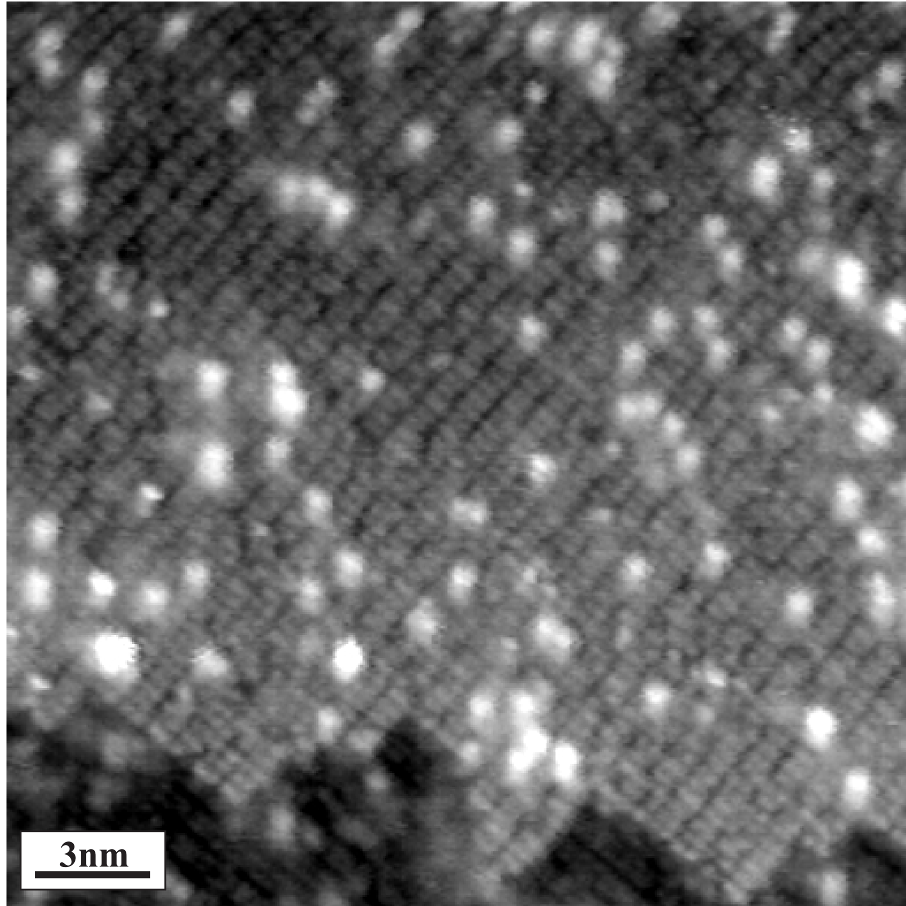


Figure 4.21: Upper terrace image of Fig.4.16. This image shows no buckled dimer features.

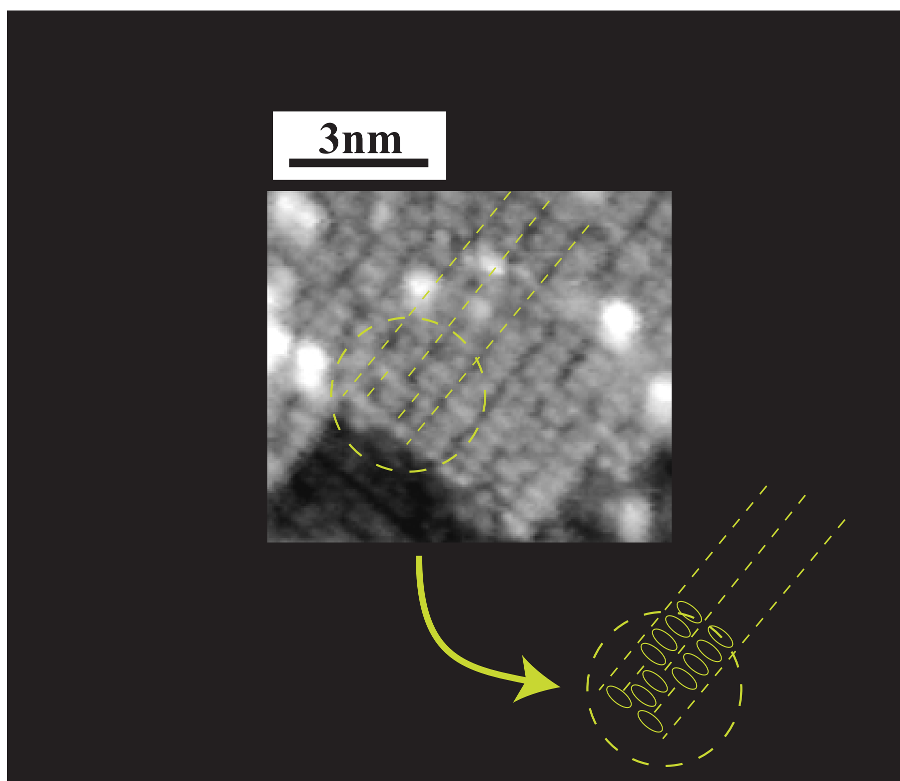


Figure 4.22: Other expanded image of 4.16. Recomposing of dimer bonds occured in the dashed line circle.

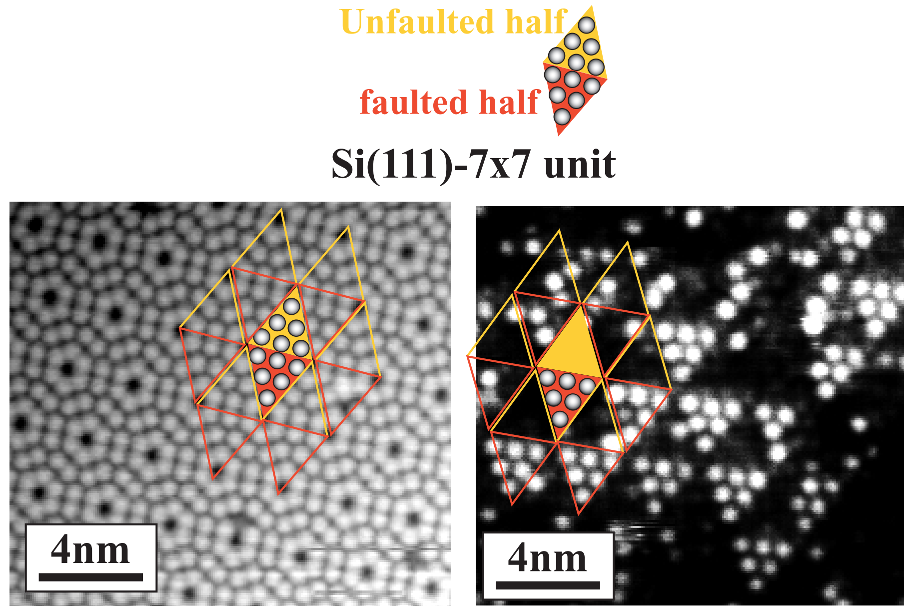


Figure 4.23: An example of the ordered oxidation. (a) The surface of Si(111)-7×7 has two different subunits; the faulted and unfaulted halves. (b) This oxidized surface of Si(111) was observed after the exposure of ozone (O_3) gas at room temperature. It is found that the unfaulted sites are imaged as depressions at the sample bias of +1.0V, but the other sites are mostly unchanged (Ozone's triangles!!).

able to observe oxidized surface by plasma oxidation with STM, because even the W-tip of STM was oxidized by a little oxygen gas (we used W (tungsten) tip to observe the surfaces at higher temperature) and then the tunneling current between the sample and the tip could not constantly flow. At an early date, to prevent tip from being oxidized we will use Pt-tip.

On the other hands, although with other method, we could observed the ordered oxidation with other radical species which is not plasma. The results show later, and this demonstrates possibilities of the ordered oxidation.

Selective Oxidation of Unfaulted Halves of Si(111): (7×7) with Ozone For nano-scale devices, an electronic isolation technique is desired since the working current is very small and therefore leakage through the substrate should be suppressed. Here, we demonstrate a nano-scale selective reaction to realize this isolation technique. For this model case, we used an ozone radical to selectively oxidize one subunit of Si(111)-7×7. This sur-

face has two different subunits; the faulted and unfaulted halves. Due to whether or not a stacking fault exists, the electronic states of the two halves differ, especially around -0.35 and +1.4eV from Fermi level. This state is the backbond state of the adatoms. Therefore, we could expect selective oxidation with some radicals on either the faulted or the unfaulted half. In conventional thermal oxidation, this possibility has already been denied, while ozone remains a possibility. We introduced highly concentrated ozone gas onto 7×7 reconstructed Si(111) surface in a UHV chamber in a split second. At room temperature, by means of STM, it is found that the unfaulted sites are imaged as depressions at the sample bias of +1.0∼1.5 V, but the other sites are mostly unchanged (Fig.4.23). This depression could be due to back bond oxidation. The reactive site of the oxygen changed with temperature, up to 873 K. As the substrate temperature increased, the chemical reaction selectivity disappeared. This may to due to diffusion of backbond oxygen atoms, which has a barrier height around 1eV. We intend to make 2.7 nm-size nano structure (unit size of 7×7) with atomic scale isolation of backbond oxidation state. It is very useful because it does not involve processing at high temperature, which could induce damage, or disappearance, of nano structures.

Chapter 5

Conclusion

We have presented evidence for relaxation of surface stress induced in Si by electron irradiation. Upon irradiating electrons on the disordered Si surfaces, the disorder-induced compressive stress becomes relaxed completely. The stress relaxation is found to depend only on the number of the irradiated electrons and is independent of the total energy deposition. This indicates that complete relaxation is promoted not by thermal activation mechanism, but by non-thermal mechanism induced by irradiated electrons

Our STM observation of the Si(100) surface under similar electron irradiation after ion-bombardment revealed that the surface where the surface stress was relaxed restored to highly and atomically ordered surface from the disordered surface. In demonstrating that surfaces can have the capacity of self-healing, these results suggest new opportunities for atomic scale surface engineering.

Furthermore, the STM observation after electron irradiation demonstrated athermal recrystallization of the surface atoms. In general, the structure generated by thermal annealing has the 2×1 symmetry due to Si dimers. The fine structure of the reordered surface after the electron irradiation is different from the typical thermal annealed surface, slightly closer to positions of a 'bulk terminated' 1×1 structure. It suggests that non-thermal genesis of a new surface structure was achieved by very low energy electron irradiation.

Bibliography

[1] *e.g.,* N. C. Bartelt, R. M. Tromp and E. D. Williams, Phys. Rev. Lett., **73**, 1656 (1994).

[2] S. C. Khare and T. L. Einstein, Phys. Rev. **B 54**, 11752 (1996).

[3] K. Morgenstern, G. Rosenfeld and G. Comsa, Phys. Rev. Lett., **76**, 2113 (1996).

[4] G. Binning, H. Rohrer, Ch. Gerber, and E. Weibel, Phys. Rev. Lett. **50**, 120 (1983).

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

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

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

[8] R. S. Becker, J. A. Golovchenko, E. G. McRae and B. S. Swartzentruber, Phys. Rev. Lett. **55**, 2028 (1985).

[9] R. S. Becker, J. A. Golovchenko, D. R. Hamann and B. S. Swartzentruber, Phys. Rev. Lett. **55**, 2032 (1985).

[10] R. M. Feenstra, A. P. Fein and W. A. Thompson, Phys. Rev. Lett. **56**, 608, (1986).

[11] K. C. Pandey, Phys. Rev. Lett. **47**, 1913 (1981); Phys. Rev. Lett. **49**, 223 (1982).

[12] R. M. Tromp, R. J. Hamers and J. E. Demuth, Phys. Rev. Lett. **55**, 1303 (1985).

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

[14] J. J. Lander and J. Morrisonm, J. Chem. Phys. **37**, 729 (1962).

[15] T. D. Poppendieck, T. C. Ngoc and M. B. Webb, Surf. Sci. **75**, 287 (1978).

[16] M. J. Cardillo and G. E. Becker, Phys. Rev. Lett. **40**, 1148 (1978); Phys. Rev. **B 21**, 1497 (1980).

[17] W. A. Harrison, Surf. Sci. **55**, 1 (1976).

[18] R. Seiwartz, Surf. Sci. **2**, 473 (1964).

[19] J. A. Appelbaum, G. A. Baraff and D. R. Hamann, Phys. Rev. Lett. **35**, 729 (1975); Phys. Rev. **B 11**, 3822 (1975); ibid Phys. Rev. **B 12**, 5479 (1975); ibid Phys. Rev. **B 14**, 588 (1976); ibid Phys. Rev. **B 15**, 2408 (1977).

[20] G. P. Kerter, S. G. Louie and M. L. Cohen, Phys. Rev. **B 17**, 706 (1978).

[21] J. E. Rowe, Phys. Lett. **A 46**, 400 (1974).

[22] F. Jona, H. D. Shith, A. Ignatiev, D. W. Jepson and P. M. Marcus, J. Phys. **C 10**, L67 (1977).

[23] S. J. White and D. P. Woodruff, Surf. Sci. **64**, 131 (1977).

[24] J. A. Appelbaum and D. R. Hamann, Phys. Rev. Lett. **74**, 459 (1978).

[25] S. Y. Tong and A. L. Maldondo, Surf. Sci. **78**, 459 (1978).

[26] F. J. Himpsel and D. E. Eastman, J. Vac. Sci. Technol. **16**, 1297 (1979).

[27] D. J. Chadi, Phys. Rev. Lett. **43**, 43 (1979).

[28] M. T. Yin and M. L. Cohen, Phys. Rev. **B 24**, 2403 (1981).

[29] K. C. Pandey, p.55, Proc. 17th Int. Conf. Phys. Semicond. in D. J. Chadi and W. A. Harrison (eds.), Springer, New York, 1985.

[30] M. Aono, Y. Hou, C. Oshima and Y. Ishizawa, Phys. Rev. Lett. **49**, 567 (1982).

[31] I. Stensgaard, L. C. Feldman and P. J. Silverman, Surf. Sci. **102**, 1 (1981).

[32] R. M. Tromp, R. G. Smeenk and F. W. Saris, Phys. Rev. Lett. **46**, 9392 (1981).

[33] Y. Chabal, Surf. Sci. **168**, 594 (1986).

[34] S. J. White, D. C. Frost and K. A. R. Mitchell, Solid State Commun. **42**, 763 (1982).

[35] W. S. Yang, F. Jona and P. M. Marcus, Phys. Rev. **B 28**, 2049 (1983).

[36] B. W. Holland, C. B. Duke and A. Paton, Surf. Sci. **140**, 1269 (1984).

[37] R. J. Hamers, R. M. Tromp and J. E. Demuth, Phys. Rev. **B 34**, 5343 (1986).

[38] R. J. Hamers, R. M. Tromp and J. E. Demuth, Surf. Sci. **181**, 346 (1987).

[39] R. J. Hamers, Ph. Avouris and F. Bozso, Phys. Rev. Lett. **59**, 2071 (1987).

[40] J. J. Boland, Phys. Rev. Lett. **67**, 1539 (1991).

[41] E. Artacho and F. Yndurain, Phys. Rev. Lett. **62**, 2491 (1989).

[42] G. K. Wertheim, D. M. Riffe, J. E. Rowe and P. H. Citrin, unpublished.

[43] M. C. Payne, N. Roberts, R. J. Needs, M. Needles and J. D. Joannopoulos, Surf. Sci. **211**, 1 (1989).

[44] N. Roberts and R. J. Needs, Surf. Sci. **236**, 112 (1990).

[45] J. Ihm, M. L. Cohen and D. J. Chadi, Phys. Rev. **B 21**, 4592 (1982).

[46] P. Kruger, A. Mazur, J. Pollmann and G. Wollgarten, Phys. Rev. Lett. **57**, 1468 (1986).

[47] J. Pollmann, P. Kruger and A. Mazur, J. Vac. Sci. Technol. **B 54**, 945 (1987).

[48] G. P. Kochanski and J. E. Griffith, Surf. Sci. **249**, L293 (1991).

[49] P. Badziag, W. S. Verwoerd and M. A. Van Hove, Phys. Rev. **B 43**, 2058 (1991).

[50] D. J. Chadi, Phys. Rev. Lett., **59**, 1691 (1987).

[51] D. E. Aspens and J. Ihm, Phys. Rev. Lett., **57**, 3054 (1986).

[52] J. E. Griffith, G. P. Kochanski, J. A. Kubby nad P. E. Wierenga, J. Vac. Sci. Technol. **A 7**, 1914 (1989).

[53] R. M. Feenstra, J. A. Stroscio, Phys. Rev. Lett. **59**, 2173, (1987).

[54] M. Henzler nad J. Clabesm "Structural and electronic properties of stepped semiconductor surfaces", in Proceedings of the 2nd Int. Conf. on Solid Surfaces, Jpn. J. Appl. Phys., Suppl. **2**, 389 (1974).

[55] O. L. Alerhand, A. N. Berker, J. D. Joannopoulos, D. Vanderbilt, R. J. Hamers and J. E. Demuth, Phys. Rev. Lett., **64**, 2406 (1990).

[56] D. Salonar, J. A. Martin, M. C. Tringides, D. E. Savage, C. E. Aumann and M. G. Lagally, J. Appl. Phys., **61**, 2884 (1987).

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

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

[59] J. E. Griffith and G. P. Kochanski, Critical Reviews in Solid State and Materials Science, **16**, 255 (1990).

[60] B. S. Swartzentruber, Y. W. Mo, R. Kariotis, M. G. Lagally and M. B. Webb, Phys. Rev. Lett., **65**, 1913 (1990).

[61] J. W. Gibbs, The Scientific Papers of J. Willard Gibbs, Vol.**1** (Longmans-Green, London, 1906) p.55.

[62] J. S. Vermaak, C. W. Mays and D. Kuhlmann-Wilsdorf, Surf. Sci., **12**, 128 (1968); C. W. Mays, J. S. Vermaak and D. Kuhlmann-Wilsdorf, Surf. Sci., **12**, 134 (1968); H. J. Wassermann and J. S. Vermaal, Surf. Sci., **22**, 164 (1970); H. J. Wassermann and J. S. Vermaal, Surf. Sci., **32**, 168 (1972).

[63] H. Helmann, Einfḧrung in die Quantentheorie (Deuticke, Leipzig, 1937) p.285; R. P. Feynman, Phys. Rev., **56**, 340 (1939).

[64] K. P. Bohnen and K. H. Ho, Surf. Sci. Rep., **19**, 99 (1993).

[65] O. H. Nielsen and R. M. Martin, Phys. Rev. Lett., **50**, 697 (1983).

[66] O. H. Nielsen and R. M. Martin, Phys. Rev., **B 32**, 3780 (1985); O. H. Nielsen and R. M. Martin, Phys. Rev., **B 32**, 3792 (1985); Matter. Res. Soc. Proc., **141**, 451 (1989).

[67] P. J. Feibelman, Phys. Rev., **B 50**, 1908 (1994).

[68] R. J. Needs, Phys. Rev. Lett., **58**, 53 (1987); R. J. Needs and M. J. Goodfrey, Phys. Scr., **T19**, 391 (1987).

[69] R. J. Needs, M. J. Goodfrey and M. Mansfield, Surf. Sci., **242**, 215 (1991).

[70] P. R. Watson, M. A. Van Hove and K. Hermann, NIST Surface Structure Database Ver.2.0 (National Institute of Standards and Technology, Gaithersburg, M.D, 1996).

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

[72] C. Stampfl, M. Scheffler, H. Over, J. Burchhardt, M. Nielsen, D. L. Adams and W. Moritz, Phys. Rev., **B 49**, 4959 (1994).

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

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

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

[76] A. Grossmann, W. Erley and H. Ibach, Surf. Sci., **337**, 183 (1995).

[77] H. Ibach, Surf. Sci. Rep., **29**, 193 (1997).

[78] A. Grossmann, Thesis, RWTH Aachen D80 (1996).

[79] P. Gumbsch and M. Daw, Phys. Rev. **B 44**, 4263 (1991).

[80] P. Feibelman, Phys. Rev. **B**, in press.

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

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

[83] M. C. Payne, N. Roberts, R. J. Needs, M. Needels and J. D. Joannopoulos, Surf. Sci. **211**, 1 (1989), V. Fiorentini, M. Methfessel and M. Scheffler, Phys. Rev. Lett., **71**, 1051 (1993).

[84] Koch and Abermann, Thin Solid Films, **129**, 63 (1985).

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

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

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

[88] G. G. Stoney, Proc. R. Soc. London, Ser. **A82**, 172 (1909).

[89] E. Kobeda, E.A. Irene, J. Vac. Sci. Technol. **B 6**, 574 (1988).

[90] A. N. Itakura, T. Narushima, M. Kitajima, K. Teraishi, A. Yamada and A. Miyamoto, Appl. Surf. Sci. **159 - 160**, 62 - 66 (2000).

[91] H. Ibach, Surf. Sci. Rep. **29** (5 - 6), 193 - 263 (1997).

[92] R. Koch, D. Winau, K.H. Rieder, Physica Scripta **49**, 539 (1993).

[93] P. Y. Yu and M. Cardona, 'Fundamentals of Semiconductors', Springer-Verlag, (1996).

[94] T. Narushima, A.N. Itakura, T. Kurashina, T. Kawabe and M. Kitajima, Appl. Surf. Sci. **159 - 160**, 25 - 29 (2000).

[95] G. D. Watkins, 'Lattice Defects in Semiconductors', Inst. Phys. Conf. Ser. **23**, 1 (1974).

[96] W. Frank, 'Lattice Defects in Semiconductors', Inst. Phys. Conf. Ser. **23**, 23 (1974).

[97] J. B. Pendry, Low Energy Electron Diffraction, Academic Press, London, (1974).

[98] C. J. Powell and M. P. Seah, J. Vac. Sci. Technol. **A8**, 735 (1990).

[99] R. D. Ramsier and J. T. Yates, Jr., Surf. Sci. Rep. **12**, 243 (1991).; T. E. Madey and J. T. Yates, Jr., J. Vac. Sci. Technol. **8**, 525 (1971).

[100] K. Nakayama and J. H. Weaver, Phys. Rev. Lett. **82**, 980 (1999).

[101] M. A. Lieberman and A.J. Lichtenberg, 'Principles of plasma discharges and materials processing', John Wiley & Sons, Inc., (1994).

[102] M. P. Seah and W. A. Dench, Surf. Interf. Anal. **1**, 2 (1979).

[103] M. Thompson, M. D. Baker, A. Christe and J. F. Tyson, 'Auger Electron Spectroscopy', John Wiley & Sons, New York, (1985).

[104] J. F. Ziegler , J. P. Biersack and U. Littmark, 'The Stopping and Ranges of Ions in Solids', Pergamon, New York, (1984).

[105] G. G. Stoney, Proc. R. Soc. London, Ser. **A82**, 172 (1909).

[106] E. Kobeda, E.A. Irene, J. Vac. Sci. Technol. **B 6**, 574 (1988).

[107] A. N. Itakura, T. Narushima, M. Kitajima, K. Teraishi, A. Yamada and A. Miyamoto, Appl. Surf. Sci. **159 - 160**, 62 - 66 (2000).

[108] K. -H. Hellwege and O. Madelung, 'Landolt-Börnstein Numerical data and functional relationships in science and technology, New Series, Group III:Crystal and Solid State Physics, Vol.**17a**, Semiconductors', p.3, Springer-Verlag, New York.

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

[110] F. Liu and M. Lagally, Phys. Rev. Lett. **76**, 3156 (1996).

[111] R. L. Novak, Bull. Am. Phys. Soc. **8**, 235 (1965).

[112] H. M. James and K. Lark-Horovitz, Z. physik. Chem. **198**, 107 (1951).

[113] C. Kittel, 'Introduction to Solid State Physics', 6th ed., John Wiley & Sons, Inc., (1986).

[114] M. Kitajima, Crit. Rev. Solid State Mater. Sci. **22**, 275 (1997).

[115] Carslaw, H. S. and Jaeger, J. C., 'Conduction of heat in Solids' Clarendon Press, Oxford (1991).

[116] http://www.nanosensors.com/gendesc.htm; J. K. Gimzewski, Ch. Gerber, E. Meyer, and R. R. Schlittler, Chemical Physics Letters, **294**, 589 (1994).

[117] J. A. Van Vechten and M. Wautelet, Phys. Rev. **B23**, 5,543 (1981).

[118] M. Takai, Laser kenkyu **11**, 649 (1983).(in Japanese)

[119] J. C. Bourgoin and J. W. Corbett, 'Lattice Defects in Semiconductors' Inst. Phys. Conf. Ser. **23**, 149 (1974).

[120] R. Car, P. J. Kelly, A Oshiyama and S. T. Pantelides, Phys. Rev. Lett. **52**, 1814 (1984).

[121] R. Ditchfield, D. Llera-Rodriguez, and E. G. Seebauer, Phys. Rev. **B61**, 13,710 (2000).; H. Y. H. Chan and E. G. Seebauer, American Vacuum Society 47th International Symposium in Boston, **SS3-MoA9**, (2000).

[122] C. Kittel, 'Introduction to Solid State Physics', 6th ed., John Wiley & Sons, Inc., New York, (1986).

[123] K. Nakayama and J. H. Weaver, Phys. Rev. Lett. **82**, 980 (1999).

[124] M. Nishizawa, T. Yasuda, S. Yamasaki, K. Miki, M. Shinohara, N. Kamakura, Y. Kimura, and M. Niwano; to be presented in AVS 2001; 'Generation of type-C defects on Si(100) by bimolecular adsorption of water: a FT-IR, STM, AES and QMS study'.

[125] R. J. Hamers and U. K. Kohler, J. Vac. Sci. Technol. **A 7**, 2854 (1989).

[126] J. A. Appelbaum, G. A. Baraff, and D. R. Hamann, Phys. Rev. **B14**, 588 (1976).

[127] Jun Xu, S. H. Overbury, and J. F. Wendelken, Phys. Rev. B-Rapid Communications, Vol. **53**, R4245 (1996).

[128] K. C. Pandey, in Proceedings of the 7th International Conference on the Physics of Semiconductors, edited by D. J. Chadi and W. A. Harrison (Springer, New York, 1985),p.55.

[129] N. Roberts and R. J. Needs, Surf. Sci. 236,112 (1990).

Appendix A

Scanning Tunneling Microscopy

Scanning tunneling microscopy (STM) provides three-dimensional real-space images of surfaces at high spatial resolution. When the surface is flat and clean, even atoms can be imaged. Its extreme usefulness has led it to near instantaneous acceptance as a characterization tool. This subsection covers fundamental concepts of STM theory, operation, image interpretation.

A.1 Basic principles

A.1.1 Vacuum tunneling

In vacuum tunneling, the potential in the vacuum region acts as a barrier to electrons between the two metal electrodes, in this case the surface and the tip. This barrier is shown schematically in Fig.A.1. The transmission probability for a wave incident on a barrier in one dimension is easily calculated. For STM we typically need consider only the limit of weak transmission, corresponding to the most common range of barrier height and widths. This limit gives a very simple behavior.

The solutions of Schödinger's equation inside a rectangular barrier in one dimension have the form

$$\Psi = e^{\pm\kappa z}. \tag{A.1}$$

Thus the crucial parameter is κ, where

$$\kappa^2 = 2m(V_B - E)/\hbar^2. \tag{A.2}$$

Here E is the energy of the state, and V_B is the potential in the barrier. In general, as shown Fig.A.1, V_B may not be constant across the gap; but for the moment in our discussion, it will be adequate to replace the potential in the barrier with average value, so that we need only consider a rectangular

barrier. In the simplest case V_B is simply the vacuum level; so far states at the Fermi level, V_B is just work function.

The transmission probability, or the tunneling current, thus decays exponentially with barrier width d as

$$I \propto e^{-2\kappa d}. \tag{A.3}$$

The generalization to a real three-dimensional surface is given below.

For a tunneling between two metals with a voltage difference V across the gap, only the states within V above or below the Fermi level can contribute to tunneling, with electrons in states within V below the Fermi level on the negative side tunneling into empty states within V above the Fermi level on the positive side. As shown in Fig.A.1, other states cannot be contribute either because there are no electrons to tunnel at higher energy, or because of the exclusion principle at lower energy. (Of course we should say "within energy eV above or below the Fermi level," where e is the electron charge. But throughout this paper, V will be written in such instances, to avoid confusion with the abbreviation for "electron volts.")

Since most work functions are around $4-5$eV, from Eq.(A.2) we find that typically $2\kappa \sim 2\text{Å}^{-1}$. Thus the tunneling current drops by nearly an order of magnitude for every 1Å of vacuum between the electrodes. Such tunneling can therefore only be observed in practice for very small separations. Achieving such small separations, and keeping the current even moderately stale, requires very precise control of the positions of the electrodes, limiting vibrations to much less than angstrom.

While such vacuum tunneling has long been understood in principle, the first report[2] of the direction observation of vacuum tunneling did not come until 1970s, and in 1982 Binnig *et al.* demonstrated well-controlled vacuum tunneling in their first step towards STM. They used a piezoelectric driver to accurately control the height of a metal tip above a surface, and these were largely restricted to tunneling through a static barrier, consisting of a layer of oxide sandwiched between metal electrodes[3].

A.1.2 Scanning tunneling microscopy

The basis idea behind STM is quite simple,[1] as illustrated in Fig.A.2. A sharp metal tip is brought close enough to the sample surface that electrons can tunnel quantum mechanically through the vacuum barrier separating tip and sample. As discussed earlier, this tunneling current is extremely sensitive to the gap, i.e., to the height of the tip above surface.

The position of the tip in three dimensions is accurately controlled by piezoelectric drivers. The tip is scanned in the two lateral dimensions, while

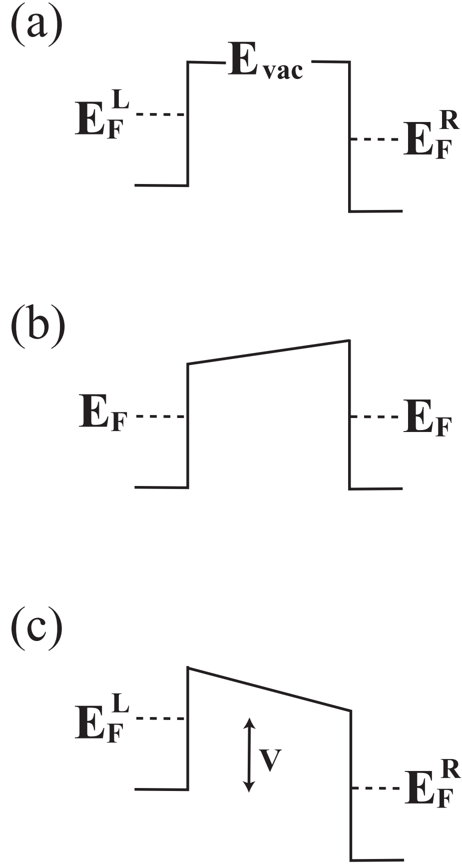


Figure A.1: Schematic of potential barrier between electrodes for vacuum tunneling. (a) Two non-interacting metal electrodes, separated by vacuum. The Fermi levels E_F of the two materials differ by an amount equal to the work function difference. (E_F^L and E_F^R denote the Fermi levels of the left and right electrode respectively, in the cases (a) and (c) where the two are not in equilibrium.) (b) The two electrodes are allowed to come into two electrical equilibrium, so that there is a unique common fermi level. The difference in work functions is now manifested as an electric field in the vacuum region. (c) A voltage is applied. There is a voltage drop V across the gap, i.e., the Fermi levels differ by eV. The field in the barrier includes contributions from both the applied voltage and the work function. The arrows indicate the range of energy over which tunneling can occur. At higher energies, there are no empty states to tunnel into.

a feedback circuit constantly adjusts the tip height, to keep the current constant. A constant current yields roughly a constant tip height, so the shape of the surface is reproduced by path of the tip, which can be inferred directly from the voltage supplied to the piezoelectric drivers.

It is also possible to use a slower feedback for the tip height, so that the height remains constant above the averaged surface, and small features are reflected in fluctuations of the current, rather than in the tip height[4]. However, this "constant height" mode of imaging is only practical in special cases where the surface is extremely flat, and is not fundamentally different from the usual "constant current" mode, so it is not discussed separately here.

Yet another mode of imaging is occasionally used with interesting results.[5] By modulating the tip height slightly (at a frequently above the response of the feedback), and measuring the resulting current modulation, one can obtain the local value of $d \ln I/dz$, and hence κ (or equivalently what is sometimes called the apparent barrier height or "effective work function," $\hbar^2\kappa^2/2m$) across the whole surface. This yields a picture in which the local variations of κ presumably reflect differences in chemical composition ot such, rather than topography.

Of course, since the tip has a finite radius, the surface topography is determined only with a finite resolution. One can make a simple estimate of the resolution as follows.[1] At a given lateral position x, relative to the center of the tip, the height of the corresponding point on the tip is $d+x^2/2R$, assuming a parabolic tip with radius of curvature R, and distance of closest approach d. The corresponding current in one-dimensional model is $I(x) \propto \exp(-\kappa x^2/R)$. Thus the current has a Gaussian profile, with root-mean-square width $\sim 0.7(R/\kappa)^{1/2}$. Since κ is typically $\sim 1\text{Å}^{-1}$, even a large tip radius such as 1000Å gives a rather sharp (though not atomic) resolution of 50Å.

It is possible to make metal tips with a radius of curvature of a few hundred Å, but not much less. However, because the tunneling current is so sensitive to distance, if the tip is a bit rough, most of the current will go to whatever atomic-scale asperity approaches closest to the surface. It is generally believed that the best STM images result from tunneling to a single atom, or at most a few atoms, on the tip.

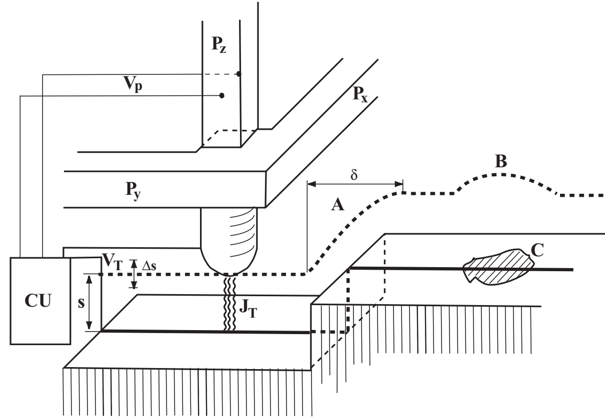


Figure A.2: Principle of operation of the scanning tunneling microscope (schematic, not to scale). The piezodrives P_x and P_y scan the metal tip over the surface. The control unit (CU) applies the necessary voltage V_P to the piezodrive P_z to maintain constant tunnel current J_T at bias voltage V_T. The broken line indicates the z displacement in a y scan at (A) a surface step, and (B) a spot C with lower work function.

A.2 Theory of STM imaging

A.2.1 Beyond Topography

So long as the features resolved are on the nanometer scale or larger, interpretation of the STM image as a surface topograph (complicated by local variations in barrier height) is generally adequate. But soon after the invention of STM,[1] Binnig *et al.* reported the first atomic-resolution images.[6] On the atomic scale, it is not even clear what one would mean by a topograph.

The most reasonable definition would be that a topograph is a contour of constant surface charge density. However, there is no reason why STM should yield precisely a contour of constant charge density, since only the electrons near the Fermi contribute to tunneling, whereas all electrons below the Fermi levels contribute to the charge density. Thus on some level, the interpretation of STM images as surface topographs must be inadequate. The following sections describe a more precise interpretation of STM images, applicable even in the case of atomic resolution.

One can calculate directly the transmission coefficient for an electron incident on the vacuum barrier between a surface and tip.[7] (We discuss such calculations later.) However, such a calculation is fairly complex for a realistic model of the surface. Fortunately, for typical tip-sample separations

(of order 9Å nucleus-to-nucleus[8]) the coupling between tip and sample is weak, and the tunneling can be treated with first-order perturbation theory. Since the problem is otherwise rather intractable except for simple models of the surface,[7, 8, 9] we restrict ourselves to this weak-coupling limit in the first part of our discussion.

A.2.2 Tunneling Hamiltonian approach

In first-order perturbation theory, the current is

$$I = \frac{2\pi e}{\hbar} \sum_{\mu,\nu} \{f(E_\mu)[1 - f(E_\nu)] - f(E_\nu)[1 - f(E_\mu)]\} |M_{\mu\nu}|^2 \delta(E_\nu + V - E_\mu), \tag{A.4}$$

where $f(E)$ is the Fermi function, V is again the applied voltage, $M_{\mu\nu}$ is the tunneling matrix element between states Ψ_μ and Ψ_ν of the respective electrodes, and E_μ is the energy of Ψ_μ. For most purposes, the Fermi functions can be replaced by their zero-temperature values, i.e., unit step functions, in which case one of the two terms in braces becomes zero. In the limit of small voltage, this expression further simplifies to

$$I = \frac{2\pi}{\hbar} e^2 V \sum_{\mu\nu} |M_{\mu\nu}|^2 \delta(E_\mu - E_F) \delta(E_\nu - E_F). \tag{A.5}$$

These equations are quite simple. The only real difficulty is in evaluating the tunneling matrix elements, Bardeen[10] showed that, under certain assumptions, these can be expressed as

$$M_{\mu\nu} = \frac{\hbar^2}{2m} \int d\boldsymbol{S} \cdot (\Psi_\mu^* \nabla \Psi_\nu - \Psi_\nu \nabla \Psi_\mu^*), \tag{A.6}$$

where the integral is over any surface lying entirely within the barrier region and separating the two half-spaces. If we chose a plane for the surface of integration, and neglect the variation of the potential in the region of integration, then the surface wave function at this plane can be conveniently expanded in the generalized plane-wave form

$$\Psi = \int d\boldsymbol{q} a_q e^{-\kappa_q z} e^{iq\cdot x} \tag{A.7}$$

where z is height measured from a convenient origin at the surface, and

$$\kappa_q^2 = \kappa^2 + |\boldsymbol{q}|^2. \tag{A.8}$$

A similar expansion applies for the other electrode, replacing a_q with b_q, z with $z_t - z$, and $\boldsymbol{x}$ with $\boldsymbol{x} - \boldsymbol{x}_t$. Here $\boldsymbol{x}_t$ and z_t are the lateral and

vertical components of the position of the tip. Then, substituting these wave functions into Eq.(A.6), one obtains

$$M_{\mu\nu} = -\frac{4\pi^2\hbar^2}{m}\int d\boldsymbol{q} a_q b_q^* \kappa_q e^{-\kappa_q z_t} e^{-iq\cdot x_t}. \tag{A.9}$$

Thus, given the wave functions of the surface and tip separately, i.e., a_q and b_q, one has a reasonably simple expression for the matrix element and tunneling current.

A.2.3 Modeling the tip

In order to calculate the tunneling current, and hence the STM image or spectrum, it is first necessary to have explicitly the wave functions of the surface and tip, for example in the form of Eq.(A.7) for use in Eq.(A.9). Unfortunately, the actual atomic structure of the tip is generally not known.[11] Even if it were known, the very low symmetry makes accurate calculation of the tip wave functions difficult.

One can therefore adopt a reasonable but somewhat arbitrary model for the tip. To motivate the simplest possible model for the tip, consider what would be the ideal STM.[12] First, one wants the maximum possible resolution, amd therefore the smallest possible tip. Second, one wants to measure the properties of the bare surface, not of the more complex interacting system of surface and tip. Therefore, the ideal STM tip would consist of a mathematical point source of current, whose position we denote $\boldsymbol{r}_t$. In that case, Eq.(A.5) for the current at small voltage reduces to[12]

$$I \propto \sum_{\nu} |\Psi_{\nu}(\boldsymbol{r}_t)|^2 \delta(E_{\nu} - E_F) \equiv \rho(\boldsymbol{r}_t, E_F). \tag{A.10}$$

Thus the ideal STM would simply measure $\rho(\boldsymbol{r}_t, E_F)$. This is a familiar quantity, the local density of states at E_F (LDOS), i.e., the charge density from states at the Fermi level. Note that the LDOS is evaluated for the bare surface, i.e., in the absence of the tip, but at the position which the tip will occupy. Thus within this model, STM has quite a simple interpretation as measuring a property of the bare surface, without reference to the complex tip-sample system.

It is important to see how far this interpretation can be applied for more realistic models of the tip. Reference [12] showed that Eq.(A.10) remains valid, regardless of tip size, so long as the tunneling matrix elements can be adequately approximated by those for an s-wave tip wave function. The tip position r_t must then be interpreted as the effective center of curvature of

the tip, i.e., the origin of the s-wave which best approximates the tip wave functions.

One can also to some extent go beyond the s-wave tip approximation, while still getting useful analytical results. A discussion of the contribution of wave function components of higher angular momentum was given by Tersoff and Hamann,[12] who showed that these made little difference for the observable Fourier components of Typical STM images. This issue was raised again by Chung *et al.*,[13] and by Chen,[14] who extended the analysis of Ref.[12].

Chen noted that in more recent images of close-packed metals,[15, 16] the relevant Fourier components are high enough that higher angular momentum components of the tip wave function could indeed affect the image substantially. However, it is important to recognize that such deviations from the behavior expected for an s-wave tip would be large only on this special class of surfaces. If a tip were to have a purely d_z wave function, then then corrugation of Al(111), for example, would be drastically increased; but the effect on surfaces observable with typical STM resolution would be relatively modest. Whether such tips exist is in any case an open question.

Sacks *et al.*[17] have also introduced a model that includes all components of the wave functions of both tip and surface. The essential approximation there was to treat both electrodes as rather flat, so that any deviations from planarity could be treated in perturbation theory.

Most recently, it was shown[18] that for a free-electron-like tip, the s-wave tip model should accurately describe STM images except in the case of tunneling to band-edge states, e.g., to semiconducting surfaces at low voltage; and that even then, none but the lowest Fourier component of the image should differ much from the s-wave result. This of course neglects the obvious effects of tip geometry, e.g., double tips.[19]

To model the tip more realistically, one must turn to numerical calculations of wave functions for a specific tip. Several studies in this vein, which support the simple LDOS interpretation of Eq.(A.10), are described in a later section.[20, 21, 22, 23]

A.3 Metal surfaces: STM as surface topography

A.3.1 Calculation of the LDOS

In order to interpret STM images quantitatively, just as with other experimental techniques, it is often necessary to calculate the image for a proposed

structure or set of structures, and compare it with the actual image. Such calculations for STM are still rather rare. We will discuss some examples and general features of such calculations, as well as the qualitative interpretation of STM images without detailed calculations.

For simple metals, there is typically no strong variation of the local density of states or wave functions with energy near the Fermi level. For purposes of STM, the same is presumably true for noble and even transition metals, since the d shell apparently does not contribute significantly to the tunneling current.[22] It is therefore convenient in the case of metals to ignore the voltage dependence, and consider the limit of small voltage of Eq.(A.10) (Effects of finite voltage are discussed in subsubsection A.4.)

This is particularly convenient, since we then require only the calculation of the LDOS $\rho(\boldsymbol{r}_t, E_F)$, a property of the bare surface. Nevertheless, even this calculation is quite demanding numerically. One case however of the STM image being calculated for a real metal surface, and compared with experiment, is that of Au(110) 2×1 and 3×1, treated in Ref.[12]. The LDOS calculated for these Au surfaces id shown in Fig.A.3.

A.3.2 Atom-superposition modeling

The real strength of STM is that, unlike diffraction, it is a local probe, and so can be readily applied to large, complex unit cells, or even to disordered surfaces or to isolated features such as defects. However, while the accurate calculation of $\rho(\boldsymbol{r}_t, E_F)$ is difficult even for Au(110) 3×1, it is out of the question for surfaces with very large unit cells, and *a fortiori* for disordered surfaces or defects. It is therefore highly desirable to have a method, however approximate, for calculating STM images in these important but intractable cases.

Such a method has been suggested and treated in Ref.[12]. It consists of approximating the LDOS (A.10) by a superposition of spherical atomic-like densities,

$$\rho(\boldsymbol{r}_t, E_F) \propto \sum_{\boldsymbol{R}} e^{-2\kappa|\boldsymbol{r}-\boldsymbol{R}|}/(E_0\kappa|\boldsymbol{r}-\boldsymbol{R}|). \quad \text{(A.11)}$$

Here each term is an atomic-like density centered on the atom site. The choice of an s-wave Hankel function allows convenient analytical manipulations, and provides an accurate description even at large distance.[12] E_0 is an energy which relates the charge to the density of states, and is typically of order 0.5-1.0eV.[12]

This approach is expected to work very well for simple and noble metals, ans was tested in detailed for Au(110).[12] The success of the method relies

on the fact that the model density, by construction, has the same analytical properties as the true density for a constant potential. Thus if the model is accurate near the surface, it will automatically describe accurately the decay with distance.

As one example of where this approach can be useful, the image expected for Au(110) 3×1 was calculated for two plausible models of the structure, differing only in the presence or absence of a missing row in the second layer.[12] The similarity of the model images at distances of interest suggested that the structure in the second layer could not be reliably inferred from experimental images. Quantifying the limits of valid interpretation in this way is an essential part of the analysis of STM data.

While this method is intended primarily for metals, Tromp *et al.*[24] applied it to Si(111) 7×7 with remarkable success. They simulated the images expected for a number of different proposed models of this surface, and compared them with an experimental image. This comparison is shown in Fig.A.4. The so-called " Dimer-Adatom-Stacking Fault" model of Takayanagi[25] gives an image which agrees almost perfectly with experiment, and a simple adatom model[6] is also rather close. Other models, although intended to be consistent with the STM measurements, lead to images with little similarlity to experiment. Thus the usefulness of such image simulations must not be underestimated, although few such applications have been made to date.

A.4 Semiconducting surfaces: role of surface electronic structure

A.4.1 Voltage dependence of images

At very small voltages, the s-wave approximation for the tip led to the very simple result in Eq.(A.10). At larger voltages, one might hope that this could be easily generalized to give a simple expression such as

$$I \sim \int_{E_F}^{E_F+V} \rho(\boldsymbol{r}_t, E_F) dE. \tag{A.12}$$

This is not strictly correct for two reasons. First, the matrix elements and the tip density of states are somewhat energy dependent, and any such dependence is neglected in Eq.(A.12). Second, the finite voltage changes the potential, and hence the wave functions, outside the surface. Nevertheless, Eq.(A.12) is a reasonable approximation for many purposes,[23] so long as

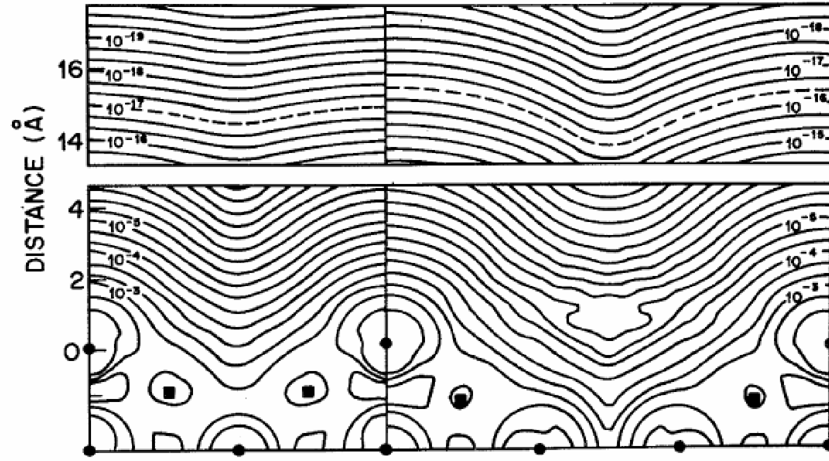


Figure A.3: Calculated $\rho(\boldsymbol{r}_t, E_F)$ for Au(110) 2×1 (left) and 3×1 (right) surfaces. This figure shows $(1\bar{1}0)$ plane through outermost atoms. Positions of nuclei are indicated by circles (in plane) and squares (out of plane). Contours of constant ρ are labeled in units of a.u.$^{-3}$eV^{-1}. Note break in vertical distance scale. Assuming a 9Å tip radius in the s-wave tip model, the center of curvature of the tip is calculated to follow the dashed line. (From Ref. [12])

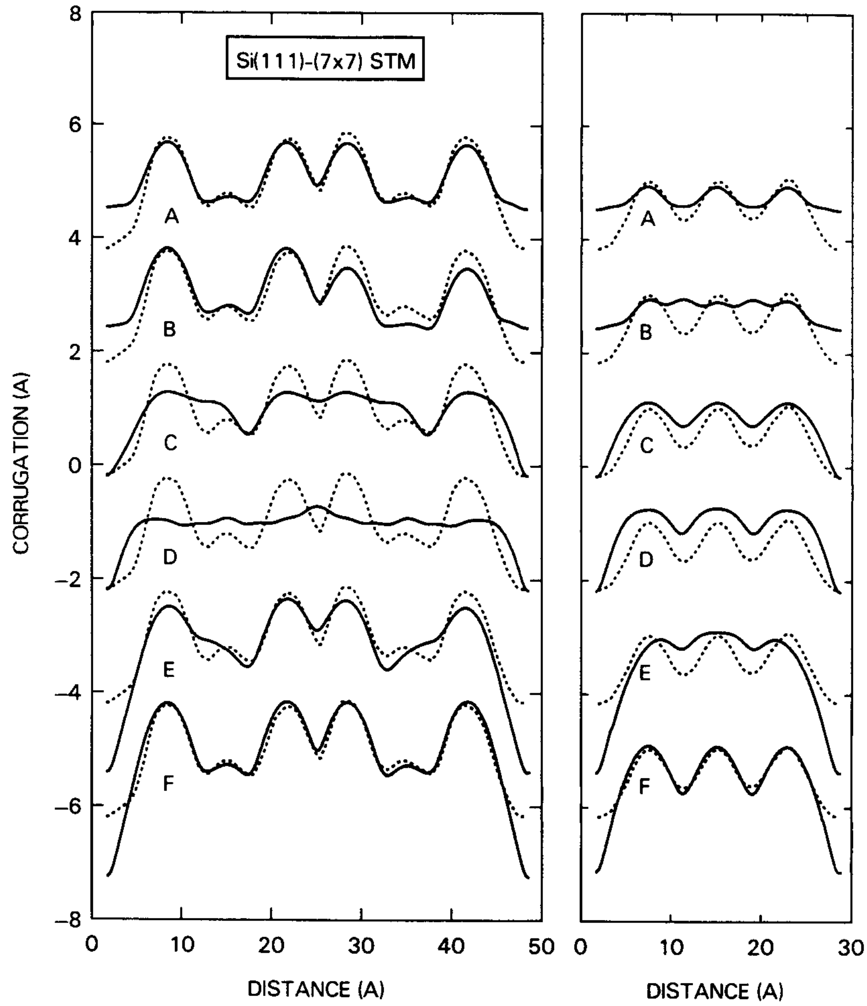


Figure A.4: Comparison between experimental STM images of Si(111) 7×7 (dotted line), and the images calculated with the atom-superposition model (Eq. A.11), for six different proposed structures of this surface. Left panel shows line scan along long diagonal of 7×7 cell, right panel along short diagonal. Curve A is for the adatom model proposed by Binnig *et al.*,[6] Curve F is for the model of Takayanagi *et al.*,[25] other curves are discussed in Ref.[24]. (From Ref.[24])

the voltage is much smaller than the work function. We shall therefore use Eq.(A.12)in discussing STM images of semiconductors at modest voltages.

Unlike metals, semiconductors show a very strong variation of $\rho(\boldsymbol{r}_t, E_F + V)$ with voltage. In particular, this quantity changes discontinuously at the band edges. With negative sample voltage, current tunnels out of the valence band, while for positive voltage, current tunnels into the conduction band. The corresponding images, reflecting the spatial distribution of valence and conduction-band wave functions respectively, may be qualitatively different.

A particularly simple and illustrative example, which has been studied in great detail, is GaAs(110). There, it was propose[12] that since the valence states are preferentially localized on the As atoms, and the conduction states on the Ga atoms, STM images of GaAs(110) at negative and positive sample bias should reveal the As and Ga atoms respectively. Such atom-selective imaging was confirmed[12] by direct calculation of Eq.(A.12), and was subsequently observed experimentally.[26]

In a single image of GaAs(110), whether at positive or negative voltage, one simply sees a single "bump" per unit cell, as shown in Fig.A.5. In fact, the images at opposite voltage look quite similar. It is therefore crucial to obtain both images *simultaneously*, so that the dependence of the absolute position of the "bump" on voltage can be determined. While neither image alone is very informative, by overlaying the two images the zig-zag rows of the (110) surface can be clearly seen[26]. Thus in this case, as with many non-metallic surfaces, voltage-dependent imaging is essential for the meaningful interpretation of STM images on an atomic scale.

Even in this simple case, however, the interpretation of the voltage-dependent images as revealing As or Ga atoms directly is a bit simplistic. Fig.A.6 shows a line-scan from a measured GaAs image, as well as theoretical "images" (i.e., contours of constant LDOS) for two cases: the ideal surface formed by rigid truncation of the bulk; and the real surface, where the As atom buckles upward, and the Ga downward. In each case, two curves are shown, corresponding to positive and negative bias.

For the ideal surface, at both biases the maxima in Fig.A.6a are almost directly over the respective atoms, supporting the simple interpretation of the image. For the buckled surface, though, the apparent positions of the atoms in the images deviate significantly from the actual positions. The separation between the Ga and As atoms in the (001) direction after buckling is less than 1.3Å, while the separation in this direction of the maxima in the image (Fig.A.6b) is 2.0Å. Thus, the distance between maxima in the image differs from that between atoms by over 50%. Qualitatively, one might say that the maxima correspond to the positions of the respective dangling bonds; but such an interpretation is difficult to quantify.

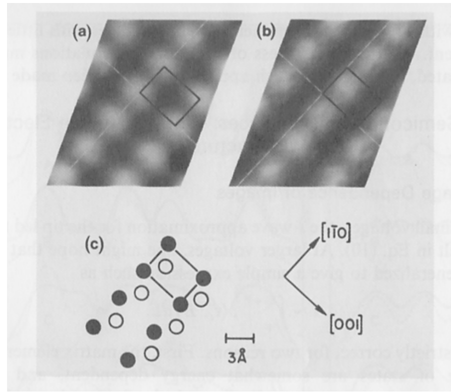


Figure A.5: Grey-scale STM images of GaAs(110) acquired at sample voltages of (a) +1.9 and (b) -1.9V. (c) Top view of GaAs surface structure. The As and Ga atoms are shown as open and filled circles respectively. The rectangle indicates a unit cell, whose position is the same in all three figures. (From Ref.[26])

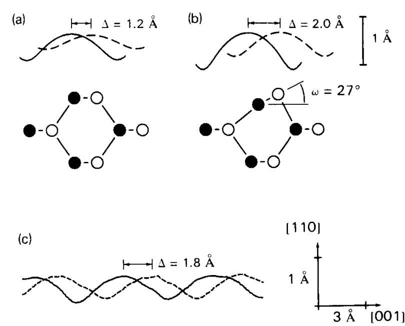


Figure A.6: Contour of constant LDOS in (1$\bar{1}$0) cross-section, for occupied (dashed line) and unoccupied (solid line) states. Absolute vertical positions are arbitrary. Δ is the lateral distance between peak positions for occupied and unoccupied states. (a) Theoretical results for ideal (unbuckled) surface. (b) Theoretical results for surface with 27° buckling. Side view of atomic structure is also shown, with open and filled circles indicating As and Ga. (c) Experimental results. (From Ref.[26])

This deviation of the bumps from the atom positions could be viewed as an undesirable complication, since it makes the image even less like a topograph. Alternatively, it is possible to take advantage of this deviation. The apparent positions of the atoms turn out to be rather sensitive to the degree of buckling associated with the (110) surface reconstruction. As a result, it is possible to infer the surface buckling quantitatively from the apparent atom positions. Thus the images are actually quite rich in information; but the quantitative interpretation requires a more detailed analysis than is often feasible. Even for a qualitative analysis, we cannot sufficiently emphasize the importance of voltage-dependent imaging, to help separate electronic from topographic features.

In tunneling to semiconductors, there are additional effects not present in metals. There may be a large voltage drop associated with band-bending in the semiconductor, in addition to the voltage drop across the gap. This means that the tunneling voltage may be substantially less than the applied voltage, complicating the interpretation. More interestingly, local band-bending associated with defects or adsobates on the surface can lead to striking non-topographic effects in the image, and localized states in the band gap lead to fascinating voltage-dependent images. These effects are of great interest for anyone involved in STM of semiconductor surface.

Bibliography

[1] G. Binning, H. Rohrer, Ch. Gerber, and E. Weibel, Phys. Rev. Lett. **49**, 57 (1982); G. Binning, H. Rohrer, Rev. Mod. Phys. **59**, 615 (1987).

[2] G. Binnig, H. Rohrer, Ch. Gerber, and E. Weibel, Appl. Phys. Lett. **40**, 178 (1982). For earlier reports of vacuum tunneling see E. C. Teague, Ph.D. Thesis, North Texas State University, 1978, reprinted in J. Res. National Bureau of Standards **91**, 171 (1986); R. Young, J. Ward and F. Scire, Phys. Rev. Lett. **27** 922 (1971); and W. A. Thompson, Rev. Sci. Instr. **47**, 1303 (1976).

[3] See for example C. B. Duke, p.1 in Tunneling in Solids, Suppl. 10 of Solid State Physics, (F. Seitz and D. Turnbull, eds.) Academic Press, New York, 1969, and references therein.

[4] A. Bryant, D. P. E. Smith, and C. F. Quate, Appl. Phys. Lett. **48**, 832 (1986).

[5] G. Binnig and H. Rohrer, Surf. Sci. **126**, 236 (1983); R. Wiesendanger, L. Eng, H. R. Hidber, P. Oelhafen, L. Rosenthaler, U. Staufer and H. J. Güntherodt, Surf. Sci. **189/190**, 24 (1987); B. Marchon, P. Bernhardt, M. E. Bussell, G. A. Somorjai, M. Salmeron and W. Siekhaus, Phys. Rev. Lett. **60**, 1166 (1988).

[6] G. Binning, H. Rohrer, Ch. Gerber, and E. Weibel, Phys. Rev. Lett. **50**, 120 (1983).

[7] N. Garcia, C. Ocal and F. Flores, Phys. Rev. Lett. **50**, 2002 (1983); E. Stoll, A. Baratoff, A. Selloni and P. Carnevali, J. Phys. C **17**, 3073 (1984).

[8] N. D. Lang, Phys. Rev. **B 36**, 8173 (1987).

[9] N. D. Lang, A. Yacoby and A Imry Phys. Rev. Lett. **63**, 1499 (1989).

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

[11] For a unique exception, see Y. Kulk, P. J. Silverman and H. Q. Nguyen, J. Vac. Sci. Technol. **A 6**, 524 (1988).

[12] J. Tersoff and D. R. Hamman, Phys. Rev. **B 31**, 805 (1985); and Phys. Rev. Lett. **50**, 1998 (1983).

[13] M. S. Chung, T. E. Feuchtwang and P. H. Cutler, Surf. Sci. **187**, 559 (1987).

[14] C. J. Chen, J. Vac. Sci. Technol. **A 6**, 319 (1988); and Phys. Rev. Lett. **65**, 448 (1987).

[15] V. M. Hallmark, S. Chiang, J. F. Rabolt, J. D. Swalen and R. J. Wilson, Phys. Rev. Lett. **59**, 2879 (1987).

[16] J. Wintterlin, J. Wiechers, H. Brune, T. Gritsch, H. Höfer and R. J. Behn, Phys. Rev. Lett. **62**, 59 (1989).

[17] W. Sacks, S. Gauthier, S. Rousset and J. Klein, Phys. Rev. **B 37**, 4489 (1988).

[18] J. Tersoff, Phys. Rev. **B 41**, 1235 (1990).

[19] S. Park, J. Nogami and C. F. Quate, Phys. Rev. **B 36**, 2863 (1987); H. A. Mizes, S. Park and W. A. Harrison, Phys. Rev. **B 36**, 4491 (1987).

[20] N. D. Lang, Phys. Rev. Lett. **56**, 1164 (1986).

[21] N. D. Lang, Phys. Rev. Lett. **55**, 230 and 2925 (E) (1985).

[22] N. D. Lang, Phys. Rev. Lett. **58**, 45 (1987).

[23] N. D. Lang, Phys. Rev. **B 34**, 5947 (1986).

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

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

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

Appendix B

Basic plasma equations and equilibrium

The plasma medium is complicated in that the charged particles are both affected by external electric and magnetic fields and contribute to them. The resulting selfconsistent system is nonlinear and very difficult to analyze. Furthermore, the interparticle collisions, although also electromagnetic in character, occur on space and time scales that are usually much shorter than those of the applied fields or the fields due to the average motion of the particles.

To make progress with such a complicated system, various simplifying approximations are needed. The interparticle collisions are considered independently of the larger scale fields to determine an *equilibrium distribution* of the charged-particle velocities. The velocity distribution is averaged over velocities to obtain the *macroscopic motion*. The macroscopic motion takes place in external applied fields and in the macroscopic fields generated by the average particle motion. These self-consistent fields are nonlinear, but may be linearized in some situations, particularly when dealing with waves in plasmas. The effect of spatial variation of the distribution function leads to pressure forces in the macroscopic equations. The collisions manifest themselves in particle generation and loss processes, as an averaged friction force between different particles species, and in energy exchanges among species. In this section we consider the basic equations that govern the plasma medium, concentrating attention on the macroscopic system. The complete derivation of these equations, from fundamental principles, is beyond the scope of the text. We shall make the equations plausible and, in the easier instances, supply some derivations in appendices.

In next subsection B.1 we introduce the macroscopic field equations and the current and voltage. In subsection B.2 we introduce the fundamental

equation of plasma physics, for the evolution of the particle distribution function, in a form most applicable for weakly ionized plasmas. We then define the macroscopic quantities and indicate how the macroscopic equations are obtained by taking moments of the fundamental equation. Although the macroscopic equations depend on the equilibrium distribution, their form is independent of the equilibrium. To solve the equations for particular problems the equilibrium must be known. In subsectionB.3 we introduce the equilibrium distribution and obtain some consequences arising from it and from the field equations. The form of the equilibrium distribution will be shown to be a consequence of the interparticle collisions, in appendices.

B.1 Field equations, current

Maxwell's equations

The usual macroscopic form of Maxwell's equations are

$$\nabla \times \boldsymbol{E} = -\mu_0 \frac{\partial \boldsymbol{H}}{\partial t} \tag{B.1}$$

$$\nabla \times \boldsymbol{H} = \epsilon_0 \frac{\partial \boldsymbol{E}}{\partial t} + \boldsymbol{J} \tag{B.2}$$

$$\epsilon_0 \nabla \cdot \boldsymbol{E} = \rho \tag{B.3}$$

and

$$\mu_0 \nabla \cdot \boldsymbol{H} = 0 \tag{B.4}$$

where $\boldsymbol{E}(\boldsymbol{r}, t)$ and $\boldsymbol{H}(\boldsymbol{r}, t)$ are the electric and magnetic field vectors and where $\mu_0 = 4\pi \times 10^{-7}$H/m and $\epsilon_0 \approx 8.854 \times 10^{-12}$F/m are permeability and permittivity of free space. The source of the fields, the charge density $\rho(\boldsymbol{r}, t)$ and the current density $\boldsymbol{J}(\boldsymbol{r}, t)$, are related by the charge continuity equation:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \boldsymbol{J} = 0. \tag{B.5}$$

In general,

$$\boldsymbol{J} = \boldsymbol{J}_{cond} + \boldsymbol{J}_{pol} + \boldsymbol{J}_{mag} \tag{B.6}$$

where the conduction current density $\boldsymbol{J}_{cond}$ is due to the motion of the free charges, the polarization current density $\boldsymbol{J}_{pol}$ is due to the motion of bound charges in a dielectric material, and the magnetization current density $\boldsymbol{J}_{mag}$ is due to the magnetic moments in a magnetic material. In a plasma in vacuum, $\boldsymbol{J}_{pol}$ and $\boldsymbol{J}_{mag}$ are zero and $\boldsymbol{J} = \boldsymbol{J}_{cond}$.

If Eq.(B.3) is integrated over a volume V, enclosed by a surface S, then we obtain its integral form, Gauss's law:

$$\epsilon \oint_S \boldsymbol{E} \cdot d\boldsymbol{A} = q \tag{B.7}$$

where q is the total charge inside the volume. Similarly, integrating Eq.(B.5), we obtain

$$\frac{dq}{dt} + \oint_S \boldsymbol{J} \cdot d\boldsymbol{A} = 0 \tag{B.8}$$

which states that the rate of increase of charge inside V is supplied by the total current flowing across S into V, i.e., that charge is conserved.

In Eq.(B.2), the first term on the RHS is the displacements current density flowing in the vacuum, and the second term is the conduction current density due to the free charges. We can introduce the total current density

$$\boldsymbol{J}_T = \epsilon_0 \frac{\partial \boldsymbol{E}}{\partial t} + \boldsymbol{J} \tag{B.9}$$

and taking the divergence of Eq.(B.2), we see that

$$\nabla \cdot \boldsymbol{J}_T = 0. \tag{B.10}$$

In one dimension, this reduces to $dJ_{Tx}/dx = 0$, such that $J_{Tx} = J_{Tx}(t)$, independent of x. Hence, for example, the total current flowing across a spatially nonuniform one-dimensional discharge is independent of x, as illustrated in Fig.B.1. A generalization of *Kirchhoff's current law*, which states that the sum of the currents entering a *node*, where many current-carrying conductors meet, is zero. This is also shown in Fig.B.1, where $I_{rf} = I_T + I_1$.

If the time variation of the magnetic field is negligible, as is often the case in plasmas, them from Maxwell's equations $\nabla \times \boldsymbol{E} \approx 0$. Since the curl of a gradient is zero, this implies that the electric field can be derived from the gradient of a scalar potential,

$$\boldsymbol{E} = -\nabla \Phi. \tag{B.11}$$

Integrating Eq.(B.11) around any closed loop C gives

$$\oint_C \boldsymbol{E} \cdot dl = -\oint_C \nabla \Phi \cdot dl = \oint_C d\Phi = 0. \tag{B.12}$$

Hence, we obtain *Kirchhoff's voltage law*, which states that the sum of the voltages around any loop is zero. This is illustrated in Fig.B.1, for which we obtain

$$V_{rf} = V_1 + V_2 + V_3 \tag{B.13}$$

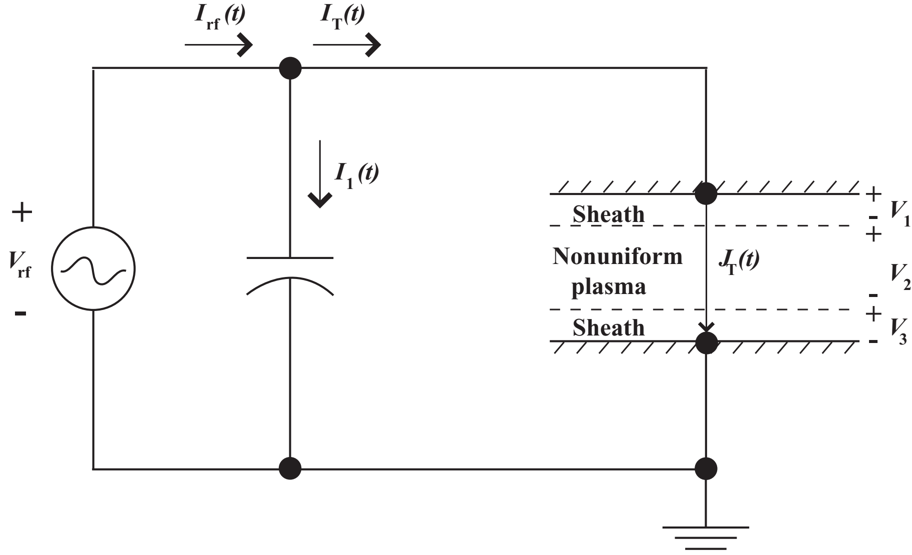


Figure B.1: Kirchhoff's circuit laws: The total current $\boldsymbol{J}_T$ flowing across a nonuniform one-dimensional discharge is independent of x; the sum of the currents entering a node is zero ($I_{rf} = I_T + I_1$); the sum of voltages around a loop is zero ($V_{rf} = V_1 + V_2 + V_3$).

i.e., the source voltage V_{rf} is equal to the sum of the voltages V_1 and V_3 across the two sheaths and the voltage V_2 across bulk plasma. Note that currents and voltages can have positive or negative values; the directions for which their values are designated as positive must be specified, as shown in the figure.

If Eq.(B.11) is substituted in Eq.(B.3), we obtain

$$\nabla^2 \Phi = -\frac{\rho}{\epsilon_0}. \tag{B.14}$$

Eq.(B.14), *Poisson's equation*, is one of the fundamental equations that we shall use. As an example of its application, consider the potential in the center ($x = 0$) of two grounded ($\Phi = 0$) plates separated by a distance $l = 10$cm and containing a uniform ion density $n_i = 10^{10}\text{cm}^{-3}$, without the presence of neutralizing electrons. Integrating Poisson's equation

$$\frac{d^2\Phi}{dx^2} = -\frac{en_i}{\epsilon_0} \tag{B.15}$$

using the boundary conditions that $\Phi = 0$ at $x = \pm l/2$ and that $d\Phi/dx = 0$

at $x = 0$ (by symmetry), we obtain

$$\Phi = \frac{1}{2}\frac{en_i}{\epsilon_0}\left[\left(\frac{l}{2}\right)^2 - x^2\right] \tag{B.16}$$

The maximum potential in the center is 2.3×10^5V, which is impossibly large for real discharge. Hence, the ions must be mostly neutralized by electrons, leading to a quasineutral plasma.

Electric and magnetic fields exert forces on charged particles given by the Lorentz force law:

$$\boldsymbol{F} = q(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}) \tag{B.17}$$

where $\boldsymbol{v}$ is the particle velocity and $\boldsymbol{B} = \mu_0 \boldsymbol{H}$ is the *magnetic induction vector*. The charged particles move under the action of the Lorentz force. The moving charges in turn contribute to both ρ and $\boldsymbol{J}$ in the plasma. If ρ and $\boldsymbol{J}$ are linearly related to $\boldsymbol{E}$ and $\boldsymbol{B}$, then the field equations are linear. As we shall see, this is not generally the case for a plasma. Nevertheless, linealization may be possible in some cases for which the plasma may be considered to have an *effective dielectric constant*; that is, the "free charges" play the same role as "bound charges" in a dielectric.

B.2 The conservation equations

B.2.1 Boltzmann's equation

For a given species, we introduce a *distribution function* $f(\boldsymbol{r}, \boldsymbol{v}, t)$ in the six-dimensional *phase space* $(\boldsymbol{r}, \boldsymbol{v})$ of particle positions and velocities, with the interpretation that

$$f(\boldsymbol{r}, \boldsymbol{v}, t)d^3rd^3v \;=\; \text{number of particles inside a six-dimensional phase space volume } d^3rd^3v \text{ at } (\boldsymbol{r}, \boldsymbol{v}) \text{ at time t.} \tag{B.18}$$

The six coordinates $(\boldsymbol{r}, \boldsymbol{v})$ are considered to be independent variables. We illustrate the definition of f and its phase space in one dimension in Fig.B.2. As particles drift in phase space or move under the action of macroscopic forces, they flow into or out of the fixed volume $dxdv_x$. Hence the distribution function f should obey a continuity equation which can be derived as follows. In a time dt,

$f(x, v_x, t)dxa_x(x.v_x, t)dt$ particles flow into $dxdv_x$ across face 1

$f(x, v_x + dv_x, t)dxa_x(x.v_x + dv_x, t)dt$ particles flow out $dxdv_x$ across face 2

$f(x, v_x, t)dv_xv_xdt$ particles flow into $dxdv_x$ across face 3

$f(x + dx, v_x, t)dv_xv_xdt$ particles flow out $dxdv_x$ across face 4

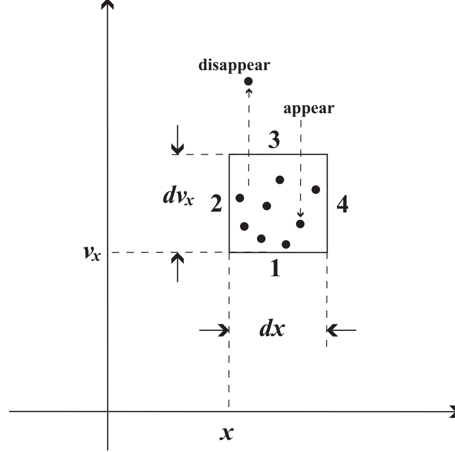


Figure B.2: One dimensional $v_x - x$ phase space, illustrating the derivation of the Boltzmann equation and the change in f due to collisions.

where $a_x \equiv dv_x/dt$ and $v_x \equiv dx/dt$ are the flow velocities in the v_x and x directions, respectively. Hence

$$\begin{aligned}
&f(x, v_x, t + dt)dxdv_x - f(x, v_x, t)dxdv_x \\
&= [f(x, v_x, t)a_x(x, v_x, t) - f(x, v_x + dv_x, t)a_x(x, v_x + dv_x, t)]\, dxdt \\
&+ [f(x, v_x, t)v_x - f(x + dx, v_x, t)v_x]\, dv_x dt. \qquad \text{(B.19)}
\end{aligned}$$

Dividing by $dxdv_xdt$, we obtain

$$\frac{\partial f}{\partial t} = -\frac{\partial}{\partial x}(fv_x) - \frac{\partial}{\partial v_x}(fa_x). \tag{B.20}$$

Noting that v_x is independent of x and assuming that acceleration $a_x = F_x/m$ of the particles does not depend on v_x, them Eq.B.20 can be rewritten:

$$\frac{\partial f}{\partial t} + v_x\frac{\partial f}{\partial x} + a_x\frac{\partial f}{\partial v_x} = 0. \tag{B.21}$$

The three dimensional generallization,

$$\frac{\partial f}{\partial t} + \boldsymbol{v} \cdot \nabla_{\boldsymbol{r}} f + \boldsymbol{a} \cdot \nabla_{\boldsymbol{v}} f = 0 \tag{B.22}$$

with $\nabla_{\boldsymbol{r}} = (\hat{x}\partial/\partial x + \hat{y}\partial/\partial y + \hat{z}\partial/\partial z)$ and $\nabla_{\boldsymbol{v}} = (\hat{x}\partial/\partial v_x + \hat{y}\partial/\partial v_y + \hat{z}\partial/\partial v_z)$ is called the *collisional Boltzmann equation* or *Vlasov equation.*

In addition to flows into or out of the volume across the faces, particles can "suddenly" appear in or disappear from the volume due to very

short times scale interparticle collisions, which are assumed to occur on a time scale shorter than the evolution time of f in Eq.B.22. Such collisions can practically instantaneously change the velocity (but not the position) of a particle. Examples of particles suddenly appearing or disappearing are shown in Fig.B.2. We account for this effect, which changes f, by adding a "collision term" to the right-hand side of Eq.B.22, thus obtaining the *Boltzmann equation*:

$$\frac{\partial f}{\partial t} + \boldsymbol{v} \cdot \nabla_{\boldsymbol{r}} f + \frac{\boldsymbol{F}}{m} \cdot \nabla_{\boldsymbol{v}} f = \left. \frac{\partial f}{\partial t} \right|_c . \tag{B.23}$$

The preceding heuristic derivation of the Boltzmann equation can be made rigorous from various points of view.

B.2.2 Macroscopic quantities

The complexity of the dynamical equations is greatly reduced by averaging over the velocity coordinates of the distribution function to obtain equations depending on the spatial coordinates and the time only. The averaged quantities, such as species density, mean velocity, and energy density are called macroscopic conservation equations. To obtain these averaged quantities we take *velocity moments* of the distribution function, and the equations are obtained from the moments of the Boltzmann equation.

The average quantities that we are concerned with are the particle density,

$$n(\boldsymbol{r}, t) = \int f d^3 v, \tag{B.24}$$

the particle flux

$$\Gamma(\boldsymbol{r}, t) = n\boldsymbol{u} = \int \boldsymbol{v} f d^3 v \tag{B.25}$$

where $\boldsymbol{u}(\boldsymbol{r}, t)$ is the mean velocity, and the particle kinetic energy per unit volume

$$w = \frac{3}{2} p + \frac{1}{2} m u^2 n = \frac{1}{2} m \int v^2 f d^3 v \tag{B.26}$$

where $p(\boldsymbol{r}, t)$ is the isotropic pressure, which we define below. In this form, w is sum of the *internal energy density* $\frac{3}{2}p$ and the *flow* energy density $\frac{1}{2}mu^2 n$.

B.2.3 Particle conservation

The lowest moment of the Boltzmann equation is obtained by integrating all terms of Eq.B.23 over velocity space. The integration yields the macroscopic *continuity equation*:

$$\frac{\partial n}{\partial t} + \nabla \cdot (n\boldsymbol{u}) = G - L. \tag{B.27}$$

The collision term in Eq.B.23, which yields the right-hand side of Eq.B.27, is equal to zero when integrated over velocities, except for collisions that create or destroy particles, designated as G and L, respectively (e.g., ionization, recombination). In fact, Eq.B.27 is transparent since it physically describes the conservation of particles. If Eq.B.27 is integrated over a volume V bounded by a closed surface S, then Eq.B.27 states that the net number of particles per second generated within V either flows across the surface S or increases the number of particles within V. For common low-pressure discharges in the steady state, G is usually due to ionization by electron-neutral collisions:

$$G = \nu_{iz} n_e \tag{B.28}$$

where ν_{iz} is the ionization frequency. The volume loss rate L, usually due to recombination, is often negligible. Hence,

$$\nabla \cdot (n\boldsymbol{u}) = \nu_{iz} n_e \tag{B.29}$$

in a typical discharge. However, note that the continuity equation is clearly not sufficient to give the evolution of the density n, since it involves another quantity, the mean particle velocity $\boldsymbol{u}$.

B.2.4 Momentum conservation

To obtain an equation for $\boldsymbol{u}$, a first moment is formed by multiplying the Boltzmann equation by $\boldsymbol{v}$ and integrating over velocity. The details are complicated and involve evaluation of tensor elements. The calculation can be found in most plasma theory texts. The results is

$$mn\left[\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u}\right] = qn(\boldsymbol{E} + \boldsymbol{u}\times\boldsymbol{B}) - \nabla\cdot\Pi + \boldsymbol{f}\mid_c . \tag{B.30}$$

The left-hand side is the species mass density times the convective derivative of the mean velocity, representing the mass density times the acceleration. The convective derivative has two terms: first term $\partial\boldsymbol{u}/\partial t$ represents an acceleration due to an explicitly time-varying $\boldsymbol{u}$; the second term "inertial" term $(\boldsymbol{u}\cdot\nabla)\boldsymbol{u}$ represents an acceleration even for a steady fluid flow ($\partial/\partial t \equiv 0$) having a spatially varying $\boldsymbol{u}$. For example, if $\boldsymbol{u} = \hat{x}u_x(x)$ increases along x, then the fluid is accelerating along x. The second term is nonlinear in $\boldsymbol{u}$ and can often be neglected in discharge analysis.

The mass times acceleration is acted upon, on the right-hand side, by the body forces, with the first term being the electric and magnetic force

densities. The second term is the force density due to the divergence of the pressure tensor, which arises due to the integration over velocities

$$\Pi_{ij} = mn\langle (v_i - u)(v_j - u)\rangle_v \tag{B.31}$$

where the subscript i, j give the component directions and $\langle (v_i - u)(v_j - u)\rangle_v$ denotes the velocity average of the bracketed quantity over f. We assume f is normalized so that $\langle f\rangle_v = 1$. For weakly ionized plasmas it is almost never used in this form, but rather an isotropic version is employed:

$$\Pi = \begin{pmatrix} p & 0 & 0 \\ 0 & p & 0 \\ 0 & 0 & p \end{pmatrix} \tag{B.32}$$

such that

$$\nabla \cdot \Pi = \nabla p \tag{B.33}$$

a pressure gradient, with

$$p = \frac{1}{3} mn\langle (v - u)^2\rangle_v \tag{B.34}$$

being the scalar pressure. Physically, the pressure gradient force density arises as illustrated in Fig.B.3, which shows a small volume acted upon by a pressure that is an increasing function of x. The net force on this volume is $p(x)dA - p(x + dx)dA$ and volume is dAdx. Hence the force per unit volume is $-\partial p/\partial x$.

The third term on the right in Eq.B.30 represents the time rate of momentum transfer per unit volume due to collisions with other species. For electrons or positive ions the most important transfer is often due to collisions with neutrals. The transfer is usually approximated by a Krook collision operator

$$\boldsymbol{f}\ |_c = -\sum_\beta mn\nu_{m\beta}(\boldsymbol{u} - \boldsymbol{u}_\beta) - m\boldsymbol{u}(G - L) \tag{B.35}$$

where the summation is over all other species, with $\boldsymbol{u}_\beta$ the mean velocity of species β and $\nu_{m\beta}$, the momentum transfer frequency for collisions with species β. The last term in Eq.B.35 is generally small and gives the momentum transfer due to the creation or destruction of particles. For example, if ions are created at rest, then they exert a drag force on the moving ion fluid because they act to lower the average fluid velocity.

A common form of the average force (momentum conservation) equation is obtained from Eq.B.30 for slow time variation, neglecting the inertial ($\boldsymbol{u} \cdot$

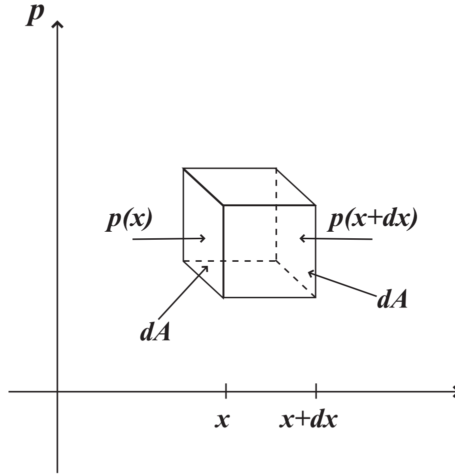


Figure B.3: The force density due to the pressure gradient.

$\nabla \boldsymbol{u}$) and magnetic forces, and taking $\boldsymbol{u}_\beta = 0$ in Krook collision term for collisions with one neutral species. The result is

$$0 = qn\boldsymbol{E} - \nabla p - mn\nu_m \boldsymbol{u} \tag{B.36}$$

where only the electric field, pressure gradient, and frictional forces appear. However, for fast time variation, we will also consider the acceleration term $mn\partial\boldsymbol{u}/\partial t$ on the left-hand side of Eq.B.30.

Eq. B.27 and B.30 together still do not form a closed set, since the pressure tensor Π (or scalar pressure p) is not determined. The usual procedure to close the equations is to use a thermodynamic *equation of state* t relate p to n. The *isothermal* relation for an equilibrium Maxwellian distribution is

$$p = nkT \tag{B.37}$$

so that

$$\nabla p = kT\nabla n \tag{B.38}$$

where T is the temperature in kelvins and k is Boltzmann's constant ($k = 1.381 \times 10^{-23}$J/K). This holds for slow time variations, where temperatures are allowed to equilibrate. In this case, the fluid can exchange energy with its surroundings, and we also require an energy conservation equation (see below) to determine p and T. For a room temperature (297K) neutral gas having density n_g and pressure p, EqB.37 yields

$$n_g[\text{cm}^{-3}] \approx 3.250 \times 10^{16} p[\text{torr}]. \tag{B.39}$$

Alternatively the *adiabatic* equation of state is

$$p = Cn^{\gamma} \tag{B.40}$$

such that

$$\frac{\nabla p}{p} = \gamma \frac{\nabla n}{n} \tag{B.41}$$

where γ is the ratio of specific heat at constant pressure to that at constant volume. The specific heats are defined; $\gamma = 5/3$ for a perfect gas; for one dimensional adiabatic motion, $\gamma = 3$. The adiabatic relation holds for fast time variations, such as in waves, when the fluid does not exchange energy with its surroundings to discharge analysis, we use the isothermal equation of state.

B.2.5 Energy conservation

The energy conservation equation is obtained by multiplying the Boltzmann equation by $\frac{1}{2}mv^2$ and integrating over velocity. The integration and some other manipulation yield

$$\frac{\partial}{\partial t}\left(\frac{3}{2}p\right) + \nabla \cdot \frac{3}{2}(p\boldsymbol{u}) + p\nabla \cdot \boldsymbol{u} + \nabla \cdot \boldsymbol{q} = \frac{\partial}{\partial t}\left(\frac{3}{2}p\right) |_c . \tag{B.42}$$

Here $\frac{3}{2}p$ is the energy density [J/m^3], $\frac{3}{2}(p\boldsymbol{u})$ is the macroscopic energy flux [W/m^2], representing the flow of the internal energy density at the fluid velocity $\boldsymbol{u}, p\nabla \cdot \boldsymbol{u}$ [W/m^3] gives the heating or cooling of the fluid due to compression or expansion of its volume, $\boldsymbol{q}$ is the heat flow vector [W/m^2], which gives the microscopic energy flux, and the collision term includes all collisional processes that charge the energy density. These include ionization, excitation, elastic scattering, and frictional (ohmic) heating. The equation is usually closed by setting $\nabla \cdot \boldsymbol{q} = 0$ or by letting $\boldsymbol{q} = -\kappa_T \nabla T$, where κ_T is the thermal conductivity. For most steady-state discharges the macroscopic energy flux is balanced against the collisional processes, giving the simpler equation

$$\nabla \cdot \left(\frac{3}{2}p\boldsymbol{u}\right) = \frac{\partial}{\partial t}\left(\frac{3}{2}p\right) |_c . \tag{B.43}$$

Eq.B.43, together with the continuity equation Eq.B.29, will often prove sufficient for our analysis.

B.2.6 Summary

Summarizing our results for the macroscopic equations describing the electron and ion fluids, we have in their most usually used forms the continuity

equation

$$\nabla \cdot (n\boldsymbol{u}) = \nu_{iz} n_e \tag{B.44}$$

the force equation, with the acceleration term on the left-hand side,

$$mn\frac{d\boldsymbol{u}}{dt} = qn\boldsymbol{E} - \nabla p - mn\nu_m \boldsymbol{u} \tag{B.45}$$

the isothermal equation of state

$$p = nkT \tag{B.46}$$

and the energy-conservation equation

$$\nabla \cdot \left(\frac{3}{2}p\boldsymbol{u}\right) = \frac{\partial}{\partial t}\left(\frac{3}{2}p\right)|_c \,. \tag{B.47}$$

These equations hold for each charged species, with the total charges and currents summed in Maxwell's equations. For example, with electrons and one positive ion species with charge Ze, we have

$$\rho = e(Zn_i - n_e) \tag{B.48}$$

$$\boldsymbol{J} = e(Zn_i\boldsymbol{u}_i - n_e\boldsymbol{u}_e). \tag{B.49}$$

These equations are still very difficult to solve without simplifications. They consist of 18 unknown quantities $n_i, n_e, p_i, p_e, T_i, T_e, \boldsymbol{u}_i, \boldsymbol{u}_e, \boldsymbol{E}$,and $\boldsymbol{B}$, with the vectors each counting for three. Various simplifications used to make the solutions to the equations tractable will be employed as the individual problemes allow.

B.3 Equilibrium properties

Electrons are generally in near-thermal equilibrium at temperature T_e in discharges, whereas positive ions are *almost never* in thermal equilibrium. Neutral gas molecules may or may not be in thermal equilibrium, depending on the generation and loss processes. For a single species in thermal equilibrium with itself (e.g., electrons), in the absence of time variation, spatial gradients, and accelerations , the Boltzmann equation Eq.B.23 reduces to

$$\frac{\partial f}{\partial t}|_c = 0 \tag{B.50}$$

where the subscript c here represents the collisions of a particle species with itself. We show that the solution of Eq.B.50 has Gaussian speed distribution of the form

$$f(v) = Ce^{-\xi^2 mv^2}. \tag{B.51}$$

The two constants C and ξ can be obtained by using the thermodynamic relation

$$w = \frac{1}{2}mn\langle v^2\rangle_v = \frac{3}{2}nkT \tag{B.52}$$

i.e., that average energy of a particle is $\frac{1}{2}kT$ per translational degree of freedom, and by using a suitable normalization of the distribution. Normalizing $f(v)$ to n, we obtain

$$C\int_0^{2\pi} d\phi \int_0^{\pi} \sin\theta d\theta \int_0^{\infty} \exp\left(-\xi^2 mv^2\right)v^2 dv = n \tag{B.53}$$

and using Eq.B.52, we obtain

$$\frac{1}{2}mC\int_0^{2\pi} d\phi \int_0^{\pi} \sin\theta d\theta \int_0^{\infty} \exp\left(-\xi^2 mv^2\right)v^4 dv = \frac{3}{2}nkT \tag{B.54}$$

where we have written the integrals over velocity space in spherical coordinates. The angle integrals yield the factor 4π. The v integrals are evaluated using the relation

$$\int_0^{\infty} e^{-u^2}u^{2i}du = \frac{(2i-1)!!}{2^{i+1}}\sqrt{\pi} \tag{B.55}$$

$$\text{where } i \text{ is an integer} \geq 1. \tag{B.56}$$

Solving for C and ξ we have

$$f(v) = n\left(\frac{m}{2\pi kT}\right)^{3/2} \exp\left(-\frac{mv^2}{2kT}\right) \tag{B.57}$$

which is the *Maxwellian distribution*.

Similarly, other averages can be performed. The average speed $\bar{v}$ is given by

$$\bar{v} = (m/2\pi kT)^{3/2} \int_0^{\infty} v\left[\exp\left(-\frac{v^2}{2v_{th}^2}\right)\right] 4\pi v^2 dv \tag{B.58}$$

where $v_{th} = (kT/m)^{1/2}$ is the thermal velocity. We obtain

$$\bar{v} = \left(\frac{8kT}{\pi m}\right)^{1/2}. \tag{B.59}$$

The directed flux Γ_z in (say) the $+z$ direction is given by $n\langle v_z\rangle_v$, where the average is taken over $v_z > 0$ only. Writing $v_z = v\cos\theta$ we have in spherical coordinates

$$\Gamma_z = n\left(\frac{m}{2\pi kT}\right)^{3/2} \int_0^{2\pi} d\phi \int_0^{\pi/2} \sin\theta d\theta \int_0^{\infty} v\cos\theta \exp\left(-\frac{v^2}{2v_{th}^2}\right)v^2 dv. \tag{B.60}$$

Evaluating the integrals, we find

$$\Gamma_z = \frac{1}{4} n\bar{v}. \tag{B.61}$$

Γ_z is the number of particles per square meter per second crossing the $z = 0$ surface in the positive direction. Similarly, the average energy flux $S_z = n\langle \frac{1}{2}mv^2 v_z\rangle_{\boldsymbol{v}}$ in the $+z$ direction can be found: $S_z = 2kT\Gamma_z$. We see that the average kinetic energy W per particle crossing $z = 0$ in the direction is

$$W = 2kT. \tag{B.62}$$

It is sometimes convenient to define the distribution in terms of other variables. For example. we can define a distribution of energies $W = \frac{1}{2}mv^2$ by

$$4\pi g(W)dW = 4\pi f(v)v^2 dv. \tag{B.63}$$

Evaluating dv/dW, we see that g and f are related by

$$g(W) = \frac{v(W)f[v(W)]}{m} \tag{B.64}$$

where $v(W) = (2W/m)^{1/2}$.

B.3.1 Boltzmann's relation

A very important relation can be obtained for the density of electrons in thermal equilibrium at varying positions in plasma under the action of a spatially varying potential. In the absence of electron drifts ($\boldsymbol{u}_e \equiv 0$), the inertial, magnetic, and frictional forces are zero, and the electron force balance is from Eq.B.36,

$$en_e\boldsymbol{E} + \nabla p_e = 0. \tag{B.65}$$

Setting $\boldsymbol{E} = -\nabla\Phi$ and assuming $p_e = n_e kT_e$, Eq.B.65 becomes

$$-en_e\nabla\Phi + kT_e\nabla n_e = 0 \tag{B.66}$$

or, rearranging,

$$\nabla(e\Phi - kT_e \ln n_e) = 0. \tag{B.67}$$

Integrating, we have

$$e\Phi - kT_e \ln n_e = \text{const} \tag{B.68}$$

or

$$n_e(\boldsymbol{r}) = n_0 e^{e\Phi(\boldsymbol{r})/kT_e} \tag{B.69}$$

which is *Boltzmann's relation* for electrons. We see that electrons are "attracted" to regions of positive potential. We shall generally write Boltzmann's relation in more convenient units

$$n_e = n_0 e^{\Phi/T_e} \tag{B.70}$$

where T_e is now expressed in volts, as is Φ.

For positive ions in thermal equilibrium at temperature T_i, a similar analysis shows that

$$n_i = n_0 e^{-\Phi/T_i}. \tag{B.71}$$

Hence positive ions in thermal equilibrium are "repelled" from regions of positive potential. However, positive ions are almost never in thermal equilibrium in low-pressure discharges because the ion drift velocity $\boldsymbol{u}_i$ is large, leading to inertial or frictional forces in Eq.B.36 that are comparable to the electric field or pressure gradient forces.

B.3.2 Quasineutrality

The characteristic length scale in a plasma is the electron Debye length λ_{De}. As we will show, the Debye length is the distance scale over which significant charge densities can spontaneously exist. For example, low-voltage (undriven) sheaths are typically a few Debye lengths wide. To determine the Debye length, let us introduce a sheet of negative charge having surface charge density $\rho_S < 0\text{C/m}^2$ into an infinitely extended plasma having equilibrium densities $n_e = n_i = n_0$. For simplicity we assume immobile ions, such that $n_i = n_0$ after the sheet is introduced. However, the negative sheet "repels" nearby electrons, leading to a reduced electron density near the sheet. To determine the potential and density variation, we use Poisson's equation, which in one dimension can be written

$$\frac{d^2\Phi}{dx^2} = -\frac{e}{\epsilon_0}(n_i - n_e). \tag{B.72}$$

Setting $n_e = n_0 \exp(\Phi/T_e)$, from the Boltzmann relation Eq.B.70, and taking $n_i = n_0$, Poisson's equation becomes

$$\frac{d^2\Phi}{dx^2} = \frac{en_0}{\epsilon_0}(e^{\Phi/T_e} - 1). \tag{B.73}$$

Expanding $\exp(\Phi/T_e)$ in a Taylor series for $\Phi \ll T_e$, Eq.B.73 becomes, to lowest order in Φ/T_e,

$$\frac{d^2\Phi}{dx^2} = \frac{en_0}{\epsilon_0}\frac{\Phi}{T_e}. \tag{B.74}$$

The symmetric solution of Eq.B.74 that vanishes at $x = \pm\infty$ is

$$\Phi = \Phi_0 e^{-|x|/\lambda_{De}} \tag{B.75}$$

where

$$\lambda_{De} = \left(\frac{\epsilon_0 T_e}{e n_0}\right)^{1/2}. \tag{B.76}$$

In practical units, we find

$$\lambda_{De}[\mathrm{cm}] \approx 743\sqrt{T_e/n_e} \tag{B.77}$$

with T_e in volts and n_e in cm^{-3}. We find for $T_e = 4\mathrm{V}$ and $n_e = 10^{10}\mathrm{cm}^{-3}$ that $\lambda_{De} = 0.14\mathrm{mm}$. It is on space scale larger than a Debye length that the plasma will tend to remain neutral.

The potential variation across a plasma of length $l \gg \lambda_{De}$ can be estimated from Poisson's equation Eq.B.14:

$$\nabla^2\Phi \sim \frac{\Phi}{l^2} \sim \left|\frac{e}{\epsilon_0}(Zn_i - n_e)\right|. \tag{B.78}$$

We generally expect that

$$\Phi \lesssim T_e = \frac{e}{\epsilon_0} n_e \lambda_{De}^2 \tag{B.79}$$

where the equality on the right follows from the definition of λ_{De}. Combining Eq.B.78 and we have

$$\frac{|Zn_i - n_e|}{n_e} \lesssim \frac{\lambda_{De}^2}{l^2}. \tag{B.80}$$

For $\lambda_{De}^2/l^2 \ll 1$, EqB.80 implies that

$$|Zn_i - n_e| \ll n_e \tag{B.81}$$

such that we can set

$$Zn_i = n_e \tag{B.82}$$

except when used in Poisson's equation. Relation Eq.B.81 is the basic statement of *quasineutrality* of a plasma, and is often called the *plasma approximation.* The plasma approximation is violated within a plasma sheath, in proximity to a material wall, either because the sheath thickness $s \approx \lambda_{De}$, or because $\Phi \gg T_e$.

Appendix C

Discharge and diagnostics of cold plasmas

C.1 Introduction to plasmas

Low-pressure plasma, cold plasma, nonequilibrium plasma, and glow discharge plasma are some of the synonymously used terms to designate the same types of process. The technologies using these plasma assisted processes are generally referred to as plasma processing and include such diverse approaches as plasma assisted chemical vapor deposition (PACVD), plasma enhanced chemical vapor deposition (PECVD), ionitriding, and plasma etching. Plasma assisted process is a critical technology used in the production of advanced microelectronics and in the production of present and future generations of large-scale integrated circuits. It would not be possible to manufacture very-large-scale integrated circuits (VLSI) computer chips without deposition of thin films by plasma assisted vapor deposition or without plasma assisted etching, which enables etching of submicronsize features with vertical walls in silicon, metals, and dielectrics. Plasma processing also made possible the development of special materials with unique properties such as amorphous silicon or diamondlike carbon.

Plasma chemistry takes place under nonequilibrium conditions, and the reactions can occur while the gas or parts exposed to it remain at relatively low temperatures. The advantages of plasma processing are being exploited in various areas besides microelectronics. For example, the plasma assisted chemical vapor deposition technique called *ionitriding* allows replacement and upgrade of a conventional technology for surface hardening of metals, done by thermal nitriding, and thus achieves much more efficient surface hardening. The ionitriding technique enables control and adjustment of the

properties of the hardened surface layers, not manageable while using the conventional method nitriding.

At the base of the mentioned technologies is the cold plasma, a phenomenon similar to that occuring in fluorescent bulbs or neon lights, that is, an electrical discharge in a gas at low pressure. The phenomena occuring in cold plasma are very complex and not yet fully understood. However, it is possible with the present knowledge of plasma physics and chemistry to adjust and control the composition of the gas mixtures and the parameters of the discharge to achieve required results in terms of processing and materials properties. The plasma assisted techniques allow increased production rates, precise production, and devising of materials with unique properties which evolve from the chemistry of cold plasmas.

The present paper aims to provide a broad introduction to the physics and chemistry of cold plasmas, to present simple diagonostic techniques used for studies, of plasma, for monitoring and optimization of plasma operations, and to discuss the applications of cold plasmas.

C.2 Definition of plasma

Taking into consideration the energy of the particles consisting it, plasma is energetically the fourth state of the matter, apart from the solid, liquid, and gas states. Fig.C.1 presents schematically the ranges of temperature, or particle energy, in which each of the four forms of matter occur in nature. For the plasma state, the temperature range reflects only the energy of the heavy particles (not of electrons) for reasons to be explained later.

Langmuir and collaborators were the first to study phenomena in plasma in the early 1920's while working on the development of vacuum tubes for large currents, and it was Langmuir who in 1929 used the term "plasma" for the first time to described ionized gases.[6]

In a more rigorous way, a plasma can be defined as a *quasi-neutral gas* of charged and neutral particles characterized by a *collective behavior*.

Let us first define the collective property of the plasma. The behavior of a neutral gas is described by the kinetic theory of gases. According to this theory, in an ordinary neutral gas no forces act between the molecules of the gas (gravitational forces are considered negligible), and the particles travel in straight lines, with a distribution of velocities. The motion of the molecules is controlled by the collision among themselves and with the walls of the container. As a result of these collisions, the molecules of a neutral gas follow a random Brownian motion, as illustrated in Fig.C.2(a).

Assuming the particles of the neutral gas to be rigid spheres of radius r

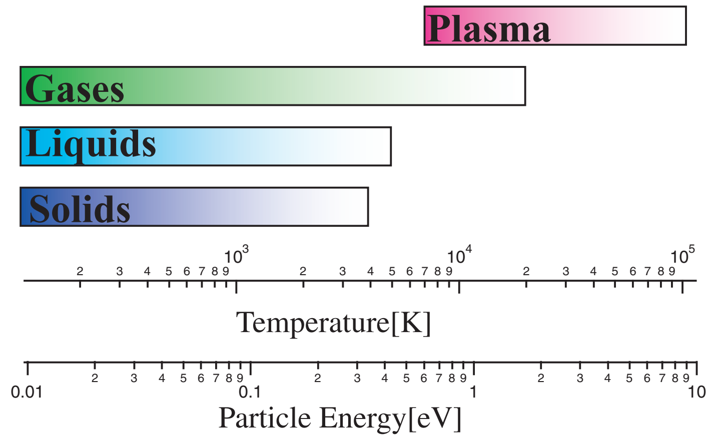


Figure C.1: State of matter versus temperature.

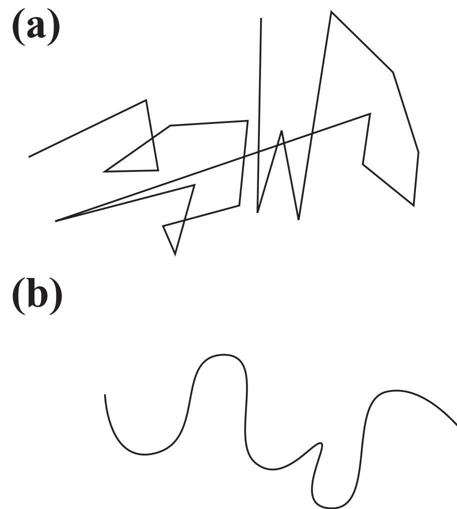


Figure C.2: Particle path in a neutral gas and under collective behavior in a plasma: (a) Brownian motion of a neutral gas molecule; (b) motion of a charged particle in a plasma.

and their density n, the kinetic theory of the gases defines the *cross section for collision*, σ, and mean free path, λ, as

$$\sigma = \pi r^2 \tag{C.1}$$

$$\lambda = \frac{1}{\sigma n}. \tag{C.2}$$

The average number of collision per second, called the *collision frequency*, ν, and the it mean time between collisions, τ, are given by

$$\nu = \frac{\bar{v}}{\lambda} \tag{C.3}$$

$$\tau = \frac{1}{\nu} = \frac{\lambda}{\bar{v}} \tag{C.4}$$

where $\bar{v}$ is the *average velocity* of the molecules in the gas which is determined by its temperature, T:

$$\bar{v} = \left(\frac{kT}{M}\right)^{1/2}. \tag{C.5}$$

M is the mass of the molecule, and k is the Boltzmann constant. If the temperature of the gas is constant, the *collisional mean free path* is inversely proportional to the pressure in the system:

$$\lambda = \frac{ct}{p} \tag{C.6}$$

where $ct =$ a constant depending on the gas, $p =$ pressure in the gas.

In a plasma, contrary to the preceding description, the motion of the particles can cause *local concentrations of positive and negative electric charges*. These charge concentrations create long-ranged Coulombic fields that affect the motion of charged particles far away from the charge concentrations. Thus elements of the plasma affect each other, even at large scale separations, giving the plasma its characteristic collective behavior. A charged particle in a plasma moves along a path which on average follows the electric field. Such a path is illustrated in Fig.C.2(b). In some conditions, at low pressure, the effect of the long-range electromagnetic forces on the motion of the particles can be much stronger than the effect of the collisions between the particles. In such cases the plasma is called a *collisionless plasma*.

Local concentrations of charges in a plasma are confined to volumes of small dimensions of size λ_{De}, where λ_{De} is a characteristic dimension of the plasma, called the Debye length. This term will be explained in detail in subsection C.3. For a plasma to be stable, it is necessary that the dimensions

of the system be much larger than the Debye length. Outside of these small volumes, of orders of tens of micrometers, the *charge density* of ions, is equal to the density of electrons, making the plasma electrically neutral. As a result, the plasma is considered a quasi-neutral gas.

A plasma is usually obtained when sufficient energy, higher than the ionization energy, is added to atoms of a gas, causing ionization and production of ions and electrons. Parallel and concomitant to the ionization occurs the opposite process of recombination of electrons with ions to form neutral atoms or molecules. Although plasmas are commonly produced by electrical discharges in gases, a plasma can also be obtained when sufficient energy is provided to a liquid or a solid to cause its vaporization and ionization. Such plasma can be obtained when high-density energy is supplied to a solid or a liquid by a laser. In a gas, a plasma is usually excited and sustained by providing to the gas electromagnetic energy in different forms: direct current, radio frequency, microwaves, and so on. Plasmas are often referred to as *gas discharges* because the most common way to produce plasma is by passing an electrical discharge through the gas.

C.3 Plasma parameters

A plasma, especially one sustained in a mixture of molecular gases, contains a multitude of different neutral and charged particles. A group of identical particles in a plasma is commonly referred to as a *species*.

The plasma is broadly characterized by the following basic parameters:

- The *density of the neutral particles*, n_s.
- The *densities of the electrons and ions*, n_e and n_i. In the quasi-neutral state of plasma the densities of the electrons and of the ions are usually equal, $n_i = n_e = n$ and n is called the *plasma density*.
- The *energy distributions* of the neutral particles, $f_n(W)$; ions, $f_i(W)$; and electrons, $f_n(W)$.

The plasma density is an important parameter in plasma processing because the efficiency of the processes occuring in the plasma and their reaction rates are generally dependent directly on the density of the charged particles. The electrons are main factor responsible for the transfer of the energy from the external field to the discharge gas. Being electrically charged, both electrons and ions interact with the applied external electric field and accelerated by absorbing energy from it. Because the electrons are the lightest particles in the plasma, they are easiest accelerated and absorb the largest amount of

energy from the external field. The electrons then transfer through collisions energy to the molecules of the gas and cause their ionization and dissociation. The effectiveness of these processes increases with increasing electron density. Ions also play a significant role in the chemical reactions taking place in the plasma. Many of the reactions occuring in a plasma are controlled, or affected, by ion chemistry. It is therefore important to achieve high ion densities to increase the rates of reactions involving the ions.

As in any gaseous system, particles in the plasma are in continuous motion, inducing collisions between them. The collisions which take place between the particles in the plasma are two types, elastic or inelastic. Collisions between electrons and heavy targets (i.e., neutral or charged particles) that do not result in an excitation of the target are called *elastic collisions*, whereas those collisions that leave the target in an excited state are called *inelastic collisions*.

The *energy transfer* W_{Tr} in an elastic collision between an electron and a heavy target is determined by mass ratio of the particles

$$W_{Tr} = \frac{4m_e M}{(m_e + M)^2} \approx \frac{4m_e}{M} W \tag{C.7}$$

where M = mass of the heavy particle, W = energy of the electron, m_e = mass of electron. For an elastic collision of an electron with an argon atom, the fraction of transferred energy is therefore very small, about

$$\frac{W_{Tr}}{W} \approx \frac{1}{18400}. \tag{C.8}$$

On the other hand, a significant amount of energy is transferred in a collision between two electrons.

The electrons gain energy through acceleration by the electric field, which sustains the plasma and transfers that energy by inelastic collisions with the neutral gas molecules. The inelastic collisions between energetic electrons and the heavy species of the plasma result in excitation, ionization, or dissociation of the target if it is multiatomic. Energy transfer in an inelastic collision is not controlled by the mass ratio of the colliding particles. In an inelastic collision between two particles, the fraction of transferred energy is given by

$$\frac{W_{Tr}}{W} = \frac{4m_{in} M}{(m_{in} + M)^2} \tag{C.9}$$

where m_{in} is the mass of particle losing energy.

According to Eq.C.9, in an inelastic collision between an electron and a heavy particle ($m_{in} = m_e \ll M$), the electron can transfer almost all its

energy to the heavy particle, creating an energetic plasma species. The inelastic collisions therefore sustain the plasma by producing the particles that form it and giving the plasma its special features. Inelastic collisions involve energy transfer in amounts that vary from less than 0.1eV (for rotational excitation of molecules) to more than 10eV (for ionization).

Electron-electron collisions can also play a significant role in the energy transfer processes in the plasma. Their importance depends on the degree of ionization prevalent in the plasma. For degrees of ionization below 10^{-10}, the contribution of the electron-electron collisions to the energy transfer is negligible. However, in *electron cyclotron resonance (ECR)* plasmas, where the degree of ionization can be above 10^{-3}, electron-electron oscillations dominate.[7]

The relative contribution of each type of collision to the processes taking place in the plasma depends on additional plasma parameters, which will be discussed next and which derive from the previously described parameters.

C.3.1 The degree of ionization

The parameter that defines the density of the charged particles in the plasma is the *degree of ionization* of the gas. It specifies the fraction of the particles in the gaseous phase which ionized. The degree of ionization, α, is defined as

$$\alpha = \frac{n_i}{n}. \tag{C.10}$$

For plasmas sustained in low-pressure discharges, the degree of ionization is typically 10^{-6} to 10^{-3}. However, if the electrical discharge is assisted and confined by an additional magnetic field, the degree of ionization can reach values of 10^{-2} or higher, as for example, in an ECR plasma. Table C.1 presents the range of values of the degree of ionization encountered in different low-pressure plasmas used for processing of solids.

Plasma type	Pressure(torr)	Ion density(cm^{-3})	Degree of ionization
Deposition/etching	< 10	$< 10^{10}$	10^{-6}
Reactive ion etching	10^{-2}-10^{-1}	10^{10}	10^{-6}-10^{-4}
Magnetron sputtering	10^{-3}	10^{11}	10^{-4}-10^{-2}
Electron cyclotron resonance	10^{-4}-10^{-2}	10^{12}	$< 10^{-1}$

Table C.1: Ranges of parameters for various low-pressure plasmas

The degree of ionization in a plasma is a function of the elements contained in the plasma. For example, in plasmas used in magnetron sputtering,

the degree of ionization of the sputtered metal is higher than that of the process gas employed for the sputtering.

The value of the *critical ionization* is defined by[8]

$$\alpha_c \approx 1.73 \times 10^{12} \sigma_{ea} T_e^2 \tag{C.11}$$

where σ_{ea} = electron-atom collision cross section at the average electron velocity, expressed in cm^2, T_e = electron temperature of the plasma, expressed in eV. The electron temperature will be defined later.

If the degree of ionization is much bigger than the critical ionization value, the charged particles behave as in a fully ionized gas.

C.3.2 Plasma temperature

Out of the physical parameters defining the state of a neutral gas in thermodynamic equilibrium is its temperature, which represents the *mean translational energy* of the molecules in the system. A plasma contains a mixture of particles with different electric charges and masses. At a first approximation, the plasma may be considered, thermally, as consisting of two systems: the first containing only electrons and the second containing the heavy species, that is, neutral atoms or molecules, ions, and neutral molecular fragments.

The electrons gain energy from the electric field, which energizes the plasma, and lose part of it by transfer to the second system through elastic or inelastic collisions. The system of heavy particles loses energy to the surroundings, either by radiation or by heat transfer to the walls of the vessel containing the plasma. The electrons and the heavy species in the plasma can be considered approximately as two subsystems, each in its own thermal quasi-equilibrium.

The ions and electrons in the plasma can therefore be characterized by their specific different average temperatures: the *ion temperature*, T_i, and the *electron temperature*, T_e. Actually in some cases additional temperatures may characterize the particles in plasma. For example, in the presence of magnetic field, even a single plasma species, for example, the ions, is characterized by two different temperatures, one representing the translation of the ions parallel to the magnetic field, $T_{||}$, and one representing the translational perpendicular to the magnetic field, $T_{\perp}$. This is caused by the fact that the forces acting perpendicular to it.

The situation is even more complicated, as the heavy species in the plasma can be characterized by several temperatures at the same time, even in the absence of a magnetic field: *the temperature of the gas*, T_g, which characterizes the translatory energy of the gas; the *excitation temperature*, T_{ex}, which

characterizes the energy of the excited particles in the plasma; the *ionization temperature*, T_{ion}; the *dissociation temperature*, T_d, which characterizes the energy of ionization and dissociation; and the *radiation temperature*, T_r, which characterizes the radiation energy. Thermodynamic equilibrium will exist in the plasma only if the following equation is satisfied:

$$T_g = T_{ex} = T_{ion} = T_d = T_r = T_e. \tag{C.12}$$

Complete thermodynamic equilibrium cannot be achieved in the entire plasma because the radiation temperature, T_r, at the envelope of the plasma cannot equal the temperature in the plasma bulk. However, under certain laboratory conditions, it is possible to achieve local thermodynamic equilibrium in plasma in volumes of order of the mean free path length. If this happens, the plasma is called a *local thermodynamic equilibrium (LTE)* plasma. In low pressure plasmas, produced by direct current glow discharge or radio frequency excitation, the LTE conditions are generally not achieved. These plasmas are therefore called *non-LTE* plasmas.

In non-LTE plasmas the temperatures of the heavy particles are normally too small to promote chemical reactions in thermodynamic equilibrium. The electron temperature is therefore the most important temperature in non-LTE plasmas, among all those different temperatures mentioned previously. The fraction of electrons that will cause the different reactions in the plasma, the overall efficiency of the plasma processes, and the processing rates increase with increasing electron temperature. The electron temperature is discussed in further detail in the next paragraph.

Electron temperature The *velocity distribution* function $f(v)$ for a system of particles is defined as the density of particles in the velocity space that satisfies the equation

$$n = 4\pi \int_0^\infty f(v) v^2 dv \tag{C.13}$$

where v = velocity, $f(v)$ = velocity distribution function (density in velocity space), n = the density of the particles in the geometrical space.

If it is assumed that the velocity distribution of the electrons in the plasma is isotropic, that the effects of inelastic collisions act only as a perturbation to the isotropy, and that the effects of the electric fields are negligible, then the velocity distribution is Maxwellian. The *Maxwellian distribution* assumes that the temperature of the electrons equals the temperature of the gas, $T_e = T_g$.

If the distribution of the electron velocities can be considered Maxwellian, then it can be described by

$$f(v) = n_e \left(\frac{m_e}{2\pi k T_e}\right)^{3/2} \exp\left(-\frac{m_e v^2}{2kT_e}\right) \tag{C.14}$$

The *electron energy distribution function* $f(W)$ is related to the velocity distribution function $f(v)$ through the relation

$$f(W) = \frac{4\pi}{m_e} v f(v). \tag{C.15}$$

Therefore, the Maxwellian energy distribution function for the electrons is given by

$$f(W) = 2.07 W_{av}^{-3/2} W^{1/2} \exp\left(-\frac{1.5W}{W_{av}}\right) \tag{C.16}$$

where W_{av} is *the average energy of electrons.* It can be shown that average energy of the electrons is related to their temperature by

$$W_{av} = \frac{3}{2} k T_e. \tag{C.17}$$

Due to the simplifying assumptions, the Maxwellian distribution provides only a first approximation of the electron energy (or velocity) distribution in plasma.

The assumptions made for the Maxwellian distribution can be replaced in low-pressure plasmas by the following assumptions:

1. The electric field strength in the plasma is sufficiently low such that one can neglect the inelastic collisions, but large enough for the electron temperature to be much higher than the ion temperature, $T_e \gg T_i$.

2. The electric field is of sufficiently low frequency, that is, it is of a frequency ω much lower than the frequency of collisions ν.

3. The collision frequency is independent of the electron energy.

Under these assumptions the distribution of electrons in the plasma is given by a *Druyvesteyn distribution.*[9]

At low pressures, the electron temperature is much greater than the temperature of the gas, $T_e \gg T_g$. When the pressure in the plasma increases, the energy transfer from electron to neutrals increases, causing an increase in the temperature of the gas and decrease fo the electron temperature. The electron and gas temperature converge to similar values at a pressure between

10 torr and 100 torr and plasma becomes arclike. In arcs at atmospheric pressure the two temperature are equal, $T_e = T_g$. When the two temperatures are about the same, the distribution of the species in the plasma can be described by equilibrium relations, while in the case when $T_e \gg T_g$, the distribution of active species is best represented by the electron temperature, T_e.

Although the temperatures of the electrons in the types of plasmas to be discussed in the following subsections are several times 10^4 °K (several electron volts), it does not imply that these plasmas are hot. Typical electron densities in these plasmas are about 10^{10}cm^{-3}, as compared to the density of particles in a gas at atmospheric pressure which is about $\sim 2.7 \times 10^{19}\text{cm}^{-3}$. Due to the very low density and the very low heat capacity of the electrons, the amount of heat transferred by the electrons to the gas (heavy neutral and ionized particles) and to the walls of the container is very small. Thus, the term "cold plasma" derives its meaning from the small amount of heat transferred to the gas or solid surfaces in contact with it.

C.3.3 Debye length

Another important parameter of a plasma is its *Debye length.* If an electric field is created in the plasma, the charged particles will react to reduce the effect of the field. The lighter, more mobile, electrons will respond fastest to reduce the electric field. If a plasma had an excess of positive or negative particles, this excess would create an electric field and the electrons will move to cancel the charge.

The response of charged particles to reduce the effect of local electric fields is called *Debye shielding* and the shielding gives the plasma its quasi-neutrality characteristic. Let assume that an electric potential is applied between two surfaces immersed in a plasma. The concentration of charged particles near the two surfaces will shield the charged surfaces from the plasma bulk, which will remain neutral. The applied electrical potential will therefore develop mostly near the surfaces, over a distance λ_{De}, called the *Debye length* and defined by

$$\lambda_{De} = \left(\frac{\epsilon_0 k T_e}{n_e e^2}\right)^{1/2} \tag{C.18}$$

where ϵ_0 = the permittivity of the free space, e = the charge of the electron.

To clarify the meaning of the Debye length, let's assume that a positive electric charge, q, is inserted in a plasma that is initially quasi-neutral. The

charge will create an electric potential, which in free space would be

$$V_0 = \frac{q}{4\pi\epsilon_0 d} \tag{C.19}$$

where d is the distance from the charge. In the plasma, potential is affected by the plasma electrons and ions and its value can be obtained by solving Poisson's equation,

$$\nabla^2 V = -\frac{\rho}{\epsilon_0} \tag{C.20}$$

where ρ is the total charge density in the plasma and is given by:

$$\rho = e(n_i - n_e) + q\delta(d) \tag{C.21}$$

where $\delta(d)$ is the Dirac δ function, indicating that q is a point charge.

The potential V changes the electron density, and assuming that the electrons are in thermodynamic equilibrium at temperature T, the density can be calculated as:

$$n_e = n \exp\left(\frac{eV}{kT}\right). \tag{C.22}$$

As it can be assumed that $eV \ll kT$, the Poisson's equation can be rewritten using Eq.(C.21) as

$$\nabla^2 V = \frac{V}{\lambda_{De}} + q\delta(d) \tag{C.23}$$

with λ_{De} defined by Eq.(C.18). The solution of Eq.(C.23) is

$$V(d) = \frac{q}{4\pi\epsilon_0 d} \exp\left(-\frac{d}{\lambda_{De}}\right) = V_0 \exp\left(-\frac{d}{\lambda_{De}}\right). \tag{C.24}$$

Eq.(C.24) shows that the plasma changes the potential of free space, V_0, causing its attenuation with a decay length equal to the Debye length, λ_{De}. This attenuation of the potential produced by a local charge in the plasma is the Debye shielding effect.

For estimation purposes and taking into account that $T[°\text{K}] = 11,600T[\text{eV}]$, it is convenient to calculate the Debye length from

$$\lambda_{De}[\text{cm}] = 6.93 \left[\frac{T_e[°\text{K}]}{n_e[\text{cm}^{-3}]}\right]^{1/2} = 743 \left[\frac{T_e[\text{eV}]}{n_e[\text{cm}^{-3}]}\right]^{1/2}. \tag{C.25}$$

An example of typical value found in a cold plasma is

$$T_e = 1\text{eV}, n_e = 10^{10}\text{cm}^{-3}, \text{and} \;\; \lambda_{De} = 74\mu\text{m} \tag{C.26}$$

As indicated in by Eq.(C.18), the Debye length decreases with increasing electron density. An ionized gas is considered a plasma only if the density of the charged particles is large enough such that $\lambda_{De} \ll L$, where L is the dimension of the system. If this condition is satisfied, local concentrations of electric charges which may occur in the plasma are shielded out by the Debye shielding effect over distances smaller than the Debye length. Outside these volumes of charge concentrations the plasma bulk is quasi-neutral. The Debye length, λ_{De}, is therefore the characteristic dimension of the regions in which breakdown of neutrality (formation of local concentrations of charges) can occur in a plasma.

Another plasma parameter related to the Debye length is the *number of particles*, N_D, in a *Debye sphere*, that is, in a sphere of radius equal to λ_{De}. The solution of Poisson's equation given by Eq.(C.24) can be obtained only by assuming that the shielding effect is produced by a large number of electrons, or in other words, the shielding effect can occur only if the Debye sphere contains a large number of electrons. Due to the exponential decay of the potential, it can be assumed that the shielding is caused by the electrons in the Debye sphere, whose number is given by

$$N_D = \frac{4\pi}{3} n_e \lambda_{De}^3 = \frac{1.38 \times 10^3 T_e^{3/2} [^\circ\mathrm{K}]}{n_e^{1/2}} = \frac{1.718 \times 10^9 T_e^{3/2} [\mathrm{eV}]}{n_e^{1/2}}. \tag{C.27}$$

N_D has to be therefore much larger than unity to fulfill the collective characteristic of the plasma. For electron temperatures $T_e > 1$eV and densities $n_e < 10^{12}$ about 10^4 to 10^7 electrons in a Debye sphere.

C.3.4 Plasma frequency

Although the plasma bulk is quasi-neutral, local perturbations from neutrality can occur in volumes smaller than the Debye sphere. Due to their low mass, the electrons will respond faster than the ions to the electric forces generated by the perturbation from neutrality. The response to the perturbation will be through oscillations. The frequency of these electron oscillations is called the *plasma*, or *Langmuir frequency*, ω_p, and is given by the relation

$$\omega_p = \left(\frac{n_e e^2}{m_e \epsilon_0} \right)^{1/2} = 18,000 \pi n_e^{1/2} \mathrm{Hz} \tag{C.28}$$

for n_e expressed in cm^{-3}. For a typical plasma density of $10^{10}\mathrm{cm}^{-3}$, the plasma frequency is 9×10^8Hz, a frequency much higher than the 13.56 MHz generally used to sustain a radio frequency (RF) discharge.

It can be deduced from Eq.(C.18) and Eq.(C.28) that the plasma frequency, ω_p, is related to the Debye length by the relation

$$\lambda_{De}\omega_p = \left(\frac{kT_e}{m_e}\right)^{1/2} \approx \bar{v}_e. \tag{C.29}$$

Here, $\bar{v}_e$ is the *average thermal velocity* of the electrons in the plasma.

Eq.(C.29) shows that the electrons can move over a distance of one Debye length during a period of the plasma oscillation. This indicates that if a perturbation of frequency $\omega < \omega_p$ is occuring in the plasma, the electrons can respond sufficiently fast to maintain the neutrality of the plasma. The oscillation frequency of the ions being much smaller due to their larger mass, only the electrons are able to maintain their own thermal equilibrium in the plasma. Plasma perturbations of frequencies higher than ω_p will not be shielded out through the response of the electrons.

C.4 Conditions for plasma existence

In subsection C.3.3 it was concluded that to sustain a plasma in a gas, two conditions have to be met:

$$\lambda_{De} \ll L \tag{C.30}$$

$$N_D \gg 1. \tag{C.31}$$

However, an additional third condition has to be fulfilled by a gas to become plasma. This condition is related to the frequency of collisions in the plasma. If the charged particles collide too frequently with neutral atoms, their motion is controlled by ordinary hydrodynamic forces rather than by electromagnetic forces. Under these circumstances the collective behavior condition is not satisfied, and the gas is not behaving as a plasma. If τ is the mean time between collisions of charged particles with neutral atoms, the product $\omega\tau$ has to be bigger than 1 for the gas to behave like a plasma rather than a neutral gas.

The three conditions that have to be satisfied by a plasma are therefore given by Eq.(C.30), Eq.(C.31), and Eq.(C.32),

$$\omega\tau > 1. \tag{C.32}$$

C.5 Plasma DC sheath calculations

C.5.1 Basic concept and equations

At the edge of a bounded plasma a potential exists to contain the more mobile charged species. This allows the flow of positive and negative carriers to the wall to be balanced. In the usual situation of an electropositive plasma, consisting of equal numbers of positive ions and electrons, the electrons are far more mobile than the ions. The plasma will therefore charge positively with respect to a grounded wall. The nonneutral potential region between the plasma and the wall is called a "*sheath*".

In a weakly ionized plasma the energy to sustain the plasma is generally heating of the electrons by the source, while the ions are at near equilibrium with the background gas. The electron temperature is then typically of few volts, while the ions are cold. In this situation we may think of monoenergetic ions being accelerated through the sheath potential, while the electron decreases according to a Boltzmann factor, as described in Subsection B.3. The electron density would then decay on the order of a Debye length λ_{De}, to shield the electrons from the wall. However, we cannot linearize the Poisson equation, as we did in deriving λ_{De} in Subsection B.1, if we wish to obtain the exact flux balance. Furthermore, we will show that a transition layer or *presheath* must exist between the neutral plasma and the nonneutral sheath in order to maintain the continuity of ion flux, giving rise to an ion velocity at the plasma-sheath edge known as the *Bohm velocity* u_B. The need for this sheath presheath will arise naturally in our derivation in next subsubsection C.5.2.

If a potential is placed between bounding electrodes, then, while the overall flux balance is maintained, each electrode may separately draw current. The most straightforward analysis is of a boundary with a large negative potential with respect to the plasma. The simplest example is a uniform ion charge density, or *matrix sheath*. This occurs in the cathode sheath of a dc discharge, for example. A matrix sheath is also created transiently with a pulsed negative electrode voltage in which the electrons are expelled from a plasma region, leaving a uniform ion density behind. This occurs naturally in *plasma immersion ion implantation*. We consider the matrix sheath in subsubsectionC.5.3.

For a high voltage-sheath, the current to the electrode is almost all ion current. Provided the ion motion in the sheath is collisionless, then the steady self-consistent ion density is not uniform, but rather is described by the *Child-Langmuir law* of *space-charge-limited current* in a planer diode. We also discuss this situation in subsubsectionC.5.3.

The idealized conditions described in subsubsection C.5.2 and C.5.3 are not always met. The temperature of the ions cannot always be ignored with respect to the electron temperature. This situation arises, for example, in highly ionized plasmas. In this case more complicated kinetic treatments are required. In a similar vein, the electron distribution may not be Maxwellian. This may arise due to particular heating or loss mechanisms, which ocurr, for example, in low-pressure capacitive rf plasmas. In this situation the decrease in electron density in the sheath is not given by a Boltzmann factor but must be obtained kinetically. If the neutral gas is electronegative, such that electron attachment is significant, then the negative charges divide between electrons and negative ions. If the fraction of negative ions present becomes large, the mobility of the negative charge can be greatly reduced, changing the conditions at the sheath edge. We consider these various topics, which, in fact, have some unity of analysis. Electronegative plasmas are of considerable importance in processing applications.

Other situations that differ from the basic theory arise due to collisional effects in the sheath region. In this case the ion flow is impeded by collisional processes with neutrals, and the transport is mobility rather than inertia limited. We discuss two simple limiting collisional cases in this subsection. A full treatment, including both inertial and collisional effects, is very complicated, requiring numerical solution of the kinetic equations.

This section deals with sheaths that are constant in time. Two other interesting cases are sheaths formed in oscillating rf potentials and sheaths formed transiently by pulsed potentials. In both situations approximate solutions can be obtained if there is a separation of time scales such that electrons respond rapidly to the time variation while ions respond slowly. This separation is characterized by the inequalities

$$f_{pe} \gg \frac{1}{\tau} \gg f_{pi} \tag{C.33}$$

where τ is the time scale of field variation ($\tau = 2\pi/\omega$ for an oscillatory variation) and f_{pe} and f_{pi} are the electron and ion plasma frequencies, respectively. An oscillatory potential applied to an electrode is characteristic of a capacitively excited rf discharge. But we will not consider this sheath, since an effect of the oscillatory potential is negligible in this paper.

The collisionless sheath We use the assumptions (1) Maxwellian electrons at temperature T_e, (2) cold ions ($T_i = 0$), and (3) $n_e(0) = n_i(0)$ at the plasma-sheath interface (interface between essentially neutral and non-neutral regions) at $x = 0$. As shown in Fig.C.3, we define the zero of the

potential Φ at $x = 0$ and take the ions to have a velocity u_s there. Ion energy conservation (no collisions) then gives

$$\frac{1}{2}Mu^2(x) + e\Phi(x) = \frac{1}{2}Mu_s^2(x). \tag{C.34}$$

The continuity of ion flux (no ionization in the sheath) is

$$n_i(x)u(x) = n_{is}u_s \tag{C.35}$$

where n_{is} is the ion density at the sheath edge. Solving for u from Eq.(C.34) and substituting in Eq.(C.35) we have

$$n_i = n_{is}\left(1 - \frac{2e\Phi}{Mu_s^2}\right)^{-1/2}. \tag{C.36}$$

The electron density is given by the Boltzmann relation

$$n_e(x) = n_{es}e^{\Phi(x)/T_e}. \tag{C.37}$$

Setting $n_{es} = n_{is} \equiv n_s$ at the sheath edge and substituting n_i and n_e into Poisson's equation

$$\frac{d^2\Phi}{dx^2} = \frac{e}{\epsilon_0}(n_e - n_i). \tag{C.38}$$

We obtain

$$\frac{d^2\Phi}{dx^2} = \frac{en_s}{\epsilon_0}\left[\exp\left(\frac{\Phi}{T_e}\right) - \left(1 - \frac{\Phi}{E_s}\right)^{-1/2}\right] \tag{C.39}$$

where $eE_s = \frac{1}{2}Mu_s^2$ is the initial ion energy. Eq.(C.39) is the basic nonlinear equation governing the sheath potential and ion and electron densities. However, as we shall seen in the next subsection, it has stable solutions only for sufficiently large u_s, created in an essentially neutral *presheath* region.

C.5.2 The Bohm sheath criterion

A first integral of Eq.(C.39) can be obtained by multiplying Eq.(C.39) by $d\Phi/dx$ and integrating over x:

$$\int_0^\Phi \frac{d\Phi}{dx}\frac{d}{dx}\left(\frac{d\Phi}{dx}\right)dx = \frac{en_s}{\epsilon_0}\int_0^\Phi \frac{d\Phi}{dx}\left[\exp\left(\frac{\Phi}{T_e}\right) - \left(1 - \frac{\Phi}{E_s}\right)^{-1/2}\right]dx. \tag{C.40}$$

Canceling the dx's and integrating with respect to Φ, we obtain

$$\frac{1}{2}\left(\frac{d\Phi}{dx}\right)^2 = \frac{en_s}{\epsilon_0}\left[T_e\exp\left(\frac{\Phi}{T_e}\right) - T_e + 2E_s\left(1 - \frac{\Phi}{E_s}\right)^{1/2} - 2E_s\right] \tag{C.41}$$

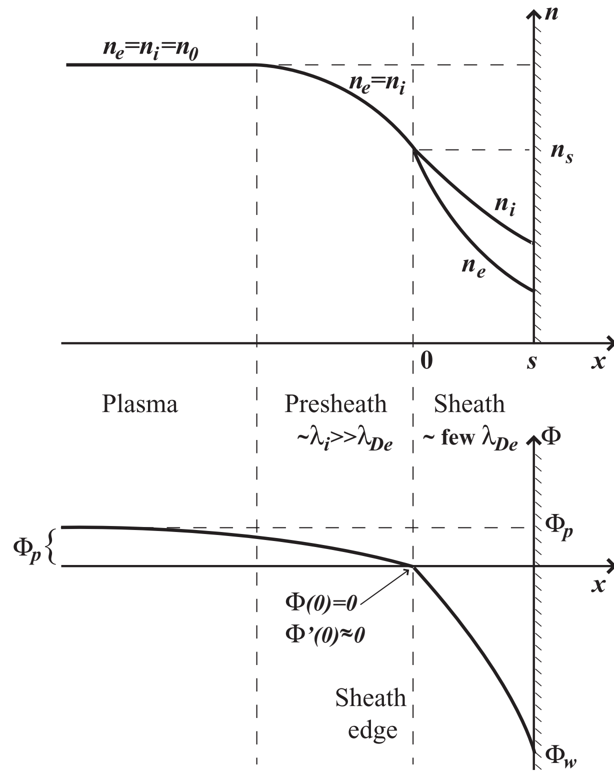


Figure C.3: Qualitative behavior of sheath and presheath in contact with a wall.

where we have set $\Phi = 0$ and $d\Phi/dx = 0$ at $x = 0$ corresponding to a field free plasma. Eq.(C.41) can be integrated numerically to obtain $\Phi(x)$. However, it is apparent that the RHS of Eq.(C.41) should be positive for a solution exist. Physically this means that the electron density must always be less than the ion density in the sheath region. Since we expect this to be a problem only for small Φ, we expand the RHS of Eq.(C.41) to second order in a Taylor series to obtain the inequality

$$\frac{1}{2}\frac{\Phi^2}{T_e} - \frac{1}{4}\frac{\Phi^2}{E_s} \geq 0. \tag{C.42}$$

We see that Eq.(C.42) is satisfied for $E_s \geq T_e/2$ or, substituting for E_s,

$$u_s \geq u_B = \left(\frac{eT_e}{M}\right)^{1/2}. \tag{C.43}$$

This result is known as the *Bohm sheath criterion*. To give the ions this directed velocity u_s, there must be a finite electric field in the plasma over some region, typically much wider that the sheath, called the *presheath* (see Fig.C.3). Hence the presheath region is not strictly field free, although E is very small there. Since the field at the edge between the sheath and the presheath is not precisely defined, only approximate solutions are obtained by mutching sheath to presheath solutions. Nevertheless, if we can make Eq.(C.43) sharper, by using the equality on the right, this relation is sufficient to obtain quantitative solutions for the plasma equilibrium. The procedure for doing this is to examine the solution in the presheath region of the plasma in the quasineutral approximation $n_i = n_e$, to see how the presheath solution joins with that of the sheath region. We sketch the calculation.

Presheath requirements Setting

$$n_i = n_e \tag{C.44}$$

within the presheath and taking the of the logarithm of Eq.(C.44) we have

$$\frac{1}{n_i}\frac{dn_i}{dx} = \frac{1}{n_e}\frac{dn_e}{dx}. \tag{C.45}$$

Substituting on the left for the ion current, through the relation $n_i = J_i/eu_i$, this becomes

$$\frac{1}{J_i}\frac{dJ_i}{dx} - \frac{1}{u_i}\frac{du_i}{dx} = \frac{1}{n_e}\frac{dn_e}{dx}. \tag{C.46}$$

Figure C.4: Φ/T_e versus position within the presheath, showing (a) the geometric presheath, (b) a planer collisional presheath, and (c) a planer ionization presheath. The sheath-presheath edge is at right. (Ref. from [5].)

Assuming the Boltzmann form of n_e and rearranging, Eq.(C.46) becomes

$$\frac{1}{u_i}\frac{du_i}{dx} + \frac{1}{T_e}\frac{d\Phi}{dx} = \frac{1}{J_i}\frac{dJ_i}{dx}. \tag{C.47}$$

In the presheath we expect the flow to be subsonic, with $u_i < u_B$, and applying this condition to Eq.(C.47) we have

$$\frac{1}{u_B}\frac{du_i}{dx} + \frac{1}{T_e}\frac{d\Phi}{dx} < \frac{1}{J_i}\frac{dJ_i}{dx}. \tag{C.48}$$

This satisfied for either

$$\frac{1}{u_B}\frac{du_i}{dx} + \frac{1}{T_e}\frac{d\Phi}{dx} < 0, \quad \frac{1}{J_i}\frac{dJ_i}{dx} = 0 \tag{C.49}$$

or

$$\frac{1}{u_B}\frac{du_i}{dx} + \frac{1}{T_e}\frac{d\Phi}{dx} > 0, \quad \frac{1}{J_i}\frac{dJ_i}{dx} > \frac{1}{u_B}\frac{du_i}{dx} + \frac{1}{T_e}\frac{d\Phi}{dx}. \tag{C.50}$$

Since the ion energy conservation Eq.(C.49) imply ion friction in the presheath, whereas taking the equality on the right implies current conservation. Relation Eq.(C.50) imply ionization or geometric contraction, the presheath equations can be solved analytically. This has been done, for example, for (a) a geometric presheath with current contraction onto a spherical probe, (b) a plane parallel collisional presheath, and (c) an ionizing presheath with the ionization proportional to n_e. These solutions are plotted in Fig.C.4 They show quite different behavior in the plasma region: The geometric presheath relaxes to the undisturbed (field free) plasma, the collisional presheath (b) tends to a logarithmic potential shape (indicating that the ion transport requires a residual plasma field), and the ionizing presheath (c) ends with zero field at a finite point representing the midplane of a symmetric plasma. For (b) or (c) the presheath width is of order the mean free path for ion-neutral collisions or for electron-neutral ionization, respectively. Despite the differences, all solutions run quite similarly into the singularity $u_i = u_B$ at the sheath edge. The growing field inhomogeneity approaching this singularity indicates the formation of space charge and the breakdown of the quasineutral approximation. Matching the ion velocity across the sheath-presheath

interface then gives us the equality $u_i = u_B$ for the sheath region. Of course, the true behavior is quite complicated at this interface, thus needing a more sophisticated treatment. For more details, including a kinetic treatment, it is discussed in Ref.[5].

The potential drop across the presheath, which accelerates the ions to the Bohm velocity, is given by

$$\frac{1}{2}Mu_B^2 = e\Phi_p \tag{C.51}$$

where Φ_p is the plasma potential with respect to the potential at the sheath-presheath edge. Substituting for the Bohm velocity from Eq.(C.43), we find

$$\Phi_p = \frac{T_e}{2}. \tag{C.52}$$

This is shown as the dashed line in Fig.C.3. The ratio of the density at the sheath edge to that in the plasma is then found from the Boltzmann relation

$$n_s = n_b e^{-\Phi_p/T_e} \approx 0.61 n_b \tag{C.53}$$

where n_b is the density where the presheath and bulk plasma join.

Sheath potential at a floating wall It is quite straightforward to determine the potential drop within the sheath between a plasma and floating wall. We equate the ion flux (assumed constant through the sheath),

$$\Gamma_i = n_s u_B \tag{C.54}$$

to the electron flux at the wall,

$$\Gamma_e = \frac{1}{4}n_s \bar{v}_e e^{\Phi_w/T_e} \tag{C.55}$$

where $\bar{v}_e = (8eT_e/\pi m)^{1/2}$ is the mean electron speed and Φ_w is the potential of the wall with respect to the sheath-presheath edge. We have, after substituting for the Bohm velocity from Eq.(C.43),

$$n_s\left(\frac{eT_e}{M}\right)^{1/2} = \frac{1}{4}n_s\left(\frac{8eT_e}{\pi m}\right)^{1/2} e^{\Phi_w/T_e}. \tag{C.56}$$

Solving for Φ_w, we obtain

$$\Phi_w = -T_e \ln\left(\frac{M}{2\pi m}\right)^{1/2}. \tag{C.57}$$

The wall potential Φ_w is negative and is related linearly to T_e with a factor proportional to the logarithm of the square root of the mass ratio. For hydrogen, for example, $\ln(M/2\pi m)^{1/2} \approx 2.8$, where for argon ($M = 40$ amu) the factor is 4.7. Thus argon ions with initial energy $E_s = T_e/2$ at the sheath-presheath edge that fall through a collisionless dc sheath to a floating wall would bombard the wall with an energy of $E_s \approx 5.3T_e$. Of course, electrodes that have potentials on them, either dc or rf, can be bombarded with much higher energy, but these electrodes must draw a substantial net current, as we will discuss in subsubsectionC.5.3.

The sheath width s is found by integrating Eq.(C.41) to obtain $\Phi(x)$ ans setting $\Phi(s) = \Phi_w$, with Φ_w given by Eq.(C.57). The integral must be done numerically. Typical sheath widths are a few electron Debye lengths λ_{De}.

C.5.3 The high voltage sheath

Matrix sheath Sheath voltages are often driven to be very large composed to T_e. The potential Φ in these sheaths is highly negative with respect to the plasma-sheath edge; hence $n_e \sim n_s e^{\Phi/T_e} \to 0$ and only ions are present in the sheath. The simplest high-voltage sheath, with a uniform ion density, is known as *matrix sheath*. Letting $n_i = n_s =$ const within the sheath of thickness s and choosing $x = 0$ at the plasma-sheath edge, then from Maxwell's equations???,

$$\frac{dE}{dx} = \frac{en_s}{\epsilon_0} \tag{C.58}$$

which yields a linear variation of E with x:

$$E = \frac{en_s}{\epsilon_0}x. \tag{C.59}$$

Integrating $d\Phi/dx = -E$, we obtain a parabolic profile

$$\Phi = -\frac{en_s}{\epsilon_0}\frac{x^2}{2}. \tag{C.60}$$

Setting $\Phi = -V_0$ at $x = s$, we obtain the matrix sheath thickness

$$s = \left(\frac{2\epsilon_0 V_0}{en_s}\right)^{1/2}. \tag{C.61}$$

In terms of the electron Debye length $\lambda_{De} = (\epsilon_0 T_e/en_s)^{1/2}$ at the sheath edge, we see that

$$s = \lambda_{De}\left(\frac{2V_0}{T_e}\right)^{1/2}. \tag{C.62}$$

Hence the sheath thickness can be tens of Debye lengths.

Child law sheath In the steady state, the matrix sheath is not self-consistent since it does not account for the decrease in ion density as the ions accelerate across the sheath. In the limit that the initial ion energy E_s is small compared to the potential, the ion energy and flux conservation equations Eq.(C.34) and Eq.(C.35) reduce to

$$\frac{1}{2}Mu^2(x) = -e\Phi(x) \tag{C.63}$$

$$en(x)u(x) = J_0 \tag{C.64}$$

where J_0 is the constant ion current. Solving for $n(x)$, we obtain

$$n(x) = \frac{J_0}{e}\left(-\frac{2e\Phi}{M}\right)^{-1/2}. \tag{C.65}$$

Using this in Poisson's equation, we have

$$\frac{d^2\Phi}{dx^2} = -\frac{J_0}{\epsilon_0}\left(-\frac{2e\Phi}{M}\right)^{-1/2}. \tag{C.66}$$

Multiplying Eq.(C.66) by $d\Phi/dx$ and integrating from 0 to x, we have

$$\frac{1}{2}\left(\frac{d\Phi}{dx}\right)^2 = 2\frac{J_0}{\epsilon_0}\left(\frac{2e}{M}\right)^{-1/2}(-\Phi)^{-1/2} \tag{C.67}$$

where we have chosen $d\Phi/dx = -E = 0$ at $\Phi = 0$ ($x = 0$). Taking the (negative) square root (since $d\Phi/dx$ is negative) and integrating again, we obtain

$$-\Phi^{3/4} = \frac{3}{2}\left(\frac{J_0}{\epsilon_0}\right)^{1/2}\frac{3}{2}\left(\frac{2e}{M}\right)^{-1/4}x. \tag{C.68}$$

Letting $\Phi = -V_0$ at $x = s$ and solving for J_0, we obtain

$$J_0 = \frac{4}{9}\epsilon_0\left(\frac{2e}{M}\right)^{1/2}\frac{V_0^{3/2}}{s^2}. \tag{C.69}$$

Eq.(C.69) is the well-known Child law of space-charge-limited current in a plane diode. With fixed spacing s it gives the current between two electrodes as a function of the potential difference between them, and has been traditionally used for electron diodes. However, with J_0 given explicitly as

$$J_0 = en_s u_B \tag{C.70}$$

in Eq.(C.69), we have a relation between the sheath potential, the sheath thickness, and the plasma parameters, which can be used to determine the

sheath thickness s. Substituting Eq.(C.70) in Eq.(C.69) and introducing the electron Debye length at the sheath edge, we obtain

$$s = \frac{\sqrt{2}}{3}\lambda_{De}\left(\frac{2V_0}{T_e}\right)^{3/4}. \tag{C.71}$$

Comparing this to the matrix sheath width, we see the Child law sheath is larger by a factor of order $(V_0/T_e)^{1/4}$. Hence the Child law sheath can be of order of 100 Debye lengths ($\sim$ 1cm) in a typical processing discharge. Since there are no electrons within the sheath to excite the gas, the sheath region appears dark when observed visually.

Inserting Eq.(C.69) into Eq.(C.68) yields the potential within the sheath as a function of position

$$\Phi = -V_0\left(\frac{x}{s}\right)^{4/3}. \tag{C.72}$$

The electric field $E = d\Phi/dx$ is

$$E = \frac{4}{3}\frac{V_0}{s}\left(\frac{x}{s}\right)^{4/3} \tag{C.73}$$

and the ion density $n = (\epsilon_0/e)dE/dx$ is

$$n = \frac{4}{9}\frac{\epsilon_0}{e}\frac{V_0}{s^2}\left(\frac{x}{s}\right)^{-2/3}. \tag{C.74}$$

We see that n is singular as $x \to 0$, a consequence of the simplifying assumption in Eq.(C.63) that the initial ion energy $E_s = 0$. The analysis can be carried through for a finite $eE_s = \frac{1}{2}Mu_B^2$, using Eq.(C.34), resolving the singularity and yielding $n \to n_s$ as $x \to 0$.

The Child law solution is valid if the sheath potentials are large compared to the electron temperature. It is therefore not appropriate for use where the sheath potential is the potential between a plasma and a floating electrode.

C.6 Diagnostics of plasma: electrostatic probe diagnostics

We used electrostatic probe to diagnose plasma. A metal probe, inserted in a discharge and biased positively or negatively to draw electron or ion current, is one of the earliest and still one of the most useful tools for diagnosing a plasma. These probes, introduced by Langmuir and analyzed in considerable detail by Mott-Smith and Langmuir[1] are usually called *Langmuir probes.*

As with any other electrode, the probe is surrounded by a sheath, such that its analysis naturally fits into the present subsection. However, unlike large electrode surfaces that are used to control a plasma, probes are usually quite small and under suitable conditions, produce only minor local perturbations of the plasma.

The voltage and current of a probe defined in Fig.C.5 lead to a typical probe voltage-current characteristic as shown in Fig.C.6. At the probe voltage $V_B = \Phi_p$, the probe is at the same potential as the plasma and draws mainly current from the more mobile electrons, which is designated as positive current flowing from the probe into the plasma. For increasing V_B above this value, the current tends to saturate at the electron saturation current, but, depending on the probe geometry, can increase due to increasing effective collection area. For $V_B < \Phi_p$, electrons are repelled according to the Boltzmann relation, until at Φ_f the probe is sufficiently negative with respect to the plasma that the electron and ion currents are equal such that $J = 0$. Φ_f is known as the *floating potential*, because it is the potential at which an insulated probe, which cannot draw current, will float. For $V_B < \Phi_f$, the current is increasingly ion current (negative into the plasma), tending to an ion saturation current that may also vary with voltage due to a change of the effective collection area. The magnitude of the ion saturation current is, of course, much smaller than the electron saturation current due to the much greater ion mass.

The basic theory for a plane collection area, based on the sheath calculations, is quite simple. However, to minimally disturb the plasma and also for ease of construction, Langmuir probes are often thin wires with the wire radius $a < \lambda_{De}$. The trajectories of charged particles in the sheath then become important in determining the collected current, and the analysis becomes quite complicated. As the voltage is raised, either to large positive or large negative values with respect to the plasma, the sheath thickness s increases according to Child's law, and consequently the effective collecting area also increases. If $T_i \sim T_e$, then additional complications arise to make calculations very involved. There are also difficulties if the momentum transfer mean free paths $\lambda_i, \lambda_e \lesssim s$, which can occur in high-pressure discharges. A review of the analysis, including many of these complications, is given by Chen.[2] The extension to $T_i \sim T_e$, which is not usually of great interest in processing discharges, is given in a report by Laframboise.[3]

Probe theory has generally been developed for plasmas in which the electron distribution is approximated by a Maxwellian. Plasma densities obtained from the ion saturation current to probes have been compared with other measurement techniques such as microwaves. Generally, the comparisons have indicated that probe-predicted densities, using ion saturation cur-

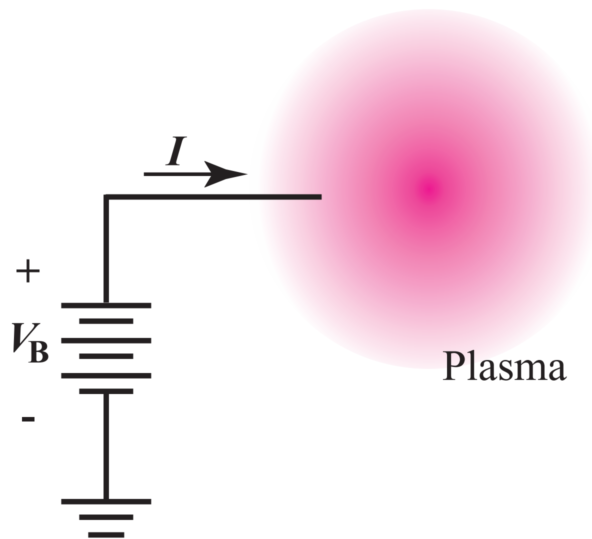


Figure C.5: Definition of voltage and current for a Langmuir probe.

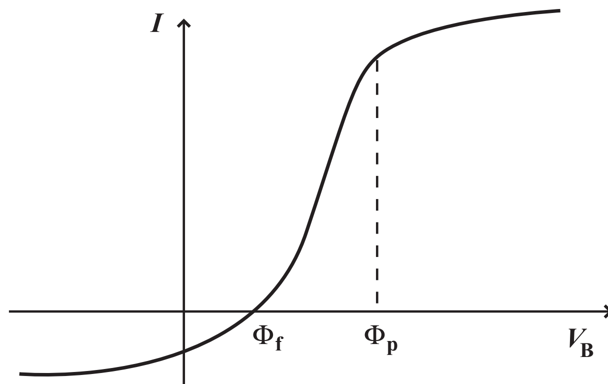


Figure C.6: Typical $I - V_B$ characteristic for a Langmuir probe.

rent, are somewhat high when compared under conditions for which the microwave predictions are expected to be high accurate. This result would generally agree with the arguments presented by Godyak.[4] However in many situations, the densities obtained by probe and microwave techniques are quit close. The accuracy of using the ion saturation current to measure the plasma density depends on the closeness of the electron distribution to an assumed Maxwellian at the probe sheath edge, and therefore to the type of plasma being diagnosed. Hence, for convenience of the experimental setup, we measured the plasma density using the ion saturation current.

Next, we shall briefly discuss practical probes and circuits for their use.

C.6.1 Planer probe with collisionless sheath

Consider a flat plate probe with the (two-sided) physical probe area $A \gg s^2$, where s is the sheath thickness, such that the collecting area A is essentially independent of s. As we discussed in previous section, if a large voltage is applied to the probe, then $s \gg \lambda_{De}$, and we find that A is quite large to satisfy the above condition. For this reason we expect that biasing the probe strongly positive to collect only electron current would strongly perturb plasma. Consider therefore that the probe is biased sufficiently negatively to collect only ion current. From Eq.(C.70) the current "collected" (see direction in Fig.C.5) by the probe is

$$I = -I_i = -en_s u_B A \tag{C.75}$$

where, as in Eq.(C.43) with $T_i \ll T_e$, the Bohm velocity u_B is given by

$$u_B = \left(\frac{eT_e}{M}\right)^{1/2}. \tag{C.76}$$

If we know T_e, then the density at the sheath edge n_s is determined from the measurement of I_i. As in Eq.(C.53), the plasma density in the probe neighborhood is then obtained as

$$n_0 \approx \frac{n_s}{0.61}. \tag{C.77}$$

Since the electron temperature in most discharge is clamped in the range of 2-5V by particle balance, a reasonable estimate of density can be obtained without knowing T_e. However, by varying the probe voltage, it is also straightforward to measure T_e. Considering that the probe potential is retarding with respect to the plasma potential, then, using Boltzmann's relation, the electron component of the probe current is

$$I + I_i = I_e = \frac{1}{4} en_s \bar{v}_e A \exp\left(\frac{V_B - \Phi_p}{T_e}\right) \tag{C.78}$$

where $\bar{v}_e = (8eT_e/\pi m)^{1/2}$, and $V_B - \Phi_p < 0$ is the potential between the probe and the plasma. There is an exponential increase in I_e with increasing V_B in this range. Defining an electron saturation current

$$I_{esat} = \frac{1}{4} e n_s \bar{v}_e A \tag{C.79}$$

and taking the logarithm of Eq.(C.78), we have

$$\ln\left(\frac{I_e}{I_{esat}}\right) = \left(\frac{V_B - \Phi_p}{T_e}\right). \tag{C.80}$$

From Eq.(C.80) we see that the inverse slope of the logarithmic electron probe current with respect to V_B (in volts) gives T_e directly in volts.

The above simple interpretation is limited by the dynamic range over which Eq.(C.78) holds. For I_e too small, adding the measured I_i to I can introduce errors in the determination of I_e. For V_B too large, the Boltzmann exponential no longer is accurate, as electron saturation is approached. The nominal useful range of voltages over which the slope can be measured is then

$$\frac{|\Delta V_B|}{T_e} \approx \ln\left(\frac{\bar{v}_e}{4u_B}\right) = \ln\left(\frac{M}{2\pi m}\right) \tag{C.81}$$

which is approximately 4.7 for argon. This range is sufficient, provided there are no geometric complications.

The floating potential Φ_f and the plasma potential Φ_p are often of interest in discharge operation. The floating potential is the potential at which the probe draws equal electron and ion currents. If the plasma is mainly surrounded by grounded conducting surface, then we would expect the floating potential to lie near this ground, as shown in Fig.LangmuirIV. This follows because the ground is usually not, itself, drawing significant net current, and thus at $V_B = \Phi_f$ the probe behaves as part of the ground. The plasma (space) potential, given by Eq.(C.81) with $\Phi_p - \Phi_f = \Delta V_B$, can be approximately determined from the knee (point of maximum first derivative) of the electron saturation portion of the $I - V_B$ characteristic of Fig.C.6. For planer probes the knee is easily recognizable, but the current drawn may be too large, either modifying the plasma or destroying the probe.

C.7 Cold plasma generation

The cold plasma are usually excited and sustained electrically by direct current (DC), radio frequency (RF), or microwave (MW) power applied to a gas. In fact, I used these three methods of generation of plasma. Plasma

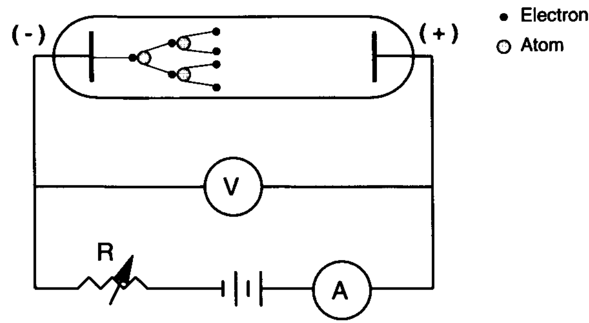


Figure C.7: DC glow discharge setup.

chemistry in cold plasma is controlled mainly by electron energies and gas temperatures. Therefore, as far as identical energies and temperatures can be achieved, the type of discharge used to create the plasma is of little importance. The choice of a specific method and equipment to produce discharge is determined by requirements of flexibility, process uniformity, cost, and spended time. The different methods used for the generation of cold plasmas are described in the following subsubsections.

C.7.1 DC grow discharges

A DC glow discharge is produced by applying DC voltage between two conductive electrodes inserted into a gas at low pressure as illustrated in Fig.C.7. A high impedance power supply is used to provide the electrical field.

A small amount of free electrons is always present in the gas, as a result of ionization by naturally occuring radioactivity or cosmic lays. Free electrons can also be produced by photoionization or field emission. As the voltage applied to the gas in the discharge tube is gradually increased, the available free electrons are accelerated in the electric field, thereby gaining kinetic energy. Concomitantly, the electrons lose energy in collisions with the atoms or molecules of the gas. These atoms or molecules will be referred to also as collision targets.

Initially, when the energy of the electron is too low to excite or ionize a target, the collisions will necessarily be elastic. The average fraction of the electron energy lost in an elastic collision with a gas atom or molecule us $-2m_e/M$, where M is the mass of the target (Eq.(C.7)). Thus only a very small fraction of the total kinetic energy of the electron, typically only 10^{-5}, is lost per elastic collision. Meanwhile, the electron continues

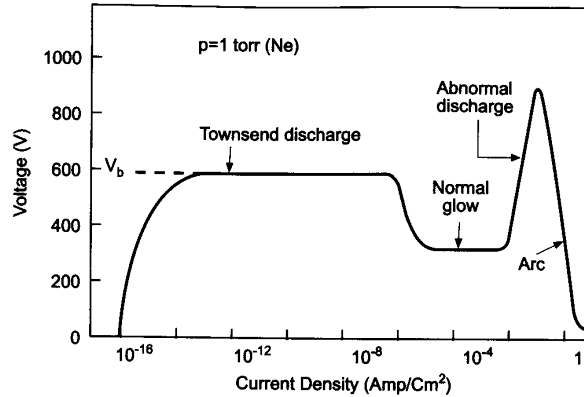


Figure C.8: The $I-V$ characteristic of a DC glow discharge.

to gain energy between collisions until it attains sufficient energy to cause ionization of the targets through inelastic collisions. Large amounts of energy are transferred to the targets in the inelastic collisions, making those collisions a more efficient mean of energy transfer. The new electrons produced in the ionization process are in turn accelerated by the electric field and produce further ionization by impact with the neutral atoms or molecules of the gas.

An electron multiplication process thus takes place. This process can be characterized by a macroscopic coefficient, σ_T, which represents the mean number of ion-electron pairs formed by an electron along a path of 1cm. The coefficient σ_T is called the *first Townsend coefficient*, and is dependent on the electric field (E), the pressure (p), and the nature of the gas as well. Experimentally, it has been shown that for a given gas, σ_T/p depends on E/p only.

The changes that take place in the gas as a function of the applied voltage are described by the typical dependence of discharge current, I on the applied voltage, V. This dependence is called the $I-V$ *characteristic* of the discharge. A representative typical characteristic of a DC glow discharge is shown in Fig.C.8. When the applied voltage is low, the current through the tube is produced by the collection of the free charges and is negligibly small. When the voltage increases and more charged particles are created by the ionization of the gas, the current increases steadily, while the voltage reaches a limit determined by the output impedance of the power supply. This region of the $I-V$ characteristic is known as the *townsend discharge* and indicated in Fig.C.8.

When the applied voltage reaches a certain threshold value, marked V_b in Fig.C.8, an avalanche process occurs as a result of three simultaneous

process:

1. Ions accelerated by the electric field strike the cathode with sufficient energy to cause the *emission of secondary electrons* from it; the secondary electrons in turn form more ions by collision with the neutral atoms of the gas.

2. The newly formed ions are accelerated to the cathode where collisions produce more electrons, which in turn more ions.

3. In parallel, electrons created in ionizing collisions and by secondary processes are removed from the plasma by drift and diffusion to the walls, by recombination with positive ions, and in certain gases, by formation of negative ions. At pressures lower than a few mtorr, the recombination occurs mainly at the wall of the reactor.

When the number of electrons is sufficient to produce just enough ions to regenerate the number of lost electrons, a steady state is reached in which an equilibrium is established between the rate of formation of ions and the rate of their recombination with electrons. At this stage the discharge is self-sustaining. Extensive breakdown occurs in the gas and the *glow discharge* is thus established. The gas begins to glow, the voltage drops, and the current rises abruptly. The mode of the discharge at this point is called the *normal glow* (see Fig.C.8). The energy decay of the electronically excited states of molecules and atoms account for much of the luminous glow of the gaseous discharge.

The minimal threshold voltage required to produce the glow discharge, V_b, is called the *breakdown voltage*.

The sustaining glow discharge is controlled by the emission of secondary electrons. That emission is characterized by the *electron emission coefficient*, defined as the ratio between the number of the emitted secondary electrons and the number of impacting ions. This coefficient is often called the *yield of secondary electrons*. Because the electron emission coefficient is of the order of 0.1 for most materials, much more than one ion must strike the cathode to produce one electron. Then, depending on losses, each secondary electron must be responsible for 10-20 ionizations to sustain a continuous stable plasma.

Although ion impact on the cathode is the main source of secondary electrons in a DC glow discharge, other mechanisms can also contribute to secondary electron emission. Among those, the most dominant are:

1. Ionization of gas atoms or molecules by the photons emitted during ion-electron recombination (*photoionization*)

2. Emission of secondary electrons by the cathode due to impact of photons (*photoemission*)
3. Emission of secondary electrons by the cathode due to impact of excited atoms in a metastable state

The contribution of each of these effects to the emission of secondary electrons is strongly dependent on the nature of the gas and its pressure.

After the voltage reaches the breakdown value V_b and the glow discharge is established, initially at low power, the discharge covers only the area near the rim of the cathode. When the power is raised, the current increases and the discharge spreads to cover the whole surface of the cathode. When the current of the discharge increases past the point of complete cathode coverage, the voltage begins to rise. This corresponds to the *abnormal discharge* region in Fig.C.8, which is the mode used in glow discharge discharge processing[10]. At this stage, the discharge current is limited by the surface area of the electrodes and by the resistance of the power supply and electrical circuit. A further increase in power will cause the heating of the cathode, which will, at a certain stage, result in thermionic emission. When this happens, the voltage decreases and the glow changes into an arc.

Paschen's law

The breakdown voltage in a DC discharge is determined by the discharge gas, the gas pressure, and the tube dimension. The dependence of the breakdown voltage on gas pressure and inter-electrode distance can be expressed as:

$$V_b = \frac{C_1(pd)}{C_2 + \ln(pd)} \tag{C.82}$$

where $d =$ distance between electrodes, C_1 and $C_2 =$ constants that change with the nature of the gas. According to Eq.C.82, for large pd values the breakdown voltage is proportional to pd. This dependence is called *Paschen's law*, which reflects in fact the dependence of the DC glow discharge on the secondary electrons.

If the distance between the electrodes is small or the pressure is small, the secondary electrons emitted from the cathode can reach the anode while undergoing only a very small number of collisions not creating a sufficient number of ions required for the regeneration of the secondary electrons. On the other hand, if the pressure is too high, the electrons cannot acquire sufficient energy between collisions to produce enough ions. If the distance between electrodes is too large, only a small fraction of the produced ions will succeed in reaching the cathode and create secondary electrons.

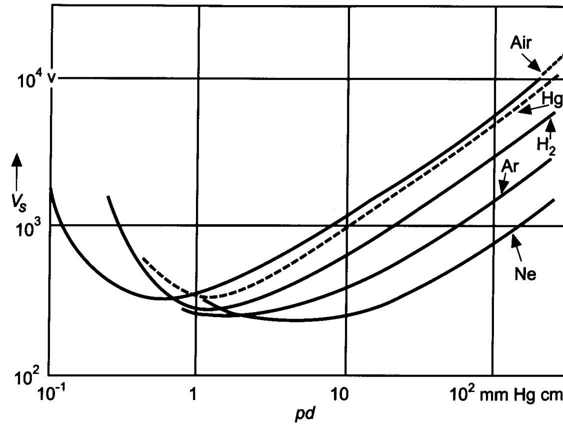


Figure C.9: Paschen's curves for various gases. From [11].

At both extremes of the value of the pd product, the probability of ionization - and/or ion collection - is small and the breakdown voltage required to sustain the discharge high. The breakdown voltage reaches a minimum between the two extremes. This behavior is described by Paschen's curves, which are shown for a few gases in Fig.C.9.

For most gases the minimum breakdown voltage is between 100 and 500V and occurs for pd in the range of $10^{-1} \sim 10$ torr cm. At 1 torr, typical values of the electric field are $10 \sim 100$ V cm^{-1}. Impurities inn the can reduce the breakdown voltage, either through Penning ionization or when a low ionization potential.

Characteristics of DC glow discharge

Fig.C.10 shows a diagram of the regions of the discharge as they appear in the discharge tube together with the distribution of the potential and electric field along tube. As can be seen in Fig.C.10(a), the discharge exhibits several bright and dark region along the discharge tube. The dark regions are called the *cathode or Crooke's dark space, the Faraday dark space,* and the *anode dark space.* The luminous regions are called the *cathode glow*, the *negative glow*, and the *positive column.*

The luminosity of the cathode glow results from the decay of the excitation energy of the positive ions during neutralization with electrons.

The cathode dark space is a positive charge plasma sheath. At a given applied discharge voltage the product between the thickness of the cathode dark space, d_{cd}, and the pressure in the discharge, p, is approximately constant.

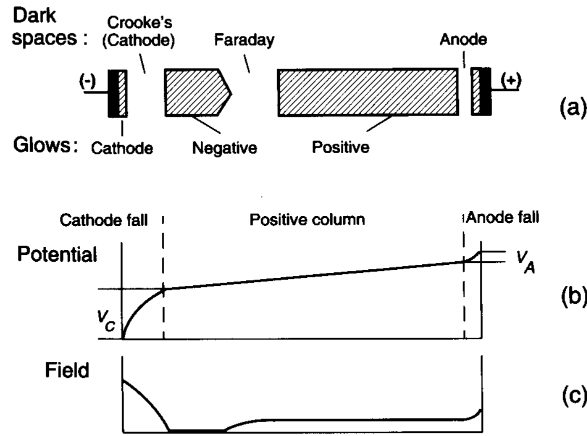


Figure C.10: Regions and characteristics of a DC glow discharge : (a) discharge regions; (b) potential distribution in discharge tube; (c) distribution of electric field in discharge tube. From [11].

For example, for a discharge in argon at 2000V, $p \cdot d_{cd} = 50$ mtorr cm. THe secondary electrons produced at the cathode by positive ion bombardment gain energy through the cathode dark space.

Most of the glow discharge volume is occupied by the positive column.

The voltage drop occurs over three regions indicated in Fig.C.10(b): the *cathode fall*, the *positive column*, and the *anode fall*. Most of the voltage drop occurs in the cathode fall region, whose dimension is a function of the material used for the cathode and the nature of the discharge gas. Values range of the voltage drop in the cathod fall are from 59V to 500V for different combinations of cathod material and discharge gas.

Since at constant voltage the product of the thickness of the dark space with pressure is fairly constant, the distance over which the cathode fall extends decreases with increasing pressure. For the pressures normally used in DC glow discharges, these value of the product $p \cdot d_{cd}$ for several combinations of cathode material and discharge gas corresponds to dark space thickness, d_{cd}, of order of milimeters to centimeters.

The voltage drop at the anode fall is of the same order as the ionization potential of the discharge gas and usually extends over a few millimeters. In the anode region secondary electrons are created primarily from electron bombardment. These are accelerated in the anode fall and are a source of particles and energy.

Most of the ionization needed to supply the ions that are accelerated to

the cathode takes place in the cathode sheath. However, high electric fields extend into the negative glow, causing ionization also in the negative glow region.

Chemical reactions occur mainly in the cascade fall and positive column zone.

When the distance between the electrodes in a DC discharge decreases, the Faraday dark space and positive column shrink and finally disappear, leaving only the negative glow and cathode dark space. This is generally the situation encountered in most glow-discharge processing reactors. The minimum distance between the electrodes has to be about twice the thickness of the cathode dark space; at smaller electrode separation, the discharge is extinguished.

C.7.2 Radio frequency discharges

To sustain a DC discharge, electrically conductive electrodes have to be inserted inside a reactor and be in direct contact with the plasma. If the discharge is used for deposition of dielectric films, the electrodes exposed to the plasma gradually become covered with an insulator. Therefore, although a DC discharge may be initiated, it will quickly extinguish as the electrons accumulate on the insulator and recombine with the available ions. In some cases it is preferred to have the electrodes outside the reactor, to avoid or minimize contamination of the process by material removed from the electrodes. Such problems can be solved by alternating the polarity of the discharge.

When an alternating electric field of low frequency (< 100Hz) is applied between two electrodes of the discharge tube, each electrode acts alternately as cathode or anode. Once the breakdown potential is surpassed on each half cycle, a temporary DC glow discharge is obtained. When the voltage drops during the cycle below the breakdown value, the discharge is extinguished, and for sufficiently low frequencies, the space charge decays before the discharge is reinitiated with inverse polarity.

When the frequency of the electric field increases above a *critical ion frequency*, f_{ci}, defined by Eq.C.83, the time taken by the positive ions to move between electrodes becomes larger than half the period of the electric field. Ions created near a momentary anode cannot reach the cathode before the field is reversed. In this situation, the distance traveled by the ions in the electric field becomes smaller than the thickness of the plasma sheath. At such frequencies the positive space charge is partly retained between the two half cycles of the alternating electric field and facilitates the reinitiation of the discharge. The critical ion frequency, also called the *ion transition*

frequency, is defined by:

$$f_{ci} = \frac{< v >_{di}}{2L} \tag{C.83}$$

where $< v >_{di}$= average drift velocity of the ion, L = the distance between the two electrodes.

A similar critical frequency can be defined for electrons:

$$f_{ce} = \frac{< v >_{de}}{2L} \tag{C.84}$$

where $< v >_{de}$= average drift velocity of the electrons.

Due to the much larger mobility of the electrons as compared to that of the ions, f_{ce}, is much higher than f_{ci}. For frequencies higher than f_{ce}, both positive and negative space charges are retained between cycles, and as a result, the voltages required to initiate and maintain the alternating current discharge decrease strongly in comparison to a DC glow discharge.

The frequencies used in the high-frequency discharges are in the range of radio transmission giving the high-frequency discharges the name of *radio frequency*, or shortly, *RF discharges.*

The elastic collision frequency, ν, in gases at glow discharge conditions is normally between 10^9 and 10^{11} collisions/sec. Thus the collisions frequency is much higher than the applied radio frequency even for 13.56MHz discharges and electrons will experience many collisions during each applied field cycle. They will be lost by diffusion to the reactor walls and will be regenerated by impact ionization in the body of the plasma. Therefore, the loss of electrical carriers from the RF discharge is controlled by ambipolar diffusion and homogeneous recombination (recombination in the gas phase) and not by the electric field. New charged particles are produced mainly through electron impact ionization of the neutral gas atoms and molecules.

The *power absorption* by the RF discharge can be either collisional ot collisionless. The collisional absorption of high-frequency power in the plasma is due to collisions of electrons with ions, at a frequency ν_{ei}, and with neutral particles, at a frequency ν_{en}. In plasmas at pressures above 10^3Pa (7.5torr), the degree of ionization is usually very low ($< 10^{-4}$), the density of neutrals is much higher than that of ions, and the electron-neutral collisions are predominant. At pressures below 1Pa (7.5mtorr), the degree of ionization can reach values higher than 10^{-2}, and the electron-ion collisions are predominant. At intermediate pressures, when the collision frequency decreases and $\nu/\omega \ll 1$, collisionless absorption becomes dominant in the plasma.

In a collisionless situation, an electron would oscillate in the RF field and would reach maximal velocity $\dot{x}$, amplitude x, and energy W, given by

$$\dot{x} = \frac{eE_0}{m_e\omega} \tag{C.85}$$

$$x = \frac{eE_0}{m_e \omega^2} \tag{C.86}$$

$$W = \frac{m_e \dot{x}^2}{2} \tag{C.87}$$

where E_0 is the amplitude of the electric field.

At a typical RF frequency of 13.56MHz and a field strength of 10V/cm, this corresponds to

$$\text{amplitude} = 2.42\text{cm};\ \ \text{velocity} = 2.1 \times 10^8 \text{cm s}^{-1};\ \ \text{energy} = 11.3\text{eV}. \tag{C.88}$$

This indicates that, for an electron to reach the ionization energy of argon (15.7eV) in a collisionless plasma, a field somewhat higher than 10V/cm is required. However, the collision with the atoms of the gas cause a random motion of the electrons and the electrons acquire additional energy from the external field during each collision with the atoms. If an electron makes an elastic collision with an atom, reversing its motion at the time the electric field changes direction, it will continue to gain speed and energy. Electron in a RF discharge could thus accumulate enough to cause ionization even at low electric fields. As a result of this behaviour, the RF discharge is more efficient than the DC discharge in promoting ionization and sustaining the discharge.

The *mean power absorbed* by an electron, $\bar{P}$, is given by

$$\bar{P} = \frac{e^2 E_0^2}{2m_e} \frac{\nu_{ea}}{\nu_{ea}^2 + \omega^2} \tag{C.89}$$

where ν_{ea} is the elastic collision frequency of an electron with atoms or molecules of the gas. The dependence of $\bar{P}$ on E_0^2 indicates that the absorbed power is independent of the sign of the electric field and the electron gains energy both when it moves with the field or against it. The quantity

$$E_{eff} = \frac{E_0}{\sqrt{2}} \left(\frac{\nu_{ea}^2}{\nu_{ea}^2 + \omega^2} \right)^{1/2} \tag{C.90}$$

is called the *effective electric field strength.*

The *average RF power* transferred from the outside electric field to the *unit volume of gas*, $\bar{P}_\nu$, is

$$\bar{P}_\nu = \frac{n_e e^2 E_0^2}{2m_e} \left(\frac{m}{\nu^2 + \omega^2} \right). \tag{C.91}$$

For RF frequency of $\omega \approx 10^7$Hz, and with $\nu > 10^9 \text{sec}^{-1}$, the collision frequency is much higher than the frequency of the field $\nu \ll \omega$. In this case

the transferred power, $\bar{P}_\nu$, is practically not affected by the driving frequency, ω.

At frequencies higher than 50kHz, the oscillating electrons acquire sufficient energy to cause ionizing collisions, thus reducing the dependence of the discharge on secondary electrons and lowering the breakdown voltage. The discharge can thus be sustained independent of the yield of secondary electrons from walls and electrodes. Because of the low mobility of the ions, their distribution is essentially stationary at high frequencies. In the same time, the electrons are swept between the electrodes by the electric field and their distribution is a function of position and time.

The secondary electrons emitted from the electrodes or walls of the reactor of an RF discharge are accelerated across the plasma sheath and add to the ionization process but are less important in sustaining the discharge. At a given pressure, the electrical impedance of the discharge decreases with increasing frequency, making it possible to drive more current through the discharge at the same voltage. The yield of atoms and free radicals in a molecular discharge is also increased by the use of RF excitation as compared to a DC discharge of the same field strength and pressure.

Although frequency of power supplies used for the excitation of RF plasmas can be as high as 100 MHz, lower frequencies are generally used. This is due to the fact that sophisticated arrangements have to be made to sustain uniform plasma in large volumes when the wavelength becomes comparable to, or shorter than, the dimensions of the reactor.

The RF discharge can be operated at pressures as low as 1mtorr because the efficiency of ionizing collisions is enhanced by the electron oscillations. This is useful in sputtering, where it is undesirable to have sputtered material reflected back to the etched surface as a result of collisions with gas molecules, or in etching or deposition, when directionality of the ions is required.

Self-bias in RF plasmas

Let's consider a RF plasma generated between two parallel electrodes, as shown in Fig.C.11, and assume that one electrode has an area much larger than the other. The electrodes will be at negative potentials, V_1 and V_2, relative to the plasma and the sheaths of thickness d_{s1} and d_{s2} will develope near the two electrodes fo area A_1 and A_2. If the RF field is connected directly to the electrodes, as illustrated in Fig.C.11(a), the two electrodes will be at the same potential relative to the plasma (Fig.C.11(b)), because the plasma is equipotential:

$$V_1 = V_2, \tag{C.92}$$

$$d_{s1} = d_{s2}. \tag{C.93}$$

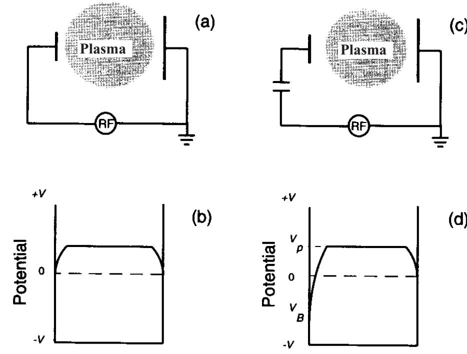


Figure C.11: RF connection to parallel plate reactor with unequal electrodes and developed self-bias: (a) direct RF connection to electrodes; (b) potential distribution between electrodes for arrangement (a); (c) RF connection through blocking capacitor; (d) potential distribution between electrodes for arrangement (c). From [11].

The situation changes, however, if a blocking capacitor is inserted, as often happens, between the RF supply and the electrodes, as shown in Fig.C.11(c). In this case, the potential distribution between the electrodes is as illustrated in Fig.C.11(d) and this nonsymmetric potential develops as explained next.

If a square-wave potential of amplitude V, as illustrated in Fig.C.12(a), is applied to the electrodes through the capacitor, the voltage across the plasma will follow the curve shown in Fig.C.12(b). Initially the voltage across the plasma will be equal to the applied voltage V. The capacitor will charge up rapidly by electron current, and the potential will drop as shown in Fig.C.12(a). When the applied voltage decays at a slower rate because the capacitor charges this time by the current of less mobile ions. This process continues until the time-averaged ion and electron currents become equal.

This steady state results in a *time-averaged negative bias* on the small electrode. The time-averaged potential distribution between the electrodes is therefore as illustrated in Fig.C.11(b). The same will happen if a sinusoidal voltage is applied to the electrodes, as illustrated in in Fig.C.13(a). This figure also illustrates the time variation of the plasma potential (curves 2 for the symmetric and nonsymmetric cases. If the electrodes have the same area, the system remains symmetric, as shown in Fig.C.13(b).

In all cases, the plasma is at higher potential than each electrode.

When a blocking capacitor is used with two electrodes of different areas,

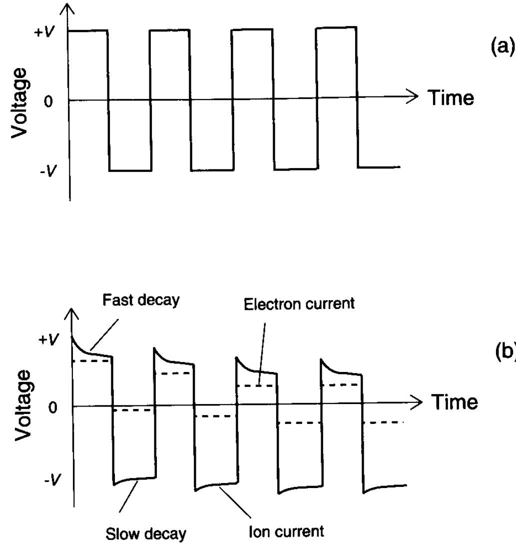


Figure C.12: Development of a self-bias in a parallel plate discharge: (a) applied voltage; (b) voltage across the discharge versus time. From [11].

the negative self-bias of the electrodes relative to the plasma is independent on the relative areas of the two electrodes. According to Koenig and Maissel[12], the following relation exists,

$$\frac{V_1}{V_2} = \left(\frac{A_2}{A_1}\right)^4, \tag{C.94}$$

$$\frac{d_{s1}}{d_{s2}} = \left(\frac{V_1}{V_2}\right)^{3/4} \tag{C.95}$$

where V_1, V_2 = negative biases on the two electrodes and A_1, A_2 = areas of the electrodes. Often, one of the electrode is grounded together with the walls of the reactor and its effective area becomes very large, causing the other electrode to become much more negative.

Eq.C.94 was obtained assuming that the ions pass through the sheath without collisions and that the current density of the ions is equal at the two electrodes. The first assumption is correct only at pressures of a few mtorr, and the second assumption is questionable. Eq.C.94 takes into consideration the total areas of the electrodes, which includes the area of the reactor for the electrode that is electrically connected to it. However, the plasma is often mostly confined between the electrodes while the walls are practically not exposed to it.

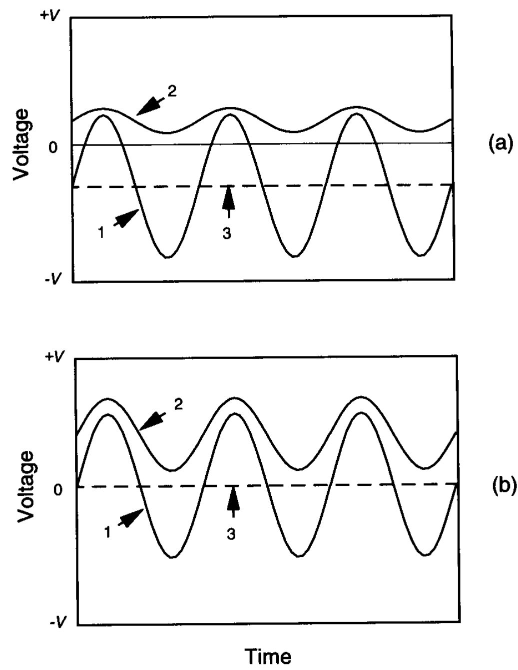


Figure C.13: Development if a self-bias with a sinusoidal wave: (a) unequal electrodes; (b) electrodes of equal area. 1, voltage on powered electrode; 2, plasma potential; 3, time-averaged voltage on powered electrode. From [11].

The experimental evidence indicates that the voltage ratio V_1/V_2 depends also on the gas used in the discharge, the peak-to-peak applied voltage, as well as the area ratio $(A_2/A_1)^n$. For area ratios between 0.6 and 0.1 the fourth power relationship is approximately obeyed. However, it is not obeyed for smaller area ratios. For an Ar discharge with an electrode area ratio of 0.3, the value of n can range from 1.2 to 2.5.

A RF power supply is connected to a plasma reactor through an *impedance matching unit*. The unit is needed to match the impedance of the plasma reactor to the output impedance of the power supply which is usually at 50Ω. Because the matching units normally do include a blocking capacitor in the planar diode reactors used in PECVD processing, the electrodes are made of equivalent areas when similar voltages relative to the plasma are required on both electrodes. However, the described effect can be exploited to obtain a higher negative bias on one electrodes.

The value of the self-bias is dependent on the RF power applied to the electrode and the pressure in the reactor. It increases with increasing power and decreasing,

$$V_B \propto \left(\frac{P_{RF}}{p}\right)^{1/2} \tag{C.96}$$

where V_B = self-bias on the powered electrode, P_{RF} = RF power.

As the frequency of the electric field increases (e.g., from 50kHz to 13.56MHz), there is increasingly less time between cycles available for the diffusion of the charged particles to the reactor walls. Therefore, according to Bohm sheath criterion (see sectionC.5.2), less negative bias has to develop across the sheath to keep the electrons in the plasma. This, in turn, means that the ion bombardment of the electrode surface decreases with increasing frequency of the electric field.

If each electrode shown in Fig.C.11 is RF powered relative to the grounded chamber, it is possible to control the DC bias on each electrode independently. Thus the plasma can be generated by powering one electrode and the degree of ion bombardment on the other electrode can be regulated independently by a second power supply.

Magnetic fields parallel to the electrode are sometimes superposed on the RF plasma. The magnetic field confines the electrons next to the electrodes and increases the ionization efficiency. This increase in ionization lowers the sheath potential and the energy of the bombarding ions.

RF versus DC plasma

The advantages of RF discharges over the DC discharges, as summarized next, explain the wider use of RF plasmas as compared to DC plasmas:

1. RF plasmas can be excited and sustained using either conductive or nonconductive electrodes, while DC discharges require the electrodes to be conductive throughout the process.

2. RF plasmas can be sustained with internal as well as external electrodes, while DC discharges require the electrodes to be inserted inside the reactor and be in direct contact with the plasma. Use of external electrodes is sometimes required when the gases of the discharge are corrosive or when one wants to reduce contamination of the plasma with the material of the electrodes.

3. RF plasmas are characterized by higher ionization efficiencies than are the DC plasmas.

4. RF plasmas can be sustained at lower gas pressures than are DC plasmas.

5. In RF plasmas the energy of the ions bombarding the sample is controlled by the negative bias, which can be adjusted over a wide range of values. Samples placed on the cathode of the DC discharge are exposed to bombardment of high-energy ions that are accelerated at voltages that have to be above the minimal breakdown voltage. This can cause damages to sensitive substrates.

Bibliography

[1] H. M. Mott-Smith and I. Langmuir, Phys. Rev., **28**, 727 (1926).

[2] F. F. Chen in Plasma Diagnostic Techniques, R. H. Huddlestone and S. L. Leonard, eds., Academic, New York (1965).

[3] Theory of spherical and cylindrical langmuir probes in a collisionless, Maxwellian plasma at rest, UTIAS Report No.100, University of Tronto (1966).

[4] V. A. Godyak in Plasma-Surface Interactions and Processing of Materials, O. Auciello et al., eds., Kluwer Academic, Boston.

[5] K-U. Riemann, J. Phys. D: Appl. Phys., **24**, 493 (1991).

[6] I. Langmuir, Phys. Rev., **33**, 954 (1929).

[7] J. L. Cecchi, in Handbook of Plasma Processing Technology, eds. Stephen M. Rossangel, Jerome J. Cuomo and William D. Westwood, p.14, Park Ridge, NJ: Noyes Publication, 1990.

[8] J. A. Thornton and A. S. Penfold, in Thin Film Processes, eds. J. L. Vossen and W. Kern, p.75. New York: Academic Press, 1978.

[9] A. T. Bell, in Tenchniques and Applications of Plasma Chemistry, eds. John R. Hollahan and Alexis T. Bell, P.1. New York, John Wiley & Sons, 1974.

[10] Vossen, J. L. and W. Kern, eds., Thin Film Processes. New York: Academic Press (1978).

[11] M. A. Lieberman and A.J. Lichtenberg, 'Principles of plasma discharges and materials processing', John Wiley & Sons, Inc., (1994).

[12] H. R. Koenig and L. I. Maissel, IBM J. Res. Dev., **14**, 276 (1970).

List of publications

1. Tetsuya Narushima, Akiko N. Itakura, Takaya Kawabe, and Masahiro Kitajima "Electron-Stimulated Surface Stress Relaxation of Si" Applied Physics Letters, **79**, 5, 605-607 (2001).

2. Tetsuya Narushima, Takayuki Kurashina, Akiko N. Itakura, Takaya Kawabe, and Masahiro Kitajima, "Effects of Surface Disorder on the Surface Stress of Si(100) during Oxidation" Appl. Surf. Sci., **159-160**, 25 (2000).

3. A. N. Itakura, T. Narushima, M. Kitajima, K. Teraishi, A. Yamada, and A. Miyamoto "Surface Stress in Thin Silicon Oxide Layer Made by Plasma Oxidation with Sample Bias" Appl. Surf. Sci., **159-160**, 62 (2000).

4. A. Kurokawa, S. Ichimura, T. Narushima, A. N. Itakura, and M. Kitajima, "Real-time Observation of Surface Stress during Initial Ozone-Oxidation of Si(100)" (ISTC2001), Shanghai, 2001 (The Electrochemical Society) 212-217.

5. A. N. Itakura, T. Narushima, and M. Kitajima "Surface stress in Silicon Oxide Layer" Trance. MRS-J **24**, 75 (1999).

Acknowledgement

I would like to thank Professor T. Kawabe of Institute of Physics, University of Tsukuba for his continual instructions, valuable discussions, his continual encouragements, critical reading of this thesis.

This study was performed under a collabolation research between University of Tsukuba, National Institute for Materials Science (NIMS) and National Institute of Advanced Industrial Science and Technology (AIST).

The experimental works of surface stress measurements were performed at NIMS and I would like to thank Dr. M. Kitajima of NIMS for introducing me to this field of research, for his valuable instructions, many fruitful discussions, advices and for continual encouragements and inspirations every day.

The experimental works of STM observaions were performed at AIST and I wish to also express gratitude to Dr. K. Miki of AIST for instructing me in the way of operation and observarion of STM from the foundation, for his collaborations and for providing me with this excellent oppotunity for career development.

I wish sincerely thank Professor A. Oshiyama, who is my thesis advisor, of Institute of Physics, University of Tsukuba for his generous help and advise. I am grateful to Professor K. Murakami of Institute of Applied Physics, University of Tsukuba and Professor T. Cho of Institute of Physics, University of Tsukuba for their valuable advice and discussion.

I would like to thank good collaborators: Dr. A. N. Itakura of NIMS, Dr. A. Kurokawa of AIST, Dr. S. Watanabe of fujitsu Co., and Dr. Ruediger Berger of IBM SSD GmbH, Mainz, Germany for their aid and advice during my work.

I also thank Dr. K. Ishioka, Dr. M. Hase, Dr. M. Yata, Dr. M. Imai, Dr. Y. Saito of NIMS for their valuable instructions, many advises and suggestions.

I am grateful to Professor Arthur G. Every of University of the Witwatersland, South Africa and Dr. James H. Owen of STA fellow reseacher of AIST for reading manuscripts and variable discussions.

I also thank Professor T. Ono of Faculty of Engineering, University of Tohoku for instructing me to fabricate especial microcantilever samples.

I wish to thank Mr. A. Kubo, Mr. Y. Ishii, Mr. T. Hayashi and Mr. M. Fukazawa of graduated student, University of Tsukuba, Dr. W. Yashiro and Dr. I. Shiraki of Post doc reseacher of AIST, Mr. K. Aoki and Mr. N. Ueda of undergraduate student, University of Tsukuba for their continual encoragements and helps.

Finally, I thank my dear parents, my sister, grandmothers and grandfathers, my sweetheart, freinds and relatives for their loves and continual supports during the study.