Semicond. Sci. Technol. **10** (1995) 215–244. Printed in the UK

TOPICAL REVIEW

# Oxidation of silicon: the VLSI gate dielectric?

**C J Sofield and A M Stoneham**

AEA Technology, Harwell, Didcot, Oxon OX11 0RA, UK

Received 20 October 1994, accepted for publication 18 November 1994

**Abstract.** Silicon dominates the semiconductor industry for good reasons. One factor is the stable, easily formed, insulating oxide, which aids high performance and allows practical processing. How well can these virtues survive as new demands are made on integrity, on smallness of feature sizes and other dimensions, and on constraints on processing and manufacturing methods? These demands make it critical to identify, quantify and predict the key controlling growth and defect processes on an atomic scale.

The combination of theory and novel experiments (isotope methods, electronic noise, spin resonance, pulsed laser atom probes and other desorption methods, and especially scanning tunnelling or atomic force microscopies) provide tools whose impact on models is just being appreciated. We discuss the current unified model for silicon oxidation, which goes beyond the traditional descriptions of kinetic and ellipsometric data by explicitly addressing the issues raised in isotope experiments. The framework is still the Deal–Grove model, which provides a phenomenology to describe the major regimes of behaviour, and gives a base from which the substantial deviations can be characterized. In this model, growth is limited by diffusion and interfacial reactions operating in series. The deviations from Deal–Grove are most significant for just those first tens of atomic layers of oxide which are critical for the ultra-thin oxide layers now demanded. Several features emerge as important. First is the role of stress and stress relaxation. Second is the nature of the oxide closest to the Si, both its defects and its differences from the amorphous stoichiometric oxide further out, whether in composition, in network topology, or otherwise. Thirdly, we must consider the charge states of both fixed and mobile species. In thin films with very different dielectric constants, image terms can be important; these terms affect interpretation of spectroscopies, the injection of oxidant species and relative defect stabilities. This has added importance now that $P_b$ concentrations have been correlated with interfacial stress. This raises further issues about the perfection of the oxide random network and the incorporation of interstitial species like molecular oxygen.

Finally, the roles of contamination, particles, metals, hydrocarbons etc are important, as is interface roughness. These features depend on pre-gate oxide cleaning and define the Si surface that is to be oxidized which may have an influence on the features listed above.

## 1. Introduction

Silicon has dominated the semiconductor industry for so long that the two have become almost synonymous. Partly this position is the result of excellent basic properties of silicon; partly it is the wealth of experience in materials control, so that obstacles can be overcome with confidence. Partly, too, it is ancilliary properties like an easily formed, stable, insulating, oxide which allows ingenuity in processing and routes to high performance. It is the oxide which gives an edge for silicon (among other factors) over, say, III–Vs, or diamond. The industrial drive to ever higher packing density of CMOS devices requires this technical edge to be maintained as ever thinner oxides are needed for gate dielectrics. Can silicon's supremacy be maintained? Will new approaches be needed when the nanometre scale is really the right characteristic length? Are there new problems, as well as new opportunities? For the major areas of semiconductor technology, a range of factors come in: breakdown and wear-out of the oxide, manufacturability with adequate process control, noise associated with traps in the oxide, charges in the oxide affecting transport near to the interface, etc.

In this review we therefore concentrate on setting out our understanding of the growth and properties of thin

0268-1242/95/030215+30$19.50

silicon dioxide layers up to several nm thick. This gate oxide thickness range—a few tens of atomic layers—is also the range in which the oxidation kinetics and the oxide growth mechanism depart from the 'standard' Deal–Grove model (in which diffusion and interface processes operate in series) and enter the range where growth models are still controversial.

### 1.1. Thickness of MOS gate structures in CMOS devices

With each new generation of microcircuits, there is a scaling down in size (Barrat 1993) of the CMOS transistors which dominate VLSI technology. In particular, the oxide thickness of the MOS gate in a CMOS transistor also has to scale down with the device dimensions (Sze 1985). For example, in 0.25 $\mu$m technology the gate oxide could be less than 10 nm thick while achieving good performance of a CMOS transistor (Sze 1985), though theoretical analysis (Schuegraf *et al* 1992) indicates that gate oxides still need to be thicker than about 4 nm to limit direct tunnelling leakage. Gate operating voltages do not scale down as fast as gate oxide thickness (Sze 1988), so that size reductions may lead to increases in electric fields across the gate dielectric. With the gate oxide thickness in the 4 to 10 nm range and with gate oxide fields increasing, there are more and more stringent demands on gate oxide integrity to wear out and resistance to breakdown.

### 1.2. What limits the performance of the gate oxide?

The gate oxide clearly needs to survive against electrical breakdown at the device operating voltage, which is usually less than 5 V. Just as important the oxide must also survive for the projected lifetime of the device under field stressing at less than the breakdown value. Breakdown strength ($V_{bd}$) has become an important indicator of reliability partly because it is correlated with charge-to-breakdown ($Q_{bd}$) at lower field stresses (Rowland *et al* 1986) and thus offers an accelerated test parameter. The lowest breakdown field across the oxide can be determined from a knowledge of oxide thickness and of $V_{bd}$, but care must be exercised to correct the measured $V_{bd}$ for any series resistance voltage drops. In addition one must account for changes in the oxide that may arise from charge injection during $V_{bd}$ or $Q_{bd}$ measurement. Verhaverbecke (1993) and Schuegraf and Hu (1994) give a detailed discussion of these measurement issues. The statistics of breakdown strength are also important: it is important to achieve a high average $V_{bd}$, but it is also important that there is no significant tail in the distribution of low $V_{bd}$ values, i.e. good quality oxide has to be produced over the whole surface of large Si wafers. Thus some important issues of oxide performance arise from this need to achieve high yields.

The breakdown strength of gate oxides has improved as the drive to smaller device dimensions has been achieved. Whereas some ten years ago (Wolters and Verwey 1986) 'standard' values of 8.5 MV cm$^{-1}$ were accepted as good quality, current practice yields values of 13.5 MV cm$^{-1}$. The breakdown is believed to occur at defects in the oxide arising, for example, from particle, metal or other impurity-related defects. The trend of rising breakdown strength with device generation may be correlated with the trend to reductions in contaminant levels. These breakdown strength measurements are usually carried out on relatively large MOS capacitors (50 $\mu$m diameter and upwards) to make probe contacts easy to apply and to obviate further processing to establish contacts to smaller capacitors. There is no well accepted theory of the intrinsic breakdown strength of gate oxides, and so there are no clear guidelines as to what degree of improvement might be feasible. Recent studies (Schwerin and Heyns 1991, Di Maria *et al* 1993, Marwick *et al* 1993, Schuegraf and Hu 1994) show that sufficiently energetic hot electron injection during field stressing can give rise to defects both by interaction with the $SiO_2$ and with H contaminants trapped at the interface. These defects (particularly holes and electron traps) degrade the oxide until breakdown occurs. When the oxide is thin enough (< 5 nm) hot electrons are not produced but direct tunnelling starts to occur. Thus to obtain 'ideal' dielectric behaviour both the quality of the oxide and its resistance to degradation under stress requires attention.

Experimental observations of $SiO_2$ breakdown strength suggest that further improvements (beyond 13.5 MV cm$^{-1}$) may be achievable. Early laser induced breakdown studies of bulk fused silica (Smith *et al* 1977) gave breakdown strengths of 50 MV cm$^{-1}$, much higher than the values for thin gate oxides quoted above. Further laser induced breakdown studies in bulk silica using laser pulse widths down to 150 fs demonstrates that electric field driven avalanche breakdown arises at threshold fields of a few tens of MV cm$^{-1}$ and higher (Du *et al* 1994).

Recently a conducting cantilever atomic force microscope experiment was devised (Murrell *et al* 1993) to examine oxide breakdown and the onset of Fowler–Nordheim tunnelling prior to breakdown with a spatial resolution of about 20 nm. In this study 10 nm thick gate oxide was directly examined and small regions of oxide were found capable of withstanding 40 MV cm$^{-1}$. These same gate oxides suffer breakdown at 13.5 MV cm$^{-1}$ when tested with sets of 50 $\mu$m diameter capacitors. Since some of this difference in performance must arise from extrinsic defects, the importance of this new technique arises in part from the ability to examine the dielectric on a small scale and search for such defects.

### 1.3. Related processing issues and constraints

In the fabrication of CMOS devices the oxidation of Si to form a gate dielectric in a transistor is only one stage of fabrication, albeit a critical stage. The influence of the pre-gate oxide cleaning process on the Si surface that is to be oxidized is crucial. The contact to the oxide to form an MOS gate electrode must also be carried out in a manner to achieve low contact and interconnect resistivity without perturbing the dielectric. These additional constraints may limit the way the oxide can be grown and hence whether or not, for example, an oxynitride might be a better choice of dielectric. Some of the processing issues that must be considered as an integral part of the gate oxidation process to achieve a high yield of VLSI devices are briefly discussed below. A review of the process challenge presented by the development of 64 Mbit DRAM highlights the complexity of integrating all the various interrelated processes (e.g. Lowrey *et al* 1993).

**1.3.1. Pre-oxidation cleaning.** The oxidation of Si to produce a satisfactory gate dielectric cannot be achieved without due attention to earlier process steps. Prior to gate oxidation, the condition of the Si surface must be controlled to some statistically satisfactory degree (Ohmi 1990) as regards particle, metal, organic and 'native oxide' contamination if satisfactory control of oxide thickness and defect levels are to be achieved. This is a very demanding task (e.g. Hattori 1990). Particles, for example, are required to have an areal density of less than 0.001 $\text{cm}^{-2}$ for sizes less than 0.003 $\mu$m for 0.25 $\mu$m geometry devices (Ruzyllo 1990). Currently (Singer 1993) such particle sizes cannot be measured with such low areal density using standard light scattering wafer inspection methods. Some metal have worse effects than others on oxide breakdown (Verhaverbeke *et al* 1992) and need to be removed to levels lower than the current detection limits (about $10^9$ $\text{cm}^{-2}$). The favoured method for such analysis is vapour phase decomposition of the native oxide to collect all the metal contamination into a small droplet on the wafer, followed by TXRF (total internal reflection x-ray fluorescence) assaying (Prange 1989); this technique has poor spatial resolution but a detection limit of about $10^9$ $\text{cm}^{-2}$. Organic contamination (Budde *et al* 1993) must also be limited to very low levels. The cleaning procedures used to control the above contaminants usually leave the Si surface with a 'native oxide' or H termination which has to be kept stable prior to the growth of the gate oxide. Transport from the cleaning stage to the oxidation stage must thus be carried out in an ambient that maintains this surface and does not recontaminate it.

The Si surface morphology also requires stringent control. For example, a single atomic step on a Si(100) surface is about 7% of the thickness of a 4 nm thick gate oxide and thus may give rise to a significant field enhancement even for 'uniform' oxide coverage. The step density on the Si surface may, as a result, be an important factor in gate oxide integrity. We shall discuss later recent correlations between Si surface roughness on an atomic scale and gate oxide yields which have been observed.

The trend to larger wafers in VLSI fabrication and the need to control Si surface morphology and chemistry places considerable demands on surface inspection, surface cleaning and oxidation technology extending from the atomic scale to the wafer scale.

**1.3.2. Oxidation technology.** The need for thinner gate oxides places stringent demands on the oxidation technology used to grow the oxide layer. For example a 7 nm thick oxide which may be required for 0.25 $\mu$m device geometry (Fair and Ruggles 1990) needs to be controlled to the level of 0.2 nm over a large wafer to achieve satisfactory operating variability. Standard furnace oxidation techniques at atmospheric oxygen pressure preclude such precision. Thus other techniques such as low-pressure and/or low-temperature oxidation, UV/ozone oxidation and or rapid thermal annealing techniques may be required to achieve adequate process control. This has to be achieved without sacrificing dielectric quality and defect levels.

**1.3.3. Gate contact.** Although not an oxidation problem *per se*, the interaction of the gate oxide contact with the oxide becomes more critical with the very thin gate oxides now required for VLSI. One cannot use materials that react even slightly with the oxide at temperatures encountered in subsequent process stages (e.g. Al). Even doped polysilicon contacts, which are very stable, present problems because the B dopant used may diffuse through the oxide structure, and hence interfere with the transistor channel. There is interest therefore in growing thin oxynitride structures to achieve both the desired dielectric properties and to give improved B diffusion barrier properties (Wong *et al* 1983). Thus the kind of dielectric used, oxide or, for example, oxynitride, may not be chosen independent of this other process consideration.

## 2. Basic ideas of oxidation

The basic concepts in the understanding of oxidation kinetics were laid down over half a century ago (Mott and Gurney (1940); for a more recent review in the context of defect processes see Hayes and Stoneham (1985)). The first basic idea is transport: either Si ions must move out, or oxygen ions must move in, or perhaps both may occur. If ions are transported, rather than neutral atoms, then electronic charges must move too. The transport phenomenon has many possible complications, but clearly the simplest assumption is that the chemical potential difference which drives transport is independent of film thickness, so the flux (proportional to the gradient of that chemical potential) is inversely proportional to the oxide thickness, $x$. This, when diffusion is rate-determining, leads directly to the famous 'parabolic' law of oxidation, with the square of the thickness linear in time ($x^2 \propto At$).

The second key idea is an interfacial rate process, whether the adsorption or reaction of gas at the outer oxide surface or some interface process between silicon and its oxide. It is tempting to say this rate is independent of oxide thickness (so giving a linear rate, $x \propto Bt$), but again there are possible complications. One exception is the Mott–Cabrera mechanism: here electrons tunnel from substrate (e.g.

Si) to adsorbed species at the outer oxide surface (e.g. adsorbed oxygen molecules give oxygen ions) until a particular potential difference is established. This potential difference provides a thickness-dependent field which affects injection of ions at the Si/oxide interface.

All models of this type lead to particular kinetics $x(t)$. There is considerable—usually unjustified—optimism in attempting to deduce subtleties of mechanism from measurements of oxidation rate. The thickness $x(t)$ itself is not alway measured directly; usually it is the number of oxygens per unit area which is known. One of the aims of our review is to bring in information from other types of experiment, especially when these relate to oxide quality.

### 2.1. A framework for silicon oxidation: basic kinetic studies

Here the classic description is that of Deal and Grove (1965), whose model forms the framework within which almost all generalizations are analysed. As described above, oxidation is controlled by the two main processes (diffusion and some interface step) operating 'in series'. If the Deal–Grove description were complete, with the two processes determined by diffusion and interface reaction parameters independent of oxide thickness, oxidation could be described by an equation in which the inverse oxidation rate ($\mathrm{d}t/\mathrm{d}x$) is linear in thickness, $x$. The slope is controlled by the reciprocal diffusion rate, and the intercept at small thicknesses $x$ by the mechanism independent of thickness, perhaps the reciprocal interfacial reaction rate. The predicted linearity of ($\mathrm{d}t/\mathrm{d}x$) with $x$ is observed for thick oxides, but there is a substantial deviation for thin oxides (for dry oxidation, this means thicknesses less than 10–20 nm). It is this deviation and its implications which account for much of the interest and the plethora of models; it is this ambiguity of interpretation which is a feature of kinetic models.

Comparisons with simple phenomenological models nearly all make an implicit assumption, namely that the surface is flat, so that oxide thickness $x$ has meaning. This is not always appropriate. Yakovlev et al (1992) argue from ellipsometry data that, in high-temperature annealing of an oxide layer (750–1100 °C) there are two modes: a rapid change corresponding to the disappearance of protrusions and a slower change corresponding to the loss of suboxide. Suboxide reduction is most evident at lower temperatures; viscous relaxation dominates at higher temperatures.

The kinetics (i.e. $x(t)$) for a real growing oxide put limits on the processes involved, but do not indicate unique microscopic processes. To go further, other detailed calculations and experiments are necessary, going beyond kinetics alone. In particular, isotope experiments have been especially revealing, and a number of experiments which measure composition close to the interface indicate features of interest. It is the combination of deviations from Deal–Grove and the results of isotope experiments which point clearly to potential problem areas for any application of thin oxides.

### 2.2. General features of silicon oxidation

At the risk of repeating much that is well known, the results which are of direct interest here are summarized below. A fuller discussion, with extensive references is given by Mott *et al* (1989); that review is based, in turn, largely on the papers collected in volume B55 of the *Philosophical Magazine*.

First, the oxide formed is largely silicon dioxide. There are serious qualifications for the region close to the Si/oxide interface, notably from measurements with the pulsed laser atom probe (PLAP) which show a limited region of stoichiometric SiO. With such limited exceptions, the oxide appears to form a continuous random network, with each Si linked to four further silicons via an oxygen 'bridge'. The oxide is usually amorphous, again with the exception of local regions at the interface where steps favour epitaxy, or impurity features like SiC, or else on very flat Si, where some crystalline form of silica might be stable. Any local regions of crystallinity appear to have no effect on oxidation kinetics.

Secondly, there is a major difference between dry and wet oxidation; even minor levels of water enhance the rate. Just as in quartz, where there are at least six forms of hydrogeneous species, there may be many distinct hydrogenic species. It is known from theory and from muon and other resonance studies that atomic hydrogen is stable in quartz (contrary to widely expressed views) and H atoms are presumably stable in thermal oxide away from defects. Clearly, however, a mobile reactive species like atomic hydrogen will rapidly seek defects. Water can have a range of effects on oxide properties, especially on mechanical properties and plasticity. Some of these differences are a natural consequence of the continuous random network, in that less energetic processes can split a water molecule and insert a proton and hydroxyl than are needed to split an oxygen molecule and insert it into the network (possibly at a defect site). A fuller analysis of atomic mechanisms is given by Heggie and Jones (1986, 1987) and Jones *et al* (1992).

Thirdly, in the diffusion-limited regime (thicker oxides) the transport mechanism is mainly interstitial molecular oxygen (dry case) or interstitial molecular water (wet case). Other species can contribute, especially in special regimes or in the presence of electronic excitation. Silicon motion is negligible (Murrell *et al* 1991) contrary to the oxidation of silicides and alloys of Si. In most cases, diffusion of oxygen into the silicon substrate is not rate-determining, though the consequences (like enhanced impurity diffusion from Si interstitial injection) can be important. The dominant transport processes identified lead to oxidation rates which show good agreement between theory and other experiments (Hagon *et al* 1987). In particular, the parabolic rate constant agrees well with independent measurement of diffusion rates in macroscopic silica (Atkinson 1987), and the activation energies observed agree well with those predicted theoretically. Theory also makes it clear that the water molecule diffusing in quartz would have a lowest-energy path with rotation, and this is presumably more structure-sensitive than the oxygen molecule motion for which barriers will be relatively similar from one oxide structure to another. A recent calculation (Jones *et al* 1992) shows that in quartz the oxide is not a simple rigid medium, but one in which the water molecule causes an Si–O–Si unit to flip as it passes.

### 2.3. Deviations from Deal–Grove

A fit to the Deal–Grove model does not say whether Si or O is mobile, nor indicate its charge state or the mechanism. Yet there is the important message that deviations from Deal–Grove kinetics tell one that the model is insufficient. We give three examples.

First, we consider the general form for $dx/dt$ versus time for dry oxidation. Data from several sources collated by Mott *et al* (1989) show deviations from the linear form for small oxide thicknesses (below about 20 nm). These deviations can be interpreted with equal success as an effect on the interface reaction rate (which would need to be enhanced) or on the diffusion rate (which should be reduced).

As a second example, we follow the approach of Stoneham and Tasker (1987) and examine the dimensionless parameter

$$g(x) = -\mathrm{d}[\log(\mathrm{d}x/\mathrm{d}t)]/\mathrm{d}(\log x)$$

evaluated from data close to some reference time and corresponding thickness $x$. This allows us to use a dimensionless parameter $g$ to relate the rate of oxidation ($dx/dt$) to the oxide thickness, rather than time of exposure (and so is closer to the spirit of oxidation kinetics models). In the Deal–Grove model, $g$ must lie between 0 and 1. Experimentally, this is not always so (see figure 1(*a*)): deviations become substantial for low oxygen pressures and for low temperatures. Oxidation kinetics studies confirm too that there are deviations from Deal–Grove, and emphasize that these deviations are greatest in the thin-oxide regime of most interest for thin gate fabrication.

At high temperatures and low pressures a reaction has been observed that leads to the decomposition of the $SiO_2$; this is not included in oxidation theory. Annealing an oxide after growth in low $O_2$ partial pressures can lead to low-field breakdown (Rubloff *et al* 1987). Recent observations of oxidation and 'etching' in STM and TEM (see section 4.1) are usually carried out in the regime of $g \gg 1$. A summary of these oxidation pressure/temperature regimes is given in figure 1(*b*).

The third example concerns enhanced oxidation, such as the effect of an electron beam. Here, Deal–Grove might still apply, provided the effects were merely to change the parameters of the model. Yet the effect is quite different (see figure 2 in Stoneham *et al* (1987)). For very low oxide thicknesses, even very low energy electrons have a striking effect: for thin oxide (less than 2–3 nm) the oxidation rate falls as temperature rises! The controlling parameter is in fact the sticking probability for oxygen molecules at the surface. However, there is another effect: the dependence on electron energy implies dissociation of interstitial molecular oxygen. Since work on interstitial oxygen in silica glasses (Awazu and Kawazoe 1990) shows that dissociation reactions like this can be controlled by ArF laser excitation (the probable reactions give ozone and atomic oxygen) there are links to other oxidation approaches based on, for example, ozone (Boyd *et al* 1993). Dissociation requires more than about 5.1 eV. It is not clear whether an electron in the oxide conduction band can supply enough energy to cause dissociation, so as to give a basis for thermionic and similar models (Irene and Lewis 1987).

### 2.4. What are the alternative mechanisms?

Many discussions retain the Deal–Grove description; they also retain the view that only the standard diffusion mechanism occurs (e.g. oxygen molecules diffuse interstitially); they then introduce some modest change, like a stress gradient, and hence have weak dependences of the diffusion and interface reaction parameters in the Deal–Grove expressions. We list here some of the proposals which go beyond this.

**2.4.1. The Mott–Cabrera mechanism.** The Mott–Cabrera model is based on two key ideas. First, there is an electron tunnelling process which transfers charge across the oxide until the potential generated prevents further transfers. Secondly, the electric field set up in this process affects the injection of mobile ions into the oxide; there are no effects on the process of transport across the oxide in this model. For silicon oxidation, for instance, the transfer would normally be of electrons from the silicon substrate onto adsorbed oxygen molecules (giving a negative molecular ion at the oxide–gas interface); the field due to these negative adsorbed ions then aids the injection of a mobile species (perhaps the molecular oxygen ion $(O_2)^-$; perhaps a mobile oxygen ion $O^-$ if the charge transfer leads to an O atom and an O ion, possibly even a silicon ion if it is believed that the Si is mobile). Recent experiments indicate the presence of the molecular oxygen ion as a precursor in the initial monolayer oxidation of clean Si surfaces (Yu and DeLouise 1994 and references therein). The molecular oxygen ion is also very common on oxide surfaces (Che and Tench 1982, 1983).

**2.4.2. Image effects on transport by charged species.** It is widely believed (and probably true) that most transport is by neutral species, but there are special cases (e.g. in the presence of electron beams) when charged species may be important; some workers strongly favour charged species. The electrostatic image interaction will affect the kinetics of oxidation involving charged species (Stoneham and Tasker 1987).

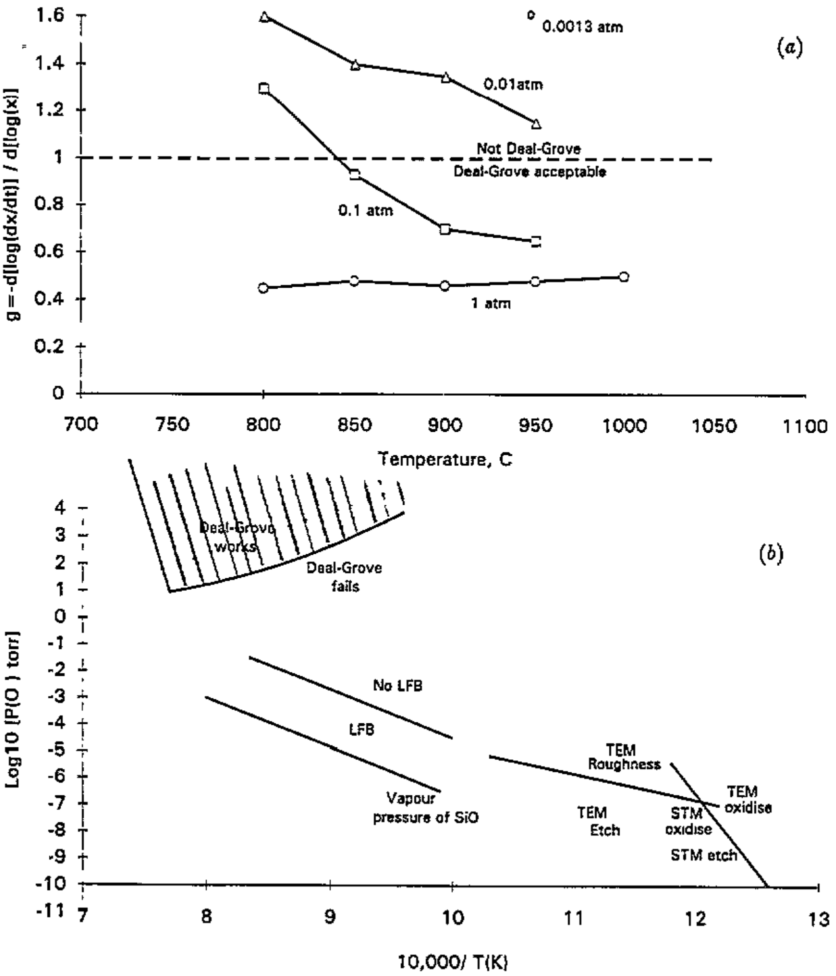

**Figure 1.** (*a*) Oxidation kinetics data plotted as a function of *g*. Departures from Deal–Grove occur where the rate is not determined by a combination of thickness independent diffusion and interface reactions in series. (Data from Wolters and Zegers-van Duynhoven 1989, Massoud *et al* 1985 and Adams *et al* 1980.) (*b*) Pressure and temperature regimes of interest to oxidation studies. The regime of non-Deal–Grove growth ($g > 1$) is shown and the regime where 'etching' occurs instead of oxide growth, i.e. where SiO is volatile. The pressure/temperature range in which treatment of an oxide gives rise to low-field breakdown (LFB) of capacitors is indicated. The regime in which initial oxidation has been directly observed by TEM and STM are also shown (see discussion in section 4).

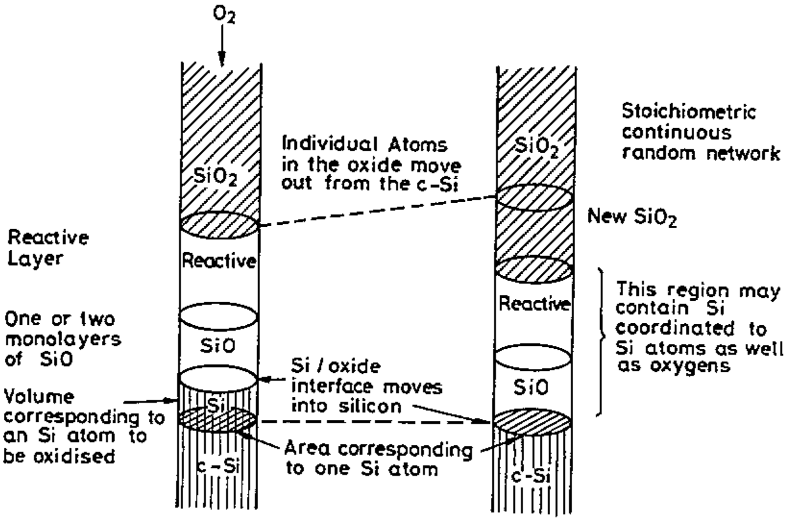

**Figure 2.** A schematic diagram of the progress of oxidation. A column whose area corresponds to one Si atom at the interface is shown before and after oxidation by one oxygen molecule.

**2.4.3. Composition changes near the interface.** As oxidation proceeds, the interface between oxide and silicon moves into the silicon. Figure 2 shows how this can be envisaged if there is an intermediate layer of oxide. The figure is general, in that it need not presume any special properties of the intermediate layer. One proposal is of a 'reactive layer', where interstitial oxygen reacts with the network to incorporate oxygen. Clearly Si–Si bonds might be attacked in this way, and indeed there is some indirect evidence for such bonds within the oxide near to the Si/oxide interface. In this case (Stoneham *et al* 1987) one would expect the initial oxide (to a thickness similar to that of the reactive layer) to grow by reaction at the outer surface. When the oxide is much thicker than the layer, oxygen molecules will diffuse to the outside of the 'reactive layer', and there react, exchanging with network oxygens. This simple picture is sufficient to explain some of the isotope experiments (which are not to be understood by most models based on kinetics alone) and to suggest further isotope oxidation sequences. It may not be necessary, for better isotope data do not seem to need a reactive layer. Nevertheless, some level of Si–Si bonds seems to be needed to understand the PLAP data and the relation of stress to the $P_b$ centre density. A small degree of oxygen exchange (about 0.3%) with the silica network has been observed (Murrell 1991). This may arise from such reactive Si–Si bond sites.

There is evidence for some sort of 'intermediate layer' between the silicon and the amorphous silicon dioxide. Its apparent thickness ranges from 0.3 to 3 nm, depending on preparation and on the type of experiment used to detect it. Thus the layer can differ in composition, for the pulsed laser atom probe shows several layers of stoichiometric SiO; it may differ in local coordination, for spectroscopic methods appear to indicate charge states of Si other than those expected from Si/silicon dioxide alone; it may appear to differ simply because the different dielectric constants of Si and oxide lead to image terms affecting energies of transitions with change in localized charge states (Browning *et al* 1989). The intermediate layer may differ too in refractive index, suggesting altered local density, or perhaps local topology. Differences in structure are deduced by Brugemann *et al* (1990).

## 2.5. Early stages of silicon oxidation

The formation of the first few layers of oxide presents puzzles. The major puzzle is this: essentially complete layers of oxide appear to grow before the next layer starts (e.g. Borman *et al* 1991), yet oxidation does not occur preferentially at steps (see the discussion in section 4.1.4). We exclude here consideration of H-terminated surfaces for which island growth of the oxide is proposed (Hattori 1994). We shall describe new ideas which appear to explain this (Torres and Stoneham 1993, unpublished work). Any explanation of this major puzzle has implications for oxidation kinetics (oxide thickness $x$ as a function of time $t$). The kinetics of these first stages are certainly not Deal–Grove (determined by diffusion and interface steps in series). However, there are also difficulties with the Mott–Cabrera process often suggested. The process we identify is distinct from Mott–Cabrera oxidation, but is more closely related to the ideas of Stoneham and Tasker (1987), and appears to be consistent with observed kinetics.

We can address some of these issues using a model far simpler than the real silicon/oxide system. This is possible provided we can ignore precise structural issues for the oxide and many details of the transport processes. We shall not worry whether or not there is a reactive layer (Stoneham *et al* 1987), and we shall take it for granted that the species which moves in the oxidation process is oxygen, not silicon. What we want to do is identify the types of controlling process which would give the qualitative behaviour observed. We can link these model processes to real processes which might happen in the real, complex system. This will define processes by which layer-by-layer growth can be reconciled with terrace oxidation.

As oxidation proceeds, the number of layers of oxide above each substrate silicon will vary from place to place. Assume for ease of discussion that this number is well defined, so that there are $N(x, y)$ layers above the surface silicon atom with coordinates $(x, y)$. If there is to be terrace growth without growth at steps, we shall need a process whose rate falls off rapidly as $N(x, y)$ increases. Two possible processes are these:

(i) Process I: The sticking probability depends on $N$ (diffusion laterally across the oxide surface is neglected). The importance of the sticking probability of some oxygen species, whether on the bare silicon surface or on the oxide, has been remarked on before (Mott *et al* 1989, Miotello and Toigo 1987). Why should the sticking probability depend on the oxide thickness? One reason could be charge transfer. Note that in oxidation of bare Si surfaces there is experimental evidence of a molecular oxygen ion precursor (see section 2.4.1). The energy to take a delocalized electron (whether from the valence band or from the conduction band) from the Si substrate to localize it on an oxygen species at the oxide/gas interface depends on the oxide thickness, due to the image interaction. If so, we might expect a dependence on number of layers which proceeded as $\exp(-AN)$ or even faster. If the critical steps were image potential-driven diffusion after sticking, a power-law behaviour could be more appropriate. If it is the mechanical properties of the oxide, as opposed to the silicon substrate, then one might guess some further dependence. However, the power-law and exponential forms are helpful reference cases.

(ii) Process II: There is an adsorbed state in which lateral motion across the oxide surface is easy, but the 'injection' step, in which the oxygenic species moves from adsorbed state into the oxide, is strongly dependent on $N$.

In Mott–Cabrera oxidation, it is this injection step which is rate-determining. Why should either step depend on $N$? The major factor, we believe, is the image interaction. Unlike thick dry oxide growth, for which the moving species is the neutral interstitial oxygen molecule, we suspect that the oxygen species for thin oxide growth is charged, following electron transfer from the silicon substrate. This would be true for Mott–Cabrera oxidation as well as for the mechanisms proposed by Stoneham and Tasker (1987). Electron transfer would be negligible for thick oxides, but can affect the kinetics of growth for thin oxides. If so, this offers electronic ways to influence the oxidation process.

**2.5.1. Probabilistic model.** The basic computer-based model is mathematical, rather than physical. It aims to show which forms of behaviour (e.g. sticking probability $S$ as a function of number of layers $N$) lead to the behaviour we seek. The forms of $S(N)$ which we shall try are guided by physics, but only later will we examine in detail the link to the real system. The model does not, of itself, tell us what the governing processes are, but it imposes constraints which can guide us. At this level, a very simple model suffices. We can take an idealized geometry and describe the oxidation steps as the addition of an 'atom' above one or other of these sites. Obviously, the real process is more complex: the oxide is amorphous; it may be a molecular species moving; this molecular species may penetrate the oxide to react at the Si/oxide interface; each 'atom' represents at least half a silicon atom as well as an oxygen, etc. However, these issues can be addressed separately. Here we need to know only what dependences of process on $N$ will give layer-by-layer growth, and what these dependences imply for the kinetics.

If it is the sticking probability which is critical, then we shall assume some form $S(N)$. The fraction sticking will be a function of $N$, falling as the oxide gets thicker. Thus there is a preference for sticking to occur in gaps in the innermost layers, rather than starting a new layer on top of an incomplete one. Then follows a Monte Carlo calculation, in which atoms incident on the surface either stick or are reflected. The calculation is performed as if they arrived one at a time but, provided the sticking probability depends only on the site of impact, this assumption has no effect. From the models, we learn how fast the sticking probability $S(N)$ must depend on $N$ for layer-by-layer growth to be observed. It is this dependence which must be linked to physical mechanisms appropriate to the real system. The results from the model can be used too to analyse the roughness predicted in various ways, and we may also be able to relate the values of parameters to the possible physical phenomena like charge transfer process.

**2.5.2. Kinetics.** Since the number of incident atoms is effectively a measure of time elapsed, the results also say something about the apparent oxidation kinetics. The number of atoms incident is proportional to time $t$, whereas the average number of layers is proportional to the oxide thickness $x$. We shall follow Stoneham and Tasker (1987) in using the logarithmic derivative of oxidation velocity ($v = \mathrm{d}x/\mathrm{d}t$) as a function of thickness $x$,

$$g = -\mathrm{d}[\ln(\mathrm{d}x/\mathrm{d}t)]/\mathrm{d}(\ln x).$$

For Deal–Grove oxidation, $g$ lies between 0 (surface control) and 1 (diffusion control), whereas data for thin oxides can have $g$ as high as 10 or 20. Behaviour expected from the model can be indicated (no more) analytically by a simple assumption. The growth velocity is proportional to the rate of particles meeting the surface multiplied by an average sticking probability. If this average sticking probability is taken to be the sticking probability for the average number of layers (an assumption false in detail, but acceptable for trends, at least when the layers are nearly flat) then one finds:

$$g = -N(\mathrm{d}S/\mathrm{d}N)/S(N).$$

Thus for $S = A + B\exp(-CN)$ we find $g = NBC/[A\exp(CN) + B]$; for $S = A + B/N^D$ we find $g = BD/(AN^D + B)$.

Results with the exponential form ($A = 0, B = 1$) show that values of $C = 3$ would lead to layers which are apparently flat (values of $C = 2$ might be acceptable too; values of $C$ much larger than 4 lead to negligible growth after very few layers). If $C = 3$ is correct, then, since the layers of oxide are perhaps 0.2 nm thick, the model is indicating a sticking probability varying as $\exp(-x/0.7\ \text{Å})$. The scale length of 0.7 Å might reasonably be related to the characteristic length in a wavefunction overlap defining a tunnelling step. Whether this is true or not, the argument shows how a model can be related to some more complete description.

The standard Mott–Cabrera argument is as follows. Electrons transfer from the substrate to oxygen at the oxide/gas interface (where there are assumed to be enough adsorbed oxygenic species) until the reverse field built up prevents further transfer. There is then a field $E$ across the oxide, with $E = V/x$ with $V$ defined by the relative energies of the electron in the silicon and on the oxygen. This field drives the injection process, here of a charged oxygenic species into the oxide (the later Deal–Grove oxidation involves uncharged oxygen molecules for dry oxidation). How this is possible for a completely flat oxide is not clear, since the injected molecule will be in the same plane as the charges supposed to drive it. The field-enhanced injection, if valid, enhances the rate by a factor $\exp(Va/xkT)$, with $a$ the length of the incorporation jump along the field.

An alternative is the Stoneham–Tasker model. Here the driving force is the image interaction between the localized charge on the oxygenic species and the substrate silicon, which is substantially more polarizable than the oxide. For this model, the oxygenic species act independently. Very roughly, there is a term in the force driving diffusion proportional to the inverse square of the distance of the charge from the silicon/oxide interface. The enhancement will give a stronger dependence on $x$. There is also an image contribution to the incorporation energy.

When it is the incorporation step, rather than sticking, the argument is similar. One assumes that atoms first attach to the surface in a mobile form which may move across the oxide surface freely. The first calculations assumed that this motion was independent of the local value of $N$, i.e. that the atom can move easily onto higher or lower planes, but this is easily generalised to make motion to lower planes (smaller $N$) easier and harder to move up onto regions of higher $N$ (so resisting step growth). Again, the likely dependences on $N$ of $I(N)$, the ratio of incorporation probability to surface jump probability, are either exponential or inverse power laws. To get a rough order of magnitude, we (Stoneham and Harker to be published) have used continuum calculations based on the expressions given by Stoneham and Tasker. These show that (for a substrate with the dielectric constant of Si, a film of thickness $L$ with the dielectric constant of silica, and a point charge of one electron in magnitude moving from 1 Å outside to 1 Å inside the gas/oxide interface) the incorporation energy for the film of thickness $L$ is less than that for an infinitely thick film by an expression which is of the order of $k \times 170$ K for $L = 2$ Å, and which falls off slightly faster than $L^{-3}$ or as $\exp(-L/0.6\ \text{Å})$, these fits being approximate. Since the classical continuum electrostatic expression tends to underestimate the energy, it is likely that simple image enhancement of the incorporation energy can explain why the oxide layers grow a layer at a time without a step mechanism.

## 3. Oxide properties near the Si/oxide interface

### 3.1. Silicon oxides

Silicon dioxide can be produced in many ways. Its properties are similar or systematically related for many of these several forms. Thus we can use for guidance on thermal oxides the data on the various amorphous silicas and even for the crystalline silicas like quartz. This is especially important for stress response, plasticity and viscosity. In almost all these oxides, the basic network topology holds (oxygens twofold coordinated to silicons, silicons fourfold coordinated to oxygens), even in amorphous forms. Navrotsky (1987) has summarized data on enthalpy and volume per (Si$+$2O) unit for many such silicon dioxides (figure 3 is based on this work). An examination of her data summary shows two main features:

(i) for crystalline forms, the lowest energy corresponds to quartz, and more compact forms have an energy higher by an amount comparable with that expected from the elastic properties of quartz;

(ii) for amorphous forms, the lowest energy corresponds to a larger unit volume (by about 25%) and a higher energy (by about 0.25 eV/unit) than quartz, consistent, inter alia, with the large stored energies reported for radiation-amorphized quartz (Tinivella 1980).

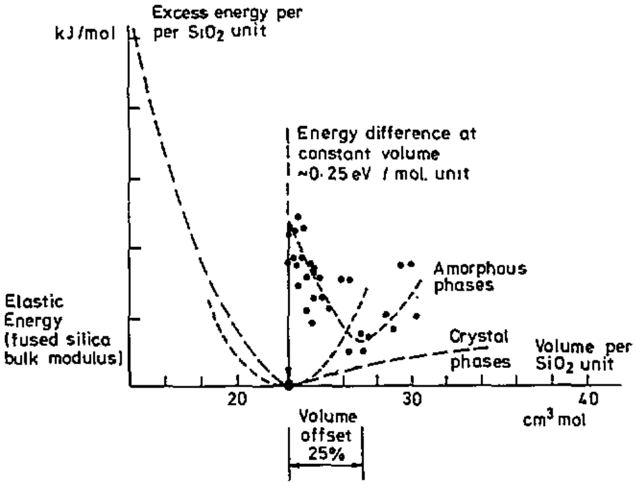

**Figure 3.** Enthalpy versus $SiO_2$ molecular volume of various forms of silica.

Whilst thermal oxides are generally amorphous, in special circumstances the oxide can be crystalline (Rochet *et al* 1986, Ourmazd *et al* 1987, Fuoss *et al* 1988, Bevk *et al* 1990) but this seems not to be important for the present discussion. Even in amorphous oxides it is an over-simplification to say there is no longer-range order greater than one or two bond lengths. Many glasses (Stevels 1944, Tinivella and Stoneham 1980) can be regarded as close-packed oxygens with interstitial cations, and the close packing has effects amounting to some longer-range order (Elliot 1991).

### 3.2. Origins and effects of stress on oxidation rates

Some properties of the thermal oxide result from interfacial stress, and this explains, at least in part, the deviations from Deal–Grove kinetics and the differences between dry and wet oxidation. The existence of stress is no surprise, since the number of silicons per unit area of a surface of silicon is nearly twice that for its oxide (this can be seen more easily by concentrating on the bonds which cut an imagined geometric plane through the two solids; clearly, the precise ratio depends on the plane and on the form of the oxide).

One significant early result (Pantelides and Long 1978) is that it is possible to join an $SiO_2$ network to a silicon substrate with all bonds complete and an interfacial energy which is not grossly high. What this result does not show, however, is (i) whether there are other, lower-energy, interfaces involving defects—in many systems, for instance, mismatch dislocations form, or (ii) whether the oxide network adjacent to the interface is similar to the oxide further out—if the oxide were crystalline, for instance, there might be more than one phase present; if amorphous, there might be more than one topology.

This early model also concentrated on bond energies (molecular mechanics, omitting what could be a substantial Coulomb energy). If the Si has a planar surface, and if all the oxide were to start with oxygens attached to it (rather than Si), then one would have a plane of charged oxygen ions, with a divergent energy for an infinite interface; such a polar interface would be a good example of the type which does not occur without extra conditions (e.g. facetting into anionic and cationic regions, or impurity compensation). Here possibilities are steps at the interface and a nearby region deficient in oxygen. The excess Coulomb energy puts constraints on the way the interface develops and on its structure; indeed, it may be this term which drives the creation of an intermediate layer more than stress itself. The same Coulomb term may well influence the roughness of the interface and step formation; crystalline interfaces are only noted for very small zones.

The plasticity of the silicon substrate itself (e.g. from interstitial generation) is important, not just the response to stress of the oxide. One view is that the oxide has an altered local topology in response to the stress, for example smaller Si–O rings. There is a wealth of papers on these aspects (surveyed by Mott *et al* 1989), though the term stress is used in several distinct ways, and the consequences claimed are sometimes naively large. Relaxation of the stress with time has been inferred from a variety of experiments. Landsberger and Tiller (1987) find two characteristic times are needed: one appears to correspond to the rapid local relaxation of compacted silica and the other to the viscous relaxation of silica, similar to that seen in bulk experiments.

The influence of stress is especially clear when small spheres (Hayashi *et al* 1990) or pillars (Liu *et al* 1992) are oxidized. Since the oxide has a far larger volume per silicon than elemental silicon itself, oxidizing a sphere leads to stresses opposing the in-diffusion and reaction of oxygen. The oxidation of gas-evaporated Si particles down to 10 nm diameter shows thermal oxide growing to about 2 nm thickness after anneal in air at 400 °C, with an O:Si ratio of about 2:1 (Hayashi *et al* 1990); this is in line with data from the pulsed laser atom probe, and also suggests an influence of small-scale roughness. The oxidation of sub-50 nm Si columns for light emission studies also shows stress retardation of oxidation, together with stress-induced generation and diffusion of Si interstitials and other phenomena.

## 3.3. Stress and the interface region

**3.3.1. Defect-influenced interfaces: the role of the $P_b$ centre.** Are there lower-energy interfaces than the direct match of defect-free dioxide and silicon? Here we may look at two new results concerning $P_b$ centres. In all models the $P_b$ centre is a dangling bond pointing from the Si into the oxide. The first important result (Stesmans 1993) is that the density of $P_b$ centres correlates with interfacial stress for a wide range of growth conditions. In this count of $P_b$ centres, those saturated by H must be included too. There is therefore the possibility that $P_b$ centres are the Si/$SiO_2$ analogue of mismatch dislocations. This has some support from a recent theoretical treatment of the $P_b$ centre (Ong *et al* 1993), suggesting strongly that there is an Si–Si bond in the oxide above the dangling bond. In such a model there are implications for control of noise and oxide perfection.

**3.3.2. Distributed stress: an interface 'phase'.** The amorphous structure and the possible intermediate layer between silicon and fully oxidized silica may be another consequence of stress. Suppose there is a layer of thickness $L$ and higher energy $\varepsilon$ per unit volume, and that this is a consequence of a mismatch $\delta$ between the silicon and its standard dioxide. The average strain in the layer will be $\delta/L$, leading to a strain energy proportional to $c(\delta/L)^2$ per unit volume, or to $cL(\delta/L)^2$ per unit area, with $c$ an elastic constant. The sum of this energy (proportional to $1/L$) and the energy $\varepsilon L$ has a minimum which defines the layer thickness. Of course, if the interfacial energy obtained omitting the layer is low enough, there will be no layer at all.

**3.3.3. Other models.** Battistiq *et al* (1994) have suggested that clusters of Si persist in the growing oxide film, a cluster consisting of tens to hundreds of atoms. This has an effect on the kinetics and also on the distribution of isotopes in experiments in which different isotopes of oxygen are used sequentially. To the extent that oxygen can react and become part of the network before reaching the continuous silicon phase, the model has elements in common with the reactive layer model. To the extent that the $P_b$ centre has an Si–Si bond above it, which could persist as the oxide grows further into the substrate, there is some link too to the $P_b$ centre model of stress relief.

## 3.4. Oxide defects: general remarks

Since an aim of understanding oxide growth must be how one can control charged defect concentrations in the oxide near to the Si, we turn to some aspects of these defects. We see later (section 6) that, for example, H is implicated in oxide wear-out behaviour. It cannot be stressed too strongly how large are the gaps in our knowledge. Even in quartz, where optical and spin resonance studies have clarified the nature of many defects, there are some major uncertainties. The many forms of hydrogen are one instance; the excited states involved in optically enhanced oxidation, or in the consequences of injected charges, are another. Such difficulties are compounded by the way in which the oxide changes close to the growing interface. Thus, when one turns to a problem like noise, or breakdown, it would be an achievement to know what are or what are not controlling factors!

**3.4.1. Fixed charge.** Fixed charges may be intrinsic or extrinsic. The extrinsic charges are often associated with Na or with H, though one should not forget that the cleaning process can leave surface charged species. The random potential has a range of effects, one relating to mobility close to Si/oxide interface (see Jain *et al* 1990 for a review).

**3.4.2. $P_b$ centres and similar defects.** Spin resonance and other methods show the existence of singly occupied silicon dangling bonds. These are presumed to be strictly at the interface; their number is small (less than one per hundred silicons; for a recent analysis see Stesmans and Van Gorp (1990)) and there is no reason to believe they play a major role in oxidation. The number depends on temperature, but varies little with oxygen pressure. These defects account for all the fast interface states (for recent views on these centres and their properties, see the special issue of *Semiconductor Science and Technology* vol 4, no 12, 1989).

Hydrogen passivation appears to work for $P_b$ centres. There are certainly other traps (some surely intrinsic) which can be passivated to a useful degree. One striking result is the significant difference in capture cross section (a factor 10) between H-associated and D-associated defects (Gale *et al* 1988). The larger value for H suggests a multiphonon process in which fewer, larger, local-mode phonons are involved.

**3.4.3. Noise and associated traps.** The $1/f$ noise observed in capacitor structures is associated with charge trapping. Systematic studies (Kirton and Uren 1989, Cobden *et al* 1990) show there are both 'fast' and 'slow' trap states. What is less clear is the nature of the traps involved. For example, is it the case that the electron and hole transitions (which can be written formally as:

Hole transition :

valence electron + [P] becomes [Pe] (1)

Electron transition :

[Ne] becomes [N] + conduction electron (2)

with [N], [P] representing defects of unspecified charge state which can capture electrons (so [P] could be a neutral acceptor and [Ne] a neutral donor). However, the trap is presumably in the oxide, and the carriers may well be in the silicon, so there is charge transfer across the interface. In consequence, the recent measurements giving both enthalpies and entropies associated with the traps raise interesting features.

Both the forward reactions ((1), (2) above) above have substantial positive entropies, of order $5k$. It is unlikely that degeneracies could give such large values; it is not very plausible that soft (low-frequency) vibrations could be effective either (this issue also arises in diffusion problems; note that experiment finds a constant-pressure entropy, not a constant-volume entropy—the two can even have opposite signs—and thermal expansion will ruin any precise compensation of force constants needed to give a high entropy). However, there is a simpler explanation. Suppose that the energy difference between the middle of the oxide bandgap and the middle of the Si gap is almost independent of temperature (this is not necessary, but simplifies explanation). Then the known strong temperature dependence of the Si bandwidth (which falls by 0.6 eV when $kT$ rises by about 0.1 eV) means that there will be a significant positive entropy (perhaps $3k$ or more) from a deep level (assumed to stay at mid-gap) for both the electron and hole channels.

What might the slow traps be? They appear to be intrinsic; they show large lattice relaxation after capture by tunnelling. The traps appear to be discrete, that is as well defined as is expected for an oxide which is non-crystalline. Could they be associated with self-trapping?

**3.4.4. Self-trapping in silicas.** In self-trapping, a free carrier—unassociated with any defect—can lower its ground-state energy by causing a lattice distortion; the localized carrier is then self-trapped. The phenomenon was first predicted by Landau, who also realized that there could be a thermal barrier to the self-trapped state. The first unambiguous observations were on KCl, where Kanzig observed the formation of Cl molecular ions, essentially a hole localized on two chloride ions. Later observations (some inferred from transport measurements) suggest self-trapping is widespread, and the barrier to self-trapping was verified too in some cases (for a review see Stoneham (1990) and Shluger and Stoneham (1993)). In the several forms of silica, there is evidence for self-trapping. In quartz, the well known blue luminescence comes from the self-trapped exciton: in effect, the hole component of the exciton is localized on an oxygen, which is substantially displaced, and the electron is localized partly at that oxygen and partly at a neighbouring silicon (Fisher *et al* 1990). Since vitreous silica shows similar luminescence, self-trapped excitons are presumably involved there too.

Whilst the evidence for self-trapped excitons is clear, the position for single carriers (electrons or holes) is less certain. Indeed, in quartz neither seem to be trapped, though a modest perturbation, like substitutional Ge, leads to trapping of both electrons and holes. In amorphous oxide, where there is a site-to-site variation of properties, the situation is different, and Griscom (1989) has spin resonance evidence for self-trapping in silica glasses. Mott and Stoneham (1977) suggested delayed self-trapping as a partial explanation of transport in thin oxide films on silicon. What is the situation in thin oxide films? Clearly, self-trapping cannot be far away: not only is it present or close in other silicas, but the non-crystalline nature of the oxide will help too. The situation is further complicated by the fixed charges known to be present, since these will also encourage charge localization (e.g. an electron would trap preferentially close to an $Na^+$ interstitial).

Yet there is a further term favouring localization too: self-trapping could occur through the extra stabilization because of the electrostatic image terms, for Si has a relatively large dielectric constant (Stoneham and Tasker 1987). If self-trapping does occur, it would be favoured very close to the silicon, so that self-trapped electrons or holes would be prime candidates for the slow states giving $1/f$ noise. This possibility, still conjecture, has implications for radiation effects. The first is for radiation-enhanced adhesion, for charged defects lead to an extra term in the interfacial energy when adjacent to a medium with higher dielectric constant. The second aspect is radiation stability. Ionization damage is well known to occur in silicas, when purely electronic excitation suffices to create vacancies and interstitials. To create damage, energy localization is needed. In cases where mechanisms are understood, it is normal for one intermediate stage to be a self-trapped species. Thus four key questions arise. First, do carriers self-trap at the Si/thermal oxide interface? Secondly, are these self-trapped forms related to the slow traps giving $1/f$ noise? Thirdly, are these same species formed during exposure to ionizing radiation? Fourthly, what defect reactions are possible involving these self-trapped states? This fourth question brings us back to the oxidation process, and to the substantial series of experiments on optical and electronic excitation. The many models and rationalizations still leave an incomplete picture. This is, of course, not the only aspect of charge transfer between silicon and its oxide which remains unsettled. Thermionic emission ideas (Irene and Lewis 1987) suggest electron flux can be a rate-limiting step. Other open questions relate to the energy needed for an electron to dissociate an interstitial oxygen molecule.

## 4. Watching oxidation happen

Recent experiments to 'view' oxidation as it happens have given new insights into interfacial reaction rates and the mobility of Si atoms during oxidation. Recall in the discussion above (section 2) that these issues need to be considered when trying to understand the departure from Deal–Grove kinetics for thin film growth. These studies also give clear insights to the initiation of the decomposition reaction of $SiO_2$ which we noted was not included in current oxidation theory and was also of technological importance (section 2.3). Any extension of oxidation theory to the thin film regime will also have to include these new observations.

### 4.1. Scanning tunnelling microscopy (STM) of Si oxidation

The STM works by bringing a sharp metal tip into close proximity (< 3 nm) with the surface being examined. A voltage is applied across the resulting gap and a tunnelling current is measured; thus only conductors can be examined. An image of the surface is built up, usually by translating the tip in $x$ and $y$ directions while recording the feedback signal required to maintain a constant tunnelling current. The fine control of $x, y, z$ coordinates is achieved with various arrangements of piezoelectric motors, the rigid tube scanner being currently favoured. With steady development of control electronics and software, and vibration isolation, coupled to UHV surface science techniques the STM has become a powerful tool for the examination of chemical reactions at submonolayer coverage with atomic resolution.

The images of surfaces obtained with an STM can be complex. An appreciation of the subtlety involved can be gained from a simplified expression for the tunnel current, $i$, given by Tersoff and Hamman 1983. For a tip approximated by a single point

$$i \propto \sum_{s} |\psi_s(r_0)|^2 \delta(|E_s - E_F|)$$

where $|\psi_s(r_0)|^2$ is the local density of states (LDOS) with energy $E_s$ evaluated at the point $r_0$ and $E_F$ the Fermi energy of the sample. The sign and magnitude of the tip voltage determine whether empty or filled surface states are sampled and their energy relative to $E_F$. We see from this that part of the image contrast comes from the variation in LDOS and that LDOS energy level spectroscopy is possible. Furthermore Tersoff and Hamman show, using the Bardeen perturbation calculation of tunnelling from metal (the tip) through an insulator (the vacuum gap in this case) to a metal (the sample surface), that

$$i \propto \sum_{s,T} |\psi_s(r_0)|^2 |\psi_T(r_0)|^2 \exp(-2kd)$$

where $|\psi_T(r_0)|^2$ is the density of states of the tip, $k \propto \phi$, the work function of the surface, and $d$ is the width of the tunnel gap (i.e. the height $z$ of the tip above the surface). Thus we see that the tunnel current is exponentially dependent on the tip to surface distance, typically a 0.1 nm change in $d$ gives an $e^2$ change in $i$. It is this sensitivity in $d$ that leads to the closest atom of the tip acting as the microscope 'aperture' and giving atomic resolution in the image.

The combination of the sensitivity of the tunnel current to the LDOS and the atomic scale variations in height on the surface can lead to complex images, the interpretation of which can require more sophisticated models of the tunnelling process than given above. Good examples of the problems encountered and there resolution can be seen from STM studies of GaAs (Feenstra *et al* 1994), and dichalcogenides (e.g. Crossley *et al* 1994a). Note also that at typical tip operating voltages of a few volts and gap currents of 1 nA, the gap field and current density can be large and may perturb the surface.

Fortunately, in the case of Si surfaces the dangling bonds ($p_z$ in character) are localized at the positions of the Si atoms and hence the LDOS variations and atomic 'corrugations' are predominantly occurring at the positions of the Si atoms when imaged at the appropriate tip voltage. It is also the dangling bond sites that are the most likely to be 'attacked' by adsorbates and reactants. Alternatively there are energies where the LDOS variation from one atom site to another is small and the images are dominated by the atomic height variations. An example of the now well known Si(111) 7 × 7 reconstructed surface imaged in empty states at −2 V, where the LDOS is similar for all atom sites (Bolland 1993), is shown in figure 4.

The STM has thus proved a very useful tool for imaging the atomic sites which are involved in surface reactions at least for submonolayer coverage. A more detailed discussion including tunnelling current/tip

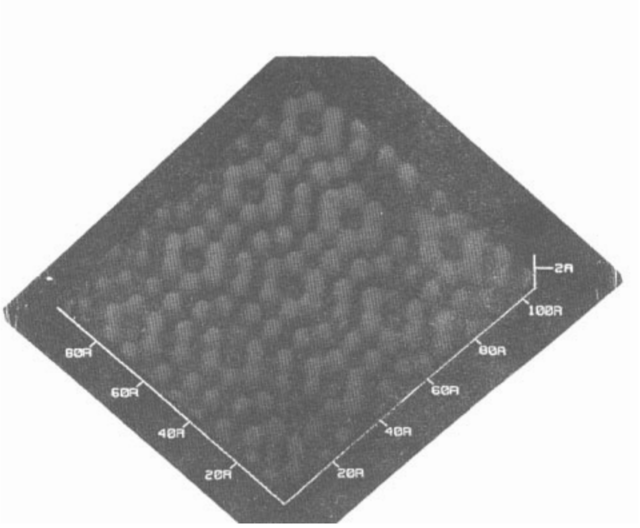

**Figure 4.** The STM image of the reconstructed Si(111) surface imaged at −2 V where variations in LDOS are small and contrast is dominated by the atomic corrugations of Si atoms in the surface.

voltage spectroscopic data of such systems can be found elsewhere for H reactions (Bolland 1993), and for oxygen reactions (Leane 1990, Golen 1992).

In the following we will concentrate primarily on data indicating the relative rate of oxygen attack at differing Si atomic sites and any differences between attack at atomic steps or on terraces. The bonding of the oxygen with the Si surface has a profound effect on the LDOS and hence gives rise to clear changes in the surface images for submonolayer coverage. Once a monolayer of oxygen has covered the Si surface, the system of STM probe, oxide layer and Si substrate behaves as an MIS structure with tunnelling to bulk Si states; this prevents the attainment of atomic resolution. Thus what STM observations provide is an insight into the initial stages of oxygen attack on the ordered Si surface. The STM studies described below, however, are in an oxidation regime well outside that where Deal–Grove oxidation occurs (see figure 2).

**4.1.1. Room temperature results.** The ordered Si surfaces (100 or 111) to be examined are normally prepared by annealing at temperatures sufficiently high to remove the native oxide, to remove any SiC contaminants, and to allow sufficient Si mobility for surface reconstruction to occur. During the annealing residual gas pressures need to be kept low ($< 10^{-9}$ Torr). Several recipes for this procedure are given in the literature. The observations of oxygen attack on the ordered Si surface structure are made at room temperature by introducing oxygen gas, after first imaging the unreacted Si surface. The introduction of the oxygen can be made whilst tunnelling is in progress (hence in the presence of a high electric field from the STM tip) or with the STM probe withdrawn (and hence zero external field). The resultant oxygen 'defect' induced images are different in the two cases; presumably the high and inhomogeneous electric field from the tip can perturb the oxygen reaction with the Si surface (e.g. Ramos *et al* 1993).

*The Si(111) 7 × 7 reconstructed surface.* Submonolayer oxygen coverage produces two kinds of 'defect' in the STM images. Type 1 appears black (in empty state images) as though an adatom is missing in the surface and type 2 appears white (in filled state images) as though protruding above the surface (Leane 1990) (see figure 5). Both oxygen defect centres increase in proportion to oxygen exposure but the faulted half of the unit cell (i.e. influenced by the next layer of Si down) has a higher reaction rate as do sites adjacent to already affected Si atoms.

It is proposed that the type 1 reaction is the result of reaction by two oxygen atoms, one attacking the dangling Si bond of the adatom and the other with the Si subsurface backbond. The type 2 site is postulated as the attack of a single oxygen atom in a Si backbond site. Further work, by e.g. low-energy ion scattering, to ascertain the position of the oxygen atoms relative to the Si atoms may help to elucidate these structures. These STM observation are consistent with the observation of two reaction rates in thermal desorption studies of submonolayer oxidation reactions at similar temperatures (Gupta *et al* 1989).

It is interesting to note the STM observation of oxidation of the H-terminated Si(111) 1×1 reconstructed surface at room temperature indicates strong inhibition of the oxidation reaction (Neuwald *et al* 1992). Clearly surface 'impurities' can strongly perturb reactions at the surface.

*The Si(100) 2 × 1 reconstructed surface.* The Si(100) surfaces to be imaged with the STM are prepared by high-temperature annealing while maintaining low pressure ($< 10^{-9}$ Torr) to give a 2 × 1 reconstructed surface with large terraces and, in general, a low degree of surface defects (e.g. Golen 1992). The surface defects are most likely to be missing Si dimers. Adsorbed H, or OH pairs can give similar images (Anderson and Kohler 1993) but the preparation conditions make this unlikely. The surface is exposed to low pressures of oxygen and generally filled state images are taken to observe the change in 'defect' level arising from oxygen exposure (see figure 6). The filled state images show bright contrast at the Si dimer $\pi$ bond sites (high LDOS) and would be expected to show 'dark' defects where oxygen reacts and the LDOS is reduced.The oxygen is believed to dissociate and an oxygen atom to insert into the $\pi$ bond site.

The observations of Golen show an increase in the density of 'dark' defects with increasing oxygen exposure. These sites are randomly dispersed over the surface showing no preferential atomic sites for oxidation even at step edges. However, there is evidence of oxygen attack at sites neighbouring already oxidized sites, leading to an island growth appearance. The relative surface coverage of 'dark' sites grows linearly with oxygen exposure. There is some evidence that the oxygen can move (although low probability) on the surface before reacting.

Another study (Cahill and Avouris 1992) found apparently conflicting observations, with 'bright' spots

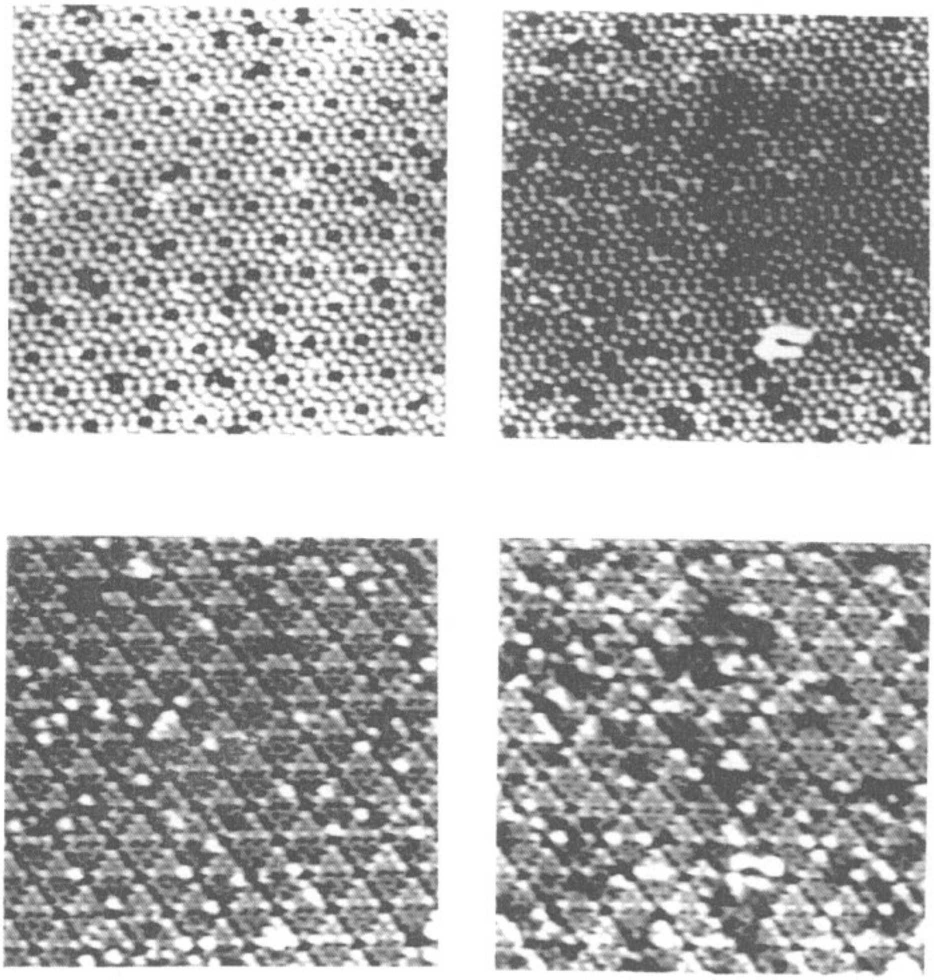

**Figure 5.** Top are empty state images of the reconstructed Si(111) surface before (left) and after (right) submonolayer oxidation. The oxidation gives rise to 'dark' defects in this image, i.e. a reduction in LDOS. On the bottom are images of filled states before and after oxidation showing the resultant (white) defects that appear to 'stick up' from the surface.

appearing on exposure to oxygen. This observation was interpreted as resulting from an oxygen reaction inducing movement of Si dimers to the top of the original Si surface. The data show these bright spot features at random over terraces, without preferential attack at step edges. Their data also show a high density of 'dark' features, which they interpret as oxygen reaction induced defects. The 'bright' features seen in this study were not observed by Golen.

These two experiments were carried out with at least one significant difference. In the first study (Golen 1992) the oxygen was introduced with the tunnelling tip withdrawn, and hence with no possible external perturbing electric field during the reaction. The study of Cahill and Avouris was carried out with the scanning probe still tunnelling while oxygen was introduced. The electric field in this case may be the reason for the different observations, especially as the tunnelling field is sufficient to perturb the buckled Si dimers rows in the unreacted Si(100) 2 × 1 reconstructed surface (Ramos *et al* 1993).

In another recent study the Si(100) 2 × 1 reconstructed surface was examined on exposure to $H_2O$ (Anderson and Kohler 1993). Their results are very similar in most respects to those of Golen. The reaction leads to 'dark' defects which are randomly spaced over the Si surface. No preferential attack is seen at step edges and attack is favoured at sites neighbouring already reacted sites. There is a low probability that $H_2O$ can move a few sites before the final reaction.

The room-temperature submonolayer oxidation results are consistent with the substoichiometric model (all be it simplistic in terms of network structure) that Engstrom *et al* (1992) derive from molecular/atomic oxygen beam oxidation and XPS examination of similar surfaces. Hattori (1994) also derives a similar model of the interafcial oxide from high resolution XPS studies of thin oxide growth but on H-terminated surfaces.

#### 4.1.2. High temperature oxidation/oxygen etching.

STM observation of the reaction of oxygen with Si at elevated temperatures presents considerable experimental difficulty due to the need to suppress differential thermal expansion between the STM tip and the Si surface. Rapid scan techniques and effective heat shielding have recently made such studies possible.

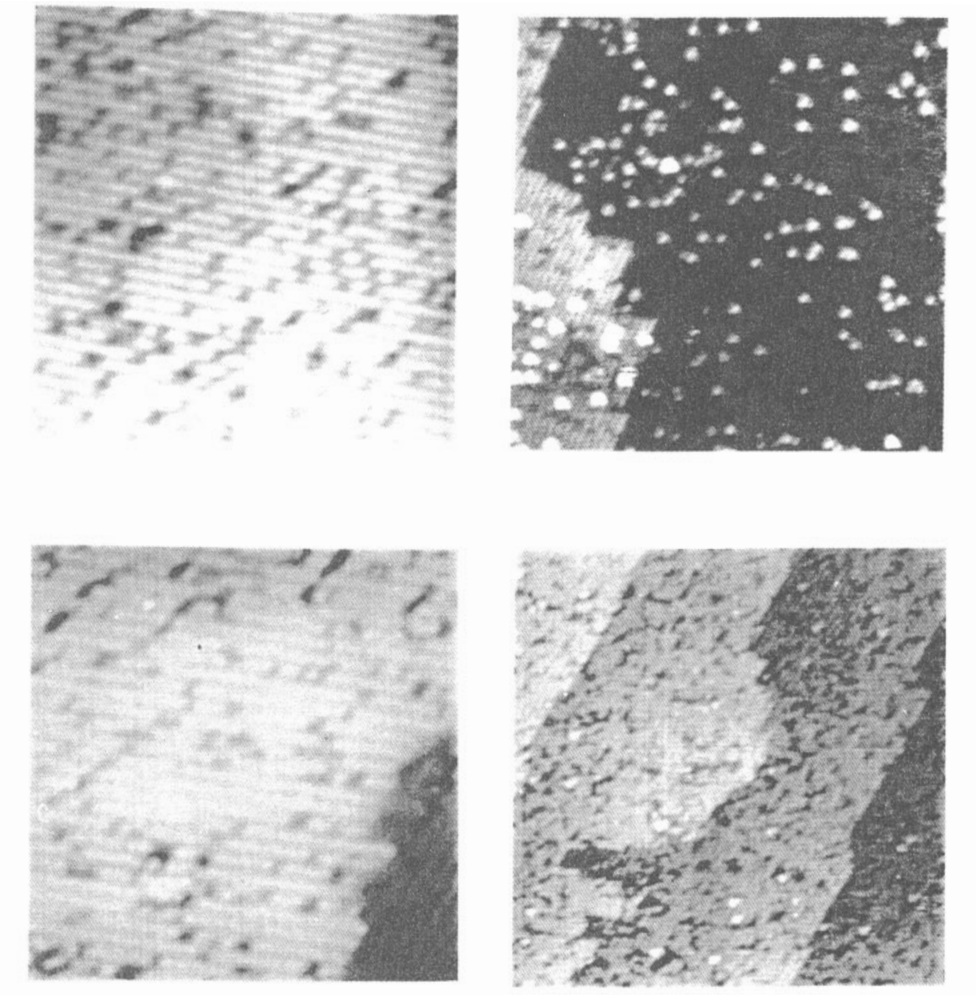

**Figure 6.** STM filled states images of the reconstructed Si(100) surface taken before (left) and after (right) submonolayer oxidation. The top images are obtained by oxygen exposure during tunnelling while the bottom images are taken with oxygen exposure with the tunnelling tip withdrawn. Clearly the difference in the defect structure is perturbed by oxidation in the presence of the strong field and/or current density accompanying tunnelling.

At elevated temperatures (above 550 °C) the oxygen reaction leads to Si etching at low oxygen pressures and oxidation at higher pressures depending on the Si temperature.

Wong and Welland (1994) studied the interaction of oxygen with the reconstructed Si(111) surface. The Si surface was heated to 550 °C and exposed to $5 \times 10^{-9}$ Torr of $O_2$ for 3 min. They observed etching of the Si (presumably to form volatile SiO) to occur preferentially at step edges, defects and dislocations. When the oxygen partial pressure was raised to $2 \times 10^{-8}$ Torr for 3 min they observed oxidation with homogeneous nucleation, hence immobile step edges (see figure 7). The $SiO_2$ growth was a single monolayer thick for coverage up to nearly 100%.

A similar experiment was carried out (Feltz *et al* 1992, 1993) on the $7 \times 7$ reconstruced Si(111) surface at 850 °C and at 1050 °C. They used an oxygen pressure of about $3 \times 10^{-8}$ Torr and saw etching with similar behaviour to that observed by Welland and Wong at lower temperature. In this study the $O_2$ pressure was too low to be in the oxidation regime (see figure 2). Other, related *ex situ* STM work (Johnson and Engel 1992) modelled to show the void formation indicated the rate limiting step is the formation of a mobile Si monomer that diffuses to void edges to rapidly interact to reduce the oxide to the volatile suboxide.

STM observations of the oxidation of $2 \times 1$ reconstruced Si(100) surfaces at a temperature of 600 °C have been reported (Udagawa *et al* 1993). The results were similar to those reported above for oxidation of the Si(111) surface. The nucleation is essentially homogeneous with no preferential nucleation at step edges and proceeds monolayer by monolayer.

Again these higher-temperature STM results are at least consistent with the substoichiometric model that Engstrom *et al* (1992) proposed from molecular/atomic beam growth studies and XPS analysis. Note, however, that in their work a transition to a denser $SiO_2$ stoichiometric layer was inferred for oxidation temperatures above 900 K; thus further STM observations at higher temperature would be interesting to see.

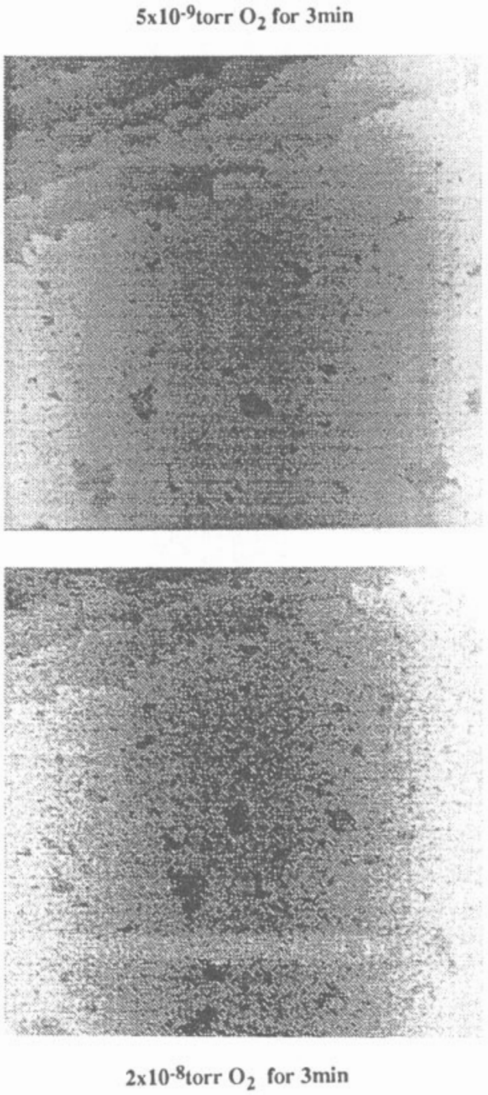

**Figure 7.** STM images of reconstructed Si(111) surface (5000 Å × 5000 Å) taken at 550 °C after exposure to oxygen. Starting at low pressure (top) the oxygen 'etches' preferentially at step edges leaving them raggy, while at higher pressure (bottom) the oxygen attacks the surface randomly, maintaining the step edge pattern.

**4.1.3. Electron microscopy through the growing oxide.** In an elegant series of transmission electron microscope (TEM) studies (Ross and Gibson 1992, Gibson *et al* 1993, Ross *et al* 1994) the observation of the attack of oxygen on the Si(111) 7 × 7 reconstructed surface has been made. A TEM was fitted with a heater stage and extra pumping to achieve the UHV regime common to the STM experiments described above. The use of forbidden reflection (FR) techniques, i.e. the observation of diffraction patterns that cannot arise from the symmetry of the bulk structure of Si, allow the observation of steps on the Si surface and of reconstructions on the Si surface terraces. These observations can be carried out through an oxide layer to 'view' the interface. Thus these studies cover submonolayer coverage through to several monolayers of oxide growth and hence extend our knowledge beyond the STM observations discussed above.

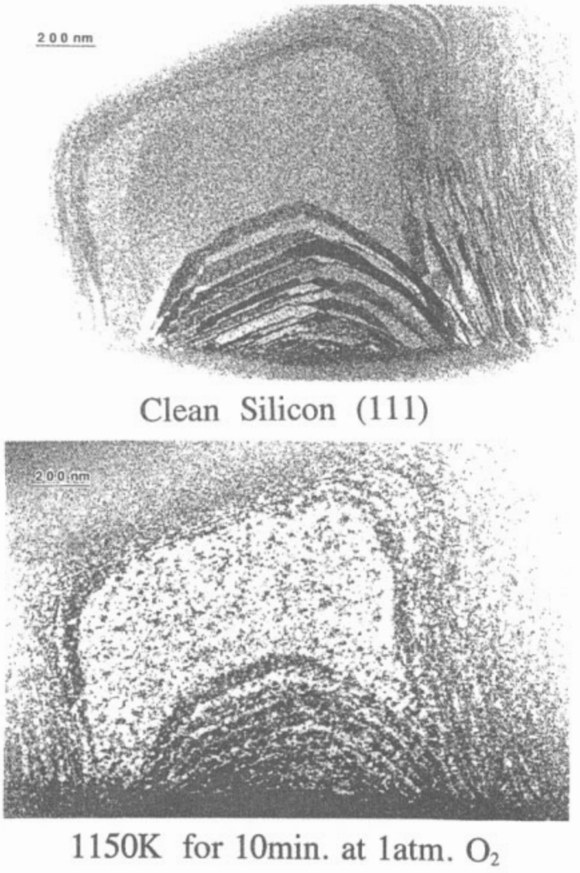

**Figure 8.** TEM images of reconstructed Si(111) surface with step and terraces before and after oxidation. The step edges are preserved under oxygen attack.

Observations were made of the oxidation of a reconstructed 7 × 7 Si surface *in situ* at temperatures up to 900 °C; the FR reflections are stronger on the Si(111) surface than on the Si(100). They observed that at $O_2$ pressures high enough to be in the oxidation regime Si step edges do not move during oxidation even for several monolayers of oxide growth (see figure 8). In an extension of these studies (Gibson *et al* 1993) to the 2 × 1 reconstructed Si(100) surface similar $SiO_2$ growth behaviour was seen. It was also concluded that the growth process proceeds monolayer by monolayer and the reconstruction is lost after about a monolayer growth. In the low-pressure regime where SiO desorption occurs the oxide was observed to preferentially attack step edges and lead to a roughened Si surface. They also note that their technique would be sensitive to the presence of crystaline $SiO_2$ (e.g. tridamite) at the Si/$SiO_2$ interface but that none was observed.

**4.1.4. What have we learned from these direct observations?** Several important new insights into Si oxidation emerge from these studies. The step edges on both Si(100) and (111) surfaces do not move during oxidation indicating that Si does not significantly diffuse laterally during oxidation. The steps do not affect lateral diffusion and reactivity. Previous sequential $^{18}O_2/^{16}O_2$ tracer experiments on Si oxidation (Murrell *et al* 1991) have shown that Si diffusion through the oxide does not occur. These observations fit well with current Si oxidation theory which do not need to invoke Si as a diffusing species.

In the case of submonolayer oxidation the reaction rate depends on the specific Si atom site for the (111) surface but not for the (100) surface. However, it is most likely that the Si surface reconstruction is modified on both these surfaces at the site of the oxidized Si atom and is totally lost as oxidation proceeds beyond a monolayer coverage. In all oxidation kinetics experiments on thin oxide growth known to us, where departures from the Deal–Grove model are observed, the Si interface or surface is not in the reconstructed state used for the STM and TEM observations described here. The surface may be hydrogen terminated, covered with 'native oxide', and/or rough with various degrees of faceting between (100) and (111). Thus the observation of reaction rate dependences on Si atom site cannot be used quantitatively to extract the expected effect on growth kinetics for these surfaces. However, following the discussion in section 2.4 above, some effect on kinetics is possible. As the Si atom positions at the interface evolve with film growth, particularly for the thin oxide regime, any consequent reactivity changes would be expected to contribute to the departure from Deal–Grove kinetics.

The low-pressure reaction rate to form volatile SiO has been shown to depend on oxygen coverage for submonolayer coverage. The reaction appears to occur more easily than implied by extrapolating from the higher temperature desorption data (Lander and Morrison 1962). This raises the question of whether or not the PLAP data (see section 2.2) showing an interface layer (about a monolayer thick) of SiO do not arise from this easier oxide decomposition for very thin oxides. There is a high-temperature/low-$O_2$ pressure regime where desorption and oxidation proceed together to give a rough etched surface (see also Feltz *et al* 1993).

The STM data also have a useful input to Si oxidation technology. If one starts from a reconstructed Si(100) surface with steps and terraces providing the dominant source of roughness (e.g. epi wafers) and the oxidation process proceeds monolayer by monolayer the roughness would not be expected to change. This appears to be so at least for the 'native oxide film thickness on epi wafers (see section 5.3.1). It is not clear what the result on roughness changes accompanying oxidation would be when starting with Si surfaces that are rough as a result of microfaceting.

Even at low partial pressures and at room temperature oxygen reacts so rapidly with a clean Si surface that the surface will not stay free of oxide after transfer from one piece of process equipment to another. UHV or ultra-clean gas transport is essential, and $O_2$ and $H_2O$ partial pressures must be kept low (about $10^{-8}$ Torr for it to take about 1 h to grow a monolayer). This is quite a stringent demand on atmospheric pressure pure inert gases, i.e. one part in $10^{11}$ impurity level.

The etching of Si preferentially at step edges and defects at low pressure and high temperature could lead to surface roughening, perhaps dramatically so if the pressure is such that there is a balance between growth and loss. For this reason ramp-up to oxidation temperatures in too low partial pressures of $O_2$ should be avoided. We consider the import and measurement of roughness below (see section 5).

Recently XPS studies (Hattori 1994) of the oxidation of H-terminated Si(100) and Si(111) surfaces have been used to propose models of the interfacial oxide structure and Si oxidation states. STM examination of such surfaces being attacked by oxygen would be of interest to obtain for comparison.

## 5. Roughness and its implications

The surface roughness of the Si that is to be oxidized is not a parameter usually included in theoretical models of oxidation (see section 2.1). The recent STM and TEM experiments (section 4) where carried out on flat Si surfaces except for a low density of atomic steps and missing atom defects. MOS capacitors fabricated on such surfaces have shown yield and breakdown strengths somewhat worse than those found for capacitors fabricated in the normal way (Leihr *et al* 1992, Sofield *et al* 1992). Partly this was due to SiC contamination arising from heating hydrocarbon contaminated H-terminated Si surfaces. Even when the hydrocarbon contamination is avoided the oxide grown on flat clean reconstructed Si(100) was found to break down at fields some 15% lower than oxides grown on 'native' oxide covered surfaces; the reasons for this are not understood. However, the surface of Si usually oxidized for most kinetics and other experiments, such as defect studies, is not the high-temperature UHV reconstructed Si surface. The 'technological' Si surface on which gate oxide is formed for VLSI CMOS devices is also not prepared in this way. We will see in the following that the roughness of a Si wafer is an important parameter in determining gate oxide performance and that the roughness depends on gate oxide precleaning methods. Obviously the requirement for thinner oxides means that the Si surface roughness is expected to become relatively more important.

### 5.1. Effect of roughness on device performance

It is to be expected that surface/interface roughness at the boundaries of layers in which electrical transport occurs, such as transistor gates and interconnects, will give rise to additional scattering processes and may degrade device performance. Atomic scale roughness at the boundaries of metallic conductors has been shown to degrade interconnect performance (e.g. Tung 1985). The effect of roughness on Schottky barrier height is known to be important (see Hasegawa and Ino 1993). Recent studies have shown roughness to be an important parameter to minimize and control at the gate oxide/Si interface in VLSI submicron device processing. A degradation of electron channel mobility in n-MOSFETs as the roughness of the Si/$SiO_2$ interface of the gate structure increases has been demonstrated (Ohmi *et al* 1991). Heyns *et al* (1992) have demonstrated that rough Si surfaces produce a decrease in the performance (breakdown yield) of MOS capacitors. In a further study of MOS capacitors (Nakanishi *et al* 1993) similar deleterious effects of roughness on breakdown and also interface state density and degradation under current stressing are reported. In some cases inappropriate cleaning processes have been shown to give large, sharp spikes which probably degrade oxide breakdown by field enhancement (e.g. Cohen *et al* 1994). A recent study of breakdown of thin oxides grown on Si(100) and Si(111) surfaces, even when roughness is small and similar in both cases, shows the Si(111) surface to be inferior (Nakamura *et al* 1993). In this case evidence was also found for an oxide structure difference. Note also that capacitors grown on flat reconstructed Si(100) material with low step density performed worse than capacitors grown on 'chemical oxide' terminated surfaces when the oxidation conditions were the same in both cases (Sofield *et al* 1992). Studies of very thin gate oxides (< 6 nm), where direct tunnelling starts to be an important current drain mechanism, have been made. In these studies the silicon surface roughness was varied from about 0.1 nm to about 0.2 nm and no evidence of an effect on gate oxide integrity was found (Hirose *et al* 1994, Depas *et al* 1994). Thus it appears that for thicker oxides and high levels of roughness (perhaps correlated with impurity levels) gate oxide integrity is compromised. However, for very smooth surfaces (single atomic steps) and for thin oxides the gate oxide integrity appears less dependent on interface roughness if it is on a small scale. The thin oxide regime where direct tunnelling is important, however, has only just started to be studied (oxide growth and thickness control are difficult to achieve) and hence it is premature to conclude that roughness will no longer matter as some effect on the variation in tunnelling barrier width would be expected.

Some of these studies used scanning probe microscope techniques to determine roughness, and clearly demonstrate the need for roughness control at the atomic dimension level. In the following the term 'roughness' is used in a relative sense, usually being quantified as, for example, root mean square (for surface profile methods such as AFM) or relative intensity (for scattering techniques). It may well be that some particular part of the roughness spectrum, for example the sharpest, highest features for breakdown studies, should be evaluated or, for example, the roughness spectrum has to be convolved with tunnelling current through a uniform oxide.

### 5.2. Contaminants and roughness

The intrinsic roughness of the Si surface prior to gate oxide growth is determined by factors such as the method of Si wafer production, CZ, FZ or epi (see e.g. Miyashita *et al* 1992), the chemomechanical polishing used to produce the finished wafer surface, and pre-gate oxide chemical cleaning. Extrinsic factors such as particulate, organic and metal contaminants may also contribute to the surface roughness before oxidation and to the interface roughness once an oxide is grown. Many of the methods of roughness measurement outlined below cannot distinguish between these extrinsic factors and intrinsic roughness.

In the device studies outlined above (section 5.1) the Si surface roughness was varied to study correlations with device performance by changing the pre-gate oxide chemical cleaning procedure. The dependence of Si surface roughness on pre-gate oxide cleaning, a key process step, is critical to studies of Si surface and $Si/SiO_2$ interface roughness. As such, this aspect of preoxidation cleaning has received much attention and is a rapidly developing field. Detailed discussions of this subject are given by Hattori (1990), Ohmi *et al* (1992), Verhaverbeke (1993) and Higashi and Chabal (1994), to select those best known to the authors. We outline the salient features below. Although this information is well appreciated by VLSI fabrication research groups it is essential that careful detailed attention be given to these matters when studies of roughness and oxidation are undertaken.

The cleaning procedures based on a HF last wash are currently of interest, partly because they leave the surface with a single monolayer H-terminated surface resistant to reoxidation. The cleaning stages prior to the HF treatment must be efficient at removing metal contaminants and particles while not 'etch' roughening the surface. It has been shown (Heyns *et al* 1993) that metal contaminants present in cleaning solutions can lead to surface roughening. The oxide grown on such a surface is degraded due to the roughness and/or the metal contamination. The Si(111) surface can be preferentially etched by BHF to smooth it, while DHF cleaning of Si(100) can be carried out to avoid further roughening. As yet no smoothing wet 'etch' has been found for Si(100). Instead, it is all too easy to roughen the Si(100) surface by wet cleaning. Modifications to the standard 'RCA' clean to yield these smoother cleaner surfaces have been developed recently. Once chemical cleaning has been carried out it is important to avoid recontamination; H-terminated surfaces are especially sensitive to hydrocarbon contamination.

Once a clean surface has been achieved and the Si transported to the oxidation furnace without recontamination it is important to avoid at least two problems that can arise during the oxidation procedure. If $SiO_2$ layers on Si, for example, thin 'native' oxide films remaining after pre-gate oxide cleaning, are heated to sufficiently high temperatures in low partial pressures of $O_2$ or $H_2O$ the oxide is removed by a $Si/SiO_2$ reaction to produce volatile SiO. The process is initiated at localized defects, at least some of which can be metal impurities (Verhaverbeke *et al* 1992). Once void formation is initiated (i.e. a hole through the $SiO_2$ is formed) it proceeds by Si migration to form a roughened Si surface at the completion of oxide desorption (Tromp et al 1985). It is therefore important in oxide growth processing schedules to avoid such heating in low partial pressures of $O_2$ or $H_2O$. It is even worse if Si is heated in a partial pressure/oxidizing regime where a balance of oxidation and oxide reduction can be struck to produce continuous etching. Even heating in the oxide desorption range for times too short to observe void formation can degrade oxide quality (low-field breakdown; Rubloff *et al* 1987). If the Si surface to be oxidized is H-terminated and contaminated with hydrocarbons, heating in low oxygen partial pressures may lead to SiC particulate formation with consequent roughening and degradation of oxide breakdown performance.

### 5.3. Observations of roughness

**5.3.1. Roughness of the Si surface prior to oxidation.** Roughness can only be controlled (systematically) if it can be measured. This is not trivial since, as we have already noted, roughness on a vertical and spatial scale of atomic dimensions has been observed to be important to MOS capacitor reliability. Moreover one has to also make measurements over the dimensions of a whole Si wafer surface if issues of variability and yield are to be addressed. However, current scanning probe methods are too time-consuming to cover the atomic length scale routinely over a whole wafer.

The most commonly used method of roughness assessment is based on laser scanning light scattering (LSLS) techniques which are used to detect haze and light point defects assumed to arise from microroughness and 'particle' light scattering respectively off the surface of a native oxide covered Si wafer. Note that the laser light can penetrate the Si surface and hence subsurface defects may also be detected. A whole wafer can be scanned in a few seconds. The technique gives very high sensitivity to small particles (greater than 0.1 $\mu$m diameter) when unpatterned highly polished wafers are examined. It is non-trivial to relate the haze light scattering intensity distribution to the roughness distribution (Vorburger *et al* 1993). This is the favoured method for assessing process equipment sources of particulates in fabrication plant (Hattori 1990, Singer 1993). Although particulates are known to be a significant source of failure in MOS gate oxides (Monkowski 1987) we will focus attention on intrinsic surface roughness. From this point of view, relative haze measurement becomes the useful parameter of interest.

Other optical methods, such as interferometer microscopy and scanning optical microscopy (SOM), may also be used for surface roughness determination. These instruments are usually slower and, in practice, cover smaller areas than those assessed by laser scanning instruments. These instruments can give subnanometre vertical resolution and submicron spatial resolution (a useful discussion of these instruments is given by Dong *et al* (1994)). An interesting variant, the laser electrochemical microscope (Kucernak *et al* 1992), can be used to image electrically active defects under an electrolyte.

The current drive to smaller MOS gate structures and thinner gate oxides gives rise to the need for roughness measurement with atomic resolution (as mentioned in section 5.1). Scanning probe microscopes, particularly the STM and AFM, offer the required vertical resolution; STM also provides the required spatial resolution when examining clean Si surfaces free of 'native oxide' or adsorbates, i.e. in ultra high vacuum conditions. The atomic force microscope gives spatial resolution of the order of the scanning cantilever tip radius (about 10 nm) when operated in contact or non-contact modes. When the surface of the Si is terminated by 'native oxide' the STM resolution is degraded from single-atom resolution to around 1 nm. In this case the local density of states variations are likely to be small, provided that the appropriate tunnelling conditions are used (see Komiyama *et al* 1993), so the STM is sensitive primarily to topographic variations. An example of a Si(100) surface examined by UHV STM (Crossley *et al* 1994b) following cleaning in HF and DI water rinsing is shown in figure 9; a 1 nm deep, 10 nm wide etch pit can be seen. Such a small feature would be difficult to detect by AFM. The relative ease of AFM operation in liquid, however, makes it an attractive tool for roughness measurement relative to the UHV STM. Our recent work (Crossley *et al* 1994c) using AFM to examine steps and plateau structures on epitaxial Si wafers under 2-butoxyethanol gives promise of useful AFM capability. The characteristic epi pattern of steps and terraces on the untreated wafer gives confidence in image quality. Examples of AFM images taken with a silicon nitride cantilever in contact with an epi-Si sample immersed in 2-butoxyethanol (contact force around 1 nN), before and after DHF and BHF cleaning and DI water rinsing, are shown in figure 10. The difference in images shows clearly that a small degree of etching has occurred during the DHF cleaning process and is worse for BHF cleaning as is expected (see spectroscopic methods below). It was found that STM images of the same surface add little if any topographic information (see figure 11). These images serve to give confidence that images of non-epi Si taken under the same conditions give reliable data, even though no discernible pattern is available to test reliability.

Studies of Si wafer roughness using a range of techniques to cover the wafer scale to the atomic scale have been carried out (Hahn *et al* 1988, Abe *et al* 1992). Hahn *et al* examined a range of roughness scales and a range of surface precleans. A strong correlation between roughness and preoxidation cleaning was found. They also determined roughness after oxidation under varying conditions (using low-energy electron diffraction) by examining the Si surface after etching off the oxide in HF. Their observations indicate that oxidation can roughen the interface and the effect is larger for thicker oxides. Abe *et al* used varying degrees of chemomechanical polishing to prepare wafers of four different surface roughnesses. They found that there was good agreement in average roughness measurements between LSLS, scanning optical microscopy (SOM), AFM and STM methods despite the very different roughness length scales sampled.

Infrared (IR) and high-resolution electron energy loss (HREELS) spectroscopic studies of the bonding

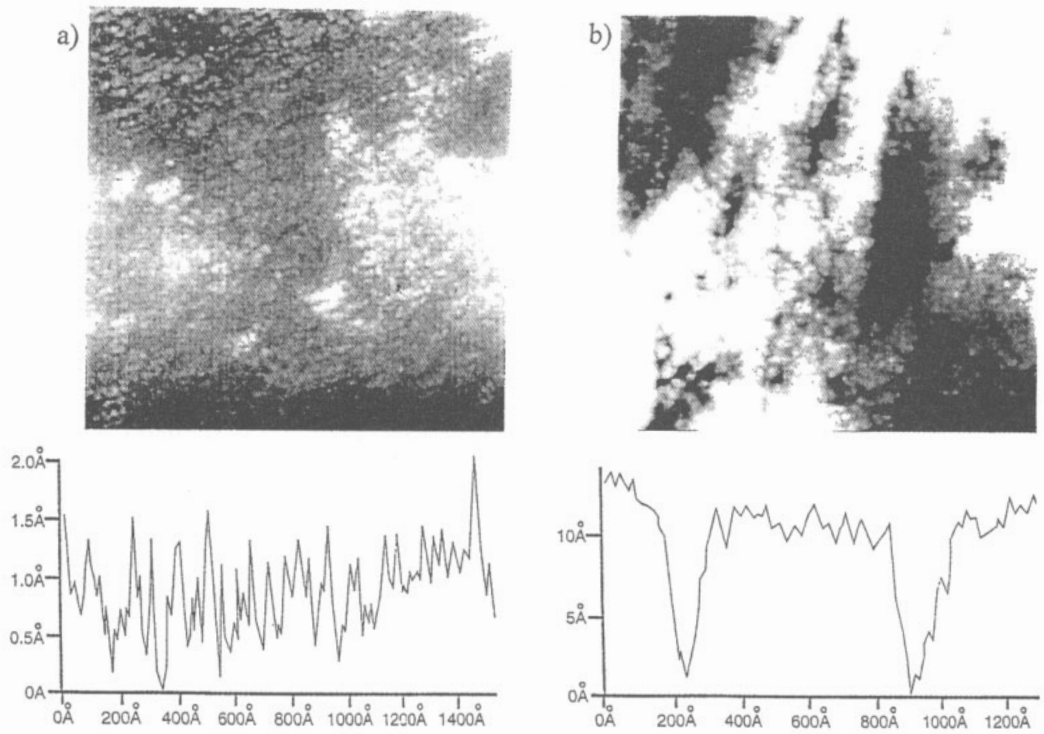

**Figure 9.** STM images, 0.1 $\mu$m by 0.1 $\mu$m, and line scans taken from Si(111) wafer cleaned in (*a*) dilute HF and 5 min DI water rinse and (*b*) 7:1 BHF and 5 min DI water rinse. The 'etching' during the latter clean is obvious.

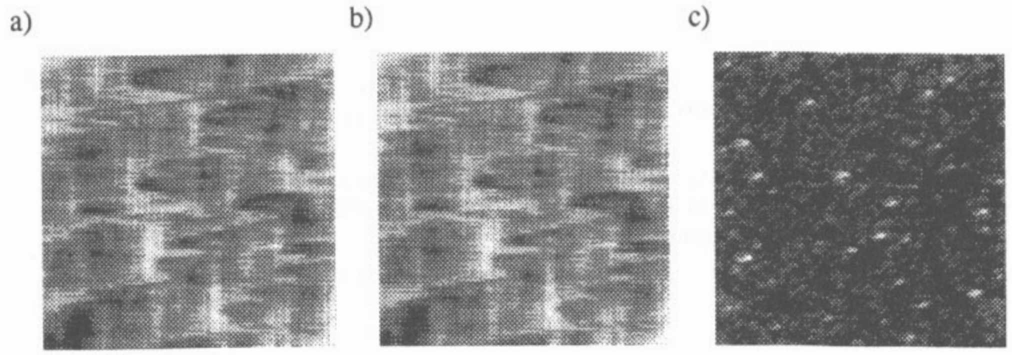

**Figure 10.** AFM images, 4 $\mu$m by 4 $\mu$m, of an epi-Si wafer surface taken at low contact force in liquid. The untreated surface (*a*), the dilute HF treated surface (*b*) and the BHF-treated surface (*c*). The characteristic atomic steps and terraces on this surface make the degree of etching during cleaning easy to interpret.

of, in particular, H to Si at the surface following HF last step chemical cleaning can also be related to surface roughness. This sensitivity to roughness, although indirect, is inferred from the relative degree of monohydride, dihydride and trihydride Si bonding observed. On a perfectly flat Si(100) surface with complete H termination only the dihydride bonding is expected, the mono and tri hydride bonds being associated with steps and Si(111) facets. These techniques, used in conjunction with roughness measurements described above, have been used to elucidate the state of the H-terminated Si surface following HF wet cleaning and water rinsing, and shown DHF rinsing roughens less than BHF rinsing. Examples of the application of these methods are given for example by Higashi *et al* (1990), Bender *et al* (1993) and Graf *et al* (1993). In a recent study of Si roughness following pre-gate oxide cleaning (Hirosi *et al* 1994) using IR and AFM, the AFM was shown to have insufficient spatial resolution to distinguish the different cleaning treatments while the IR on the Si showed clear differences in roughness. It was also shown that breakdown of oxide grown on the Si was worse for rougher surfaces.

**5.3.2. Roughness at the Si/SiO$_2$ interface.** In order to understand the relation between Si surface roughness and gate oxide integrity, and to unravel the features of

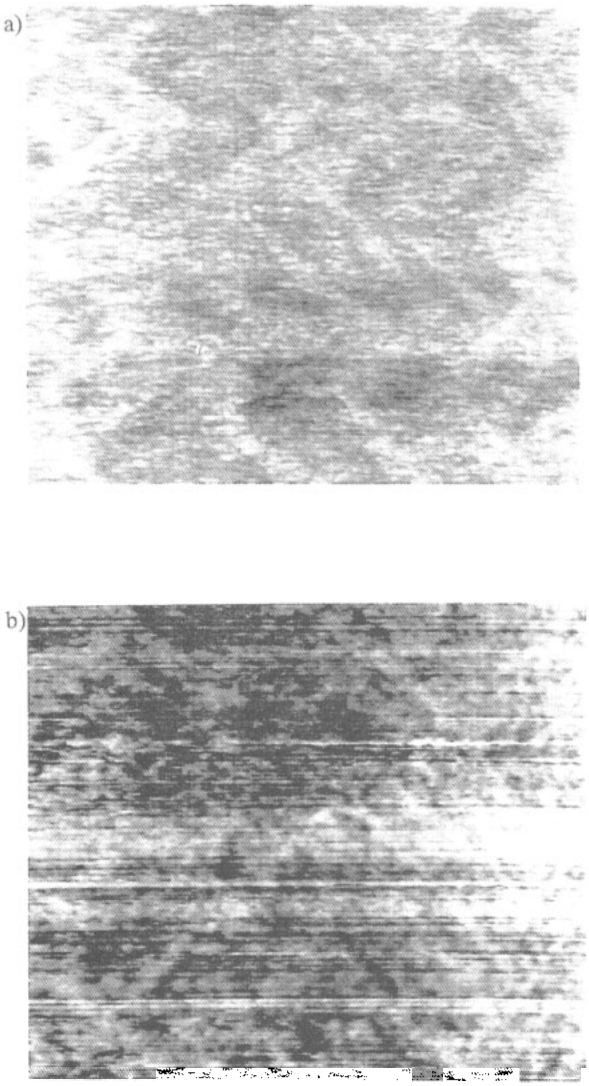

**Figure 11.** STM images, 0.6 $\mu$m by 0.6 $\mu$m, of the same material as in figure 10, (*a*) before cleaning and (*b*) after dilute HF cleaning. The image contrast at epi steps is not as good as in figure 10.

oxidation on rough surfaces, it would be useful to know whether or not the oxidation process changes the surface roughness at the growing oxide/Si interface and also how smooth is the oxide surface on which the electrical contact is deposited in an MOS structure.

High-resolution transmission and scanning transmission electron microscopy (HTEM and STEM) of thinned cross section samples is a powerful technique for observing both Si/SiO$_2$ interface roughness and variations in the oxide thickness. Sample preparation to produce samples with a thin enough cross section for electron transmission is a time-consuming, skilled process. There are also significant difficulties in aligning the electron beam perpendicular to the plane of the interface that may make the interfacial few-atom layer images difficult to interpret (Ross and Stobbs 1988), and in particular may be a confounding effect in the observation of crystalline quartz at the interface. The imaging and diffraction patterns obtained from Si(100)/SiO$_2$ interfaces which have Si(111) facets present has been shown (Akatsu *et al* 1993) to give data that could mimic crystalline quartz at the interface. In this context it is worth recalling that oxidation precleaning can give rise to roughening by producing such faceted etch pits. However, observations indicating the presence of a monolayer or two of crystalline SiO$_2$ at the interface of oxidized Si have been reported (Ourmazd and Bevk 1988).

A study of CZ wafers with intentionally introduced oxygen precipitates showed interface roughness using TEM and also a poor electrical breakdown performance of oxides grown on this material (Satoh *et al* 1994). This result agrees well with the light scattering study of Hahn *et al* mentioned above. Oxygen precipitates in CZ wafers thus appear as a special case of interface roughness that degrades breakdown performance. There are recent indications that high temperature H2 annealing can overcome this problem with CZ material and also give additional benefits to oxide performance by smoothing the interface (Bender *et al* 1994).

Goodnick *et al* (1983) observed the oxide–Si interface using TEM, and report very smooth interfaces, about 'one monolayer' in extent. However, from transport studies of MOSFETs fabricated on the same material they deduce a five-atom rough layer, a value in reasonable accord considering the limitations of their modelling. Hollinger *et al* (1988) also observed very flat interfaces using TEM.

Recent developments of STEM to include electron energy loss spectroscopy (EELS) with sample imaging can, in addition to interface roughness, give chemical bonding information. For example, Batson (1993) used a sub-0.25 nm electron beam, dark field techniques and EELS to examine the interface of a steam grown oxide on Si(100) material to elucidate the bonding states of Si across the Si/SiO$_2$ interface. This new technique offers the exciting possibility of examining both interface roughness and interface Si oxide bonding states as a function of oxidation conditions and Si oxidation precleaning methods. Such observations generally indicate a very abrupt smooth interface on there sampling length scale in contrast to the results of Hahn *et al* (see above) measured using light scattering.

Both neutron and x-ray reflectivity techniques are capable of probing the SiO$_2$ thickness and density and Si/SiO$_2$ interface roughness. Low-angle specular reflection is used to determine thickness, density profile and roughness, while in addition diffuse scattering is also used to obtain information on surface roughness. The methods are sophisticated and at least in the case of neutron diffraction, require major facilities and expensive equipment. The x-ray scattering methods are suitable for oxides as thin as 1 nm whereas the neutron scattering is probably limited to oxides thicker than 25 nm; these limits are set primarily by source intensity. Examples of the application of these techniques to Si/SiO$_2$ studies are given for example by Ashworth *et al* (1989) for neutron reflectivity and Cowley and Ryan (1987) for x-ray scattering. These methods are sensitive to the Si scattering centres and their arrangement in space, the electron density and thickness of the layer and to interface and surface roughness. The technique is not necessarily sensitive to the binding of the Si to oxygen or carbon. This can make data interpretation difficult. For example, a recent x-ray diffraction study (Brugemann *et al* 1990) of Si(100) wafers on which 20 nm to 30 nm of oxide had been grown gave evidence of the presence of crystalline $SiO_2$ at the interface. However, the wafers had been cleaned with HF solutions prior to oxidation and then heated to 1000 °C in Ar before introducing oxygen to the furnace. This preoxidation cleaned Si surface readily adsorbs hydrocarbon contaminants, and the subsequent heating in Ar may well lead to SiC precipitates being formed. It is possible therefore that the scattering data may not have distinguished crystalline $SiO_2$ from SiC. A recent study gave the surprising suggestion of the presence of substrate aligned microcrystals throughout the oxide (Takahashi *et al* 1993). Tang *et al* (1993), use high-intensity (hence good sensitivity) synchrotron x-rays to show that oxidation of Si(100) smooths the interface by a factor of two. A recent x-ray and neutron small-angle scattering study (Goff *et al* 1994), motivated by the wish to observe interface roughness and compare it with AFM results, showed the existence of a densified layer (10 to 20% increase), 0.5 nm to 1 nm thick at the $Si/SiO_2$ interface of a 30 nm thick oxide.

Nonlinear optical spectroscopy, in particular second harmonic generation (SHG) is able to probe the $Si/SiO_2$ interface structure and promising results on interface roughness, correlated with AFM measurements, have been obtained (Dadap *et al* 1994). In addition, this technique is potentially sensitive to strain and substoichiometry at the interface. A recent SHG study as a function of incident photon energy was used to deduce the presence of a monolayer or two of strained Si (in the sense of a vertical relaxation) on the Si side of the $Si/SiO_2$ interface (Daum *et al* 1993). This technique complements the channelling methods developed by Feldman and confirms the earlier channelling results (Feldman *et al* 1982).

Another method of assessing the $Si/SiO_2$ interface roughness is based on the STM/AFM profiling methods outlined above. The oxide is first removed by etching with HF solution and then a surface profile is measured. Great care has to be exercised in the control of the HF solution strength, purity and subsequent water rinsing to avoid confounding effects from Si surface etching as indicated above. A recent study (Ohmi and Aoyama 1992) has shown that the interface roughness measured by AFM and also measured by TEM were in good agreement at the level of 0.5 nm; however, from figure 11 the etching seems to perturb the surface at a level lower than this. The HF etch and AFM/STM measurements of roughness are relatively fast to make. Using this approach Nakamura *et al* (1993) observed that the $Si/SiO_2$ interface of oxides grown on Si(100) material became smoother as the oxide grew thicker while for Si(111) the interface became rougher. They also report that rougher interfaces degrade oxide quality. Homma *et al* (1992) found no evidence for oxidation-induced roughening of the $Si(111)/SiO_2$ interface. Nayar *et al* (1993), using both optical methods and AFM, find evidence for both oxidation-induced roughening and oxidation smoothing of the $Si/SiO_2$ interface depending on the length scale examined. Fukuda *et al* (1994) have used AFM to study the morphology of H-terminated Si(111) surfaces and the evolution of the morphology of the interface as it is oxidized. They find that the morphology of the original surface is preserved as oxidation proceeds. This result is entirely consistent with the immobility of Si atoms during oxidation as mentioned above (see section 4.1.4).

Rubloff *et al* (1990) have used low-energy positrons to probe the $Si/SiO_2$ interface. They varied the positron beam energy to obtain positron annihilation signals from varying depths in the $SiO_2$–Si structure. They found evidence of H at the interface and also evidence of microvoids about 1 nm in size. Other studies by Smith *et al* (1992) and Fujinami and Chiltern (1993) using these techniques to examine $Si/SiO_2$ interfaces have not observed such microvoids. This technique appears attractive for probing the interface defect structure and the presence of H.

In examining the data presented above it appears well established that the rougher the Si surface (for RMS roughness values above about 0.2 nm) prior to oxidation the worse is the oxide performance (breakdown and wear-out); the underlying reasons for this correlation, however, are still not well understood. For very thin oxides (< 6 nm) and roughness less than about 2 nm the effect of roughness appears to be negligible. However, only a few studies have been carried out on such thin oxides where direct tunnelling becomes important. A good understanding of the relation between current preoxidation cleaning procedures and Si surface roughness has been established but further improvements in reducing Si roughness appear to be warranted. The studies on the $Si/SiO_2$ interface roughness following oxidation still leave a confused picture; it not being clear under all conditions how the interface changes. A summary of the various causes of roughness is given in table 1. Further studies, preferably using more than one measurement technique, would thus appear warranted. It is clearly important to give adequate care to the details of the preoxide growth surface cleaning procedures in such studies.

The studies discussed above were all aimed primarily at measuring roughness at the $Si/SiO_2$ interface but have in some cases added to our knowledge of the structure at the interface. The observation of a densified layer at the interface of 30 nm thick oxides using x-ray scattering fits well with the substoichiometry observed in the early stage of oxidation. This is also consistent with the reactive layer model of non-Deal–Grove kinetics for thin film growth. The confirmation of the two monolayers of strained Si under the oxide also needs to be taken note of as it implies that a change in reaction rate may occur at this interface as a function of oxide thickness, at least until the densified layer grows. Again such a reaction rate dependence would be expected to contribute to the thin film growth kinetics. The existence of a

Table 1.

| Processing conditions | Roughness change | |
|---|---|---|
| | Si(100) | Si(111) |
| 1. Chemical cleaning | | |
| BHF | roughens | smooths |
| DHF | slight effect (< 0.1 nm) | smooths slowly |
| $H_2O_2$ (+metal contamination) | roughens | roughens |
| $H_2O$ (+dissolved $O_2$) | roughens | smooths |
| 2. Annealing (high temperature, low $O_2$, $H_2O$ partial pressure (see figure 1)† | | |
| 'Native' oxide | roughens | roughens |
| H-terminated (+hydrocarbons) | roughens (SiC) | roughens (SiC) |
| 3. Oxidation | | |
| Thin oxides | no effect | may roughen |
| Thick oxides | possibly smooths | ? |

† With temperature and oxidant pressure held in the regime between oxide growth and oxide decomposition severe roughening can occur.

crystalline quartz phase at the interface of thermally grown oxides may occur under special conditions but its existence after thermal oxidation of wet cleaned Si(100) is controversial.

### 5.4. Implications of roughness for oxidation

The issue of roughness brings in perhaps five major issues. First, is there a correlation between roughness and oxide structure (and hence other oxide properties) or stoichiometry? On the smallest scale, this is surely the case, and it is a virtue of the STM that it may quantify this link.

Secondly, is there a link between roughness and growth mechanism? Individual atomic steps are relatively inactive; oxidation is not dominated by terrace/ledge-kink mechanisms. But whilst steps do not dominate in the rate, the converse—for example, that island growth on the original bare Si surface might influence roughness—cannot be ruled out (Lutz *et al* 1989). It is also known that the oxidation rate on grossly non-planar surfaces (e.g. Okada and Iijima 1991) is not uniform; an effect believed to be due to strain in the growing oxide network.

Thirdly, does local roughness affect local defect densities? Might we expect reduction of the oxide (which is believed to be nucleated at some defect) to be associated with a roughness feature? Could there be altered charge states associated with the different facets at asperities? Will self-trapping occur preferentially close to Si protrusions, aided by the high Si dielectric constant (see Stoneham 1991)? And do degradation and breakdown initiate at asperities because of defects present there?

Fourthly, does roughness affect transport processes? This could be a structure effect (easy channels for oxygen motion as implied by the work of Carniato *et al* 1995) or more complex, for example, the roughness might affect interface injection processes for the thinnest oxides, where the Mott–Cabrera mechanism is likely to be important in the cases where the molecular oxygen ion precursor is incorporated into the surface.

Fifthly, are there large electrostatic effects? There can be little doubt that breakdown is provoked by field-concentrating asperities (e.g. Cohen *et al* 1994). But is more involved? 'Forming' does occur, as in the studies of many oxide thin films by Dearnaley *et al* (1970), Di Maria (1993) and Schuegraf and Hu (1994). Could fixed charge be concentrated around roughness features (e.g. could hydrogen move towards asperities) with indirect consequences? We realize that we are raising more questions than answers, but the questions reflect the probable importance of roughness and its control.

Again we see a need for atomic scale modelling and experiments on the oxidation of Si to gain a detailed knowledge of the consequences of Si surface roughness on variability of the thickness of thin gate oxides, the oxide network structure and the oxide defect structure.

## 6. Electrical breakdown

The oxide of Si grown by heating Si in oxygen (wet or dry, with or without additives, for example Cl) provides good gate dielectric with resistance to electrical breakdown and wear-out. However, the requirements for high yield and reliability as device dimensions are made smaller places stringent conditions on the acceptable quality. What are the limits to breakdown strength and wear-out?

A simple-minded view of breakdown would assume, by analogy with breakdown in gases, that the maximum field strength maintainable across the oxide is one for which electron injection does not lead to avalanche. Also by analogy to experimental observations of gas or vacuum breakdown the conditions at the cathode and anode surfaces would be expected to be important. It is also well known that more complex 'forming' processes can occur at sub-breakdown field strengths. Such forming appears to involve defect generation processes and can lead to long-term degradation. The theoretical difficulties of determining electron multiplication, energy dissipation and defect formation in $SiO_2$ are sufficiently difficult for there to be no reliable prediction of the upper limit to breakdown field strength.

### 6.1. Models of breakdown

Several models of breakdown have been noted (Hill and Dissado 1983, Rowland *et al* 1986). At present it seems likely that there is an initial degradation till a threshold, then a new conduction mechanism emerges which leads to breakdown; it seems to be a local phenomenon, and it may be associated with defect generation (Sune *et al* 1990). The low field breakdown may be the result of oxide decomposition at defects (Si + oxide going to volatile SiO) and can be suppressed by adequate oxygen. This process relates to reduction, and is the opposite to the oxidation behaviour of relevance here. Defect dynamics are thus a natural component of the complex range of phenomena which have come to be known as 'wear-out'. Whether or not there is the link one might envisage with the defect sites which nucleate oxide decomposition reactions (Rubloff *et al* 1987) is not clear. Verhaverbeke *et al* (1992) show some metals cause both breakdown and SiO nucleation but there are other defect sites for SiO nucleation as well.

Recent experimental studies and empirical models of oxide degradation by 'hot' electron injection (Schwerin and Heyns 1991, Di Maria 1993, Arnold *et al* 1994, Schuegraf and Hu 1994) give a good account of interface state and trapped charge build up as charge is injected into MOS capacitors, eventually leading to oxide degradation and breakdown. Breakdown is viewed as the end result of the degradation, and thus wear-out and breakdown are integral parts of the same process. In practical terms as high-field stressing (sufficient to drive Fowler–Nordheim tunnelling) leads to charge injection, intrinsic breakdown strength may not be easily observable. In the simpler case of electron injection into the Si substrate of the MOS device a degradation process is invoked for each of two different energy regimes of injected electrons. Charge traps are created by electrons arriving at the Si/$SiO_2$ interface with energy greater than 2 eV. The process is believed to be the breaking of H bonds to release this mobile species which then migrates to the cathode/$SiO_2$ interface to form interface states and oxide electron traps. The other process occurs for electrons with energies above 9 eV (so-called ballistic electrons) which produce holes in the oxide by impact ionization. The mobile holes are driven to the cathode by the applied field where they can become trapped or recombine with injected electrons to form interface states and traps. Eventually these traps and interface states build up to degrade the oxide and trigger catastrophic breakdown. A summary of the kinds of defects which arise from hot electrons is shown in figure 12. This view of the process leading to breakdown suggests important parameters to control are impurities that may play a part in forming traps, H being an important one (Marwick et al 1994). In addition the presence of metals and, in particular, Ca are known to degrade breakdown strength (Veraverbeke *et al* 1992). In the case of oxides sufficiently thin (about 4 nm) to give insufficient time to accelerate electrons to 2 eV with the applied field, wear-out should be reduced. This has been observed experimentally (Di Maria 1993, Depas *et al* 1994). In the latter study, although good capacitors (high charge-to-breakdown) were fabricated on 6.5 nm oxide, the yield was two orders of magnitude too low for the device dimensions requiring such thin oxide.

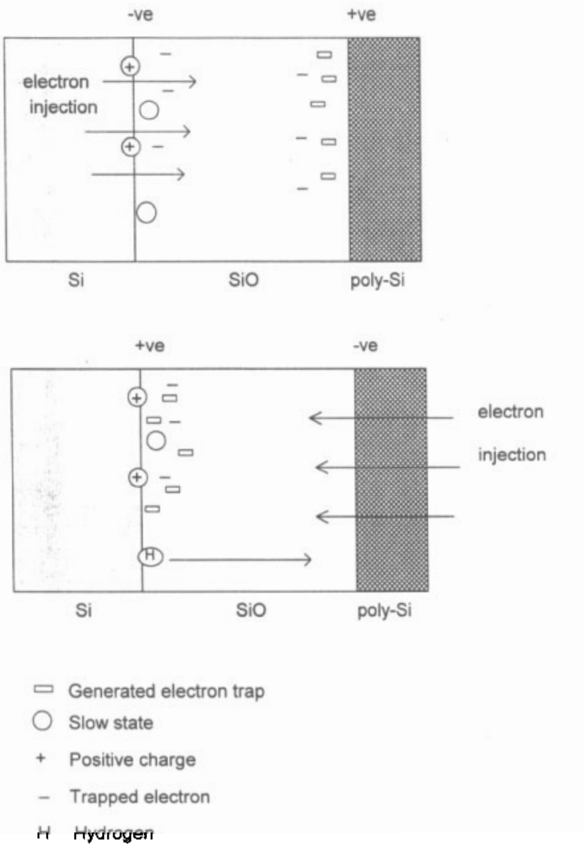

**Figure 12.** A schematic diagram of the hot electron induced defects arising from high electric field stress. The diagram follows those of Schwerin and Heyns (1991) and Di Maria (1993).

Strategies of oxidation have been devised to reduce the interfacial defects and H impurities that are invoked in hot electron degradation. Additions of F or Cl to the oxidant and oxidation with $N_2O$ to get N near the interface are believed to reduce number of defects and bound H atoms. They may also play a role in determining the network structure and perhaps counteract the effects of metal impurities. In any case they are known to 'harden' the oxide against wear-out. Whether or not such methods will be efficacious for very thin oxides where direct tunnelling is significant and hot electrons are suppressed is not yet determined. Schuegraf and Hu (1994) give a discussion of this topic in more detail.

From the above view of breakdown it would thus appear attractive to be able to measure both impurity distributions, roughness and breakdown on a small area scale, for example to search for correlations between the two. A recent AFM technique has been developed that holds some promise of achieving this in the future. Welland and Murrell (1993) describe threshold current AFM techniques to carry out wear-out and Fowler–Nordheim tunnelling measurements with spatial resolution approaching the radius (about 10 nm) of conducting AFM cantilever tips. For example, they examined 0.5 $\mu$m diameter capacitor test sets with 7 nm oxide thickness. The conducting AFM was used to map out the spatial position and topography of the test sets. At the same time they measured the voltage needed to drive a 20 pA tunnelling current through the capacitors. The current was limited to avoid destructive breakdown and a map of the voltage withstood (up to a limit of 40 V, i.e. in excess of 57 $\mathrm{MV\,cm^{-1}}$) was made. Figure 13 (top) shows such a map. During the measurement about 0.1 $\mathrm{C\,cm^{-2}}$ was injected into each capacitor. There is little variation in the high field the capacitors can withstand. Figure 13 (bottom) shows the same capacitors after three successive scans. It is now clear that current injection has led to wear-out of some capacitors (mainly in the top right-hand quadrant). From the discussions above we would expect some factor such a H distribution, surface roughness etc to vary on a similar scale. Such a correlation was not searched for in this study but could be informative in future studies. Further studies of this technique which are in progress indicate that AFM probe/surface interactions may give a contribution to the high breakdown fields observed. Further work, for example in UHV conditions, is required to ensure a reliable probe technique.

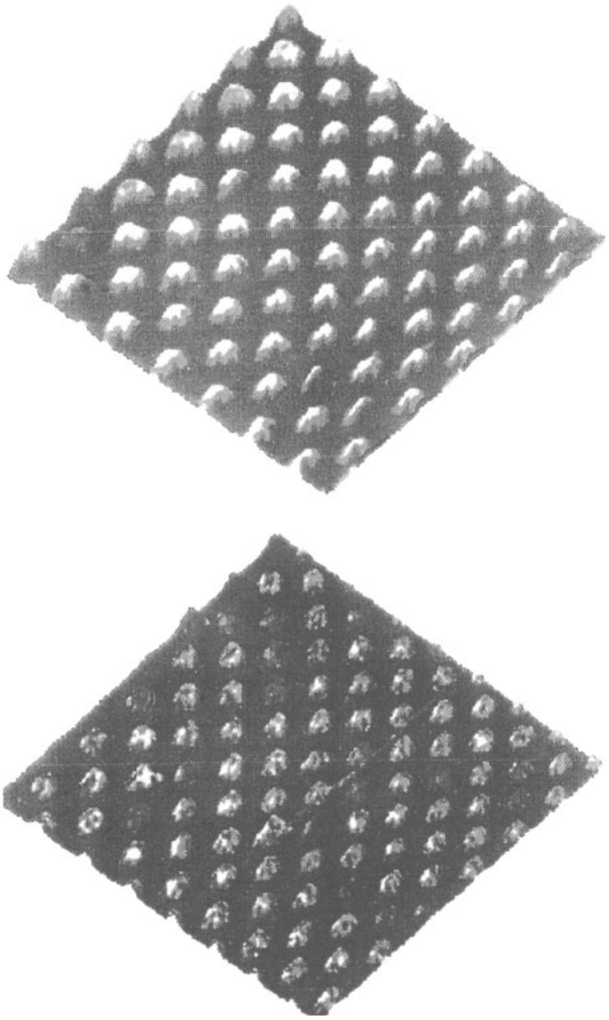

**Figure 13.** TCAFM observations of 0.5 $\mu$m diameter capacitor test sets. The top picture shows the first scan with voltage stressing; the dark to light colour scale represents 0 V to 30 V without drawing more than 20 pA. The bottom picture shows how wear-out due to current stressing has occured after several scans with voltage stress. Some capacitors no longer hold even 1 V.

Other scanning probe techniques under development are also expected to have an influence on the understanding of defect, impurity and roughness effects on gate oxide integrity. For example, ESR observations using scanning probes have been reported (McKinnon *et al* 1991, Manessen *et al* 1994) and offer the possibility of mapping for example $P_b$ centres. Another scanning probe development, the scanning Kelvin probe (Baikie *et al* 1994) may help in mapping impurity distributions and perhaps also identify species present.

## 7. Discussion

First we address the technological issues of gate oxide growth that have arisen in the previous sections. We then draw together the ideas and observations of the $Si/SiO_2$ structure to draw a picture of this system and outline the degree of uncertainty on the various aspects of the structure we present. Finally we comment on where we believe new theoretical tools could be applied to improve our understanding of the $Si/SiO_2$ system.

### 7.1. Growing a thin gate oxide: cluster tools

We have seen that it is still likely that gate oxide reliability is intimately related to degradation resulting from charge injection. It is also clear that oxide wear-out as a result of charge injection is influenced by defects and impurities. The oxide reliability also depends on the atomic-scale roughness of the Si on which oxide is grown; it seems likely this is simply a result of field enhancement at 'points' in some cases but a more complex relationship between roughness and growing oxide network structure, contamination, and defects is also likely. While breakdown is dominated by such extrinsic factors, the breakdown and wear-out are intimately connected, being facets of the same phenomenon. There is some evidence that thin oxides (about 4 nm) suffer reduced wear-out rates because electrons injected into the oxide under field stressing are too low in energy to bring about degradation. Note, however, that direct tunnelling through the oxide places a constraint on how thin a useful oxide can be. Even if sufficiently reliable oxides can be fabricated this has to be achieved with high yield. All of these aspects of oxidation are heavily dependent on the pre-gate oxide cleaning procedure and the control of the cleaned surface up to the time the oxidation is carried out.

How can one prevent changes to the H or 'chemical' oxide terminated Si surface after the pre-gate oxide wet clean? This can be tackled by transporting the cleaned wafers to the oxidation equipment in an ultra-clean environment, such as ultra-high vacuum or an ultra-pure chemically inert gas atmosphere (Ohmi 1990, Leir 1992, Sofield *et al* 1992). If these passivated layers cannot yield an effective control of gate oxide thickness variability or oxide structure or defect density their removal will pose even more stringent constraints on gas or vacuum transport systems because of the high reactivity of the Si surface to $O_2$ and/or $H_2O$.

The oxidation conditions for thin gate oxides, i.e. temperature, temperature uniformity, oxidant partial pressures and $H_2O$ contaminant levels, required to achieve acceptable process control variability place significant demands on oxidation furnace technology.

It is desirable to arrange these cleaning and oxidation process steps, as well as the gate contact formation process, in a cluster module with a central ultra-clean transport station. Such cluster tools have already shown promise in achieving better device performance for other process steps. However, the pre-gate oxide clean has so far proved difficult to integrate into a cluster tool arrangement, primarily because the best results (e.g. removal of particles and metal contamination, whilst achieving some degree of control of roughness) are achieved with wet cleaning procedures; the liquid ultra-clean gas or vacuum interface is clearly difficult to implement in a cluster tool. In addition wet cleaning leaves a passivated surface which is less reactive than clean Si. However, considerable efforts are being made to develop a pre-gate oxide dry cleaning process, partly for its compatibility with cluster processing. For example, studies have been reported of particle removal by frozen gas bombardment (e.g. Sherman *et al* 1994), organic contamination removal by UV/ozone (e.g. Lippert and Osten 1993) and 'native oxide' removal by H plasma cleaning (Ramm *et al* 1994). The removal of H or 'chemical oxide' without producing SiC from organic contaminants can be carried out by UV/ozone cleaning followed by UHV annealing; this process also can produce atomically flat surfaces, but at too high a temperature for normal furnaces. Paradoxically, for such flat surfaces, worse oxide integrity occcurs than if the 'chemical oxide' is left intact. It is still an open question whether or not pre-gate oxide dry cleaning methods as effective as wet cleaning methods can be achieved. Recent cluster tool methods, including dry cleaning have been used to fabricate III–V compound devices of very high performance (Harriot 1992). Part of this success arises from the absence of a highly stable 'native oxide' with III–V materials.

## 7.2. Oxidation and the Si/SiO$_2$ structure

The experimental evidence on the nature of the Si/SiO$_2$ structure has been added to by new STM and nonlinear optical studies. Refinements of electron microscopy, XPS and neutron and x-ray scattering techniques as well as further probing of $P_b$ centres have all contributed to a better understanding. The STM and electron diffraction experiments add considerable weight to the conclusion from isotope tracer work that Si is an immobile 'spectator' in the oxide growth process. Nevertheless the picture that emerges is still uncertain in many aspects, primarily because the most interesting features of the structure appear only a few atomic layers either side of the Si/SiO$_2$ interface and thus require very fine depth resolution to probe. Much of the information on this region also comes from probes of limited spatial resolution and hence there is still a dearth of information on any local structural variations that, for example, might arise from variations in surface 'roughness'. In figure 14 we summarize our view of this system as the oxide grows atomic layer-by-layer. First we start with a flat crystalline Si substrate, then an interfacial Si layer with vertically strained Si bonds about two monolayers thick develops as the first few monolayers of oxide grow. This strained Si is joined to a layer of 'densified' substoichiometric SiO$_2$, about 1 nm thick. Note that the term densified only refers to an excess of electron scattering centres and may be the consequence of substoichiometry alone, for example a relative excess of Si atoms or in addition a tighter oxide ring structure. This layer thus fits a description of a reactive interfacial layer containing Si–Si bonds invoked to account for departures from Deal–Grove growth kinetics for thin oxide films. This interfacial layer is a significant fraction (> 20%) of the thickness of oxides thin enough to suffer direct tunnelling conduction under electric field stressing. This layer is followed finally by stoichiometric SiO$_2$ with a lower level of defects. If the initial clean Si surface is reconstructed we see the reconstruction relaxes, primarily to the bulk structure, as oxidation proceeds over only a few monolayers.

We still have a poor understanding of the precise structure of the reactive layer and any variation in the structure as a function of distance from the interface. The number density of defects that appear to reside there is also uncertain. One aim must be to control the properties at this interface. Evidence suggests that high-temperature dry oxidation reduces the defect density ($P_b$ centres and substoichiometry) but at sufficiently high temperatures to present technical difficulties for very thin oxide growth.

The growth of thin oxides of adequate thickness uniformity and reproducibility has been achieved by conventional furnace oxidation using low partial oxygen pressure. Some evidence of improvements to the quality of thin oxides by low temperature wet growth (Ohmi 1994, Depas *et al* 1994) relative to dry oxidation have been reported. This appears to go against the trend of lower substoichiometry and defect density with temperature noted above.

Very little is known of the influence of Si surface roughness and state of surface termination on the growth rate and structure of thin oxides and on any consequences to the direct tunnelling barrier.

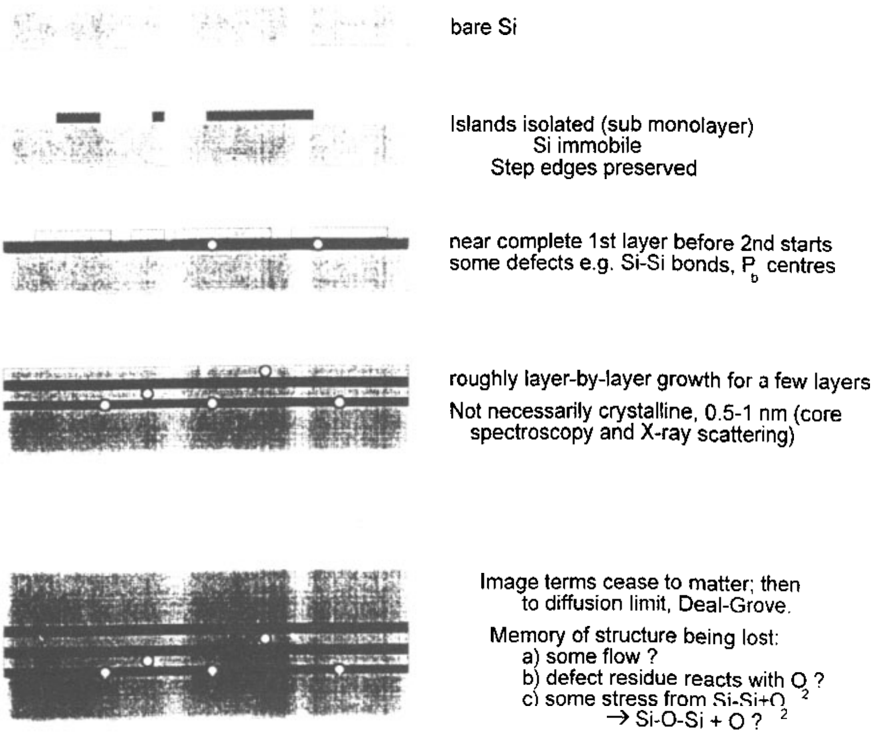

**Figure 14.** Schematic diagram of the Si/SiO$_2$ interface as it grows layer-by-layer from clean Si.

### 7.3. Theoretical models of the Si/SiO$_2$ system

It is helpful to identify those areas of theory which are well established and those where further development is essential. By well established, we mean that the theory is good enough for the questions which actually are asked to be answered. This does not mean they have all been answered to that level, but the key components are in place. Such areas include (i) the clean Si surface, (ii) the clean Si surface plus up to one monolayer of oxide, (iii) some (indeed most) of the steps in the interactions of gas-phase oxygen species with Si, (iv) diffusion of oxidizing species in the oxide, (v) kinetics, in the sense of simple models predicting some thickness $x(t)$ as a function of time (e.g. generalized Deal–Grove), (vi) some level of Monte Carlo modelling of the oxide network using interatomic potentials (e.g. see Carniato *et al* 1995), and usually requiring no network defects. There is therefore a large body of important theory established.

We may also identify another stage, when phenomenology is established. What this means is that a basic description can be given, but that important components will have to come from other observations (e.g. on bulk silicas), rather than predicted directly. Such areas include (i) the plastic response to stress, (ii) the effects of electric fields (Mott–Cabrera, image interactions), (iii) the analysis of electrical breakdown by energetic electrons (when flat surfaces assumed) and (iv) the evolution of roughness with oxidation.

Some of the other areas of theory are still at the frontiers: many of these are possible with present methods, but have not been attacked. These include (i) the processes involving excited states, such as the effects of optical and other excitation, (ii) reactions at the oxide–gas interface, including sticking and incorporation energetics, (iii) Monte Carlo modelling of the evolving oxide, to include defects (e.g. how the $P_b$ centre survives as the oxide grows) and hence how fixed charge develops etc, (iv) the electrical behaviour with realistic rough interfaces and (v) the proper description of the near interface region and its properties, including densification.

Several other problems are of increasing importance. One concerns the the role of N$_2$O, H and F in growth processes, and especially the broader understanding of reactions between interstitial species and the non-crystalline network. The other concerns tunnelling across the thinner oxides, which is one of the limits to how thin effective oxides can be. As regards tunnelling, specific issues include (i) the extent to which defects in the oxide can enhance the tunnel current and (ii) the extent to which changed defect species (including dipole layers from fixed charges) affect the potential barrier through which tunnelling occurs.

## 8. Conclusion

The breakdown strength of silicon dioxide films grown on silicon do not reach the levels set by avalanche breakdown (about 50 MV cm$^{-1}$) seen in bulk silica except for small areas as observed by threshold current atomic force microscopy. For oxides thicker than about 5 nm where electric field stressing leads to wear-out by hot electron induced defect generation the roughness of the Si surface prior to oxidation can degrade gate oxide integrity.

The picture is less clear for oxides thinner than 5 nm where direct tunnelling becomes significant. The wear-out under electric field stressing appears to be reduced as expected from the reduction in injected electron energy. Roughness of the Si/SiO$_2$ interface is expected to be important in this regime. The reactive layer, which explains departures from Deal–Grove kinetics for thin oxide growth, becomes a significant part of the oxide thickness. It will be interesting to see if this layer, where significant defects reside, will have to be engineered in a better way (as yet unspecified) to achieve control over the height of the direct tunnelling barrier. A change of dielectric can give a higher tunnelling barrier but silicon dioxide will not be abandoned lightly. This challenge will have to be met when growth rate and thickness uniformity requirements pose difficulties for oxidation furnace technology which may make high-temperature oxidation, for example, difficult to explore. The challenge to pre-gate oxide cleaning is also, in view of yield requirements, a very significant one.

## Acknowledgments

We are grateful for the support of ESPRIT Basic Research Actions 3109 and 6108 as well as AEA Technology Corporate Research programmes. Our colleagues A Crossley, P Hucknall and S Sugden gave valuable help. We wish to thank our ESPRIT Basic Research partners, M Welland, S O'Shea and J Barnes (Cambridge University, UK), M M Heyns, P Mertens, M Meuris, M Depas and S Verhaverbeke (IMEC, Leuven, Belgium), D Graf (Wacker Chemitronic, GMBH) and M P Murrell (East Coast Scientific UK) for helpful discussions, provision of data and teamwork which has added to our appreciation of the challenges of VLSI CMOS gate oxide technology.

We thank R E Harper for a critical reading of our manuscript and useful discussion on device fabrication issues. Thanks also to A H Harker, J Harding and T Hattori for discussion on the structure of the interfacial oxide layer from their theory and experiments prior to publication.

Our thanks are also given to F Ross and M Gibson for provision of their electron micrographs and helpful discussion.

## References

Abe T, Steigmeier E F, Hagleitner W and Pidduck A J 1992 *Japan. J. Appl. Phys.* **31** 721

Adams A C, Smith T E and Chang C C 1980 *J. Electrochem. Soc.* **127** 1787

Akatsu A, Sumi Y and Ohdomari I 1993 *The Physics and Chemistry of SiO$_2$ and the Si–SiO$_2$ Interface 2* ed C R Helms and B E Deal Plenum, New York

Anderson L and Kohler U 1993 *Surf. Sci.* 284, p77

Arnold D, Cartier E and D J DiMaria 1994 *Phys. Rev.* B **49** 10278

Ashworth C D, Messoloras S, Stewart R J and Penfold J 1989 *Semicond. Sci. Technol.* 4 1

Atkinson A 1985 *Rev. Mod. Phys.* 57 437

—— 1987 *Phil. Mag.* B **55** 641

Awazu K and Kawazoe H 1990 *J. Appl. Phys.* **68** 3584

Baikie I D, Bruggink G H and Rival S 1994 *Proc. 2nd Int. Symp. on Ultra Clean Processing of Silicon Surfaces* ed H Heyns *et al* (Leuven: Uitgeverij Acco)

Barrat C R 1993 *MRS Bull.* **18** 3

Battistiq G, Amsel G, Trimaille I, Ganem J J, Rigo S, Stedile F C, Baumvol I J R, Schulte W H and Becker H W 1994 *Nucl. Instrum. Methods, Phys. Res.* B **85** 326

Batson P E 1993 *Nature* **366** 727

Bender H, Li L, Mertens P, Caymax C and Heyns M M 1994 *Proc. 2nd Int. Symp. on Ultra Clean Processing of Silicon Surfaces* ed H Heyns *et al* (Leuven: Uitgeverij Acco) p 355

Bender H, Verhaverbeke S and Heyns M M 1993 *Proc. 3rd Int. Symp. Cleaning Technology In Semiconductor Device Manufacture, 184th Electrochem. Soc. Meeting, New Orleans*

Bevk J, Feldman L C, Pearsall T P, Schwarz G P and Ourmazd A 1990 *Mater. Sci. Eng.* B **6** 159

Bolland J J 1993 *Adv. Phys.* **42** 129

Borman V D, Gusev E P, Lebediaskii Yu Yu and Troyan V I 1991 *Phys. Rev. Lett.* **67** 2387

Boyd I W, Cracium V and Kazor A 1993 *Japan. J. Appl. Phys.* **32** 6141

Browning R, Sobolewski M A and Helms C R 1988 *The Physics and Chemistry of SiO$_2$ and the Si–SiO$_2$ Interface* ed C R Helms and B E Deal (New York: Plenum) p 243

Brugemann L, Bloch R, Press W and Gerlach P 1990 *J. Phys.: Condens. Matter* **2** 8869

Budde K J, Holzapfel W J and Beyer M M 1993 *Proc. 39th Annual Techn. Meeting IES, May, Las Vegas, NV* p366

Cahill D G and Avouris Ph 1992 *Appl. Phys. Lett.* **60** 326

Carniato S, Boureau G and Harding J, to be submitted to *Phys. Rev.* B

Che M and Tench A J 1982 *Adv. Catal.* **31** 77

—— 1983 *Adv. Catal.* **32** 1

Cobden D H, Uren M J and Kirton M J 1990 *Appl. Phys. Lett.* **56** 1245

Cohen S L, Syverson W, Basiliere S, Fleming J, Furman B, Gow C, Pope K, Tsai R and Liehr M 1994 *Proc. 2nd Int. Symp. on Ultra Clean Processing of Silicon Surfaces* ed H Heyns *et al* (Leuven: Uitgeverij Acco) p 35

Cowley R A and Ryan T W 1987 *J. Phys. D: Appl. Phys.* **20** 61

Crossley A, Myhra S and Sofield C J 1994a *Surf. Sci.* **318** 39

Crossley A, Sofield C J, Goff J, Lake A C I, Hutchings M T and Menelle A 1994b *J. Non-Cryst. Solids* to be published (Proc. E-MRS Spring Meeting, Strasbourg)

—— 1994c *Proc. 2nd Int. Symp. on Ultra Clean Processing of Silicon Surfaces* ed H Heyns *et al* (Leuven: Uitgeverij Acco) p 75

Dadap J I, Doris B, Deng Q, Downer M C, Lowell J K and Diebold A C 1994 *Appl. Phys. Lett.* **64** 2139

Daum W, H-Krause J, Reichel U and Ibach H 1993 *Phys. Rev. Lett.* **71** 1234

Deal B E and Grove A S 1965 *J. Appl. Phys.* **36** 3770

Dearnaley G Stoneham A M and Morgan D V 1970 *Rep. Prog. Phys.* **33** 1129

Depas M, Vermeire B. Mertens P and Heyns M M, private communication, to be published

Depas M, Vermeire B, Mertens P, Schaekers M, Meuris M and Heyns M M 1994 *Proc. 2nd Int. Symp. on Ultra Clean Processing of Silicon Surfaces* ed H Heyns *et al* (Leuven: Uitgeverij Acco) p 319

Di Maria D J 1993 *Extended Abstracts of Conf. on Solid State Devices and Materials, Makuhari, Japan*

Dong W P, Mainsah E, Sullivan P J and Stout K J 1994 *Three Dimensional Surface Topography: Measurement, Interpretation and Applications* ed K J Stout (London:

Penton)

Du D, Liu X, Korn G, Squire J and Mourou G 1994 *Appl. Phys. Lett.* **64** 3071

Elliot S R 1991 *Nature* **354** 445

Engstrom J R, Bonser O J and Engel T 1992 *Surf. Sci.* **268** 2382

Ermollett A, Martin F, Maouroux A, Marthon S and Westendorp J F M 1991 *Semicond. Sci. Technol.* **6** 98

Fair R B and Ruggles G A 1990 *Solid State Technol.* (May issue), 107

Fathy D, Holland O W and White C W 1987 *Appl. Phys. Lett.* **51** 1337

Feenstra R M, Woodall J M and Pettit G D 1994 *Mater. Sci. Forum* 143/147 1311

Feldman L C, Mayer J W and Picraux S T 1982 *Materials Analysis by Ion Channeling* (New York: Academic)

Feltz A, Memmert U and Behm R J 1992 *Chem. Phys. Lett.* **192** 271

—— 1993 *4th Int Conf on Formation of Semiconductor Surfaces, Jülich*

Fisher A J, Hayes W and Stoneham A M 1990 *Phys. Rev. Lett.* **64** 2667

—— 1990 *J. Phys.: Condens Matter* **2** 6707

Fujinami M and Chilton N B 1993 *Appl. Phys. Lett.* **62** 1131

Fukuda M, Bjorkman C H, Yamazaki T, Miyazaki S and Hirose M 1994 *Proc. 2nd Int. Symp. on Ultra Clean Processing of Silicon Surfaces* ed H Heyns *et al* (Leuven: Uitgeverij Acco) p 297

Fuoss P H, Norton W, Brenner S and Fischer A-Colbrie 1988 *Phys. Rev. Lett.* **60** 600

Gale R, Chew H, Feigl F J and Magee C W 1988 *The Physics and Chemistry of $SiO_2$ and the Si–$SiO_2$ Interface* ed C R Helms and B E Deal (New York: Plenum)

Gibson J M, Tong X, Twesten R D and Ross F M 1993 *Extended Abstracts Conf. Solid State Devices and Materials, Makuhari, Japan*

Goff J, Lake A C I, Hutchings M T, Menelle A, Crossley A and Sofield C J, to be published

Golen B M 1992 *PhD Thesis* University of Cambridge

Goodnick S M, Ferry D K, Wilmsen C W, Lilental Z, Fathy D and Krivanek O L 1985 *Phys. Rev.* B **32** 8171

Goodnick S M, Gamn R G, Sites J R, Ferry D K, Wilmsen C W, Fathy D and Krivanek O L 1983 *J. Vac. Sci. Technol.* B **1** 803

Graf D, Bauer S-Mayer, Ehlert A, Wagner P and Schnegg S 1993 *Mater. Res. Soc. Symp. Proc.* **315** 23

Griscom D 1989 *Phys. Rev.* B **40** 4224

Gupta P, Mak C H, Coon P A and George S M 1989 *Phys. Rev.* B **40** 7739

Hagon J P, Stoneham A M and Jaros M 1987 *Phil. Mag.* B **55** 241, 257

Hahn P O, Grundner M, Schnegg A and Jacob H 1988 *The Physics and Chemistry of $SiO_2$ and the Si–$SiO_2$ Interface* ed C R Helms and B E Deal (New York: Plenum) p 401

Harriot L R 1992 *Mater. Sci. Eng.* B **14** 336

Hasegawa S and Ino S 1993 *Int. J. Mod. Phys.* B **7** 3817

Hattori T 1990 *Solid State Technol.* July

Hattori T 1994 *Proc. 2nd Int. Symp. on Ultra Clean Processing of Silicon Surfaces* ed H Heyns *et al* (Leuven: Uitgeverij Acco) p 349

Hayashi S, Tanimoto S and Yamamoto K 1990 *J. Appl. Phys.* **68** 5360

Hayes W and Stoneham A M 1985 *Defects and Defect Processes in 'Non-Metallic Solids'* (New York: Wiley)

Heggie M and Jones R 1986 *Phil. Mag.* A **53** L65

—— 1987 *Phil. Mag. Lett.* **55** 47

Helms C R and Deal B E (ed) 1988 *Physics and Chemistry of Silicon Dioxide* (New York: Plenum)

Heyns M M, Hasenack C, De R Keersmaecker and Falster R 1990 *Semiconductor Cleaning Technology/1989* ed J Ruzyllo and R E Novak (Pennington, NJ: Electrochem. Soc.) PV 90–9 p 293

Heyns M M, Mertens P W, Meuris M, Verhaverbeke S, Schmidt H F, Rotondara A L and Graf D 1993 *Proc. Int. Symp on Semiconductor Manufacture, Austin, TX*

Heyns M M, Meuris M, Verhaverbeke S, Mertens P W, Philiposian A, Graf D and Schnegg A 1992 *Extended Abstracts, Int. Conf. on Solid State Devices and Materials, Tsukuba, Japan* p 187

Higashi G S and Chabal Y J 1993 *Handbook of Semiconductor Wafer Cleaning Technology* ed W Kerns (USA: Noyes)

Higashi G S, Chabal Y J, Truck G W and Raghavachari K, 1990 *Appl. Phys. Lett.* **56** 656

Hill R M and Dissado L 1983 *J. Phys. C: Solid State Phys.* **16** **2145** 4447

Hirose M, Hiroshima M, Yasaka T and Miyazaki S 1994 *J. Vac. Sci. Technol.* A **12** 1864

Hofmann K, Rubloff G W and Young D R 1987 *J. Appl. Phys.* **62** 925

Hollinger G, Saoudi R, Ferret P and Pitavol M 1988 *The Physics and Chemistry of $SiO_2$ and the Si–$SiO_2$ Interface* ed C R Helms and B E Deal (New York: Plenum)

Homma Y, Suzuki M and Yabumoto N 1992 *J. Vac. Sci. Technol.* A **10** 2055

Honda K, Takizawa R and Ohsawa A 1988 *J. Appl. Phys.* **63**

Irene E A and Lewis E A 1987 *Appl. Phys. Lett.* **51** 767

Jain S C, Winters K H and van Overstraten R 1990 *Adv. Electron. Electron Phys.* **78** 103

Johnson K E and Engel T 1992 *Phys. Rev. Lett.* **69** 339

Jones R, Oberg S, Heggie M I and Tole P 1992 *Phil. Mag. Lett.* **66** 61

Kanzig W 1955 *Phys. Rev.* **99** 1890

Kirton M J and Uren M J 1989 *Adv. Phys.* **38** 367

Komiyama M, Kirino M and Kurokawa H 1993 *Japan. J. Appl. Phys.* **32** 2934

Kucernak A R J, Peat R and Williams D E 1992 *J. Electrochem. Soc.* **139** 2337

Lander J J and Morrison J 1962 *J. Appl. Phys.* **33** 2089

Landsberger L M and Tiller W A 1987 *Appl. Phys. Lett.* **51** 1416

Leane R B 1990 *PhD Thesis* University of Cambridge

LeGoues F K, Rosenberg R and Meyerson B S 1989 *Appl. Phys. Lett.* **54** 644, 751

—— 1989 *J. Appl. Phys.* **65** 1724

Leihr M 1992 *Mater. Res. Soc. Symp. Proc.* ed Nemanich R J *et al* **259** 3

Lippert G and Osten H J 1993 *J. Crystal Growth* **127** 476

Liu H I, Maluf N I, Pease R F W, Biegelsen D K, Johnson N M and Ponce F A 1992 *J. Vac. Sci. Technol.* B **10** 2864

Lowery T, Cloud G, Kelly D, Zagar P and Seygedy M 1993 *Semicond Int.* p48, May

Lutz F, Kubler L, Bischoff J L and Bolmont D 1989 *Phys. Rev.* B 40, 11747

Manassen Y, Ter-Ovanesyan E and Shachel D, to be published

Marwick A, Buchanan D, D J DiMaria, Saunders P and Dori L 1994 *Proc. Symp. on Amorphous Insulating Thin Films* ed J Kanicki *et al* (Boston: Mater. Res. Soc.) vol 284 p 253

Massoud H Z, Plummer J D and Irene E A 1985 *J. Electrochem. Soc.* **132** 2685, 2693

McKinnon A W, Welland M E, Rayment T and Levitt M H 1991 *Proc. Int. Conf. on Scanning Tunnelling Microscopy, Interlaken, Switzerland, August*

Miotello A and Toigo F 1987 *Phil. Mag.* A **55** L53

Miyashita M, Tusga T, Makihara K and Ohmi T 1992 *J. Electrochem. Soc.* **139** 2133

Monkowski J R 1987 *Treatise on Clean Surface Technology* ed K L Mital (New York: Plenum)

Mott N F and Gurney R W 1940 *Electronic Processes in Ionic Crystals* (Oxford: Oxford University Press)

Mott N F, Rigo S, Rochet F and Stoneham A M 1989 *Phil. Mag.* B **60** 189

Mott N F and Stoneham A M 1977 *J. Phys. C: Solid State Phys.* **10** 3391

Murrell M P 1991 *DPhil Thesis* University of Oxford

Murrell M P, Sofield C and Sugden S 1991 *Phil. Mag.* B **63** 277

Murrell M P, Welland M E, O'Shea S T, Wong T M, Barnes J R, McKinnon A W, Heyns M M and Verhaverbeke S 1993 *Appl. Phys. Lett.* **62** 786

Nakamura K, Ohmi K, Yamamoto K, Makihara K and Ohmi T 1993 *Extended Abstracts, Int. Conf on Solid State Devices and Materials, Chiba, Japan*

Nakanishi T, Kishii S and Ohsawa A 1993 *Fujitsu Sci Tech. J.* **29** 154

Navrotsky A 1987 *Diffusion and Diffusion Data* **53/54** 61–6

Nayak D K, Kamjoo K, Park J S, Woo J C S and Wang K L 1990 *Appl. Phys. Lett.* **57** 369

Nayar V, Jackson R, Pidduck A J and Pickering C 1993 *The Physics and Chemistry of $SiO_2$ and the Si–$SiO_2$ Interface* vol 2 ed C R Helms and B E Deal (New York: Plenum)

Neuwald U, Hessel H E, Feltz A, Memmert U and Behm R J 1992 *Appl. Phys. Lett.* **60** 1307

Ohmi T 1990 *Summer Course on Manufacturing Science, IMEC, Leuven, Belgium*

—— 1994 *Extended Abstracts, Proc. 2nd Int. Symp. on Ultra Clean Processing of Silicon Surfaces, Brugge, Belgium*

Ohmi K and Aoyama S 1992 *Appl. Phys. Lett.* **61** 2479

Ohmi T, Kofani K, Teramoto A and Miyashita M 1991 *IElectron E E Device E Lett.* **12** 652

Ohmi T, Tsuga T, Takano J, Kogure M and Makihara K 1992 *ETrans I C E Electron.* E75-C(7), 800

Okada R and Iijima S 1991 *Appl. Phys. Lett.* **58** 1662

Ong C K, Harker A H and Stoneham A M 1993 *Interface Sci.* **1** 139

Ourmazd A and Bevk J 1988 *Mater. Res. Soc. Symp. Proc.* **105** 1

Ourmazd A, Taylor D W, Rentschler J A and Bevk J 1987 *Phys. Rev. Lett.* **59** 213

Pantelides S T and Long M 1978 *The Physics of $SiO_2$ and its Interfaces* ed S T Pantelides, (London: Pergamon) p 339

Prange A 1989 *Spectrochim. Acta* **44B** 437

Ramm J, Beck E, Dommann A, Eisele I and Kruger D 1994 *Thin Solid Films* **246** 158

Ramos M M D, Stoneham A M and Sutton A P 1993 *J. Phys.: Condens. Matter* **5** 2849

Rochet F, Rigo S, Froment M, d'Anterroches C, Maillot C, Roulet H and Dufour G 1986 *Adv. Phys.* **35** 237

Roos A, Bergkvist M and Ribbing C 1989 *Appl. Opt.* **28** 1360

Ross F M and Gibson J M 1992 *Phys. Rev. Lett.* **68** 1782

Ross F M, Gibson J M and Twesten R D 1994 *Surf. Sci.* **310** 243

Ross F M and Stobbs W M 1988 *Mater. Res. Soc. Symp. Proc.* **105** 1

Rowland S M, Hill R M and Dissado L 1986 *J. Phys. C: Solid State Phys.* **19** 6263

Rubloff G W, Hofmann K, Liehr M and Young D R 1987 *Phys. Rev. Lett.* **58** 2379

Rubloff G W, Nielsen B, Lynn K G, Welch D O and Leung T C 1990 *Vacuum* **41** 790

Ruzyllo J 1990 *Solid State Technol.* March

Satoh Y, Murakami Y, Furuya H and Shingyou T 1994 *Appl. Phys. Lett.* **64** 3

Schuegraf K F and Hu C 1994 *Semicond. Sci Technol.* **9** 989

Schuegraf K F, King C C and Hu C 1992 *Digest of Technol. Papers, 18th Symp. on VLSI Technology, Seattle, USA* June

von Schwerin A and Heyns M M 1991 *Insulating Films on Semiconductors* ed W Eccleston and M J Uren (Bristol: Adam Hilger)

Sherman R, Hirt D and Vane R 1994 *J. Vac. Sci. Technol.* A **12** 1876

Shluger A L and Stoneham A M 1993 *J. Phys.: Condens. Matter* **5** 3049

Singer P 1993 *Semicond. Int.* December, p 36

Smith D L, Smith C, Rice P-Evans, Evans J H and Evans H E 1992 *Mater. Sci. Forum* **105–110** 1451

Smith W L, Bechtel J H and Bloembergen N 1977 *Phys. Rev.* B **15** 4039

Sofield C J, Heyns M, Verhaverbeke S, Welland M E, Golan B and Barnes J R 1992 *Mater. Res. Soc. Symp. Proc.* **259** 3

Stesmans A 1993 *Phys. Rev.* B **48** 2418

Stesmans A and Van Gorp G 1990 *Appl. Phys. Lett.* **57** 2663

Stevels J M 1944 *Progress in the Theory of the Physical Properties of Glass* (Amsterdam: Elsevier)

Stoneham A M 1990 *J. Chem. Soc. Faraday Trans. II* **85** 505

—— 1991 *Insulating Films on Semiconductors 1991* ed W Eccleston and M J Uren (Bristol: Adam Hilger)

Stoneham A M, Grovenor C R M and Cerezo A 1987 *Phil. Mag.* B **55** 201

Stoneham A M and Tasker P W 1986 *Semicond. Sci. Technol.* **11** 93

—— 1987 *Phil. Mag.* B **55** 237

Sune J, Placencia I, Barniol N, Farres E, Martin F and Aymerich X 1990 *Thin Solid Films* **185** 347

Sze S M 1985 *Semiconductor Devices Physics and Technology* (New York: Wiley) p 219

—— 1988 *VLSI Technology* (New York: McGraw-Hill) p 510

Takahashi I, Shimura T and Harada J 1993 *J. Phys.: Condens. Matter* **5** 6525

Tang M T, Evans-Lutterodt K W, Higachi G S and Boone T 1993 *Appl. Phys. Lett.* **62** 3144

Tersoff J and Hamann D R 1983 *Phys. Rev. Lett.* **50** 1998

Tinivella G 1980 *Harwell Report AERE TP-877* citing e.g. Antonini A, Manara A and Lonsi P 1978 *Physics of $SiO_2$ and its Interfaces* ed S T Pantelides (New York: Academic)

Tinivella G and Stoneham A M 1980 *UKAEA Report M-3142*

Tromp R, Rubloff G W, Balk P, Legous F K and van Loenen E J 1985 *Phys. Rev. Lett.* **55** 2332

Tung R 1985 *Mater. Res. Soc. Symp. Proc.* **37** 349

Udagawa M, Niwa M and Sumita I 1993 *Extended Abstracts, Conf. Solid State Devices and Materials, Makuhari, Japan*

Verhaverbeke S 1993 *PhD Thesis* Katholieke Universitet and IMEC, Leuven, Belgium

Verhaverbeke S, Meuris M, Mertens P W, Kelleher A, Heyns M M, De Keersmaecker R F, Murrell M P and Sofield C J 1992 *Electrochem. Soc. Proc.* **92** 187

Vorburger T V, Marx E and Lettieri T R 1993 *Appl. Opt.* July

Welland M E and Murrell M P 1993 *Scanning* **15** 251

Wolters D R and Verwey J F, Chapter 6 1986 *Instabilities in Silicon Devices: Silicon Passivation and Related Instabiliteis* ed G Barbottin and A Vaiprille (Amsterdam: North-Holland)

Wolters D R and Zegers-van Duynhoven A T A 1989 *Philips Preprint*

Wong S S, Sodini C G, Ekstedt T W, Grinolds H R, Jackson K H and Evans S H 1983 *J. Electrochem. Soc.* **130** 1139

Wong T and Welland M 1995 *Surf. Sci.* to be published

Yakovlev V A, Liu Q and Irene E A 1992 *J. Vac. Sci. Technol.* A **10** 427

Yu M L and DeLouise L 1994 *Surf. Sci. Rep.* **19** (7/8)