IOPscience iopscience.iop.org

Home Search Collections Journals About Contact us My IOPscience

The oxide gate dielectric: do we know all we should?

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2005 J. Phys.: Condens. Matter 17 S2027

(http://iopscience.iop.org/0953-8984/17/21/001)

View the table of contents for this issue, or go to the journal homepage for more

Download details:
IP Address: 129.252.86.83
The article was downloaded on 28/05/2010 at 04:52

Please note that terms and conditions apply.

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **17** (2005) S2027–S2049 doi:10.1088/0953-8984/17/21/001

The oxide gate dielectric: do we know all we should?

A M Stoneham, J L Gavartin and A L Shluger

London Centre for Nanotechnology and Centre for Materials Research, Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK

Received 25 January 2005
Published 13 May 2005
Online at stacks.iop.org/JPhysCM/17/S2027

Abstract
Silicon's importance as a semiconductor owes much to its oxide. This oxide passivates, enables key process steps, and has been the gate dielectric of choice for MOSFETs for many years. Experience and know-how in using this oxide makes it hard for radical alternatives to be accepted. Yet it may not be possible for silicon dioxide to meet the stringent demands of the Semiconductor Industry Roadmap as a gate dielectric. This leads to clear technological questions. Is the oxide the best that can be grown? Could a better oxide be obtained by some modification of the growth process? What are the performance criteria that define the 'best' oxide? Are evolutionary changes, like incorporating nitrogen, to be preferred to introducing new oxides, such as those of Hf or Zr? Would there be long term problems with the new oxides? As so often in microelectronics, the technology demands new materials and new ideas from condensed matter physics. The critical role of the gate dielectric points to challenges for basic condensed matter theory, and this paper attempts to define these issues. It is certainly not sufficient to predict an equilibrium structure for oxide on silicon. The front and back regions of the oxide differ in measurable ways. Dynamical events like breakdown behaviour are certainly partly controlled by defects. Many experiments show the standard view of growth processes, 'bulk' diffusion and some interface reaction, is incomplete at best. How defects evolve as the oxide grows, and how impurities like H affect what happens could be crucial. Any analysis, even if only to create a framework of understanding, should address the various experiments exploiting different oxygen isotopes, the systematics of oxidation kinetics, and those electron microscope observations that show apparent layer-by-layer growth on terraces. The present paper aims to define an appropriate context for other detailed studies in the hope that progress in theory can contribute effectively to microelectronics futures.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

From its earliest days, the semiconductor industry has had a symbiotic relationship with solid state theory and modelling. In particular, Herman's early band structure calculations were possible because of the new computer power, and provided a pioneering example of the general-purpose computer codes we now consider the norm. In half a century, microelectronic structures have grown smaller, theoretical methods have matured, and computer power grown vastly. It is tempting to believe that significant parts of devices (such as a reasonable area of gate dielectric and its interfaces) might be modelled by state-of-the-art methods and so revolutionize the next generations of devices. This is optimistic, but not wholly outrageous, as later papers show. The optimism is perhaps misdirected, rather than misplaced. Broadly, there are three levels at which theory has impact on technology. The first is by providing a framework of understanding, so that observations—even when surprising—can be linked to basic, wide ranging ideas. The second is by scoping a system, predicting which phenomena are likely to be irrelevant, which essential, and which cannot be judged. Thirdly, there may be areas for which theory can substitute for experiments, notably when the timescales are exceptionally short or exceptionally long. For gate dielectric studies, a framework can be offered, and scoping is developing strongly.

There is little doubt that silicon will remain a dominating force in the future. If silicon is to achieve the performance demanded by users, new ideas will be necessary. The ideas will need to address heat dissipation, as well as processing and compatibility with other components in the device. One such idea may include quantum information processing (e.g., Stoneham *et al* 2003) complementing the classical approaches to silicon microelectronics. Each year, new materials are adopted, some with specialized uses; for them, one major issue is manufacturability. There may be new strategies, recognizing that device applications can have different needs (see, e.g., Houssa 2003, p 501). For portable equipment, the needs might be low operating power, for which the direct tunnel current is key, or low standby power, for which battery power loss in standby is critical. For high performance devices, like desktops, high speeds mean relatively high off-state currents; gate leakage should not exceed off-state current. There is a trade off between low switching times and low off-state leakage currents. Generally, it has been assumed in the past that there will always be some oxide dielectric that will fit the needs both of high performance and low power operation. Given the disparate needs, this assumption needs assessing.

The present paper examines some of the issues for gate dielectrics that have a condensed matter component. As such, it forms an introduction to other papers in this issue. However, it also addresses problems with which theory has still to get to grips. One such problem is structural. The importance of medium range order has been shown in previous studies (e.g., Szymanski *et al* 2001, Mukhopadhyay *et al* 2004).There is a need to represent this medium range order accurately for an oxide that is constrained at an interface, and whose growth is perhaps better described as frozen from steady state growth than an equilibrium state. Secondly, the behaviour of the oxidizing species near the constraining Si/oxide interface shows unusual features in isotope experiments. We discuss two models that have been proposed: a reactive layer model (Stoneham *et al* 1987) and a 'harpooning' model (Stoneham *et al* 2001), both of which may play a part. The mechanisms underpinning these ideas also affect oxidation kinetics. As many experiments have shown, it is very important not to accept standard reaction/diffusion models too readily for very thin oxide. Thirdly, we discuss some of the possible degradation processes. For diffusion processes, one must identify the best ways to characterize diffusion in an inhomogeneous amorphous oxide; one must also look at issues of electromigration in alternatives to silica, since the mobile species in such oxides are often charged. For electrical

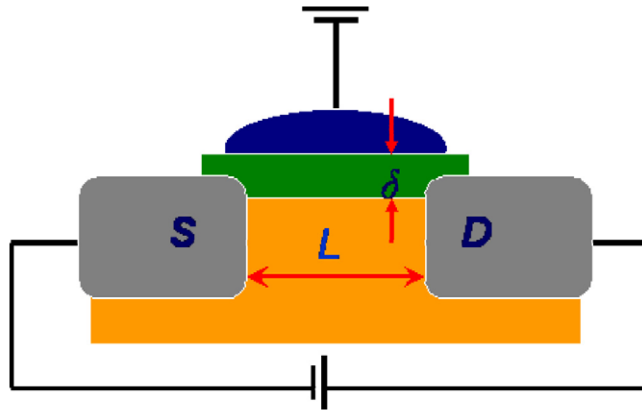

Figure 1. Schematic diagram of the MOSFET transistor with its critical parameters. L is the channel length (between the source (S) and the drain (D), δ is the thickness of the dielectric film). V_g and V_D are the gate and drain voltages respectively.

breakdown, it is clear that there are still some important ideas missing, and we give only an interim discussion.

2. What gate dielectrics need to do

Metal oxide semiconductor field-effect transistors (MOSFETs) are crucial to the success of current microelectronics. Figure 1 shows schematically a planar single-gate MOSFET, which is characterized by geometric and materials parameters. The channel is taken to have width W and length L, and the physical thickness of the gate dielectric or dielectric stack is δ. The carrier mobility in the channel is denoted μ and its dielectric constant k. When the channel is in the inverted state, the capacitance per unit area is C_{inv}, which is approximately equal to $\varepsilon_0 k/\delta$. Inversion occurs in the channel at a threshold voltage V_t. The current between source and drain in the inversion layer of an FET will depend on the gate voltage V_g, on the threshold voltage V_t, and on the drain voltage V_D. As V_D changes, this drain current saturates at a value I_{Dsat} when $V_D = V_g - V_i$. In the standard (gradual current) approximation, this saturation value is given by

$$I_{Dsat} = (W/L)\mu C_{inv}(V_g - V_i)^2,$$

which is an important figure of merit. When working with given materials, for instance Si and its oxide (so k and μ are essentially fixed), the obvious parameters to change are the channel length L and the dielectric thickness δ. There are limits to the extent these can be changed. Reducing the channel length L has to be achieved within the overall device design. Reducing δ leads to increased tunnelling and leakage currents, and also to increased propensity to breakdown. For traditional SiO_2, δ probably has to exceed 1.5 nm, which is within the range of today's technology. Moving to an alternative dielectric opens up new possibilities, the main temptation being an increase in C_{inv}. This increase is conveniently defined by an equivalent oxide thickness (EOT). If C_{inv} is $\varepsilon_0 k/\delta$, then we can write this as $\varepsilon_0 k/\delta = \varepsilon_0 k_{SiO_2}/(\text{EOT})$. Clearly, for a given EOT, a potential new dielectric with $k \sim 4k_{SiO_2}$ would be four times thicker than the SiO_2 system, and leakage currents could be much less, although these would depend on band offsets.

Change comes at a price, of course. Are the channel parameters like μ or V_t really independent of the dielectric and the Si/dielectric interface? Is the dielectric (whether silica or

a new material) really homogeneous throughout its thickness, even very close to interfaces? Can one create the dielectric in ways compatible with other aspects of silicon technology? It is encouraging that relevant binary oxides have been grown at relatively low temperatures by chemical vapour deposition or by atomic layer deposition, with EOT values of less than 1 nm. Unfortunately, channel mobilities are often unexpectedly low, and the threshold voltages V_t often high and showing thermal instability. More complex oxides can also be deposited, although issues of alloy disorder will become important.

We may group the properties necessary for a gate dielectric into three categories (Stoneham 2001a). Class I needs refer to basic properties *as a dielectric*. Class II needs ensure the dielectric can be processed without deterioration or interference with other materials. Class III needs relate to performance in a microelectronic device. There is some interdependence and scope for compromise, which offers opportunities for modelling.

Class I needs, the basic properties, provide the initial selection criteria. The dielectric must be an insulator, with an optimum dielectric constant in the range 15–25, although a wider range (say 10–30) might be acceptable. Ferroelectric materials are less desirable because of their non-linear electrical response. It should have a band gap above about 5 eV, with adequate (perhaps >1 eV) valence and conduction band offsets for use with silicon.

The polarization of the dielectric will have both electronic and ionic components. The electronic contribution by itself (related to the refractive index) can be high only if there is a low band gap, incompatible with the insulating and band offset properties needed. Thus, for any homogeneous material, it is essential that the main polarization be ionic. In principle, this could include the effects of reorientable fixed dipoles, even including H_2O. One consequence of a large ionic polarization is often that defect formation is easier (formation energies of charged defect are reduced by polarization), including possible non-stoichiometry, and this can cause other difficulties.

Many measurements of dielectric constants for oxides and related ceramics are of dubious value because of impurities or inhomogeneity, with conducting regions and non-stoichiometry. For a number of crystalline systems, it may be more accurate to calculate the dielectric constant than to measure it. For amorphous systems, there is a problem to ensure that the structure for which the dielectric constant is measured (or calculated) actually corresponds to that which will be used. It is not always clear that the dielectric will be *homogeneous*. The problems can be ones of stoichiometry or of alloy fluctuations. These are issues even for crystalline systems, e.g., in controlling Ba/Sr titanate.

Class II needs relate to processing and dielectric creation. It is taken for granted that any new material must be compatible with silicon technology to the greatest possible degree, since there is no doubt that know-how and past investment will ensure that silicon will remain the dominant semiconductor for many years.

The way the dielectric is made will, of course, depend on whether the microstructure is to be amorphous, polycrystalline, or epitaxial. It is not necessarily a problem that a dielectric oxide can exist in many closely related forms; indeed, SiO_2 is precisely such an oxide, and minor variations in structure rarely seem to have been an issue. However, it is very important that the dielectric has the intended composition. Generally, it should be non-reactive, reacting neither with Si (Hubbard and Schlom 1996) nor with the gate contact. Likewise, the interfaces should be largely stress-free. If not, the interfacial constraints will tend to force either defect creation (dislocations, possibly point defects like P_b centres) or conceivably new phases, such as amorphous forms of oxides usually crystalline (MgO, Al_2O_3, ...). Such new phases may have non-optimal properties. If the dielectric is to be amorphous, then it must *remain* amorphous at processing temperatures, and also during the operating life. This appears to be satisfactory for alumina, for Si-doped Zr aluminate, and possibly for other systems.

It is also important that homogeneity of composition and of properties must be retained during processing and in operation. Stoichiometry must be kept within precise limits. The precise definition of stoichiometry for thin layers is not always clear, and one might prefer to think in terms of a coordination criterion for a system like SiO_2. Phase separation could be a problem. Thus, Al_2O_3/Y_2O_3 alloys have two liquid phases of different density (Aasland and McMillan 1994), and it is well known that many glasses phase-separate. These demands on homogeneity must be seen in the context of an oxide whose thickness is to be accurate to a few nm over a whole wafer.

Class III needs relate to performance. For the dielectric to behave as a good insulator, there must be a low leakage current (Pr_2O_3 is said to be very good, Al_2O_3 relatively poor (Demmin 2001) and a low loss tangent. Leakage currents include contributions ranging from tunnelling to defect-related channels involving dislocations, grain boundaries, or point defects. The loss tangent describes how well the polarization follows an applied field. For DRAM, when fields change on nanosecond time scale, the loss tangent should be less than 0.005.

There should be a low effective fixed charge to cause carrier scatter in the Si; alumina is said to be relatively poor. Further issues relate to degradation and failure. For Ba/Sr titanate, there is 'resistive degradation', apparently due to O vacancy motion. Other effects seem to be related to electrode roughness, possibly analogous to those seen for silica and for other oxide films.

We discuss diffusion constraints in a later section, as these underlie some degradation mechanisms. The extent to which diffusion occurs is determined by temperature and time. For *processing conditions,* one might think of 15 s at 1050 °C; an *operating lifetime* might correspond to six years (2×10^8 s) at room temperature. For polar dielectrics, it is important that *ions* could provide the major means of atomic transport (unlike, say, O_2^0 in silica), especially in an applied field. There is a danger that the gate dielectric might electrolyse under the normal electric fields experienced in operation.

The processes underpinning electrical breakdown (Hori 1997, Stathis 1997a, 1997b, 2002) are still under debate. Whilst some changes in oxide properties can be reversible, presumably largely due to motion of trapped electrons (or possibly protons), most breakdown processes involve the creation or modification of ionic defects. Such ionic reorganizations typically require the localization of energy of the order of a few electronvolts. Such energy cannot be supplied by an electron accelerated over very short distances. For really thick oxides, there is a characteristic breakdown field; for a thin oxide, say 1 nm thick, even 10^9 V m^{-1} would only give an electron an energy of 1 eV. Different criteria emerge, notably that of charge to breakdown (Wolters 1981; see also Itoh and Stoneham 2001, p 238). There is little doubt that, whatever the process, rare vulnerable sites may cause problems. Thus, for crystalline dielectrics, the grain boundaries and dislocations normally present are likely to cause problems.

3. Homogeneity and inhomogeneity

3.1. Amorphous versus crystalline dielectrics

Three of the many basic fallacies concerning amorphous systems are these: (1) there is only a single amorphous structure for a given composition; (2) for that amorphous structure, the *mean* energies of defect formation or of trap ionization are sufficient to understand behaviour, and (3) crystals of the same composition have values for defect and trap energies which are essentially the same as those for the amorphous system. This third fallacy also applies to dielectric constants and band offsets, especially when the ionic polarization contributions are substantial. How a dielectric layer is created, manipulated or shaped can be important.

That these are fallacies is evident especially for SiO_2. Navrotsky's work (Navrotsky 1987, Itoh and Stoneham 2001, p 277) on the enthalpies and molar volumes for silicas shows

significant ranges of values. At the molar volume of α-quartz, amorphous systems have an energy per molecular unit larger by about 0.25 eV; the lowest energy amorphous 'phases' have about 30% larger molar volume than the quartz. There is similar information for other glass systems. At the very least, the density must be defined for an amorphous system, not just the composition. More generally, amorphous materials depend on the way they are created, and especially on any thermal treatment. The role of medium range order in glasses, as opposed to simple coordination pictures, has been evident for many years from the systematic trends of silica based glass densities as a function of composition. These are variously described; one simple approximate description (Stevels 1944) starts from an array of close-packed oxygens (see also Tinivella and Stoneham 1980).

3.2. Defect types

The structure of a-SiO_2 has been characterized in many studies, a common model being a continuous random network in which the crystalline connectivity is preserved and the geometric parameters are distributed in certain ranges (e.g., Wright 2000, Hobbs and Yuan 2000). Far less is known about the effects of disorder on defect properties and on relative abundances of different defect configurations. It has been widely assumed that defects in amorphous silica closely resemble their analogues in a corresponding crystalline structure, but with statistically distributed parameters manifested through inhomogeneous broadening of defect properties. If so, one way of identifying the role of disorder would be to compare the properties of corresponding defects in crystalline and amorphous phase. The situation proves less straightforward.

The neutral oxygen vacancy (NOV) in α-quartz and in a-SiO_2 is one of the simpler defects. It can be viewed as a relatively strong Si=Si bond, and is believed to exist both in α-quartz and in most types of amorphous silica (Skuja 2000). However, the NOV is diamagnetic and manifests itself mainly through the characteristic optical absorption peak and through possible transformations into other defects. The effects of disorder are not easy to extract from the shape of the absorption spectrum (Skuja 2000, Cannizzo and Leone 2004). Moreover, the NOV centres produced in different ways (during growth of thermal oxide, due to sample irradiation and via breaking Si–H bonds) may sample different sites in the amorphous network, and have different properties because of their formation steps. Another robust defect is the peroxy linkage, effectively an oxygen interstitial, as it incorporates atomic oxygen into crystalline and amorphous silicas, –Si–O–O–Si–. It is only weakly perturbed by the disorder of the amorphous network, as shown in theoretical studies (Szymanski *et al* 2001). However, its experimental properties are only starting to emerge (Kajihara *et al* 2004).

The families of paramagnetic E' centres in crystalline and amorphous SiO_2 provide qualitatively different sets of examples. An E' centre can be viewed as silicon dangling bond with an unpaired electron (Griscom 2000). This dangling bond can exist on its own as a neutral defect, as a part of positively charged oxygen vacancy, or as an even more complex defect. For example, trapping of hole by a neutral vacancy weakens or breaks the Si=Si bond, making these defects very susceptible to disorder-induced perturbations. It is not surprising that a far greater variety of E' centres has been found in amorphous silica than in quartz (Griscom 2000, Nicklaw *et al* 2002, Lu *et al* 2002, Stirling and Pasquarello 2002, Boero *et al* 2003, Mukhopadhyay *et al* 2004). E' centres are paramagnetic and some of them have assigned optical absorption bands (Griscom 2000, Skuja 2000). This makes them especially convenient for comparing the properties of defects in crystalline and amorphous environments.

Finally, some defects, such as self-trapped holes (Griscom 2000), are unique to amorphous silica and have not been found in α-quartz, although holes and electrons are trapped in α-quartz

by weak traps, such as Ge replacing Si. There have been suggestions that these defects can be formed only in some special precursor sites in a-SiO_2. Again, what constitutes a precursor site must depend on the method of defect creation, sample treatment and composition.

Thus structural disorder can lead to both a distribution of defect parameters and the creation of new defect types. Defect creation at specific sites can be driven by the release of local or medium range stress or strain in the surrounding network. In real samples, it may be hard to distinguish an integral part of a structure from a defect. In particular, many silica glass samples contain a fair amount of hydrogen terminating Si and O dangling bonds. In this case hydrogen is an integral part of the glass structure, and breaking Si–H bonds creates new defects. It is not obvious just what constitutes a defect model in amorphous material. In a crystal, a defect can be described with reference to a unique structure, and it will have a well-defined range of spectroscopic parameters, diffusion mechanisms and barriers and other properties. What would be an equivalent description in an amorphous structure where each site is unique? Can one justify 'average' or 'representative' defect models?

In an attempt to address some of these issues Szymanski *et al*, (2000a, 2000b, 2001), also Stoneham *et al*, (2000, 2001), Bongiorno and Pasquarello (2002) have used a combination of molecular dynamics and density functional methods to create realizations of amorphous SiO_2 with density similar to that of thermal oxide. They then analysed the energies of oxidizing species (O and O_2 in their neutral, negative and doubly negative charge states). They found that the structures of these defects in α-quartz and in amorphous silica would be qualitatively very similar and the dependence of formation energies, structural parameters, diffusion barriers and spectroscopic parameters of these defects on details of the local and medium range environment can be characterized theoretically by some distributions (see, for example, (Szymanski *et al* 2001, Bongiorno and Pasquarello 2002)). Although the parameters of these distributions are very difficult to verify experimentally, the qualitative results of these studies were illuminating in several respects. First, they demonstrated that the molecular species had a very different character from the atomic forms. Whereas O_2^0 was a relatively simple interstitial, O^0 formed a peroxy linkage (essentially, two oxygens in a bent structure between two silicons). This structural difference has the interesting consequence that the atomic form is very efficient at isotope exchange, whereas the molecular species only exchanges with network oxygens at special sites. Secondly, electrons with energies corresponding to the bottom of the Si conduction band could create ions, or induce reactions with ionic products. Thirdly, there was a substantial variation of insertion energies (and indeed also of diffusion barriers) from site to site in the amorphous system. This variation had a significant component associated with medium range order, not just the closest shells of ions. Fourthly, the values of key energies for the oxidation process (e.g., the insertion energy for moving an oxygen molecule from the gas phase into an interstitial site) were very different for the amorphous oxide and for quartz. The values for the amorphous oxide combined to give an activation energy for the oxidation process in good agreement with experiment.

These studies make two important points relating to modelling. The first is the need for prediction of structures (amorphous or nanocrystalline), including defects, possibly under constraints representing the interface with silicon. The second is the demonstration that earlier, simplistic ideas of key defect properties were misleading.

3.3. Further complexities

The fact that concentrations of most intrinsic defects in oxide films are on the level of 10 ppm makes it especially hard to study defects which exist only in amorphous materials. One of the problems is whether, under particular device conditions, a defect exists only at some special

precursor sites, or whether it could occur at any site in the amorphous structure, and how these sites depend on the previous treatment. To find relative defect concentrations, one needs to consider a statistically meaningful number of sites for different defects and for different types of the same defect. But in most cases one should be dealing with non-equilibrium statistics. The need to find special precursor sites in amorphous structures implies a need to consider a large number of different sites, and this makes modelling defects in amorphous materials fundamentally different from crystals, where it would suffice to make predictions for just one of the symmetry equivalent sites.

One natural way to begin has been to assume defect models developed in a crystalline phase transfer to the amorphous phase. In particular, the model of the E'_1 centre developed for α-quartz has been considered a natural prototype of the E'_γ centre in a-SiO_2 (Skuja 2000). Following this analogy, one searches the amorphous structure for local regions which can accommodate the most stable 'puckered' configuration of the E'_1 centre in quartz (Lu *et al* 2002, Mukhopadhyay *et al* 2004). This requires having a 'back oxygen' in an appropriate position to stabilize the puckering Si ion. To find all such sites in the amorphous structure one has to express this description in terms of structural parameters characterizing the location of the back oxygen with respect to the vacancy site. Two sets of such 'fingerprints' have been proposed and tested on the basis of calculated models of E' centres in a-SiO_2 and in α-quartz (Lu *et al* 2002, Mukhopadhyay *et al* 2004). From the analysis, it is clear that less than 10% of sites in the amorphous structure can accommodate puckered E' centres.

Can one predict the relative concentrations of defects? One crucial factor relates to the mechanism by which the defect is created, and will differ depending on the mechanism itself, e.g., thermal creation or creation via excitation. Another critical factor relates to the local structural characteristics of oxygen sites in amorphous material (bond lengths, ring size, dihedral angles) and how the defect-free energy varies from one site to another. This may look relatively straightforward using theoretical tools: one builds a model of an amorphous structure; one creates defects at selected sites; one calculates a distribution of defect parameters. This simple approach has proved very illuminating, and has been used to study peroxy linkages (Szymanski *et al* 2001), oxygen diffusion (Bongiorno and Pasquarello 2002), and recently to analyse different types of E' centres in silica (Lu *et al* 2002, Nicklaw *et al* 2002, Mukhopadhyay *et al* 2004). However, the meaning of statistics built on a few sites in a small periodic cell is unclear, and is by no means satisfactory on at least two accounts. First, the distributions depend on the way sites are *selected*, rather than on the defect creation mechanisms and sample history. Secondly, one needs to construct a *large* amorphous region, so that both short and medium range order regions around defects are consistent. This consistency is difficult to achieve in small periodic cells for those defects that induce a long range perturbation into the amorphous network. In particular, network distortion induced by NOV or E' centres in silica propagates up to 10 Å from the defect centre (Sulimov *et al* 2002, Mukhopadhyay *et al* 2004).

The need for consistent short and medium range order is likely to prove particularly important for predicting defect properties in thin oxide films. There both the structural constraints corresponding to interfaces and the extent of defect-induced network distortion will induce an additional parameter—the defect position with respect to the film interfaces.

3.4. More complex inhomogeneities

Thin films will usually have different properties at their inner and outer faces interfaces. For the oxide on silicon, there is a clear underlying reason. The oxidation reaction occurs at or near the Si (inner) interface, where there is a major mismatch, so there is a constraint near that growing interface. Theory is now developing detailed pictures of the structure (Bongiorno *et al*

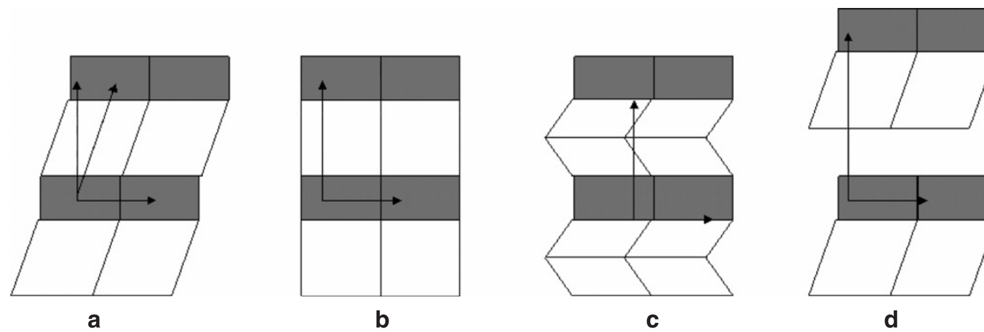

Figure 2. Illustration of the HfO_2/Si(001) (white/grey) modelled within periodic boundary conditions. (a) Upper and lower HfO_2/Si(001) interfaces are not equivalent, and thus, the model is incompatible with PBC—this incompatibility may be removed by various simplifications; (b) approximation of monoclinic HfO_2 by its cubic or tetragonal phases; (c) introduction of a twin boundary; (d) periodic slab model containing free surfaces passivated by hydrogen atoms. All (b), (c), or (d) approximations may introduce either artificial strain or surface/interface defect states which cannot be rigorously separated from effects of physical strain or defects.

2003, Bongiorno and Pasquarello 2003). The constraint at the Si/oxide interface is one reason for interfacial defects, like the P_b centre. The outer surface is largely stress-free, and novel structures may form at lower energy cost. The asymmetry shows in several ways, notably in breakdown behaviour. Some of the differences are associated with the nature of the electrode.

There is plenty of scope for complexity with other gate dielectric oxides on silicon. There will often be a thin layer of silicon oxide between the new oxide and the silicon substrate. Even when an oxide is deposited on a crystal of the same material, like $SiTiO_3$, there can be novel vacancy profiles (Muller *et al* 2004, Mannhart and Schlom 2004). In heterostructures, there is still more scope for complexity, and indeed polydomain structures have been reported, showing multiple relaxation mechanisms in $SrTiO_3$/$SrRuO_3$/$LaAlO_3$ (Ban *et al* 2004). As well as misfits, orthorhombic polydomain structures appear.

The modelling of gate dielectrics usually involves periodic boundary conditions. These boundary conditions must be treated with care, especially for disordered or partly ordered systems. For incommensurate systems, there can be strained structures with unrepresentative properties. As an example, consider monoclinic HfO_2 on Si(001), where the oxide a and b axes grow parallel to the interface, with the c axis at 9° to the interface normal.

In this case the symmetry of the interface is incommensurate with the periodic stack model. Figure 2 shows some of the ways this low symmetry system might be represented for reasonable size of cell. Important issues arise if the oxide is forced to be cubic (as by Fiorentini and Gullery 2002, Peacock and Roberston 2004) or tetragonal (as by Puthenkovilikam *et al* 2004, figure 2(b)), or if there is a twin boundary (Hobbs *et al* 2004), as in figure 2(c), or if there is a free surface and possible surface states (figure 2(d)). There is a danger that technical limitations may dominate the underlying science.

4. Challenges of growth

4.1. The phenomenology of oxidation

The rate-determining process in dry oxidation of silicon for thicker oxides (say above 10 nm) is widely accepted to be the interstitial diffusion of oxygen molecules. For thinner oxides (a few nm), this simple picture is inadequate: the Deal–Grove (reaction/diffusion) model of

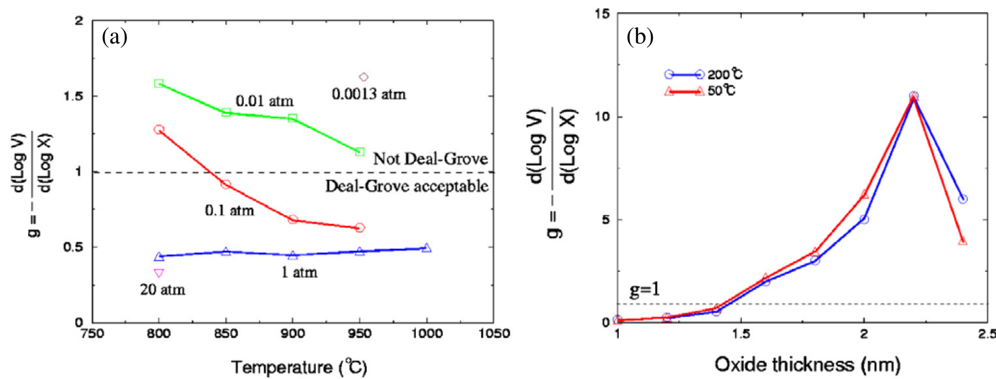

Figure 3. Values of $g = -\mathrm{d}(\log[\mathrm{d}x/\mathrm{d}t])/\mathrm{d}(\log[x])$ where x is the oxide thickness at time t, (a) for dry thermal oxidation (extended from Sofield and Stoneham (1995)) and (b) in the presence of very low energy electrons (data from Collot *et al* (1985), reanalysed following Sofield and Stoneham (1995)). The Deal–Grove model cannot be the right description whenever g exceeds unity.

oxidation kinetics fails (see, e.g., Stoneham 2001b). Oxide usually grows by an essentially layer-by-layer process, with growth at terraces, not steps, and perhaps oscillatory roughening. Isotope experiments show that interstitial and network oxygens exchange close to the Si/oxide interface and close to the oxide/gas interface.

Understanding the new regimes of silicon oxidation needs new descriptions. It is clear that diffusion is rate-determining for thicker oxides, and that the diffusion constant D is largely independent of oxide thickness. But experiments on thin oxides are not consistent with the long held Deal–Grove view of a (thickness-independent) interface process in series with diffusion, even with minor sophistications. This follows from oxidation kinetics (figure 3), although we stress that studies of thickness $x(t)$ as a function of time t are a very limited guide to mechanism. One simple way to demonstrate the failure of the Deal–Grove model is to estimate values of Stoneham and Tasker's (1987) dimensionless parameter g. If the oxide thickness is x, and the oxidation velocity is $\mathrm{d}x/\mathrm{d}t$, then there is a key dimensionless parameter $g = -\mathrm{d}[\log(\mathrm{d}x/\mathrm{d}t)]/\mathrm{d}[\log x]$. It is readily shown that $g = 0$ for reaction control and $g = 1$ for diffusion control. In any situation for which the Deal–Grove model applies, g lies between 0 and 1.

There is a wealth of information on oxidation kinetics (notably compiled by Wolters, private communication to A M Stoneham; see Sofield and Stoneham 1995) from which values of g can be estimated. Experiment (figure 3(a)) shows clearly that g is greater than 1 when the oxygen pressure is low or the temperature low. Thus, in precisely the conditions for growing *thin* films, the Deal–Grove model does not apply. The simple reaction/diffusion model only describes oxidation in a limited regime of temperature and oxygen pressure. The Deal–Grove model does not apply either under the conditions for almost all scanning probe microscope work, nor for many transmission electron microscope studies. Industrial growers of oxide do not use Deal–Grove kinetics, but adopt more empirical guidelines. Experiment shows (figure 3(b)) that g exceeds 1 by a large margin when there are low energy electrons (Collot *et al* 1985; see also Irene and Lewis (1987) for other consequences of electronic excitation). We shall note later the phenomenological model of Torres *et al* (1995) that offers one consistent interpretation of the kinetics, the layer-by-layer growth on terraces, and oscillatory roughening.

One other key experiment (Collot *et al* 1985) examined the effect of very low energy electrons (perhaps 5 eV) on oxidation. Not only do these electrons affect kinetics dramatically

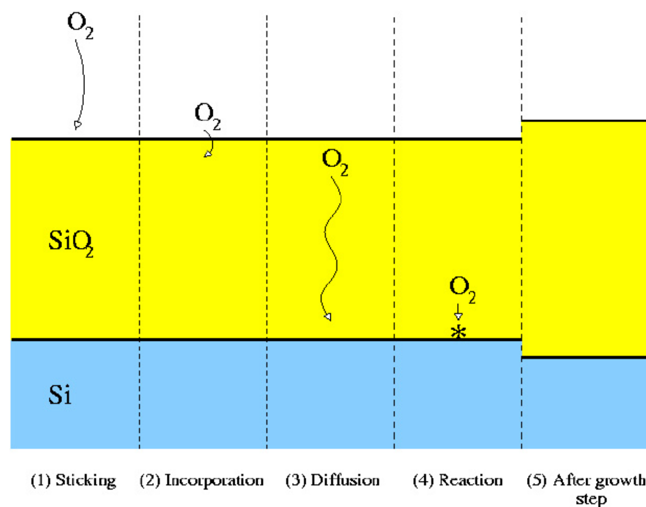

Figure 4. Oxidation processes. Four main processes are shown: (1) sticking; (2) incorporation, whether as an interstitial or as a component of the oxide network; (3) diffusion; and (4) reaction at or near the Si/oxide interface. Processes may be more complex than shown, e.g., the diffusion stage may involve change of charge state. The overall result of the oxide growth step (at the right) shows the Si/oxide interface has moved and the oxide has thickened. Some of these processes might be modified by capture of an electron from the silicon conduction band.

(with values of g up to 12) but also isotope exchange with the network occurs throughout the oxide, not merely close to the silicon/oxide interface. Since it is probable that the electrons cause dissociation of oxygen molecules, these results should be considered in the context of atomic oxygen and ozone experiments.

4.2. How do we link atomistic models to the phenomenologies?

Clearly we should consider oxidizing species other that interstitial oxygen molecules, including other possible charge states. The possible roles of other charge states should be no surprise. Species like O_2^- have been detected; telegraph noise requires change of charge state as electrons are transferred between Si and oxide. In scanning tunnelling microscopy, oxidation occurs differently if the STM tip is left in place while oxidation proceeds, rather than withdrawn and replaced only for measurement. Early studies noted effects of applied electric fields on oxidation. Breakdown processes include charge localization as well as energy localization; hole self-trapping occurs in amorphous silicas, although it is marginal in α-quartz. We remark that the low field breakdown of capacitors occurs (see figure 1 of Sofield and Stoneham (1995)) in a non-Deal–Grove regime.

If ultrathin oxides are not formed by a (thickness-independent) interface process in series with diffusion, what are the key processes? Can one understand them and control them so as to achieve better oxide (whichever oxide)? For thin oxides, one needs to understand how much is determined by chemistry, how much by electrostatics, and how much by polarizability. There are four basic stages in oxidation, each with characteristic critical processes. Surface chemistry controls the first stage (0–1 atomic layers). Rate-determining processes (figure 4) are sticking, chemical reactions and surface phase transitions or reconstructions. The ultrathin oxide (1–10 atomic layers) is controlled by sticking or incorporation steps; the Si/O reaction may be critical, but tunnelling and possible charge state changes could be critical. For thin oxide

(10–50 atomic layers), diffusion becomes increasingly important, although stress may lead to some oxide restructuring. For thicker oxides, diffusion is increasingly the rate-determining step.

Most workers assume implicitly that *structure determines process*. One common assumption is that the oxide close to the Si has a different structure because of the constraints like epitaxial misfit, so that diffusive motions and oxidation reactions are different near to the interface. In the so-called *reactive layer* model (Stoneham *et al* 1987), interstitial oxygen molecules can react directly with the oxide close to the Si/oxide interface. But there is an alternative (Torres *et al* 1995, Stoneham *et al* 2003, Stoneham and Shluger 2003) in which *process determines structure through new reactions*. Electron tunnelling from the Si to an oxidizing species occurs more easily near the Si, where also charged defects are stabilized by the image interaction with the polarizable Si. The idea is that, close to the Si/oxide interface, neutral O_{2i} capture electrons to form other oxygen species (a 'harpoon' process), and that it is these charged species which lead to differences in the oxide close to the interface. Thus the oxide network could be unreactive to oxygen molecules, yet react with the atomic oxygen species formed. Both the reactive layer and the harpoon pictures could explain isotope exchange near to the interface, and we stress that both mechanisms could occur together. We now examine the models in more detail.

The *reactive layer model* (Stoneham *et al* 1987) recognizes that the oxide near the Si/oxide interface is not an equilibrium structure, more one corresponding to steady state growth. Its structure will involve a compromise between a trend towards equilibrium, an adjustment in response to mismatch stress, and it will reflect the stochastic arrival of oxygen species. There will certainly be some Si–Si bonds on the oxide side of the average interface most to the time. Unlike bulk amorphous silica, the oxide in this region will be reactive, able to react with molecular oxygen. There are many experiments (e.g., core level spectroscopy, atom probe, and so on) that indicate an intermediate layer, and the central hypothesis of the reactive layer model is that O_2 can react near the outer side of this intermediate layer, leading to O^0 species either as peroxy or some other form.

The reactive layer model was devised to explain both the existence of the intermediate layer and why isotope exchange seems to take place a distance (0.5–1.5 nm) above the supposed interface, rather than indicating a reaction at the obvious interface. The precise structure of the reactive layer is left undefined in the model, to be determined by experiments or fuller theory. At present, theory gives many hints as to possible structures, but is not yet able to determine some key details, like the number of P_b centres expected, or to predict oxidation kinetics. P_b centres would be good reaction sites for O_2 (as would the Si–Si bond above the P_b centre in the (Ong *et al* 1993)) model that explains the absence of the lack of experimental observation of O hyperfine structure). Reaction with O_2^0 would leave a defect with an unpaired spin, and this would presumably exist in the bulk oxide once the interface has grown past the original P_b site. The reactive layer model does not address directly a number of issues, e.g., the effects of optical or low energy electron excitation, layer-by-layer growth, or roughness oscillations. Nor, without more extensive modelling, does it offer strategies to improve the gate dielectric quality.

The *harpoon* model (Stoneham *et al* 2001, Stoneham and Shluger 2003) evolved partly from detailed electronic structure calculations, and partly from Monte Carlo modelling of scenarios that might lead to layer-by-layer growth on terraces (Torres *et al* 1995). As the O_2^0 molecule nears the Si/oxide interface, it becomes energetically exothermic for an electron transfer from the Si conduction band to create O^- and O^0 species. These species will move differently in the oxide close to the Si/oxide interface and react differently, with predictable effects on oxidation kinetics. In the very early stages of oxidation, the image interaction will

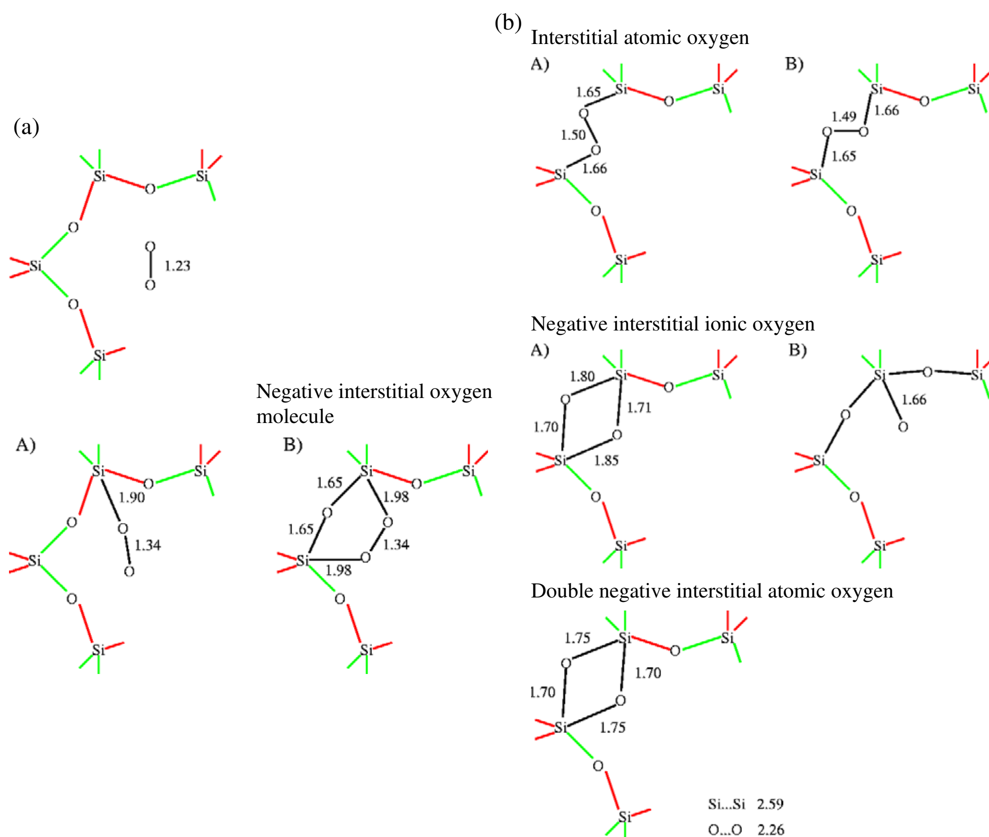

Figure 5. (a) Molecular oxidizing species. Note that the neutral molecular interstitial will not normally exchange isotopes with the network oxygens. (b) Atomic oxidizing species with structures that should lead to ready isotope exchange with the network oxygens.

influence the adsorption and incorporation steps (figure 4). The reaction can only occur within a tunnelling distance of the interface, of the order of a nanometre. Tunnelling events on this scale are shown in many experiments, including those relating to transmission (Wolters and Zegers-van 1996), breakdown, and the so-called random telegraph signal in MOSFETs, apparently due to defects near the interface and associated with a large entropy (Kirton and Uren 1989, Cobden *et al* 1990). Even a defect-free amorphous silica, unreactive to O_2^0, will form peroxy with atomic O^0 or O^- (see figure 5). The structures of these species and the range of tunnelling make it clear why isotope exchange should seem to take place a distance (0.5–1.5 nm) above the supposed interface, rather than appearing to react at the obvious interface. The harpoon model also has a natural parallel with the behaviour under excitation by 5 eV electrons, which lead to isotope exchange throughout the whole oxide volume. The observed role of optical excitation also has an explanation. With reasonable assumptions (Torres *et al* 1995), the harpoon ideas explain the observed kinetics for thin oxides (i.e., the observed values of the g parameter) and how the kinetics go over to parabolic diffusion-limited behaviour for thick oxides. Similarly, as Torres *et al* show, the model offers a possible explanation of layer-by-layer growth and roughness oscillations.

However, the harpoon description is not comprehensive. It is hard to see an explanation for the oxidation of p-Si, as opposed to n-Si, for instance. Yet the harpoon model is consistent

with the reactive layer model, so it may be one part of the complete picture. We remark on two consequences of the harpoon picture. First, the extra electron could pair with the P_b spin as that defect is eliminated (if indeed it is) during further growth of the oxide. Secondly, the model offers the possibility of modifying the component processes. As noted in SSS, applied electric fields for controlled times in each direction could encourage alternately electron transfer and O^- motion. It is conceivable that this might optimized so as to improve oxide quality, just as the oxide quality is improved, even for ultrathin oxides, by the use of a Kr/O_2 plasma to create oxygen radicals (Sekine *et al* 2001).

5. Diffusion and oxidation

5.1. Systematics of diffusion in amorphous oxides

Extensive experiments on gas migration in glasses (Shelby 1974; see also Masaryk and Fulrath 1973) identified systematic trends in diffusion for rare gases and for hydrogen. In particular, inert gas mobility and solubility proved to depend primarily on the accessible free volume of the glass structure. This simple concept rationalized dependence on composition (network former and modifier changes), thermal history and applied stress. A smaller role of phase separation was also identified. Such phase separation could be important in amorphous alloy oxides. Analogous interstitial ion diffusion and mobility processes are central to phenomena from ionic conduction to Mallory bonding (Itoh and Stoneham 2001, section 10.5.4). In some key cases of Mallory bonding, glass/metal adhesion is achieved by the removal of alkali from near the glass/metal interface under an electric field, leaving fixed charges that cause adhesion through the image interaction.

We examine here whether ideas similar to Shelby's apply to a-SiO_2, for which many different amorphous silicas observed, varying in density and formation enthalpy (Navrotsky 1987). Experimental data are sparse, mainly associated with molecular oxygen in amorphous gate dielectric oxides. However, theory (e.g., Szymanski *et al* 2001) does allow the most important features to be identified. We find that Shelby's rules need a simple but important modification, and discuss how this might be verified.

5.2. Predicting diffusion rates in amorphous oxides

In crystals, the small number of critical processes can usually be identified with certainty. In amorphous systems, there are several new issues. First, how is the amorphous system to be described? Is density enough, or do ring statistics matter? Secondly, if we predict a realization of an amorphous structure, how can we be sure that the structure is realistic? This is especially a problem in an inhomogeneous amorphous system, such as a gate oxide, where the standard structure-determining methods are harder to apply. We may also expect void and ring sizes to vary within the thickness of the oxide, i.e., with distance from the Si/oxide interface. Infrared and Raman spectroscopies claim to give ring size variations, but it is not clear what they indicate about medium range order. Thirdly, if we have a credible realization, how do is it generalized to give statistically acceptable results?

Clearly, many steps are needed to estimate the effective activation energy for diffusion in a thin oxide film. The technical issues are threefold: the creation of realizations of an amorphous silica; the calculation of key energies involving defects; and the analysis of percolative diffusion. Mott (see Mott *et al* 1989) suggested that the sites which determine solubility of interstitial species may be relatively rare, whereas percolative diffusion may have to sample sites which are energetically less favourable. A percolation approach requires

analysis of the topology of the realizations and the character of the important percolation paths. Medium range order cannot be ignored.

There have been two detailed studies of oxygen diffusion in amorphous oxides (Stoneham *et al* 2003, Bongiorno and Pasquarello 2002), both in substantial agreement. Both start with a realization of the amorphous system, and both run density functional calculations. Both fit pairwise interatomic potentials to DFT results, but do so in different ways. Pasquarello uses the pair potentials to calculate the stable sites in a bigger numbers of amorphous models, whereas Stoneham *et al* use them to only analyse the contributions from strain and interaction terms. Different approaches to the calculation of the transition states are adopted: Bongiorno and Pasquarello use their pair potentials to calculate all the transition states, whereas Stoneham *et al* calculate the transition states directly with DFT. There are further differences in modelling the diffusion process. Bongiorno and Pasquarello derive the energy distribution of stable sites and transition states, and runs Monte Carlo diffusion problem on a cubic lattice. This implicitly presumes statistical independence of energy distributions for stable sites and the transition, and so implies no medium range order. Stoneham *et al* look in detail at the transition states and the stable sites and, using theoretical results for percolative diffusion, analyse the topology of their realization. They conclude that there is percolation path through seven-membered or bigger rings and, from such a path, estimate the activation energy for diffusion. Both approaches indicate a percolative nature of diffusion and lead to rather similar energies.

There are certainly some extra insights from the Stoneham *et al*'s approach. The actual charge distributions for the transition states (Szymanski 2001) make it clear why there are differences in energy for different sizes and shapes of network rings. By using the pairwise potentials to analyse contributions to the energy, the roles of strain in and the topology of the SiO_2 network become clearer (locally soft versus hard embedding network). It also becomes more explicit that the diffusion will proceed mostly through the bigger voids and bigger rings, and hence indicates a role for medium range order.

5.3. Diffusion during processing and in operation

Is diffusion during processing significant? One should think of 15 s at 1050 °C. If the relevant diffusion distance is 2 nm, then this means that one would prefer the diffusion constant at 1050 °C to be less than 0.3×10^{-14} cm^2 s^{-1}. Data for many oxides are available (e.g., Stoneham and Smith 1991, Kingery *et al* 1976). Many of the oxides shown fail, sometimes by cation motion (even Mg in MgO and Ca in CaO; Y in Y_2O_3 probably fails), sometimes by anion (O) motion (in calcium-stabilized zirconia (CSZ), TiO_2 and Y_2O_3). Most non-stoichiometric oxides fail the criterion. Alumina seems safe so long as there are no grain boundaries, although Ag or Cu may diffuse fast enough to break the criterion (Atkinson 1993, 1985). It is possible that the criterion given is marginally too stringent, in that O in fused SiO_2 is close to failing.

Will processes controlled by thermal diffusion be sufficiently fast to cause significant degradation in operation? One might think of a six-year lifetime at 300 K. If diffusion is to occur over no more than 2 nm, the diffusion constant must be less than 2×10^{-22} cm^2 s^{-1}. For O diffusion, most oxides *except* clearly non-stoichiometric oxides are satisfactory. CSZ and Y_2O_3 fail; a-SiO_2 may fail marginally (corresponding to diffusion over about 50 nm in ten years), but the extrapolation to low temperatures is unreliable. The situation is less clear for diffusion of the ubiquitous impurity H. For some oxides ($LiNbO_3$, $LiTiO_3$, TiO_2 (both parallel and normal to the *c* axis)) the diffusion constant exceeds the critical value. For other oxides (MgO, MgO:Li, Al_2O_3, Al_2O_3:Mg, BeO) the diffusion rate is less than the critical value, and degradation of this sort is unlikely.

Estimates based on diffusion alone give an overoptimistic picture (Stoneham 2001a). Transport of charged species (say protons) will be enhanced during the device operation

under the gate bias. The important reliability issues are time-dependent dielectric breakdown (TDDP), and the less severe stress-induced leakage current (SILC) (Fleetwood 2002), and the negative bias temperature instability (NBTI) (Schroder and Babcock 2003). A critical defect for stress-induced leakage currents is credibly believed to be the H bridge (Bloechl and Stathis 1999), with H associated with an intrinsic defect. NBTI primarily affects p-channel MOSFETs. NBTI manifests itself in an increase of interface trap density and shifts of the threshold voltage associated with a build-up of positive interface charge, and it is observed in both SiO_2 and high k based devices (Houssa *et al* 2004, Zafar *et al* 2004). What causes this performance-degrading effect remains controversial. Widely accepted (yet still unproven) models for TDDP and NBTI involve the formation of Si dangling bonds by proton de-passivation of P_b centres at the Si/oxide interface. The hydrogen concentration in SiO_2 gates can reach 10^{19}–10^{20} cm^{-3}, a higher value being achieved for wet oxidation conditions (Revesz 1979). The hydrogen can have a highly non-uniform concentration profile, peaking strongly near the Si/SiO_2 interface (Krauser *et al* 1993) and to a lesser extent near the Si/HfO_2 interface (Choi *et al* 2003). In contrast, hydrogen profile is almost uniform through the bulk of $HfSiO_x$ and HfSiON (Pezzi *et al* 2004). Pezzi *et al* reported hydrogen (and deuterium) concentrations in nitrided and non-nitrided hafnium silicates that significantly exceeded those in silica or silicon oxynitride films; these concentrations are attributed to a different character and concentration of pre-existed intrinsic defects in the dielectric, a difference that requires further investigation. It is possible that the cation vacancy concentrations differ in the various films. Cation vacancies in minerals are known to adsorb up to four hydrogen atoms exothermically, forming hydrogarnet type defects (see e.g., de Leeuw 2001). The high formation energies of cation vacancies mean that their thermodynamic concentrations cannot be large, so the question remain to what extent these concentrations are controlled by the kinetics of the oxide growth. The mechanism of hydrogen migration is another unresolved issue. Statistical models of NBTI essentially assume dispersive migration (Houssa *et al* 2004, Zafar *et al* 2004 and references therein). This dispersive migration model, initially introduced for atomic hydrogen in SiO_2, does not necessarily apply to proton transport, which may differ significantly in SiO_2, HfO_2 and silicate films.

6. Interfacial barriers and interface quality

6.1. Band offsets and interface dipoles

An effective gate dielectric needs adequate band offsets from those of Si. However, the band offset is a subtle quantity (Herring and Nichols 1949, van Vechten 1985). The energy required to take a charge from one medium into another will depend on any dipole layer associated with the interface. Knowledge of the electron affinities for each material is not enough. Moreover, for MgO and perhaps Al_2O_3 it is not entirely clear what sign the electron affinity has. The results will depend on the material with which the oxide is actually in contact with. One can start from the various calculations of relative levels for infinite solids, or the several experimental measurements (Schmickler and Schultz 1986, p 371). A comparison of electron affinities is illuminating. Values from electrochemistry give very similar trends to those given by Robertson (2000): Ta_2O_5 3.7 eV (Robertson 3.3 eV), TiO_2 4.3 eV (3.9), ZrO_2 3.3 eV (2.5), HfO_2 2.9 eV (2.5), but agreement would be even better with a dipole correction of perhaps 0.4 eV. A very incomplete survey of offsets from various sources, including SiO_2, Al_2O_3 and ZrO_2, (Afanas'ev *et al* 2001, Olbrich *et al* 2001, Jungblut and Lewerenz 2000) suggests the method of preparation or subsequent processing can influence the offset, and so hints at some relatively simple means of control.

Bare silicon	*One layer oxidised; Pb centre*	*Two layers oxidised;*	*Three layers*	*Four layers, with defect*
	1111111111111	1111111111111 1111111111111	1111111111111 1111111111111 222222↑222222	1111111111111 1111111111111 222222↑222222 2222222222222
1111111111111 2222222222222 3333333333333 4444444444444 5555555555555	1111111111111 222222↑222222 3333333333333 4444444444444 5555555555555	222222↑222222 2222222222222 3333333333333 4444444444444 5555555555555	2222222222222 3333333333333 3333333333333 4444444444444 5555555555555	3333333333333 3333333333333 4444444444444 4444444444444 5555555555555

Figure 6. Schematic picture of the oxide growing past an interfacial defect, such as a P_b centre. The arrow identifies the defect with spin. The numbers identify the layer in which each Si atom starts in unoxidized silicon, with 1 the top layer, and so on. No reactive layer is shown in this figure.

Whenever one deals with highly *polar* materials, an *ionic* dipole layer affects offsets. The dipoles can be quite large. One of the very few predictions, for BaO/NiO (Stoneham and Tasker 1984) predicts a 2 eV potential barrier (corresponding to 4 eV for a 2+ ion) stabilizing positive charges in the NiO. This suggests that *interface engineering* (such as having at least one layer of ions of chosen electric charge) might resolve the offset problem, even if this is easier said than done. Sometimes even a small effect can be significant. Atoms placed on a surface which transfer some electronic charge to the substrate (like Cs, or like H on diamond) will decrease the electron affinity, even occasionally leading to negative electron affinities. Space charge, or modified near-surface probabilities of different charge states of defects or impurities can also contribute. There can also be an image charge effect, when the dielectric has a dielectric constant significantly different from that of Si.

One factor in the interfacial dipole is certainly interfacial stress. Most oxides have a relatively poor mismatch with silicon. Basically, there are five main ways to take up the strain. For very thin films, elastic strain may suffice, as in strain-layer systems. For thicker, crystalline films, misfit dislocations are expected. A third possibility is that there is a thin layer of a different phase. Point defects are another possibility. It is known that the P_b centre concentration correlates with interfacial stress (Stesmans 1993a, 1993b), and it is plausible that P_b centre creation does relieve the stress to some degree. Topology changes in an amorphous oxide are another way to reduce mismatch stress. One should bear in mind that the dielectric may be in a metastable form. The interface stress will certainly affect the dipole, especially for piezoelectric oxides.

6.2. Interface defects

The P_b centres and related $S = 1/2$ silicon dangling-bond defects form at significant densities at the Si/oxide interface, with a density that depends on the interfacial stress. It is an open question as to what happens to these defects as the oxide grows past them (figure 6). Since they have an unpaired spin, they cannot be eliminated by neutral molecular O_2 alone. There are several possibilities. Most of these possibilities leave a defect in the oxide, and so raise the important question as to whether this defect can contribute to the nucleation of electrical breakdown. No defects within the bulk oxide are seen in spin resonance. This could be because their orientations become random, so they are hard to observe, or because they are simply saturated

by hydrogen. Electron transfer is one conceivable possibility, whether from some other defect, from the substrate, or to mobile oxygens. Network reconstruction is another possibility, perhaps involving water, but any reconstruction hypothesis must address the question of what happens to the unpaired spin.

6.3. Aspects of non-stoichiometry

Oxides can be non-stoichiometric in many ways (Catlow and Stoneham 1981). Whereas SiO_2 usually has a composition with two oxygens per silicon (alternatively, O is bonded to two Si and each Si bonded to four O), other systems can deviate a lot from their nominal composition. Examples are systems for which cations can easily exist in more than one charge state (such as TiO_{2-x}) or where interstitial oxygen is readily formed (like ZrO_2). Why might non-stoichiometry matter? First, the defects which enable non-stoichiometry often have charge carriers associated with them. These can give rise to charge transport (in some cases by activated small polaron transport) or to dielectric loss. Conduction along grain boundaries or dislocations may be especially worrying (Duffy and Stoneham 1983). Non-stoichiometry may underlie the unusual magnetic results reported (Venkatesan *et al* 2004). Secondly, these defects are involved in degradation processes, such as resistance degradation. This degradation may become more important for very thin films, since the dielectric will have statistical variations in composition, and some regions will be more vulnerable than others. In certain cases, doping can help: for $SrTiO_3$, for instance, doping with Er apparently suppresses O vacancies and reduces the rate of resistance degradation (Kita *et al* 1999) Thirdly, there are likely to be sample-to-sample variations. These will arise both from the nature of the material as created and from changes during subsequent process steps.

7. Atomistic aspects of breakdown

Reliability in operation is a major issue (e.g., Hori 1997, Krautschneider *et al* 1997) and primarily involves defect processes. Are new defects created by electron–hole recombination? Gate oxides with acceptable band offsets are materials with large band gaps, so a large amount of energy is released on electron–hole recombination. Their high dielectric constants mean formation energies for defects, especially charged ones, will be relatively low. These are precisely the conditions expected to make damage likely. Defects are created in operation only by having localized energy. Indeed, since electron momenta are far too small to cause ionic restucturings, one can understand the process only by considering excited electronic states. Typically, energy of no more than the order of the band gap might be available for defect creation (Itoh and Stoneham 2001). This is often less than defect formation energies in perfect crystalline oxides. Vulnerable regions will be those, perhaps at grain boundaries or dislocations, where energy localization and modification are easy.

As most high k materials are ionic, defects and impurities are likely to be charged species. There are rare exceptions, e.g., O_2 interstitials in a-SiO_2, but ions are the norm. Even in simple oxides like MgO, transition metal ions can exist in three or perhaps four different charge states. In breakdown, is charge redistributed so as to give dielectric loss? Is the ubiquitous hydrogen a critical contributor to breakdown? Does damage build up in a systematic way, like the development of conducting channels? Is there diffusion which affects composition, or which allows electron traps to be exposed or eliminated? Will the breakdown be soft or hard? Why does the performance of thin oxide (<5 nm) depend on the sign of field? The growth models for the oxide on silicon give a natural asymmetry between oxide adjacent to Si and the oxide formed adjacent to the oxidizing medium.

Current views on breakdown raise many questions. Are failures defined by the voltage or by the strength and duration of the electric *field*? Why does breakdown depends on the *maximum* (not *average*) electron energy at the anode? These questions could be answered if the creation of key defects (those which initiate breakdown) requires a specific energy. However, it seems widely accepted that the electron energy is a controlling parameter (DiMaria 1997a, 1997b) and that the charge to breakdown is the critical parameter, since the numbers of generated defects and electron fluence are proportional. It is observed that, for many oxide thicknesses and for many voltages, the initial rate of defect generation varies inversely as the charge to breakdown (Q_{BD}) (Wolters 1981, Wu *et al* 2000, Stathis 2002, Itoh and Stoneham 2001). Breakdown is linked somehow to holes injected from the anode contact, with approximately a constant value of hole fluence to breakdown, independent of the field ($\sim$0.1 C cm^{-2} or about 600 e per molecular surface unit). It is unclear precisely what atomic processes are involved, or why holes are more effective at generating defects than electrons.

A further empirical observation comes from analysis of the statistics of breakdown, and empirical models in which random events occur in the oxide. The Weibull statistics of breakdown fit experiment well (and far better than log-normal statistics, e.g., Wu *et al* 2000, Degraeve *et al* 2000) with realizations from a model in which point defects build up into filaments across the oxide (see, e.g., Degraeve *et al* 1995, 1996, 1998, 2000, Stathis 1997a, 1997b, Bloechl and Stathis 1999). Snap-back behaviour, with well-defined conduction states in I–V characteristics after hard breakdown, also is consistent with filamentary models (Chen *et al* 2001). Such filamentary models also give a consistent picture of electrical properties of thicker dielectric films (Dearnaley *et al* 1970). For both the thicker and the thinner films, the models suggested involve a combination of an initial forming process, probably involving defects building up a conducting path, and subsequent, possibly associated with local heating, that disrupt this path.

What are the atomic processes by which defects are created when there is electronic excitation? A significant number of defect creation processes have been identified in detail for other systems, especially for halides. Energy localization in α-quartz, notably the self-trapped exciton, has been studied in depth (Itoh and Stoneham 2001), with extensive experimental validation of detailed calculations (Fisher *et al* 1990a, 1990b). In the self-trapped exciton, an oxygen moves very significantly away from its perfect-lattice site. There is a small but finite probability of defect production (oxygen vacancies and interstitials) and of the nucleation of amorphization (Itoh and Stoneham 2001, Itoh *et al* 2000). It is likely that excitation provokes similar processes in the amorphous gate oxide a-SiO_2, and that such processes initiate breakdown. What is important about this intrinsic process is that it can be efficient in the use of energy, even if its overall probability is low. For very thin oxides, a breakdown field is inappropriate: even 5×10^7 V cm^{-1} will give an electron only 10 eV on traversing a 2 nm film. Such an energy can create defects in silica, albeit with low probability. An elastic collision process would have essentially zero probability at such low energies. Other critical defects may be those formed from Si/SiO_2 interface defects (like P_b centres) after the oxide has grown past them. Elimination of such unusual sites during the oxide creation would require a deeper understanding of the oxidation process.

In filamentary models, it seems that degradation events (‘defects’) are biased somewhat to occur near pre-existing or initiating defects. This is what one might expect from macroscopic analogies (e.g., Dearnaley *et al* 1970). For instance, if the defects evolve as a conducting path, or filament, further defects might be formed in the high field region ahead of the filament tip. Formation of electron traps and conduction via these traps leads to soft breakdown (the capacitor is not destroyed thermally). Such a filament may be destroyed in a discharge event. Clearly, the noise spectrum contains information on the numbers of filaments and each spike is

a measure of charge transferred in a discharge event. Analysis of the noise spectrum suggests that each such discharge event itself could suffice to discharge the available stored charge. Analysis of several studies (Depas *et al* 1996) suggests that slow interface traps which lead to filamentary breakdown. There have been many suggestions that hole trapping which gave electron traps and which determined the breakdown behaviour. The defect processes leading to breakdown appear random across the area of the oxide, but many experiments suggest these processes vary across the thickness of the oxide. Thus, in experiments on c-Si/oxide/p-Si systems (Heyns and von Schwerin 1991) for an oxide grown in steam at lowish temperatures (650–700 °C), some defects were generated at the c-Si/oxide interface, irrespective of polarity: positive charges were created, and slow states were produced. These slow interface states are precursors in the breakdown of thin oxides. Other defect generation occurred at the anode (and at different interfaces on polarity reversal): electron traps were created, and electrons were indeed trapped.

A comprehensive study of the breakdown process thus contains three elements. First, there is an irreversible atomistic process ultimately due to localized energy release, and often described as 'bond breaking' or 'bond reorganization.' Secondly, this irreversible process occurs at statistically rare sites. Any realization of oxide structure must correctly model these rare features, which may involve medium range order. Impurities may be involved, hydrogen being a likely suspect. Thirdly, the evolution of the filamentary breakdown path should be predictable in any full analysis. The filamentary path should be understood well enough to give verifiable pictures of soft and hard breakdown, and also of the several conducting states described by several groups. At present, no such a comprehensive picture is available, although significant parts seem practical.

8. Could one design a super dielectric? Where has modelling had impact?

Perhaps the most important aspect of any oxidation model is the extent to which it enables the oxide to be improved. Improvement might be in the reduction of interface charge, or of unpaired spins, or control of interface dipoles, or generating structures less vulnerable to breakdown. In the case of the harpoon model, it is possible that beneficial changes might be achieved. The initial electron tunnelling events can be either encouraged or inhibited by an applied electric field during growth. Likewise, the subsequent motion of the ion created will be influenced by an applied field. As noted by Stoneham *et al* (2003), a sequence of applied fields of appropriate magnitudes and durations might be used to improve the oxide interface.

Short term improvements in gate dielectrics will surely come from experience and detailed optimization of relatively standard approaches. There will be enhancements from simple modifications, e.g., inclusion of N or F into silicon dioxide. There will be benefits from the scoping of many alternative oxides and the understanding of their special features. Clearly, predictive modelling will be important. Even if defect populations are far from equilibrium, it will help to know key energies of formation and motion, as well as likely charge states and the ranges of Fermi level over which they will be stable. It will be of value to know bulk band offsets and perhaps to make some estimate of interface dipoles for both the silicon/oxide interface and the oxide/polysilicon or oxide/metal interface. But, if theory is to guide radical improvements of any sort, there remain even more difficult challenges for it.

A first major challenge relates to structure. The real oxide is amorphous, formed by a growth process with a significant constraint at the silicon side. The structural model should include the medium range order of this *inhomogeneous* oxide, including variations from front to back of the oxide. To achieve this, certainly if the interface dipole is to be estimated with any precision, it may be necessary to mimic the growth process itself. In the case of oxide

alternatives to silica, the structure of the intermediate layer of oxide of silicon needs at least as much care. At some stage it will also be essential to look at large enough areas of oxide to model the creation and subsequent fates of P_b centres.

A second major challenge is the understanding of the degradation and breakdown processes. Many of the significant ingredients are known phenomenologically, and there are already a few interesting suggestions as to specific defects. What is not clear is the nature of the statistically rare features that aid the initiation of breakdown. Indeed, it is not even clear that models of amorphous structures actually contain these rare vulnerable features. This is an especially tricky problem, since some of these features may be metastable.

Finally, one should not overlook the difficulty in validating calculations. It is one thing to recognize that state-of-the-art codes give great opportunities; it is another thing to accept with confidence any results that require those codes to be used in regimes for which they are least accurate.

Acknowledgments

This work was supported by EPSRC and by Fujitsu. ALS and JLG acknowledge the support of the EU Framework 5 HIKE project and International Sematech.

References

Aasland S and McMillan P F 1994 *Nature* **369** 633
Afanas'ev V V, Houssa M, Stesmans A and Heyns M 2001 *Appl. Phys. Lett.* **78** 3073
Atkinson A 1985 *Rev. Mod. Phys.* **57** 437
Atkinson A 1993 *UKAEA Report AEAT-Intec* 1429
Ban Z-G, Alpay S P, He F, Wells B O and Xi X X 2004 *Appl. Phys. Lett.* **84** 4848
Bloechl P E and Stathis J 1999 *Phys. Rev. Lett.* **83** 372
Boero M, Oshiyama A and Silvestrelli P L 2003 *Phys. Rev. Lett.* **91** 206401
Bongiorno A and Pasquarello A 2002 *Phys. Rev. Lett.* **88** 125901
Bongiorno A and Pasquarello A 2003 *Appl. Phys. Lett.* **83** 1417
Bongiorno A, Pasquarello A, Hybertsen M S and Feldman L C 2003 *Phys. Rev. Lett.* **90** 186101
Cannizzo A and Leone M 2004 *Phil. Mag.* **84** 1651–7
Catlow C R A and Stoneham A M 1981 *J. Am. Ceram. Soc.* **64** 234
Chen T P, Tse M S, Sun C Q, Fung S and Lo K F 2001 *J. Phys. D: Appl. Phys.* **34** L95–8
Chen T P, Tse M S, Zeng X and Fung S 2001 *Semicond. Sci. Technol.* **16** 793
Choi R, Onishi K, Kang C S, Cho H-J, Kim Y H, Krishnan S, Akbar M S and Lee J C 2003 *IEEE Electron Device Lett.* **24** 144
Cobden D H, Uren M J and Kirton M J 1990 *Appl. Phys. Lett.* **56** 1245
Collot P, Gautherin G, Agius B, Rigo S and Rochet F 1985 *Phil. Mag.* B **52** 1059
Dearnaley G, Stoneham A M and Morgan D V 1970 *Rep. Prog. Phys.* **33** 1129–92
Degraeve R, Groeseneken G, Bellens R, Depas M and Maes H E 1995 *IEDM 95* (New York: IEEE) pp 863–5
Degraeve R, Groseneken G, Bellens R, Ogier J L, Depas M, Roussell P J and Maes H 1998 *IEEE Trans. Electron Devices* **45** 904
Degraeve R, Kaczer B and Groeseneken G 2000 *Semicond. Sci. Technol.* **15** 436
Degraeve R, Roussel P H, Groeseneken G and Maes H E 1996 *Microelectron. Reliab.* **36** 1939
de Leeuw N H 2001 *J. Phys. Chem.* **105** 9747
Demmin J C 2001 *Reviewing the 46th IEEE Int. Electron Devices Mtg*; *Solid State Technol.* (February) 46
Depas M, Vermeire B and Heyns M M 1996 Breakdown and defect generation in ultrathin gate oxide *J. Appl. Phys.* **80** 382
DiMaria D J 1997a *J. Appl. Phys.* **81** 3220
DiMaria D J 1997b *Solid State Electron.* **41** 957
Duffy D M and Stoneham A M 1983 *J. Phys. C: Solid State Phys.* **16** 4087
Fiorentini V and Gullery G 2002 *Phys. Rev. Lett.* **89** 266101
Fisher A J, Stoneham A M and Hayes W 1990a *Phys. Rev. Lett.* **64** 2667–70
Fisher A J, Stoneham A M and Hayes W 1990b *J. Phys.: Condens. Matter* **2** 6707–20
Fleetwood D M 2002 *Microelectron. Reliab.* **42** 523

Griscom D L 2000 *Defects in SiO_2 and Related Dielectrics: Science and Technology* (*NATO Science Series*) ed G Pacchioni, L Skuja and D L Griscom (Dordrecht: Kluwer) pp 117–59

Herring C and Nichols M H 1949 *Rev. Mod. Phys.* **21** 185

Heyns M M and von Schwerin A 1991 Charge trapping and degradation of thin dielectric layers *Insulating Films on Semiconductors* ed W Eccleston and M J Uren (Bristol: Institute of Physics Publishing)

Hobbs C C *et al* 2004 *IEEE Trans. Electron Devices* **51** 978

Hobbs L W and Yuan X 2000 *Defects in SiO_2 and Related Dielectrics: Science and Technology* (*NATO Science Series*) ed G Pacchioni, L Skuja and D L Griscom (Dordrecht: Kluwer) pp 38–71

Hori T 1997 *Gate Dielectrics and MOS ULSIs* (Berlin: Springer)

Houssa M (ed) 2003 *High k Gate Dielectrics* (Bristol: Institute of Physics Publishing) (ISBN 0 7503 0906 7)

Houssa M, De Gendt S, Autran J L, Groeseneken G and Heyns M M 2004 *Appl. Phys. Lett.* **51** 2101

Hubbard K J and Schlom D C 1996 *J. Mater. Res.* **11** 2757

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

Itoh N and Stoneham A M 2001 *Materials Modification by Electronic Excitation* (Cambridge: Cambridge University Press)

Itoh N, Stoneham A M and Tanimura K 2000 *Structure and Imperfections in Amorphous and Crystalline Silica* ed R Devine, J P Duraud and E Dooryhée (New York: Wiley) p 289

Jungblut H and Lewerenz H J 2000 *Appl. Surf. Sci.* **168** 194

Kajihara K, Skuja L, Hirano M and Hosono H 2004 *Phys. Rev. Lett.* **92** 015504

Kingery W D, Bowen H K and Uhlman D R 1976 *Introduction to Ceramics* (New York: Wiley)

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

Kita R, Matsu Y, Masuda Y and Yano S 1999 *Mod. Phys. Lett.* B **13** 983

Krauser J, Wulf F, Briere M A, Steiger J and Bräuning D 1993 *Microelectron. Eng.* **22** 65

Krautschneider W H, Kohlhase A and Terletzki H 1997 *Microelectron. Reliab.* **37** 19–37

Lu Z-Y, Nicklaw C J, Fleetwood D M, Schrimpf R D and Pantelides S T 2002 *Phys. Rev. Lett.* **89** 285505

Mannhart J and Schlom D G 2004 *Nature* **430** 620

Masaryk J S and Fulrath R M 1973 *J. Chem. Phys.* **59** 1198

Mott N F, Rigo S, Rochet F and Stoneham A M 1989 *Phil. Mag.* B **60** 189–212

Mukhopadhyay S, Sushko P V, Stoneham A M and Shluger A L 2004 *Phys. Rev.* B **70** 195203

Muller D A, Nakagawa N, Ahtomo A, Grazul J L and Hwang H Y 2004 *Nature* **430** 657

Navrotsky A 1987 *Defect. Diffus. Data* **53/54** 61

Nicklaw C J, Lu Z-Y, Fleetwood D M, Schrimpf R D and Pantelides S T 2002 *IEEE Trans. Nucl. Sci.* **49** 2667

Olbrich A, Ebersberger B, Boit C, Vancea J, Hoffmann H, Altmann H, Gieres G and Wecker J 2001 *Appl. Phys. Lett.* **78** 2934

Ong C K, Stoneham A M and Harker A H 1993 *Interface Sci.* **1** 139–46

Peacock P W and Roberston J 2004 *Phys. Rev. Lett.* **92** 057601

Pezzi R P, Miotti L, Bastos K P, Soares G V, Driemeier C, Baumvol I J R, Punchaipetch P, Pant G, Gnade B E, Wallace R M, Rotondaro A, Visokay J M, Chambers J J and Colombo L 2004 *Appl. Phys. Lett.* **85** 3540

Puthenkovilikam R, Carter E A and Chang J P 2004 *Phys. Rev.* B **69** 155329

Revesz A G 1979 *J. Electrochem. Soc.* **126** 122

Robertson J 2000 *J. Vac. Sci. Technol.* B **18** 1785

Schmickler W and Schultz J W 1986 *Mod. Aspects Electrochem.* **17** 357

Schroder D K and Babcock J A 2003 *J. Appl. Phys.* **94** 1

Sekine K, Saito Y, Hirayama M and Ohmi T 2001 *IEEE Trans. Electron Devices* **48** 1550

Shelby J E 1974 *Sandia Report* SLL-74-5210

Skuja L 2000 *Defects in SiO_2 and Related Dielectrics: Science and Technology* (*NATO Science Series*) ed G Pacchioni, L Skuja and D L Griscom (Dordrecht: Kluwer) pp 73–116

Sofield C J and Stoneham A M 1995 *Semicond. Sci. Technol.* **10** 215–40

Stathis J H 1997a *Microelectron. Eng.* **36** 325

Stathis J H 1997b *J. Appl. Phys.* **86** 5757

Stathis J H 2002 *IBM J. Res. Dev.* **46** 265

Stesmans A 1993a *Phys. Rev. Lett.* **70** 1723

Stesmans A 1993b *Phys. Rev.* B **48** 2410

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

Stirling A and Pasquarello A 2002 *Phys. Rev.* B **66** 245201

Stoneham A M 2001a Why model high-*k* dielectrics? *J. Non-Cryst. Solids* **303** 114

Stoneham A M 2001b Silicon oxidation kinetics *Encyclopaedia of Materials, Science and Technology* (*Semiconductor Processing and IC Fabrication*) (Oxford: Elsevier/Pergamon) pp 6579–86 (ISBN 0-08-0431526)

Stoneham A M, Fisher A J and Greenland P T 2003 *J. Phys.: Condens. Matter* **15** L447–451

Stoneham A M, Grovenor C R M and Cerezo A 1987 *Phil. Mag.* B **55** 201–10

Stoneham A M and Shluger A L 2003 Defects and defect-controlled behaviour in high-κ materials: a theoretical perspective *High k Gate Dielectrics* ed M Houssa (Bristol: Institute of Physics Publishing) pp 293–324 (ISBN 0 7503 0906 7)

Stoneham A M and Smith L W 1991 *J. Phys.: Condens. Matter* **3** 225 (figure 2)

Stoneham A M, Szymanski M A and Shluger A L 2000 Dynamics of silicon oxidation *Mater. Res. Soc. Symp. Proc.* **592** 3–14 (ed A Edwards, T Hattori, H von Bardeleben and D A Buchanan)

Stoneham A M, Szymanski M A and Shluger A L 2001 *Phys. Rev.* B **63** 241304 (Rapid communication)

Stoneham A M and Tasker P W 1984 *Mater. Res. Soc. Symp.* **40** 291–301

Stoneham A M and Tasker P W 1987 *Phil. Mag.* B **55** 237–52

Sulimov V B, Sushko P V, Edwards A H, Shluger A L and Stoneham A M 2002 *Phys. Rev.* B **66** 024108

Szymanski M 2001 *Thesis* University College London, unpublished

Szymanski M, Stoneham A M and Shluger A L 2000a *Microelectron. Reliab.* **40** 567

Szymanski M, Stoneham A M and Shluger A L 2000b *Solid State Electron.* **45** 1233–40

Szymanski M, Stoneham A M and Shluger A L 2001 *Phys. Rev.* B **63** 224207

Tinivella G and Stoneham A M 1980 The compaction of waste disposal glasses and its dependence on composition *AERE Report* M3142, UKAEA Harwell

Torres V J B, Stoneham A M, Sofield C J, Harker A H and Clement C F 1995 *Interface Sci.* **3** 131–44

van Vechten J H 1985 *J. Vac. Sci. Technol.* B **3** 1240

Venkatesan M, Fitzgerald C B and Coey J M D 2004 *Nature* **430** 630

See also Wolters D R 1980 *The Physics of MOS Insulators: Proc. Topical Conf. (Raleigh, NC)*

Wolters D R 1981 *Insulating Films on Semiconductors* ed M Schultz and G Pensl (Berlin: Springer) p 180

Wolters D R and Zegers-van Duijnhoven 1996 Tunnelling in thin SiO_2 *Phil. Trans. R. Soc.* A **354** 2327–50

Wright A C 2000 *Defects in* SiO_2 *and Related Dielectrics: Science and Technology* (*NATO Science Series*) ed G Pacchioni, L Skuja and D L Griscom (Dordrecht: Kluwer) pp 1–37

Wu E Y, Stathis J H and Han L-K 2000 *Semicond. Sci. Technol.* **15** 425

Zafar S, Lee B H and Stathis J 2004 *IEEE Electron Device Lett.* **25** 153