Eur. Phys. J. Appl. Phys. **28**, 265–291 (2004)
DOI: 10.1051/epjap:2004206

THE EUROPEAN PHYSICAL JOURNAL APPLIED PHYSICS

High dielectric constant oxides

J. Robertson[a]

Engineering Department, Cambridge University, Cambridge CB2 1PZ, UK

Received: 27 August 2004 / Accepted: 20 September 2004
Published online: 2 December 2004 –

Abstract. The scaling of complementary metal oxide semiconductor (CMOS) transistors has led to the silicon dioxide layer used as a gate dielectric becoming so thin (1.4 nm) that its leakage current is too large. It is necessary to replace the SiO_2 with a physically thicker layer of oxides of higher dielectric constant (κ) or 'high K' gate oxides such as hafnium oxide and hafnium silicate. Little was known about such oxides, and it was soon found that in many respects they have inferior electronic properties to SiO_2, such as a tendency to crystallise and a high concentration of electronic defects. Intensive research is underway to develop these oxides into new high quality electronic materials. This review covers the choice of oxides, their structural and metallurgical behaviour, atomic diffusion, their deposition, interface structure and reactions, their electronic structure, bonding, band offsets, mobility degradation, flat band voltage shifts and electronic defects. The use of high K oxides in capacitors of dynamic random access memories is also covered.

PACS. 85.40.-e Microelectronics: LSI, VLSI, ULSI; integrated circuit fabrication technology – 77.55.+f Dielectric thin films – 73.61.-r Electrical properties of specific thin films – 81.15.-z Methods of deposition of films and coatings; film growth and epitaxy

1 Introduction

1.1 Scaling and gate capacitance

The most important electronic device is the complementary metal oxide semiconductor (CMOS) field effect transistor (FET) made from silicon. This has arisen because the performance of CMOS devices has continued to improve over a forty year time span according to Moore's Law of scaling. This notes that the number of devices on an integrated circuit increases exponentially, doubling over 2 or 3 year period, to allow this. The minimum feature size in a transistor has decreased exponentially with year. The semiconductor Roadmap defines how each design parameter will scale in future years to continue this, as shown in Table 1 and Figure 1.

The scaling cannot go on forever, and the limits to Moore's law are often believed to be in lithography and the availability of sufficiently small wavelengths of light to pattern the minimum feature size. It turns out that materials are now also an important constraint. First, the maximum current density in interconnects between transistors recently led to copper replacing aluminium as the conductor used in interconnects. Then, the problem of RC time delays around the integrated circuit led to an effort to replace the silicon dioxide used as the inter-circuit passivant by a material of lower dielectric constant such as SiO_2F_x or SiOCH alloys. But the most serious problem in logic circuits is now in the FET "gate stack", that is the gate electrode and the dielectric layer between the gate and the silicon channel.

The thickness of the SiO_2 layer presently used as the gate dielectric is becoming so thin (under 2 nm) that the gate leakage current due to direct tunnelling of electrons through the SiO_2 will be so high, exceeding 1 A/cm^2 at 1 V (Fig. 2), that the circuit power dissipation will increase to unacceptable values [1–4]. In addition it becomes increasingly difficult to produce and measure accurately films of such small thickness. Finally, the reliability of SiO_2 films against electrical breakdown declines in thin films. Thus for these three reasons, but principally due to leakage, it is desired to replace SiO_2 as a gate oxide.

Tunnelling currents decrease exponentially with increasing distance. An FET is a capacitance-operated device, where the source-drain current of the FET depends on the gate capacitance,

$$C = \varepsilon_0 K A/t \quad (1)$$

where ε_0 is the permittivity of free space, K is the relative permittivity, A is the area and t is the SiO_2 thickness. Hence, the solution to the tunnelling problem is to replace SiO_2 with a physically thicker layer of a new material of higher dielectric constant (permittivity) K, Figure 3. This will keep the same capacitance, but will decrease the tunnelling current. These new gate oxides are called 'high K oxides'.

[a] e-mail: jr@eng.cam.ac.uk

Table 1. Summary of 2003 Roadmap. Node, gate length, equivalent oxide thickness of high power (CPU) and low standby power devices (mobile), gate oxide material, and gate electrode material.

Year	2001	2003	2005	2007	2009	2012	2016	2018
Node	130	100	80	65	45	32	22	18
ASIC 1/2 pitch	150	107	80	65	45	32	25	18
Physical gate length	65	45	32	25	20	13	9	7
T_{ox} hi power	1.5	1.3	1.1	0.9	0.8	0.6	0.5	0.5
T_{ox} lo power		2.2	2.1	1.6	1.4	1.1	1.0	0.9
Gate oxide	oxynitride			HfO_x; Si,N			$LaAlO_3$	
Gate metal	poly Si			metal gate, e.g. $TaSiN_x$				

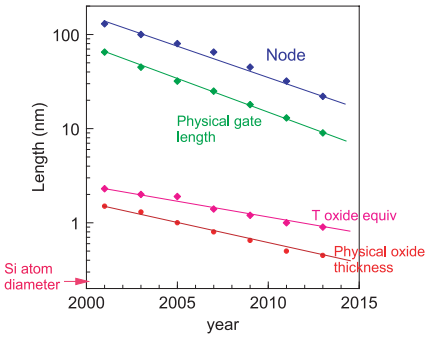

Fig. 1. The scaling of feature size, gate length, and oxide thickness according to the 2003 Semiconductor Roadmap.

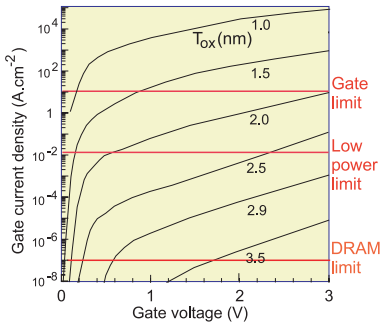

Fig. 2. Leakage current vs. voltage for various thickness of SiO_2 layers, from Lo et al. [3].

For device design, all FET dimensions scale proportionately and the precise material does not affect electrical designs, so it is convenient to define an 'electrical thickness' of the new gate oxide in terms of its equivalent silicon dioxide thickness or 'equivalent oxide thickness' (EOT) as

$$t_{ox} = \text{EOT} = (3.9/K)t_{HiK}. \quad (2)$$

Here 3.9 is the static dielectric constant of SiO_2. The objective is to develop high K oxides which allow scaling to continue to ever lower values of EOT.

The gate leakage problem has been apparent since the late 1990's [4], but then the criteria for the choice of oxide were not known. In about 2001, the choice of oxide narrowed to HfO_2, but the problems of making HfO_2

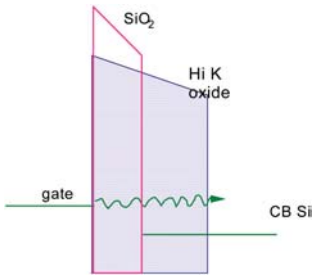

Fig. 3. Schematic of direct tunnelling through a SiO_2 layer and the more difficult tunnelling through a thicker layer of high K oxide.

into a successful electronic material appeared extremely high. It was not particularly believed that high K oxides would be used, but instead that device engineers would use a novel device design to circumvent the problem. However, the increasing importance the low-power sector of electronics, where power dissipation is a key issue, in mobile phones, lap-tops etc., meant that the problem must be confronted [1]. Low standby power CMOS requires a leakage current of below 1.5×10^{-2} A/cm^2 rather than just 1 A/cm^2. The initial problems of manufacturing high K oxide layers of sufficiently low EOT have been overcome. Recent announcements of key firms such as Intel [5] indicate that enough of the problems are now solved that high K oxides will be implemented in 2007 at the 65 nm node.

Four key problems have been identified by the industry [6]. These are (1) the ability to continue scaling to lower EOTs, (2) the loss of carrier mobility in the Si when using high K oxides, (3) the shifts of the gate voltage threshold, and finally (4) the instabilities caused by the high concentration of electronic defects in the oxides. Thus, this paper reviews the choice of oxides, their deposition, thermal stability, stability in device structures, electronic structure, interface properties, band offsets, electronic defects, carrier mobilities to understand what we have achieved so far, and how to solve these four problems.

At the same time, the scaling of the main form of memory, dynamic random access memory (DRAM), also requires a change of dielectric [7]. In DRAM information is stored as charge in a capacitor which is periodically

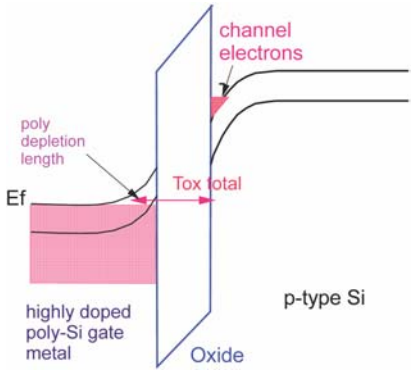

Fig. 4. The three contributions to the capacitance of the gate/electrode stack; channel, dielectric and gate depletion.

refreshed. The capacitor must retain charge during this time, so the leakage current density through the capacitor must be below 10^{-7} A/cm^2, lower than for gate dielectrics in logic circuits. The capacitance dielectric is presently Si oxy-nitride. This will have to be replaced in the same way by a material of higher K to continue the scaling. DRAMs can continue scaling by using more complex capacitor shapes with larger surface area to delay the transition, but again it will occur. Here, although the leakage current requirement is lower, the number of constraints on high K oxide are fewer, because the oxide is not in direct contact with any Si and it must only act as an insulator. The review will also cover this aspect.

1.2 EOT

In CMOS FETs, the gate capacitance is actually the series combination of three terms, the oxide capacitance, the depletion capacitance of the gate electrode, and the capacitance to the carriers in the Si channel [1], as shown in Figure 4. These three capacitances add as

$$1/C = 1/C_{\rm ox} + 1/C_{\rm gate} + 1/C_{\rm Si}. \quad (3)$$

As C varies as $1/t$, capacitances in series can be represented by a sum of effective distances. Thus we can define an 'effective capacitance thickness' (of SiO_2) as

$$\mathrm{ECT} = \mathrm{EOT} + t_{\rm gate} + t_{\rm Si}. \quad (4)$$

The channel capacitance arises because quantum delocalisation of the two-dimensional electron gas of electrons means that these electrons cannot lie infinitely close to the channel surface, but must delocalise a few Angstroms into the channel. This capacitance contribution is intrinsic and cannot easily be removed.

On the other hand, the gate electrode is presently made out of degenerately doped polycrystalline silicon, for engineering convenience. Poly-Si is a reasonable metal, but it is not the best metal. Thus, its low carrier density gives a depletion depth which is a few Å, whereas a good metal has a higher carrier density and has a depletion depth of only 0.5 Å. This depletion effect can be removed by replacing the poly-Si with a normal metal. Typical metals for this use could be TiN, TaSiN and Ru.

The metal is chosen primarily for its work function. The work function of the gate electrode determines the gate threshold voltage needed to turn the device into inversion. There are three choices [1]. In CMOS there are NMOS and PMOS devices. The first choice is to use the same metal for both NMOS and PMOS devices, in which case its work function should correspond to the mid gap energy of Si, about 4.6 eV. This is the simplest, most easily manufactured choice, but also the worst in terms of turn-on voltage. The harder choice is to use a different metal for NMOS and PMOS gates. This requires an NMOS gate metal with a work function close to the Si conduction band energy, 4.0 eV below the vacuum level. Such a metal will be quite reactive. For PMOS, this requires a metal with work function close to the Si valence band, or 5.1 eV. This metal would be very noble like Au, but such metals are difficult to etch. Thus, 'metal gates' is a separate topic, which turns out to be intimately linked to gate oxides and also requires considerable development.

2 Choice of high K oxide

Silicon dioxide is the key reason that microelectronics technology uses Si and not some other semiconductor. Si is an average semiconductor in performance, but in all other aspects SiO_2 is an excellent insulator. SiO_2 has the key advantage that it can be made from Si by thermal oxidation, whereas every other semiconductor (Ge, GaAs, GaN, SiC...) has a poor native oxide. SiO_2 is amorphous, has very few electronic defects and forms an excellent interface with Si. It can be etched and patterned to a nanometer scale. Its only problem is that when very thin it is possible to tunnel across it. Hence, we must loose these advantages of SiO_2 and start to use a new high K oxide. We can in principle choose from a large part of the Periodic table.

The requirements of a new oxide are six-fold:

1. It must have a high enough K that it will be used for a reasonable number of years of scaling.
2. The oxide is in direct contact with the Si channel, so it must be thermodynamically stable with it.
3. It must be kinetically stable, and be compatible with processing to 1000 °C for 5 seconds.
4. It must act as an insulator, by having band offsets with Si of over 1 eV to minimise carrier injection into its bands.
5. It must form a good electrical interface with Si.
6. It must have few bulk electrically active defects.

2.1 K value

The first requirement means that the oxides K should be over 10, preferably 25−30. There is a trade off with the band offset condition, which requires a reasonably large

Table 2. Static dielectric constant (K), experimental band gap and (consensus) conduction band offset on Si of the candidate gate dielectrics.

	K	Gap (eV)	CB offset (eV)
Si		1.1	
SiO_2	3.9	9	3.2
Si_3N_4	7	5.3	2.4
Al_2O_3	9	8.8	2.8 (not ALD)
Ta_2O_5	22	4.4	0.35
TiO_2	80	3.5	0
$SrTiO_3$	2000	3.2	0
ZrO_2	25	5.8	1.5
HfO_2	25	5.8	1.4
$HfSiO_4$	11	6.5	1.8
La_2O_3	30	6	2.3
Y_2O_3	15	6	2.3
a-$LaAlO_3$	30	5.6	1.8

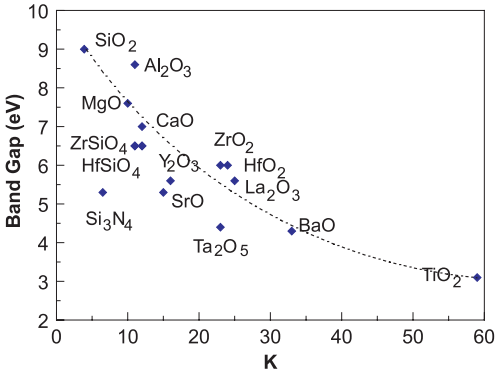

Fig. 5. Static dielectric constant vs. band gap for candidate gate oxides, after Robertson [8].

band gap. Table 2 and Figure 5 shows that the K of candidate oxides tends to vary inversely with the band gap, so we must accept a relatively low K value [8]. There are of course oxides with extremely large K's, such as ferroelectrics like $BaTiO_3$ but these have too low band gap. In fact, a huge K is undesirable in CMOS design because they cause undesirably strong fringing fields at source and drain electrodes [9].

2.2 Thermodynamic stability

The second requirement arises from the condition that the oxide must not react with Si to form either SiO_2 or a silicide according to the unbalanced reactions,

$$MO_2 + Si = M + SiO_2 \quad (5)$$

$$MO_2 + 2\ Si = MSi + SiO_2. \quad (6)$$

This is because the resulting SiO_2 layer would increase the EOT and negate the effect of using the new oxide. In addition, any silicide formed by (6) would generally be metallic and would short out the field effect.

This condition requires that the oxide has a higher heat of formation than SiO_2. Hubbard and Schlom [10,11] found that this restricts the possible oxides to very few, from columns II, III and IV of the Periodic table. These are SrO, CaO, BaO, Al_2O_3, ZrO_2, HfO_2, Y_2O_3, La_2O_3 and the lanthanides. It excludes some otherwise useful and familiar oxides such as Ta_2O_5, TiO_2 and the titanates including $SrTiO_3$ and $BaTiO_3$, which were favoured for use in capacitors in DRAMs. The group II oxides SrO, etc. are not favoured of themselves because they are very reactive with water. However, they would be acceptable as a transition layer. Hence this leaves us Al_2O_3, ZrO_2, HfO_2, Y_2O_3, La_2O_3 and various lanthanides such as Pr_2O_3, Gd_2O_3 and Lu_2O_3.

Zr and Hf are both from column IV and are generally believed to be the two most similar elements in the main Periodic table. However, it also turns out that the thermodynamic data for many oxides was not so accurate. It was subsequently found that $ZrO_2$is actually slightly unstable [12,13] and can react with Si to form the silicide, $ZrSi_2$. For this reason, HfO_2 is presently the preferred high K oxide over ZrO_2. La_2O_3 has a slightly higher K than HfO_2, but is more hygroscopic. Al_2O_3 has the disadvantage of a rather low K value. Y_2O_3 also has a lower K than La_2O_3. The other lanthanides Pr_2O_3, Gd_2O_3 and Lu_2O_3 are comparable to La [14–18].

One way to represent the stability or not of an oxide in contact with Si is on a ternary phase diagram and tie lines [1]. Figure 6 shows the ternary phase diagrams for the Ta-Si-O and Zr-Si-O systems. A given point in the diagram represents a composition and the temperature must be specified. Tie lines connect two compositions that can be in equilibrium with each other – without reaction. Tie lines cannot cross. Thus, Ta_2O_5 connects to Si via the SiO_2, not directly. On the other hand, ZrO_2 and $ZrSiO_4$

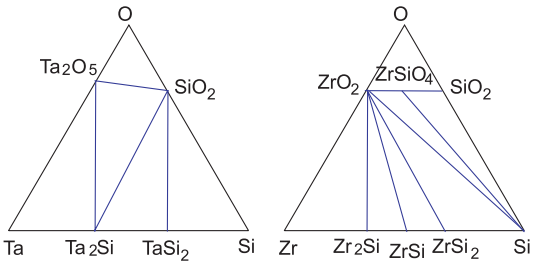

Fig. 6. Comparison of ternary phase diagrams of metastable Ta-Si-O and stable Zr-Si-O systems.

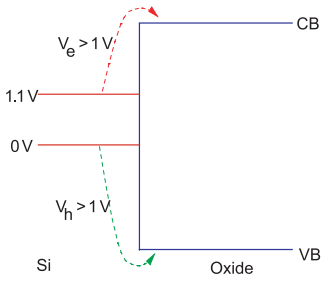

Fig. 7. Schematic of band offsets determining carrier injection in oxide band states.

and indeed any composition in $(ZrO_2)_{1-x}(SiO_2)_x$ are connected by tie-lines and are in equilibrium in contact.

2.3 Kinetic stability

The third condition is to be compatible with existing process conditions. Assuming we choose an amorphous oxide, this requires that the oxide remain amorphous when annealed to up to 1000 °C for 5 seconds. This is a strenuous condition in that SiO_2 is an excellent glass-former but most other high K oxides are not. Al_2O_3 is a reasonably good glass-former and is the next best in this respect. Ta_2O_5 is moderately good glass former, but was eliminated because it is reactive. All the other oxides crystallise well below 1000 °C.

This problem can be circumvented by alloying the desired oxide with a glass former – SiO_2 or Al_2O_3 – giving either a silicate or an aluminate [19]. This then retains the stability against crystallisation to close to 1000 °C. However, it is with the significant disadvantage of a lower K value. If this were the main condition, aluminates would be preferable to silicates, because they have a higher K. The K value roughly follows a linear rule of mixtures with composition, although there has been discussion of this aspect in a few cases. The addition of some nitrogen is found to raise the crystallisation temperature further, and so Hf silicates can just pass this criterion [20].

The other alternative is to use nano-crystalline oxides. This was originally thought to be a poor choice, because the grain boundaries would cause higher current leakage paths.

However, in practice, Lee et al. [21] found crystallised HfO_2 to have a similar leakage to amorphous HfO_2.

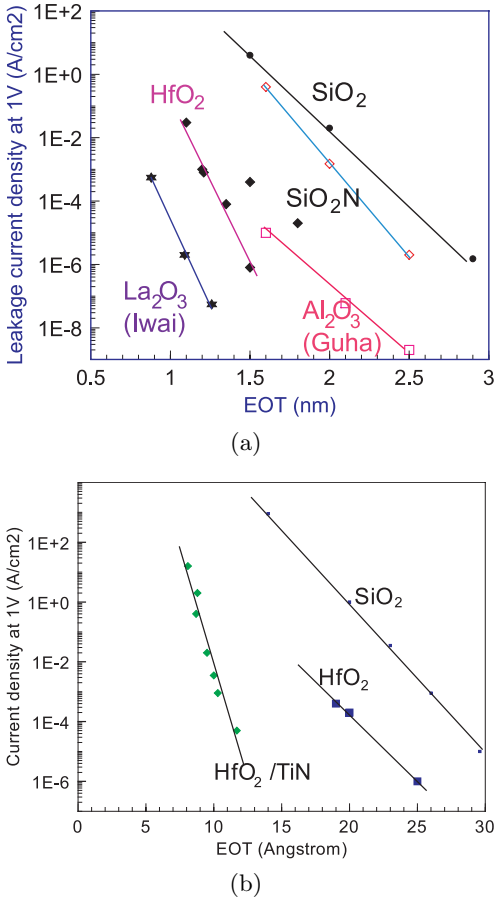

Fig. 8. (a) Leakage current density vs. EOT for various high K oxides, for HfO_2 [23], ZrO_2 [24], Al_2O_3 [23,25] and La_2O_3 [15]. (b) Leakage current density vs. EOT for HfO_2 with poly-Si gates and TiN gates, after [26].

2.4 Band offset

The high K oxide must act as an insulator. This requires that the potential barrier at each band must be over 1 eV in order to inhibit conduction by the Schottky emission of electrons or holes into the oxide bands [8,22], as shown schematically in Figure 7. SiO_2 has a wide gap of 9 eV, so it has high barriers for both electrons and holes. However, if the oxide has a narrower band gap like $SrTiO_3$, which is only 3.3 eV, its bands must be aligned almost symmetrically with respect to those of Si for both barriers to be over 1 eV. In practice, the conduction band offset is usually smaller than the valence band offset. This limits the choice of oxide to those with band gaps over 5 eV. The oxides that satisfy this criterion are Al_2O_3, ZrO_2, HfO_2, Y_2O_3 La_2O_3 and various lanthanides, and their silicates and aluminates [8]. It is interesting that these are the same oxides as pass the thermal stability criterion. This is because a high heat of formation correlates with a wide band gap, in ionic compounds.

Table 3. Comparison of deposition methods. O = good, x = bad.

	Coverage	Purity	Defects	Thickness	Large area
Sputtering	o	oo	xxx		oo
Metal dep + oxidation	o	oo	o	oo	o
MOCVD	ooo	o	oo	oo	ooo
ALD	ooo	o	oo	ooo	ooo

The leakage current for various high K oxides as a function of EOT is plotted in Figure 8. Figure 8(a) shows data for HfO_2 from Gusev [23], for ZrO_2 from Gusev [24], for Al_2O_3 of Guha [23,25], and for La_2O_3 from Iwai [15]. Figure 8(b) compares data for HfO_2 films with poly-Si electrodes and HfO_2 with TiN electrodes, from Tsai et al. [26].

Yeo et al. [27] have defined a scaling figure of Merit to compare leakage currents by combining the barrier height, tunnelling mass and K. Lanthanides have the lowest leakage in Figure 8(a) and have the highest figure of merit because they have the highest CB offset, as shown in Section 4.3. However, Hf alloys are presently preferred because La oxides are hygroscopic. Eventually La_2O_3 or a La compound such as $LaAlO_3$may be used, according to the 2003 Roadmap (Tab. 1), but this is a long way off.

2.5 Interface quality

The oxide is in direct contact with the Si channel. The carriers induced by the gate are induced within Angstroms of the Si-oxide interface. Hence, this interface must be of the highest electrical quality, in terms of roughness and the absence of interface defects. Extra defects are associated with oxide grain boundaries. Therefore, there are two ways to ensure a high quality interface, either use a crystalline oxide grown epitaxially on the Si, or use an amorphous oxide.

Using an amorphous oxide has many advantages over a poly-crystalline oxide. It is like the existing Si:SiO_2 situation. It is the lowest cost solution, most compatible with the existing process. Second, an amorphous oxide might be able to configure its interface bonding to minimise the number of interface defects. Third, it is possible to gradually vary the composition of an amorphous oxide without creating a new phase; for example as in silicate alloys, or interfacial layers, or when adding nitrogen. Fourth, an amorphous oxide and its dielectric constant is isotropic, so that fluctuations in polarisation from differently oriented oxide grains will not scatter carriers. Finally, amorphous phases have no grain boundaries. Grain boundaries in a polycrystalline oxide act as easy diffusion paths for dopants, such as B or P from a poly-Si gate electrode lying above.

The advantages of epitaxial oxides may come in the future, where their ability to create more abrupt interfaces allows us to reach lower EOTs.

2.6 Defects

Electrically active defects are defined as atomic configurations which give rise to electronic states in the band gap of the oxide. Typically these are sites of excess or deficit of oxygen or impurities. Defects are undesirable for four reasons. Firstly, charge trapped in defects causes a shift in the gate threshold voltage of the transistor, the voltage at which it turns on. Secondly, the trapped charge will change with time so the threshold voltage will shift with time, leading to instability of operating characteristics. Thirdly, trapped charge scatters carriers in the channel and lowers the carrier mobility. Fourthly, defects cause unreliability; they are the starting point for electrical failure and breakdown of the oxide.

SiO_2 is an almost ideal insulating oxide, in that it has a low concentration of defects which give rise to states in the gap. This is fundamentally because it has a low coordination number, so that its bonding can relax and rebond any broken bonds at possible defect sites. Any remaining defects are passivated by hydrogen. The high K oxides are not materials with a low intrinsic defect concentration because their bonding cannot relax as easily. Much of the present-day engineering of these oxides consists of pragmatic strategies of trying to reduce defect densities by processing control and annealing.

3 Materials chemistry of high K oxides

3.1 Deposition

The great advantage of SiO_2 is that it can be grown by thermal oxidation. In contrast, high K oxides must be deposited. Deposited oxides are never as good. The advantages and disadvantages of various deposition methods are summarised in Table 3. Sputtering is one of a number of physical vapour deposition (PVD) methods. Its advantage is that it is broadly available and can produce pure oxides. Its disadvantages are that oxides are insulators so sputtered oxides tend to have plasma-induced damage. Also, PVD methods deposit in line of sight, so they do not give good coverage.

A method for producing highly pure, thin oxides is to evaporate metal by electron beam which is highly controllable to small thickness, and to oxidise the deposited metal by ozone or UV assisted oxidation. The advantage is that this produces less damage than oxide sputtering and should produce the purest oxide. But it is not a production method. One could also ion beam sputter the metal

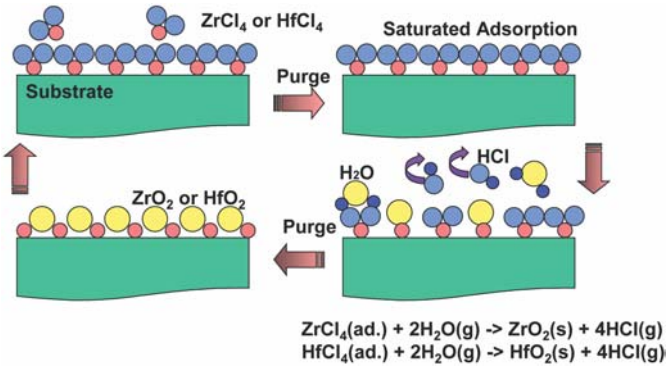

Fig. 9. Schematic of the cyclic process of Atomic Layer Deposition. Thanks to P.C. McIntyre.

– ion beam on the sputter target, not on the substrate. This does not produce damage.

The preferred industrial scale methods are chemical vapour deposition (CVD) and atomic layer deposition. CVD uses a volatile metal compound as a precursor which is introduced into the chamber and oxidised during deposition onto the substrate. The advantages of CVD are that it is already widely used in the electronics industry for insulator deposition, it gives conformal coverage over complex shapes because it is not just line of sight, and that the growth rate is controllable over a wide range from very slow to high. The CVD precursors can be metal chlorides such as $ZrCl_4$ and $HfCl_4$ or metal organics such as tetrabutoxyl Zr, in which case it is called metal organo CVD (MOCVD).

Atomic layer deposition is a method of cyclic deposition and oxidation [28,29]. As shown schematically in Figure 9, the surface is exposed to the precursor which is absorbed as a saturating monolayer. The excess precursor is then purged from the chamber by an Ar pulse. A pulse of oxidant such as H_2O, H_2O_2 or ozone is then introduced which must then fully oxidise the adsorbed layer to the oxide and a volatile by-product such as HCl. The excess oxidant is then purged by a pulse of Ar, and the cycle is repeated.

The effective chemical reactions are

$$ZrCl_4 + 2\ OH_{surface} = ZrCl_2O_2 + 2\ HCl \qquad (7)$$

$$ZrCl_2O_2 + 2\ H_2O = ZrO_2(OH)_2\ _{surface} + 2\ HCl. \qquad (8)$$

Here the existing ZrO_2 surface is assumed to be terminated by OH groups at about 300 °C. The $ZrCl_4$ chemisorbs exothermically onto the OH sites by the exothermic elimination of HCl. In the second stage, water oxidises the Cl atoms again with the elimination of HCl.

The precursor is designed so that both steps of absorption and oxidation are exothermic. The precursor must undergo self-limiting adsorption, be volatile, high purity, non-toxic, have no gas phase reactions, no self-decomposition, and no etching of the existing oxide. The first precursors for ZrO_2 and HfO_2 were the chlorides. However, these have low volatility. A wide range of new precursors in being developed [28,30].

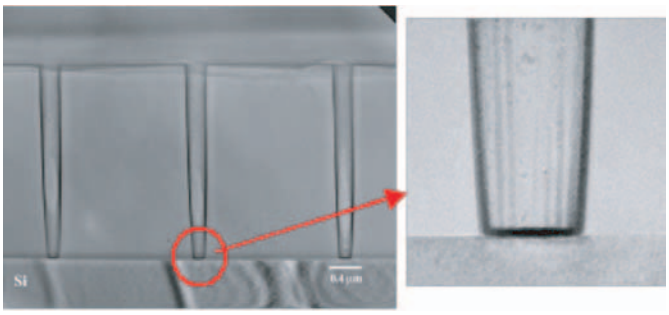

Fig. 10. Scanning electron microscope image of trench structure showing excellent coverage by ALD HfO_2. Thanks to P.C. McIntyre.

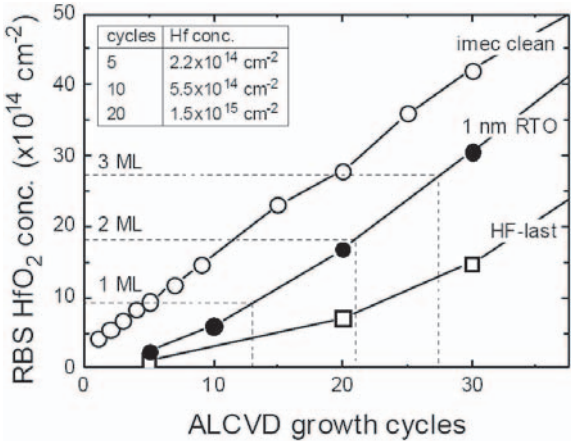

Fig. 11. Film thickness vs. number of ALD cycles, for different Si surface preparations, showing the nucleation delay on HF-last Si.

ALD was developed to produce highly conformal, pinhole-free insulating films, as seen in Figure 10. The advantages of ALD are that it is able to grow the thinnest films of all methods, and the most conformal films even into deep trenches. A disadvantage is its slow growth rate. A disadvantage of ALD and MOCVD is that they generally introduce impurities into the oxides, such as C, H or Cl, depending on precursor, whose electrical activity needs careful study. Careful annealing strategies are needed to densify the CVD and ALD oxides and remove impurities. ALD is an excellent method for producing Al_2O_3, using trimethyl-aluminium as precursor [28]. This and other reasons led to the adoption of ALD for many high K oxides.

Each cycle of ALD adds a layer of oxide which is usually much less than an atomic layer thick, despite its name. The precursor absorption saturates below one monolayer because of steric hindrance. This is not a significant disadvantage, it just takes more cycles to grow a certain thickness.

The most inert surface of Si is regarded as the H-terminated surface obtained by the HF-last cleaning procedure. In the development of the ALD, it was found that ALD of ZrO_2 and HfO_2 from chlorides or many organic precursors did not nucleate easily on HF-last Si

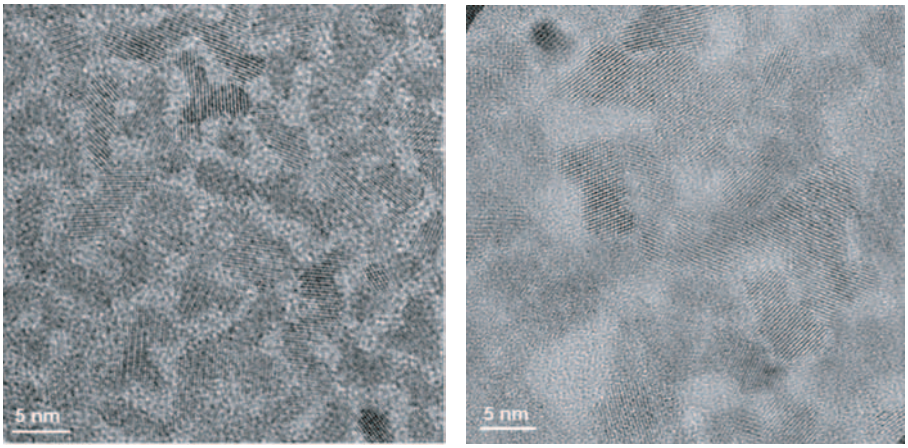

Fig. 12. Plan view TEM image of crystallisation in HfO_2/SiO_2 alloy system (a) 40% HfO_2, (b) 80% HfO_2 [39]. Thanks to S. Stemmer.

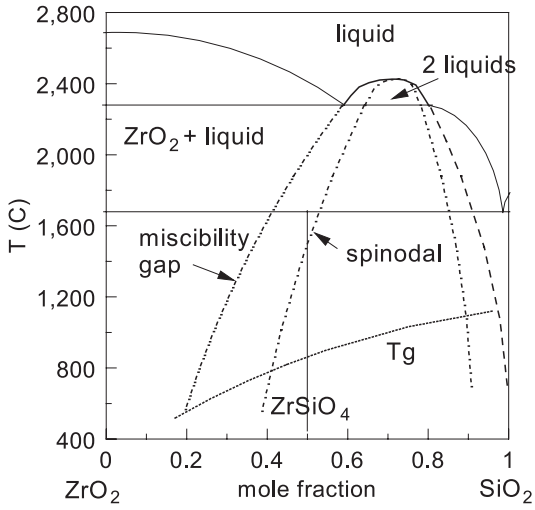

Fig. 13. Phase diagram of ZrO_2/SiO_2 showing miscibility gap. After Kim [37].

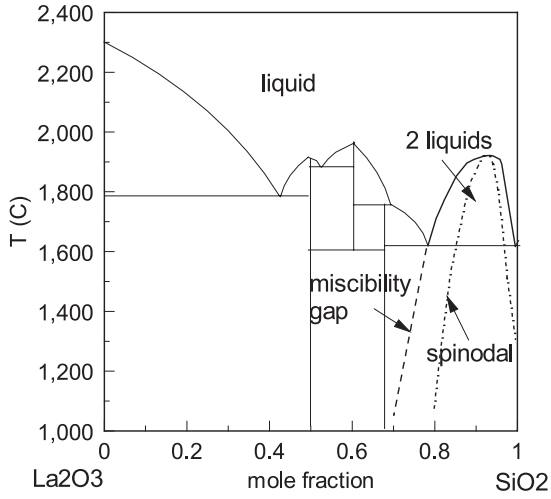

Fig. 14. Phase diagram of La_2O_3/SiO_2 with miscibility gap. After Maria [36].

surfaces and had a slow initial growth rate [31,32], as in Figure 11. This meant that oxide films even 3 ML thick were not fully covered or 'closed' but islanded [31]. It was found that nucleation occurred much more readily on a slightly pre-oxidise Si surface [31]. Thus, ALD is usually carried out on a 'chemical oxide' (SiO_2) surface formed by ozone cleaning of Si. This limits the ultimate lowest EOT that ALD can presently achieve. However, the development of ALD precursors which do nucleate on H-terminated Si and different processing strategies will overcome this obstacle when needed [33,34].

3.2 Alloy crystallisation

Silicate and aluminate alloys of Zr, Hf and La oxides are often used instead of the pure metal oxides in order to have a higher resistance to crystallisation [19,20,35]. Zr silicate has been the most widely studied. Crystallisation directly to the crystalline silicate $ZrSiO_4$ is inhibited by kinetics. Instead, Maria et al. [36] showed that crystallisation occurred by the phase separation of the ZrO_2 and SiO_2 phases followed by the crystallisation of the ZrO_2 component. This can be seen for HfO_2-SiO_2 alloys in the high-resolution transmission electron microscope images in Figure 12 for two different compositions [38].

The phase diagram of the ZrO_2-SiO_2 system is known reasonably accurately [36–38], as shown in Figure 13. That of HfO_2-SiO_2 is not know as well, but it is assumed to be similar to ZrO_2-SiO_2 because of the chemical similarity of Zr and Hf. The key factor is that ZrO_2 and SiO_2 liquids are immiscible over a small range of composition. This is attributed to the high ionic charge of Zr. This 'miscibility gap' can be continued to lower temperatures to define a miscibility gap in the solids. This also defines a spinodal region in which the alloy can spontaneous phase separate to lower its free energy [37]. The glass transition temperature is also marked in Figure 13, it reduces in ZrO_2 rich alloys. Thus, crystallisation occurs by two mechanisms. For Zr contents between 20−60 mol% ZrO_2 will crystallise by

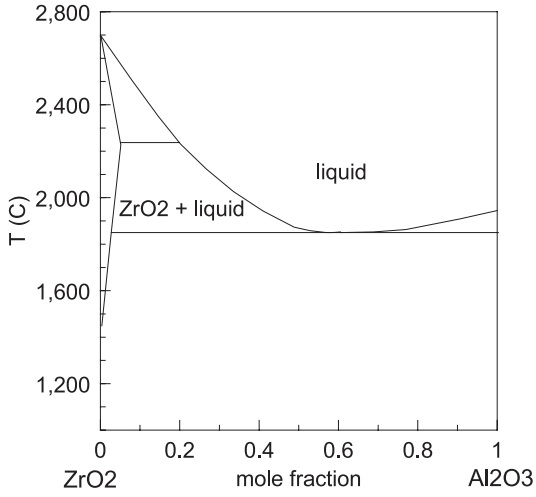

Fig. 15. Phase diagram of ZrO_2/Al_2O_3. After Zhao [40].

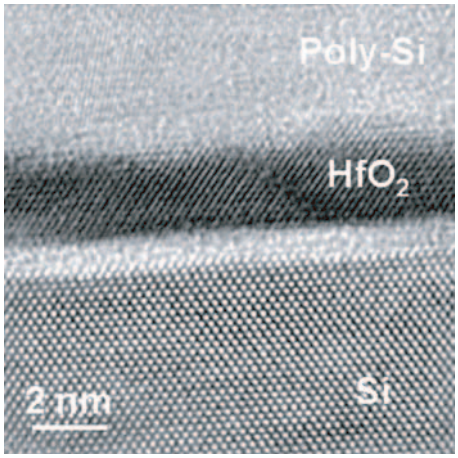

Fig. 16. HRTEM cross section showing interfacial layer of SiO_2 below the HfO_2 layer. Thanks to S. Stemmer.

spinodal decomposition followed by crystallisation. This tends to lead to small grain sizes. Films with over 60% Zr will crystallise by the kinetically limited nucleation and growth of crystalline ZrO_2. This was confirmed by extensive TEM and x-ray scattering studies on Hf silicate alloys by Stemmer et al. [38,39]. The La silicate phase diagram [36] is qualitatively similar to that of $ZrSiO_4$ except that the two-phase region is further towards SiO_2 (Fig. 14).

In contrast, the phase diagrams of aluminates such as ZrO_2-Al_2O_3 do not show miscibility gap [40], as seen in Figure 15, so they are more resistant to crystallisation [41]. However, it turns out that aluminates have higher densities of electronically active defects, so that silicates are preferred to aluminates for gate oxide applications.

Despite the use of silicates, they still cannot fully achieve the 1000 °C requirement. The final improvement in performance comes with adding a fraction of nitrogen to the alloy [15]. The N reduces the diffusion coefficient of oxygen in the alloys, and this reduces the crystallisation rate sufficiently that the alloy can now withstand 1000 °C. Lee et al. [42] have studied the effect of adding nitrogen at either interface or in the bulk.

3.3 Atomic diffusion

We noted that a gate oxide must withstand processing to temperatures of order 1000 °C without changing its state. It must also not mix with either the Si channel or the poly-Si (or metal) gate electrode, or allow components of the gate electrode through to the Si. All these aspects require the gate oxide to have low atomic diffusion coefficients. Interestingly, the proposed oxides HfO_2 and ZrO_2 belong to the class of fluorite oxides like CeO_2 which are fast oxygen ion conductors, of interest in solid state fuel cells or high temperature sensors. Clearly, for our application oxide diffusion must be inhibited.

A great advantage of alloying with SiO_2 is that the Si sites in silicates are covalently bonded to oxygen. This greatly reduces the oxygen diffusion rate. The diffusion rates of Hf, O, B and P in HfO_2 and Hf silicate have been measured after implantation by secondary ion mass spectroscopy (SIMS) and nuclear reaction profiling [43–47] to confirm these observations. The mixing of oxide and Si layers has also been studied by Medium Energy Ion Scattering (MEIS) which measures the element profile.

The basic silicate is found to perform adequately in most respects. However, alloying with nitrogen is used to lower the diffusion rates still further, as seen, which further raises the crystallisation temperature. This is a general role of N. Si_3N_4 is a much better diffusion barrier than SiO_2, because it has no open channels for molecular or ionic diffusion, and the N site has a higher coordination and thus resists network diffusion. Of course, HfO_2does not have an open lattice like SiO_2, but still N seems to lower network diffusion in $HfSiO_4$ [21].

Another key role of the oxide is to block dopant diffusion from any poly-Si gate electrode [48]. N is found to be very useful in blocking B diffusion through SiO_2 presumably because it forms bound pairs with B. In high K oxides, N is also efficient at blocking boron diffusion. A grain boundary would be a short circuit diffusion path, so here N acts to block diffusion by stopping crystallisation and the formation of any grain boundaries [21].

3.4 The interfacial layer

An interfacial layer of SiO_2 often exists between the Si channel and the high K oxide layer. Figure 16 shows a cross-sectional of an example [49]. There are advantages and disadvantages for this, as long as its presence and thickness can be controlled. The overall EOT of a layer 1 of SiO_2 and a layer 2 of high K oxide is given by the series capacitance formula,

$$1/C = 1/C_1 + 1/C_2 \tag{9}$$

which becomes

$$\mathrm{EOT} = t_{\mathrm{SiO2}} + \mathrm{EOT}_{\mathrm{hi}K}. \tag{10}$$

Thus, an extra SiO_2 layer is undesirable as it adds to the overall EOT. In fact, the K of SiO_2 (3.9) is so small that a SiO_2 layer can rapidly use up the EOT allocation. It is a severe impediment to scaling.

The SiO_2 layer often arises not because of reaction of the HfO_2 with the Si, as the HfO_2 was chosen to avoid this. It arises because O diffuses through the HfO_2 layer to oxidise the Si underneath. Indeed ZrO_2 and HfO_2 are a catalyst for this oxidation process [50]. The SiO_2 layer usually grows during the post-deposition annealing stage, not during growth. Naraynan [51] proved this for the case of Y_2O_3. This can be avoided by adding silicate or N to the HfO_2 layer to reduce atomic diffusion. However, scaling requirements will reduce the ability to use silicates in the future because they lower K.

The second reason an SiO_2 layer exists is that is it beneficial and it was deliberately put there. Firstly, a 'chemical oxide' is presently used to act as a nucleation layer for ALD growth of HfO_2 and other oxides [31,32]. With experience or the development of better ALD precursors, this need should decline.

The SiO_2 layer may also be introduced because it improves the electrical quality of the Si-oxide interface, as described later. The Si-SiO_2 interface is well understood and can be of high quality. In principle, it can be made with a very low defect concentration, and the defects can be passivated by forming gas annealing. The presence of a SiO_2 layer also spaces the Si channel from the high K oxide, which can stop the reduction in carrier mobility that high K oxides can cause, see later.

A disadvantage of this interfacial oxide is that it may not have the same quality as SiO_2 produced by thermal oxidation of Si [52,53]. It may be defective. Copel [54] has used a number of techniques such as MEIS to study the profile and composition of interfacial oxides under HfO_2. They found that they are SiO_2 despite sometimes appearing to have higher K values than thermal oxide. EELS found a similar result [55,56].

It is an advantage if we can control the thickness of the interfacial SiO_2 layer, and if necessary remove it entirely. This can be done in two ways. Firstly, Si and SiO_2 react to form volatile SiO within a range of temperatures around 900−1000 °C. The initial surface can be annealed to desorb its native oxide as SiO [57]. The SiO will also desorb from a buried layer through a high K oxide covering. The second way is to react the metal such as Hf with the SiO_2 to displace Si [58,59].

4 Bonding and electronic structure

4.1 Bonding

The oxides of interest are transition metal oxides except for Al_2O_3. Figure 17 shows the density of states (DOS) of Al_2O_3. The top of the valence band lies at 0 eV and the band gap lies from 0 to 8.8 eV. The bonding in Al_2O_3 is more ionic than in SiO_2, and its atoms have ionic coordinations. However, its electronic DOS does resemble that of

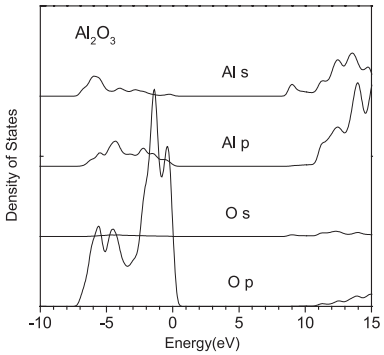

Fig. 17. Density of states of Al_2O_3 in corundum structure. Note O 2p–like valence band and 8.8 eV band gap.

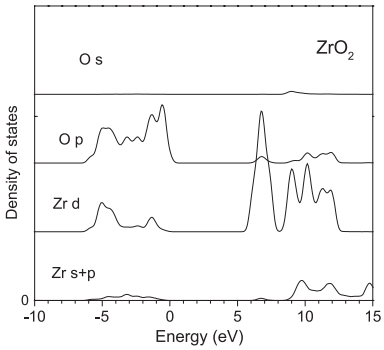

Fig. 18. Density of states of cubic ZrO_2.

SiO_2 in that the valence band consists mainly of O p states and a conduction band of mainly Al s, p states.

A more typical example is ZrO_2. ZrO_2 films are amorphous at lower temperatures, but crystallise relatively easily. ZrO_2 is stable in the monoclinic structure at room temperature, it transforms to the tetragonal structure above 1170 °C and it can be stabilised in the cubic fluorite structure by addition of Y [60]. HfO_2 is similar. In cubic and tetragonal ZrO_2, Zr has 8 oxygen neighbours and each oxygen has four Zr neighbours, while in monoclinic ZrO_2 each Zr atom has 7 oxygen neighbours. Tetragonal ZrO_2 is related to cubic ZrO_2 by displacing oxygens along the z axis towards 4 of the Zr's.

Figure 18 shows the density of states of cubic ZrO_2. It has an indirect gap of 5.8 eV, the experimental value [60]. French [60] found that the gap is narrower in the lower symmetry forms of ZrO_2. However, recent calculations find that the tetragonal phases have the widest gaps (Tab. 4) [61,62]. Our calculated band structures are similar to those found by others. The valence band is 6 eV wide, and it has a maximum at X formed from O p states. The conduction band minimum is a Γ_{12} state of Zr 4d orbitals. The Zr d states are split by the crystal field into a lower band of e states and an upper band of t_2 states 5 eV higher (at Γ). The partial DOS shows considerable charge transfer, with the valence band being strongly O p states, and conduction band on Zr d states, with 30% admixture. The band structure of HfO_2 is very similar to that of ZrO_2 except that the crystal splitting of the Hf 5d states in the conduction band is larger than Zr's (Fig. 19).

Table 4. Experimental and calculated band gaps (eV) of ZrO_2 and HfO_2 phases.

	Cubic	Tetragonal	Monoclinic
ZrO_2 (Experimental, French)	6.1	5.8	5.8
ZrO_2 (GW, Kralik)	5.55	6.4	5.42
HfO_2 (WDA, this work)	6.0	6.4	5.8

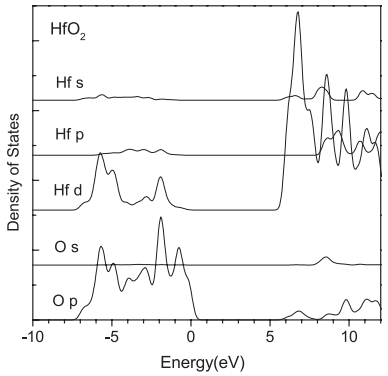

Fig. 19. Density of states of cubic HfO_2.

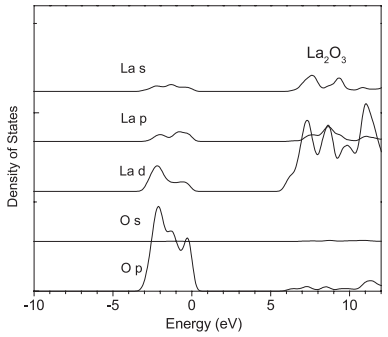

Fig. 20. Density of states of La_2O_3.

Crystalline La_2O_3, has the La_2O_3 structure in which La is 7-fold coordinated, with 4 short bonds and 3 longer bonds. The DOS of La_2O_3 in Figure 20 shows that the valence band is strongly localised on O p states and the conduction band in on La d with some La s, p states starting at 8 eV [63]. The band gap is indirect and 6 eV. The valence band is now only 3.5 eV wide, narrower than in ZrO_2. The band gap is indirect and 6 eV. The valence band is now only 3.5 eV wide, narrower than in ZrO_2. The ionicity is higher than in ZrO_2.

Of the group IIIA metal oxides, Y_2O_3 has the cubic bixbyite (defect spinel) structure. This has a large 56 atom unit cell in which there are two types of Y sites, both 7-fold coordinated. This structure occurs because Y has a smaller ionic radius than La. The band gap of Y_2O_3 is direct and is about 6 eV [63]. The valence band is again only 3 eV wide. The partial DOS shows the valence band is largely O p states. The conduction band minimum has mixed Y d, s character.

In each of these cases, the band gap is between O $2p$ valence states and metal d states. Thus the band gap is proportional to the metal atomic d orbital energy, as noted by Lucovsky [64].

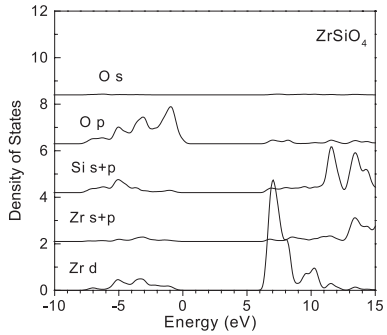

Fig. 21. Density of states of crystalline $ZrSiO_4$.

$ZrSiO_4$ is typical of the transition metal silicates. Crystalline $ZrSiO_4$ has the body-centred tetragonal structure. The Zr and Si atoms are organised in chains. Each Zr atom has eight O neighbours. Each Si has four O neighbours in a tetrahedral arrangement. These coordinations are expected to carry over to the amorphous phases and the amorphous alloys, although there has been debate about this. Its partial DOS is shown in Figure 21. The band gap is about 6.5 eV [63]. The valence band is 7 eV wide [65]. The conduction bands form two blocks. The lower conduction band is due to Zr d states and lies between 6.5 eV and 8 eV, and a second conduction band due to Si-O antibonding states lie above 9 eV. This is an important general rule that the conduction band of Zr silicates forms two non-mixing ZrO_2-like and SiO_2-like bands. The states do not mix because the Si s, p states and metal d states have different local symmetry. Thus, the CB edge of the silicates retains its Zr d character as long as Zr is present, and the band gap increases only slowly, with very strong bowing below the virtual crystal model. Experiments confirm this [66].

Another large class of possible gate oxides are the perovskites such as $SrTiO_3$. In the ABO_3 structure, the smaller transition metal ion occupies the B site, which is octahedrally coordinated by six oxygens. The oxygens are bound to two B ions, while the A ion is surrounded by twelve oxygen ions. Figure 22 shows the partial DOS of $SrTiO_3$. The band gap is direct and 3.3 eV wide. The lowest conduction bands are Ti d_{xy} t_2 states followed by the Ti d_{z2} states. The next states above 7 eV are Ti p states followed by Ba s states. Thus, the A ion states (Ba or Sr) are well away from the band gap, and the ion can be considered to be essentially fully ionised and passive. On the other hand, the Ti-O bond is polar but only about 60% ionic.

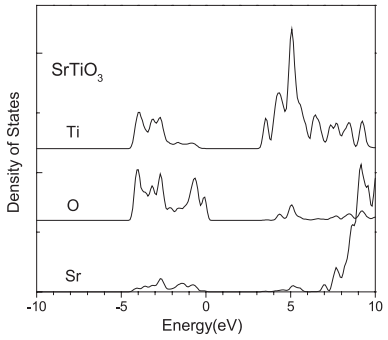

Fig. 22. Density of states of cubic $SrTiO_3$.

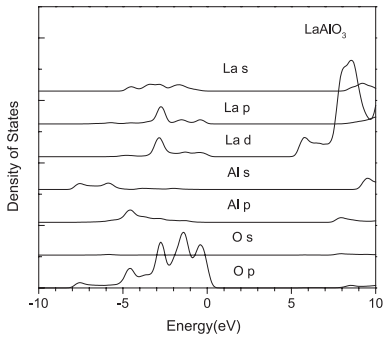

Fig. 23. Density of states of cubic $LaAlO_3$.

$LaAlO_3$ is another perovskite oxide, which is of importance as an epitaxial gate oxide because it has a large dielectric constant, and a close lattice match to Si. It is unusual in that the transition metal La occupies the A site and Al occupies the octahedral B site. The partial DOS of $LaAlO_3$ is shown in Figure 23. The band gap is taken as 5.6 eV from recent ellipsometry work [67].

4.2 Dielectric constants

The static dielectric constant of the oxides is the sum of the electronic and lattice contributions, $\kappa = \kappa_e + \kappa_l$. The electronic component κ_e is also the optical dielectric constant ε_∞ and it is given by the refractive index squared, $\kappa_e = \varepsilon_\infty = n^2$. ε_∞ values are typically 4−5 for the wide gap oxides of interest. Thus they are *not* the main source of the high K in Table 2. The large static dielectric constant arises from a large lattice contribution,

$$\kappa = n^2 + \frac{Ne^2 Z*_T^2}{m\omega_{TO}^2}. \quad (11)$$

Here, n is the refractive index, N is the number of ions per unit volume, e is the electronic charge, Z_T^* is the transverse effective charge, m is the reduced ion mass and ω_{TO} is the frequency of the transverse optical phonon. Large values of κ_l occur when Z^* is large and/or the frequency of a polar optical mode ω_{TO} is small. This means that they are incipient ferroelectrics.

The dielectric constants of the various phases of HfO_2 and ZrO_2 have been calculated in the local density formalism [68]. This is a good means to understand the differences and the anisotropies. Rignanese [69] found that the tetragonal phase has the largest and most anisotropic K, but not by as much as found earlier by Vanderbilt [68].

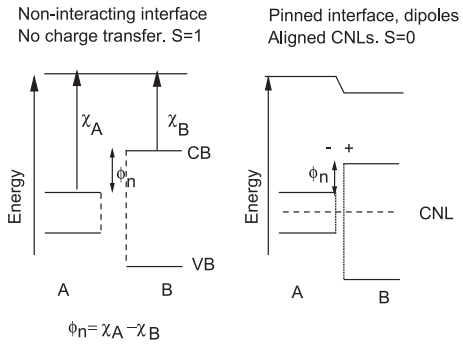

Fig. 24. Schematic of how charge transfer at semiconductor interface controls its band line up, (a) no charge transfer, (b) charge transfer.

4.3 Band offsets

The band offset between oxide and Si defines the barrier for injection of electrons or holes into the oxide bands. The electron barrier or conduction band (CB) offset tends to be the smaller of the two. The CB offset is one of the key criteria in the selection a gate oxide. It must be over 1 eV to give adequately low leakage current [8,18].

The CB offset has previously been calculated for most candidate high K oxides and it can be measured by methods such as photoemission. The band line up at an interface is controlled by a dipole formed by charge transfer across the bonds at the interface. The band offset consists of two components, a component intrinsic to the bulk oxide and Si and a component which depends specifically on the interface bonding configuration [70,71]. The intrinsic component is of interest because the specific bonding at the interface is generally not known. Usually, the intrinsic component is the main component. However, the interface specific component can be important. It means that there is no unique offset value for a given oxide on Si. This can be an advantage as it allows offsets to be controlled by varying the interface chemistry.

The band line up at an interface is controlled by a dipole formed by charge transfer across the bonds at the interface [72]. In the case of two non-interacting surfaces, the conduction band line up is given by the difference between the electron affinities (the energy of the conduction band edge below the vacuum level) (Fig. 24). This is known as the Schottky limit. If the surfaces interact, an interface dipole due to charge transfer across the interface by modifies this offset. The charge transfer acts to align an energy level in each surface. In the limit of strong coupling, known as the Bardeen limit, these levels are fully aligned. The band offset is then given by the difference of this energy level below the two conduction bands, and is independent of the vacuum levels. Most high K oxides are intermediate between the two limits.

Table 5. Comparison of the calculated conduction band offset (by LDA method) and experimental values for various gate oxides, by various authors. * = ALD.

	Calculated (eV)	Experiment (eV)	Ref
SiO_2		3.1	Alay [109]
Ta_2O_5	0.35	0.3	Miyazaki [79]
$SrTiO_3$	0.4	0	Chambers [78]
ZrO_2	1.6	1.4	Miyazaki [79]
		2.0	Afanasev [72]
		1.4	Rayner [85]
HfO_2	1.3	1.3	Sayan [84]
		2.0	Afansev [83]
Al_2O_3	2.4	2.8	Ludeke [81]
		2.2 *	Afansev [83]
a-$LaAlO_3$	1.0	1.8	Edge [87]
La_2O_3	2.3	2.3	Hattori [88]
Y_2O_3	2.3	1.6	Miyazaki [89]

A particular model is the model of metal induced gap states (MIGS) [73–76]. This model says that the reference level is the so-called charge neutrality level (CNL) of the intrinsic surface states. A semiconductor surface has gap states due to the broken surface bonds. These are spread across the energy gap. The CNL is the highest occupied surface state on a neutral surface of a semiconductor. It is like a Fermi level of the intrinsic gap states.

The MIGS model says that for a metal on the semiconductor, the MIGS are like the plane waves of the metal decaying into the semiconductor gap. The interface dipole now tries to align the semiconductor's CNL to the metal Fermi level. The Schottky barrier height, the energy of the semiconductor conduction band above the metal Fermi level, is given by

$$\phi_n = S(\Phi_M - \Phi_S) + (\Phi_S - \chi_s) \tag{12}$$

where Φ_M is the metal work function, Φ_S is the charge neutrality level of the semiconductor, and χ_S is the electron affinity (EA) of the semiconductor. S is a dimensionless pinning factor given by $d\phi_n/d\Phi_M$. S is given in the linear approximation by [77]

$$S = \frac{A}{1 + \frac{e^2 N\delta}{\varepsilon_{\varepsilon_0}}} \tag{13}$$

where e is the electronic charge, ε_0 is the permittivity of free space, N is the density of the interface states per unit area and δ is their extent into the semiconductor. In fact, this model is not strictly correct, as the whole occupied valence band states, not just those at the Fermi level contribute to S [72]. Nevertheless the MIGS model appears to give reasonably good predictions.

The model is extended to the band offsets between semiconductors. Charge transfer tends to align the charge neutrality level (CNL) of the bulk oxide with the CNL of the bulk Si. The CB offset is given by [8]

$$\phi_n = (\chi_a - \Phi_{S,a}) - (\chi_b - \Phi_{S,b}) + S(\Phi_{S,a} - \Phi_{S,b}). \tag{14}$$

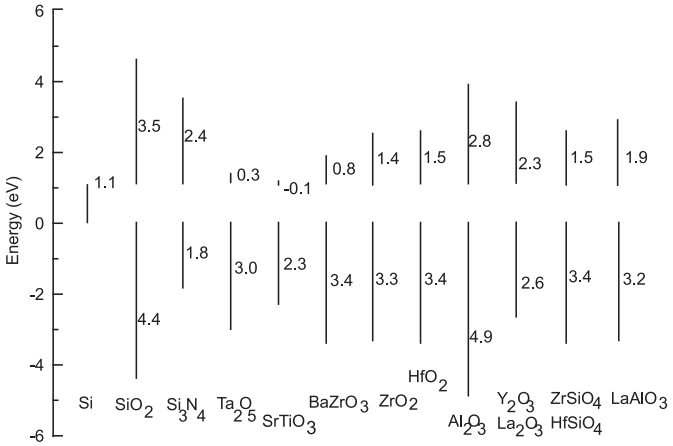

Fig. 25. Predicted barrier heights for a range of high K gate oxides, after [8].

Here, χ_a is the electron affinity (EA) of the oxide, χ_b is the electron affinity of the semiconductor, and Φ_{Sa} and Φ_{Sb} are the charge neutrality levels of the oxide and semiconductor respectively. All the energies in (14) are measured from the vacuum level, except ϕ_n which is measured from the conduction band edge. S is a constant, the Schottky barrier pinning factor, which is found by Monch [73] to vary empirically with the electronic component of the dielectric constant of the wider gap material (the oxide) as

$$S = \frac{1}{1 + 0.1(\varepsilon_\infty - 1)^2}. \tag{15}$$

The CNL model is a zeroth order but fully determined model of the band offsets, in which the CNL energy is determined by the bulk electronic structure of oxide and of Si. The local bonding at the interface does not enter in this model.

The predicted CB offsets in this model [18,63] are given in Table 5 and Figure 25 for the various oxides.

Table 5 compares these to the experimental values measured by photoemission, internal photoemission or barrier tunneling [78–89]. Photoemission measures the VB offset, and this is converted into the CB offset by subtracting the oxide and Si band gaps. Internal photoemission measures the energy from the Si valence band to the oxide conduction band, or the Si conduction band to the oxide valence band, depending on the polarity of the Si and of the applied voltage. It is seen that the predicted and experimental offsets generally agree well. Those for HfO_2 and ZrO_2 from photoemission agree well [79,84]. $SrTiO_3$ indeed has a small CB offset [78]. There is now recent data [88] for La_2O_3 which agrees well with the prediction of 2.3 eV. La_2O_3 and $LaAlO_3$ have a particularly large CB offsets [87,88] which means they could be the second generation high K oxides with lowest leakage. The largest exception is the internal photoemission of Afanasev [83] for Al_2O_3. This is because these authors used Al_2O_3 films grown by atomic layer deposition whose band gap is much less (6.8 eV) than that of the pure bulk oxide (8.8 eV).

It is seen that only Al, Y, La, Zr and Hf based oxides have CB offsets over 1 eV, which is the minimum needed to limit electron injection. The CB offsets decrease in the order of group III, IV, to IV metal oxides. This is because the CNL of the oxide rises in the gap along the sequence group III to V.

Lucovsky et al. [64,85] have observed that the x-ray absorption thresholds of the metal d states of the various oxides track the changes in CB offset. This is because the lowest conduction band of the oxide is pure metal d, and so its energy tends to follow the band offset.

4.4 Interfacial bonding

The simple MIGs model of the oxide interface has been surprisingly successful. Nevertheless, future developments will need a more detailed description of the Si-oxide interface. It is important to know the detailed bonding at the Si-oxide interfaces for two reasons. Firstly, the band offset does depend on the interface bonding. Secondly, imperfect interfaces will have defects which can give rise to states in the gap which trap charge.

It is useful to consider epitaxial oxide systems in order to understand the bonding principles in more detail [90–95]. We choose the Si:ZrO_2 system because it is a reasonably well lattice-matched interface and it has (when Y doped) the high symmetry cubic lattice. The lattice constants of Si and ZrO_2 are 5.43 Å and 5.07 Å respectively. This allows ZrO_2 to be grown epitaxially on the Si(100) cube face [96,97], with the ZrO_2 cube face lying directly on top of the Si cube face. This is expressed as ZrO_2(100)//Si(100), and with the [001] directions of Si and oxide parallel, that is ZrO_2[001]//Si[001]. The ZrO_2:Si system is representative of HfO_2 their silicates and also of other cubic oxide systems such as the bixbyite series of Y-rich oxides $(Y,La)_2O_3$ [98–101].

Our understanding of the Si:ZrO_2 interfaces can be guided by those of other fluorite compounds such as metal silicides $NiSi_2$ and $CoSi_2$, and CaF_2. They each form epitaxial interfaces with Si which have been intensively studied previously. It is possible to construct an Si:$NiSi_2$(111) interface in which the last Ni atom is 5, 7, or 8-fold coordinated [101,102]. The most stable interface of these metallic silicides can be understood in terms of the occupation of its bonding states.

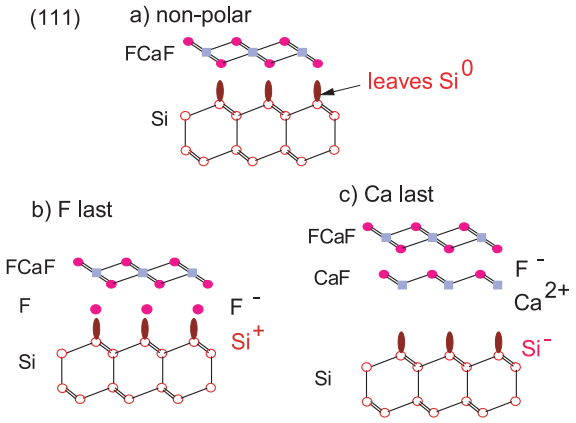

Fig. 26. Schematic of bonding at a (111)Si:CaF_2 interface for different terminations.

The CaF_2 interfaces are more complex than $NiSi_2$ interfaces because CaF_2 has no common element with Si. The ideal (100) and (111) surfaces of CaF_2 are polar, that is they contain only Ca^{2+} or F^- ions. This fixed charge makes the ideal interfaces unstable. On the other hand one can think of CaF_2 as consisting of FCaF layer units stacked along the [100] or [111] directions, in which alternate F ions are assigned to Ca above or below. These (100) or (111) faces now contain half the number of F ions, and are now non-polar (Fig. 26).

On the Si(111) surface, each Si atom has one broken or 'dangling' bond (DB), Figure 26. This state is half occupied, and it will give a metallic interface if it is left like this. We could consider making an Si:CaF_2 (111) interface by joining CaF_2 using one of these non-polar FCaF units, to give a SiFCaF layer structure. Counter-intuitively, it turns out that this would be bad! It would leave the Si DBs all half occupied, and a metallic interface [103], so it not good for a device.

What is needed is to join a *polar* FFCaF unit to the Si(111), as in Figure 26(b). The extra F of the FFCaF unit will form a strong Si-F bond with the Si DB, and this bond sweeps the DB state out of the gap. This can be considered as a $\equiv Si^+F^-F^-Ca^{2+}F^-$ unit (each dash denotes a Si-Si back-bond). An alternative is to use a polar CaF unit. This CaF unit is Ca^+F^- and the Ca^+ therefore has a spare electron. The CaF unit on the Si(111) would donate the spare electron to the Si DB to make a Si^- dangling bond. As important, the Ca s orbital energy lies above the Si gap and it will repel the Si^- DB level into the valence band, so removing all DB states from the gap. This would give a $Si^-Ca^{2+}F^-$ interface with no gap states and filled valence states – an insulating interface. In practice, experiment shows that the CaF terminated interface is formed [104].

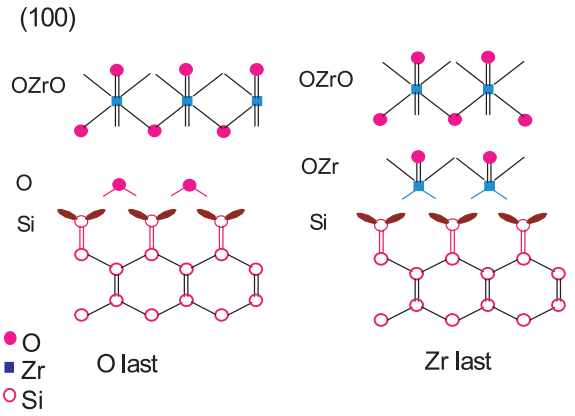

Fig. 27. Schematic of bonding at (100)Si:ZrO_2 interface. OZrO, OOZrO, ZrO.

Now extend this idea to the Si:ZrO_2 interfaces [90], as in Figure 27. As noted by Chang [105] and following the example of CaF_2, we can express bulk ZrO_2 layers as $O^{2-}Zr^{4+}O^{2-}$ (or 'OZrO') units by assigning O's alternately up or down to give non-polar faces.

Now consider the ideal Si(100) face, as shown in Figure 27. Here, each Si atom has 2 DBs which leave states in the gap. If we place a non-polar OZrO unit on this (100)Si, this will still leave Si DB states in the Si gap and we get a metallic surface.

If instead we put a polar OOZrO unit on the Si(100), the first O forms two strong Si-O bonds with each silicon. This O, being divalent, saturates the two DBs of the surface Si to form a structure like a Si-O-Si bridge. Then, the non-polar OZrO unit is added on top of this. The whole ZrO_2 lattice can then be built up on top of this interface by adding further non-polar OZrO layers.

This also works with a ZrO terminating unit. In this case, the ZrO is formally $Zr^{2+}O^{2-}$ and the Zr has two unsatisfied valences. These can be used to make two polar Zr-Si to the Si DBs. This gives an insulating interface with all valences satisfied and a chemical formula =Si=ZrO. The two examples show that epitaxial growth of ZrO_2 on (100)Si is possible, with valence satisfaction and insulating interfaces, provided that the polar faces of ZrO_2 are used.

We have carried out detailed total energy pseudopotential, local density approximation (LDA) calculations of various atomic models of (100) interfaces to test these ideas [90,92]. Some of the interfaces are shown in Figures 28 and 29. Figure 28(a) shows the ideal Si:OZrO interface, which has only one layer of 4-fold coordinated oxygen sites at the interface. We find this interface to be metallic, as expected from the above discussion. Figure 28(b) shows the ideal Si:OOZrO interface, with a double oxygen layer at the interface. Here the interfacial oxygens are 6-fold coordinated initially, bonded to two Si's and four Zr's. It is found that the interfacial oxygens relax to form the structure shown from two directions in Figure 28(c, d). Those oxygens lying in the Si-O-Si bridges relax downwards towards the silicon layer. The other two oxygens relax upwards towards the ZrO_2 layer. Hence, this replicates the discussion above. This interface is denoted the O_4.

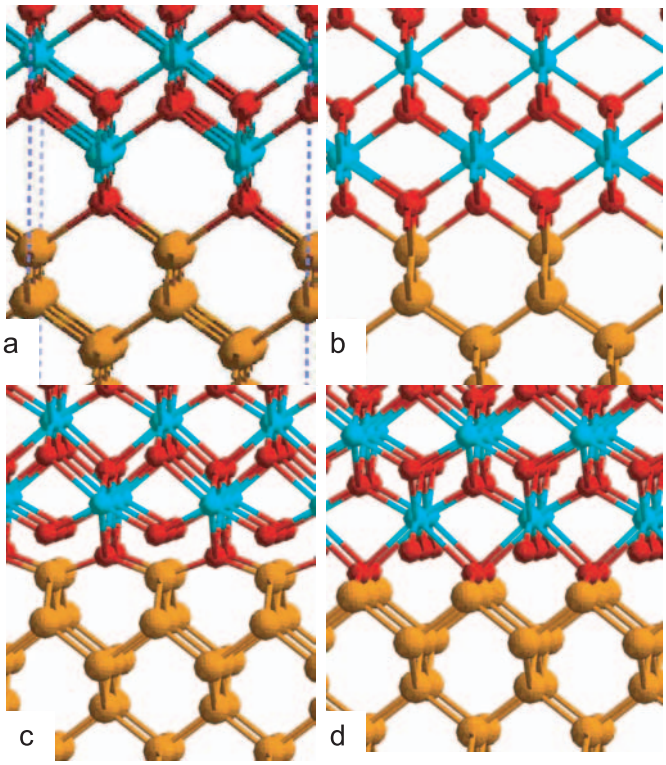

Fig. 28. Various calculated interface configurations of (100)Si. O_{4vac}, O_4, and relaxed O_4 from the [110] and [1–10] directions.

Another interface can be constructed with the oxygens being initially 3-fold coordinated, to one Si atom and two Zr atoms. This is denoted the O_3 interface. The oxygen bonding is then similar to that in $ZrSiO_4$. This interface structure relaxes to the configuration shown in Figure 29(a, b). Here, half of the oxygens are bonded to 2 Si's and 1 Zr, and the other half are bonded to 2 Zr's and one Si. The top layer Si's are each 5-fold coordinated. This interface is also insulating.

A third O-terminated interface with 3-fold coordinated oxygens is possible as shown in Figure 29(c). The ZrO_2 lattice is displaced 1/2a along [100]. It has a lower symmetry than the O_3. The interfacial O is bonded to one Si atom and two Zr atoms as in $ZrSiO_4$ but the O_3 sites are now no longer planar and this allows it to gain stability.

A fourth O-terminated structure is shown in Figure 29(d). Here, one DB of each Si is used in a lateral Si-O-Si bridge [90]. This leaves one DB to bond to the ZrO_2 layer. However, this needs an extra half monolayer of oxygen to saturate its bonding, to give overall a $Si^{+}(O^{2-})_{0.5}OZrO$ configuration. This is denoted the O_{3B} interface (B for bridge).

Finally, there is a partly covalent interface which has been studied by Fonseca et al. [106]. They created an interface where the ZrO_2 is ionic above the first Zr layer, but resembles the Si:SiO_2 interface on the Si side. We denote this as the O_{2B} interface, Figure 29(e). On the interface Si's, one of the two Si DBs is paired with its

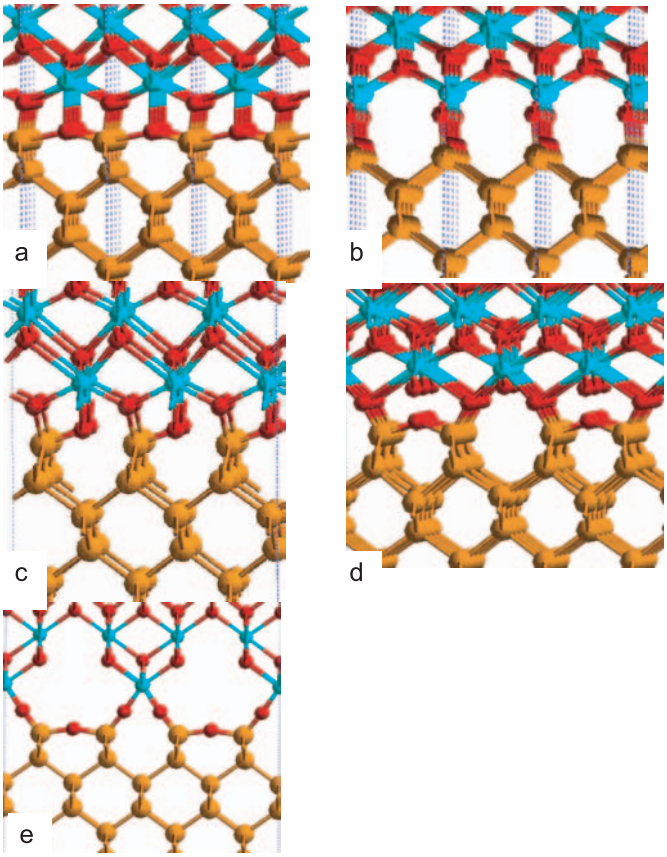

Fig. 29. Calculated configurations of O_3 interface from the [110] and [1–10] directions, the O_{3T}, O_{3B} and the O_{2B} interfaces.

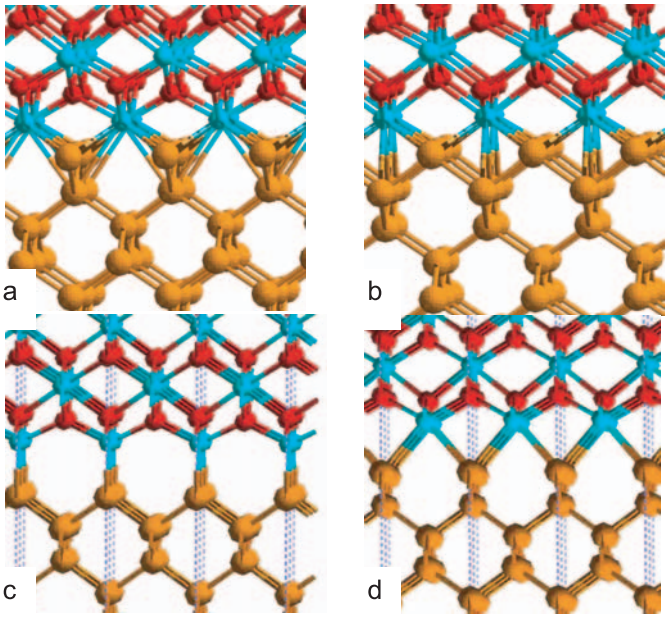

Fig. 30. Calculated metal-terminated configurations of Zr_6, Zr_{10} from [110] and [1–10] directions.

neighbour in a Si-O-Si bridge. This also occurs at the (100)Si:SiO_2 interface [107]. The other Si DB then forms a Si-O-Zr bridge to the first Zr layer. The Si-O-Zr bridge is a covalent unit. Above this Zr, the rest of the ZrO_2 bonding is ionic, as in bulk ZrO_2. This interface has 2×1 symmetry. The interesting thing here is that this interface could be formed by ALD deposition, according to molecular dynamics simulations. The precursor $ZrCl_4$ is a covalently bonded molecule, and ALD is carried out on a partly pre-oxidised Si surface. The two-step process of ALD is likely to retain the initial covalent bonding of the Si-O-Zr bridge units, and then the greater stability of ionic bulk ZrO_2 will exert itself and enforce the denser ionic structure after the first monolayer.

Overall, these interfaces have the same number of oxygen atoms at the interface. The O_3 interface is found to be the most stable structure. The O_4 interface is marginally less stable than O_3. Extensive testing finds that the O_{2B} is as stable as the O_3. This is surprising, because ZrO_2 is 2 eV less stable in the covalent quartz structure. It must arise because this interface configuration allows more structural relaxation at the interface, as the two lattices Si and ZrO_2 are not so well lattice matched.

Experimentally, Wang and Ong [97] measured the interface configuration at (100)Si:ZrO_2 by high-resolution transmission electron microscopy. They found it to have an atomic configuration like O_4, with two oxygen atoms per Si in the last O layer.

Zr-terminated interfaces are also possible. The simplest has a 6-fold coordinated Zr_6, as in Figure 30(a, b). This structure relaxes so that the terminal Zr-Si bond lengthens. Figure 30(c, d) shows another interface in which Zr is 10-fold coordinated, with the Zr bonded to four oxygens, four Si's in the top layer and to two more Si's in the layer under that. This bonding is similar to that in $ZrSi_2$. Our calculation finds that the Zr_{10} is slightly more stable of these two Zr-terminated interfaces.

The calculations find that the three interfaces, O_4, O_3 and O_{3B} and Zr_6 are insulating. They have no states in the Si band gap. However, the Zr_{10} interface is metallic. Thus, only O-terminated interfaces are useful in devices. Chang et al. [105] calculated the surface electronic structure of some Si:ZrO_2 interface configurations. However, they chose some configurations which were metallic. Similarly, Fiorentini [108] calculated the stabilities of some interfaces of Si:HfO_2 but their interface denoted M/O-vac we find to be metallic.

The band offsets have been derived from the calculations of the various interface structures of Si:ZrO_2 from the calculated alignment of the bands. This gives the offset of the valence bands. The offset of the conduction bands is not given well by the LDA calculations, as LDA underestimates the band gap. The offset of the conduction bands must be found instead by adding the experimental band gaps to the VB offset.

It is found that the VB offset is quite similar for the various O-terminated interfaces of ZrO_2. It is also similar to the bulk CNL value of VB offset of 3.3 eV. The offset for Zr-terminated interfaces is different, for Zr_{10} it is less, for Zr_6 it is more. An interface dipole has been formed which causes these differences. Thus, there is no

interface-specific interface dipole for the O-terminated interfaces, but there is for the Zr-terminated interfaces.

The constancy of the band offset for O-terminated interfaces is valuable technologically. It means that the band offset of a ZrO_2 gate oxide does not depend on the surface orientation. It is therefore constant for the polycrystalline or amorphous oxide interfaces. This is very convenient, as it means there will be a larger process window for oxide formation. It is also similar to the established case of Si:SiO_2 where the band offset is constant between Si faces [109]. On the other hand, the band offsets at the two Zr terminated interfaces differ.

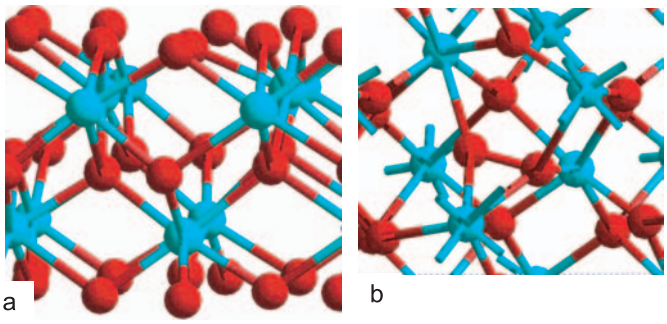

Fig. 31. The relaxed structure of (a) neutral oxygen vacancy and (b) the neutral oxygen interstitial in ZrO_2.

4.5 Electronic structure of defects

One problem with high K oxides is that they contain much higher defect concentrations than SiO_2. The SiO_2 possessed such a low concentration of defects for three reasons. First, its high heat of formation means that off-stoichiometry defects such as O vacancies are costly and so are rare. The second is that SiO_2 has covalent bonding with a low coordination. The covalent bonding means that the main defects are dangling bonds, and the low coordination allows the SiO_2 network to relax to remove any dangling bonds by rebonding the network. This occurs in particular for defects at the Si:SiO_2 interface.

The high K oxides differ in that their bonding is ionic, and they have higher coordination number [91,110]. The greater ionic character of the bonding and the higher coordination numbers mean that the high K oxides are poorer glass formers [110]. The effect of poor glass forming ability and high coordination is that the oxides have higher defect concentrations. The oxides have very high heats of formation, so the equilibrium concentration of non-stoichiometric defects should be low (except where mixed valence is possible, such as TiO_2). However, the non-equilibrium concentration of defects is high, because the oxide network is less able to relax, to rebond and remove defects.

The structure and electronic structure of the oxygen vacancy and oxygen interstitial in ZrO_2 and HfO_2 have been calculated by Foster et al. [111,112] and by Xiong [113]. Recall that the valence band of ZrO_2 consists mainly of O p states and the conduction band mainly of Zr d states. Also in the conduction band, the d_{z2} and d_{x2-y2} (e) states are the lowest conduction band and the d_{xy} (t_2) states are higher, due to crystal field splitting. This is the simple model of ZrO_2 as O^{2-} and Zr^{4+} ions. Surrounding an oxygen vacancy are the 4 metal atoms (or 3 for some sites in monoclinic). In a semiconductor like GaAs, an anion (As) vacancy is surrounded by four dangling bond orbitals on the neighbouring metal (Ga) sites. Hence, an As vacancy gives rise to states localised in the four Ga DBs and these would lie near the conduction band as the Ga forms the conduction band. In an insulator like MgO, an O vacancy again leaves metal states pointing into the vacancy. However, MgO is an insulator and the screening is poor, so the stronger vacancy potential now causes the vacancy state to lie deeper in the gap, near midgap, not close to the conduction band edge [114]. The case of ZrO_2 is closer to the MgO than the GaAs case. The vacancy leaves Zr d states on surrounding atoms, Figure 31(a). A singly degenerate state of A symmetry made of the Zr d_{z2} states lies in the gap, see Figure 32 left side. Its energy level lies moderately far down the gap. This energy depends on its charge state.

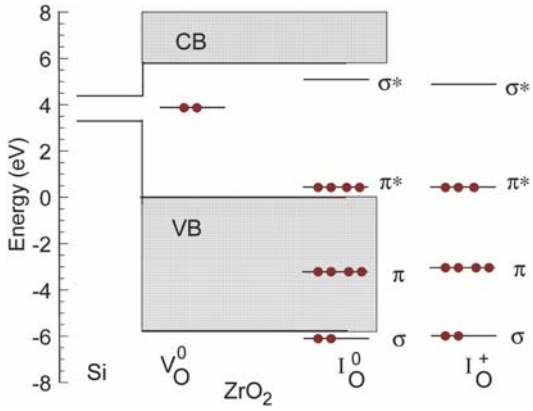

Fig. 32. Molecular orbital diagram of (a) the neutral oxygen vacancy and (b) neutral O interstitial and (c) positively charged O interstitial in ZrO_2, showing energy levels and electron occupancies.

Removing an O^{2-} ion to create a V_O^{2+} vacancy would remove a closed shell ion and leave full valence states. Thus a neutral oxygen vacancy V_O° has two more electrons and these fill the A symmetry gap state.

The energy levels of this state have been calculated by Foster et al. [111]. They calculated the band structure by LDA and they found the energy levels by calculating the ionisation energies and electron affinities of electrons in that state, rather than calculating an energy level as an eigenvalue. It is well known that LDA under-estimates the energies of unoccupied states. It also has difficulty with the localisation of defect states even when filled. Thus it is necessary to correct the value for the large underestimate of the band gap by LDA, and they did this by the scissors operator (moving the gap to fit the experimental value) and interpolation of the defect level. They found the neutral vacancy level to lie at 2.2 eV above the VB edge in ZrO_2. Aligning the bands of ZrO_2 and Si using band

offsets, this sets the neutral V_O energy level as being below the Si VB edge.

Kralik et al. [61] also calculated the energy level of the neutral O vacancy by the GW approximation, which is generally regarded as the most accurate but most expensive method to calculate empty energy states. They found the energy level of the unrelaxed vacancy to be at 3.4 eV above the VB edge in a gap of 5.4 eV, corresponding to about 3.7 eV in a 5.8 eV gap.

Xiong [113] instead used the screened exchange (sX) method [115] and the weighted density approximation (WDA) [116] to calculate the defect excitation energies more correctly than by LDA. No scissors correction is needed. A supercell of 24 atoms was used. The sX method gives the gap of ZrO_2 as 5.2 eV compared to experiment and the neutral vacancy level as 3.5 eV above the VB edge. If a small correction to the gap is applied proportionately, the level then lies at 3.9 eV above the VB edge. This is 0.6 eV above the Si VB edge, or midgap.

The WDA method is more efficient than sX and a cell of 48 atoms is used. The level is found to lie at 4.0 eV above the ZrO_2 VB edge. The structure was relaxed, the neighbouring Zr atoms were found to relax outwards by 0.1A, and the energy level moved up to 4.1 eV. This is 0.8 eV above the VB edge of Si. The latter values are closer to the recent experimental result of Takeuchi [117].

The oxygen interstitial can have a number of charge states, Figure 32. The simplest is the closed shell species O^{2-}. In this state, it lies away from other oxygen anions and it adds filled O $2p$ states just to the valence band. Removing 1 electron to give O^- leaves a hole at the VB edge. Foster [111] notes that this ion moves slightly closer to another O^{2-}, the O-O distance is 2.0 Å. The neutral O interstitial has 2 holes. This now forms the superoxy anion O_2^{2-} which has an internal O-O bond. The O-O bond length is now 1.49 Å. This bond creates a filled bonding σ orbital at -6.0 eV just below the main valence band and an empty antibonding σ^* orbital at 4.1 eV in the upper gap region. It also has filled double degenerate $p\pi$ and π^* orbitals at -3.0 in the valence band and at $+0.3$ eV just above the VB edge (Fig. 32).

The σ^* state could trap an electron, in which case this would partly break the O-O bond and the σ^* state would fall towards the VB edge. Alternatively, the π^* state could trap a further hole to give the O^+ interstitial, or superoxy radical. The hole resides in one of the π^* states, breaking their degeneracy. This radical has a characteristic g factor and has been seen by electron spin resonance in HfO_2 thin films [118].

5 Electrical quality

We have so far described the production, characterisation and bonding of high K oxides. We now continue with their use as electronic materials. It was noted that high K oxides presently perform less well than SiO_2. There are three aspects to this, mobility, gate threshold shifts and charge trapping.

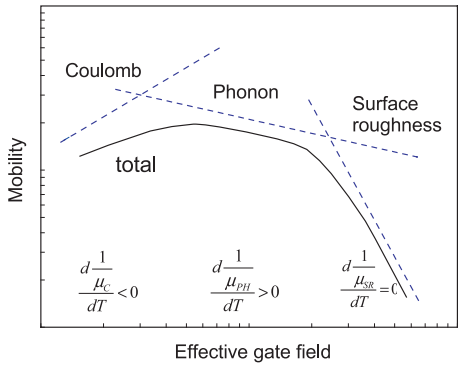

Fig. 33. Schematic carrier mobility vs. vertical field in FETs in the universal mobility model, showing the mechanisms which limit the mobility, and their temperature dependences.

5.1 Mobility degradation

The objective of device scaling is to create smaller, faster devices. Speed follows source-drain drive current, which in turn depends on the carrier mobility. Carriers in the FET behave like a two-dimensional electron gas. The carrier density is determined by the vertical gate field which induces them, by Poisson's equation. The carrier mobility in 2D gases is found to depend in a universal way on this gate field, according to the so-called universal mobility model. This idea developed from observations by Sah, Plummer [119] and others. The most recent version is by Takagi et al. [120] in which the mobility of electrons and holes depends only on the effective gate field and the Si face, [100], [110] or [111]. The individual components of mobility add according to Matthiessen's rule,

$$\frac{1}{\mu} = \frac{1}{\mu_C} + \frac{1}{\mu_{PH}} + \frac{1}{\mu_{SR}}. \tag{16}$$

The mobility is limited by different mechanisms at different fields, as each obeys a different power law with field, see Figure 33. At low fields, mobility is limited by Coulombic scattering (C) by trapped charges in the oxide and/or channel and/or the gate electrode interface; at moderate field it is limited by phonon scattering (PH), and at high fields by scattering by surface roughness (SR).

CMOS devices with a SiO_2 gate oxide have a mobility close to the universal limit. The mobility is limited mainly by interface roughness over the range of interest. The mobilities in devices with high K gate oxides presently lie well below the universal curve [6,23,32,121–125]. This is particularly true of NMOS devices. The reduction in mobility for PMOS devices is fractionally less. Figure 34 shows typical examples. A major objective of present research is to understand the cause of this lowered mobility and to try to correct it.

The cause is presently not well understood. There are two likely causes. First, there could be scattering by excessive amounts of trapped charge and interface states [6]. This is clearly true as other measurements show that high K oxides have much more trapped charge than SiO_2. Secondly, there is the possibility of remote scattering by

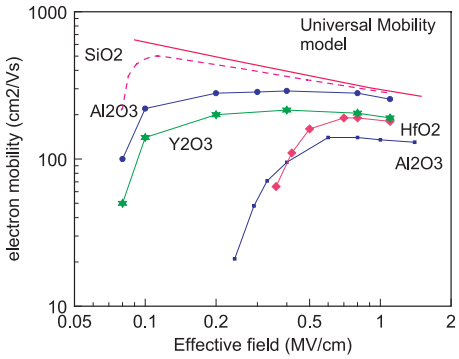

Fig. 34. Carrier mobility of n-type Si, for various gate oxides, after Gusev et al. [23].

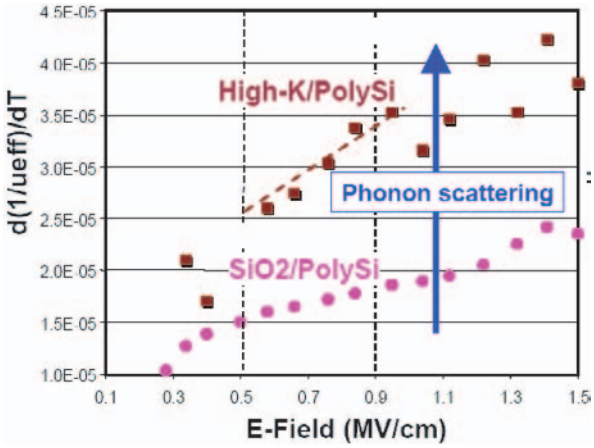

Fig. 35. Measured temperature dependence of mobility for NMOS, after Chau [5].

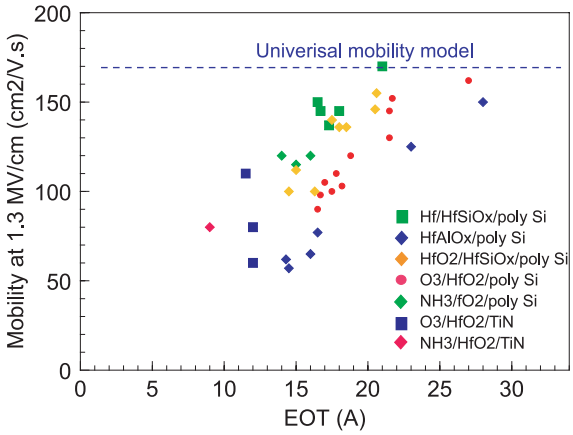

Fig. 36. Mobility vs. EOT for NMOS, showing how the mobility is reduced below the universal value for thinner oxide layers, after Murto et al. [6].

low lying polar phonon modes, as noted by Fischetti et al. [126]. The two contributions can be distinguished by their temperature and by their thickness dependence.

It is also possible that the reduced mobility is due to a reduced induced channel carrier density in inversion, due to the filling of interface traps. This effect has been analysed in detail by Ma et al. [127]. It can be excluded by direct measurements of Hall effect mobility which also shows a reduction [128].

Fischetti [126] noted that in most high K oxides of interest, the high K arises from the low-lying polar lattice vibration modes, see Section 4.2. These polar modes can be effective scatters of carriers in the Si channel – hence 'remote scattering'. The oxides are incipient ferroelectrics and these soft modes would drive the ferroelectric instability if their frequency fell to zero. On the other hand, in SiO_2 such polar modes have a much higher frequency and do not have a large coupling. Fischetti [126] modelled the effect for various oxides and SiO_2. It was found to be pronounced in ZrO_2 and HfO_2. The effect is smaller in $ZrSiO_4$ or $HfSiO_4$ which are now covalently bonding without soft modes. It is also small in Al_2O_3 which has no soft modes. The importance of the effect is that it is intrinsic for those higher K oxides such as HfO_2 and can only be moderated by using $HfSiO_4$, or by including a SiO_2 interfacial layer to separate the HfO_2 away from the channel. Both methods are undesirable as they increase EOT [6,32].

The two mechanisms can be distinguished by their temperature and their thickness dependence. Phonon scattering is the only mechanism whose mobility decreases as the temperature is raised, because the phonon numbers increase with T. Surface roughness is independent of T, and mobility limited by Coulombic scattering increases at higher temperatures (see Fig. 33). Zen et al. [129] and Chau et al. [5,130] have measured the T dependence. They found there is indeed a T dependence of 1/mobility in the mid-field range where it is expected, as seen in Figure 35. Thus, the remote phonon scattering mechanism is important. Ren et al. [129] used HfO_2 gate oxide. Chau [5] did not specify which but it is likely to be HfO_2. Ren's analysis is more complex in that they distinguish scattering by phonons in the oxide and the Si.

The second method is to plot mobility against oxide thickness, and also against thickness of any SiO_2 interlayer oxide, as in the work of Murto [6] and Ragnarsson [32]. The reduction is seen to be greatest in thin high K oxide [6], see Figure 36. Defect scattering would be dominant at lower fields and would increase with thicker oxide layers. These groups interpret their results as showing the importance of Coulombic scattering. Hence, the T-dependence and thickness data indicate that both mechanisms are operative.

Devices from some groups show only small reductions in mobility. This is after considerable processing. Generally, those devices showing small mobility reductions are because the processing has grown an extra SiO_2 interlayer which moves the HfO_2 away from the channel and reduces the remote scattering. Thus, evidence does point to some remote scattering.

Devices using Al_2O_3 gate oxide prove the importance of the charge scattering contribution. Al_2O_3 does not have soft modes, but it does have a high defect concentration. Thus, the reduced mobility seen in those devices [122] can only arise from charge scattering.

Saito et al. of Hitachi [123,131] introduced a general model including the above effects. Most of the scattering arises from charge defects in the oxide and from fluctuations in the dielectric constant from anisotropic oxide crystallites.

Chau et al. [5] suggest that metal gate electrodes would help to screen the dipole coupling of remote phonon scattering. Hence they suggest that this is a further reason for using metal gates with high K oxides.

5.2 V_T stability

The third major problem for high K oxides is the shift of flat band voltages. The flat band ($V_{\rm FB}$) voltage is derived from the capacitance-voltage curve of a CMOS capacitor. By Poisson's equation, the FB voltage measured for a range of film thickness t obeys

$$V_{\rm FB} = \Phi_{\rm ms} + \frac{Qt}{K\varepsilon_0}. \tag{17}$$

Here, $\Phi_{\rm ms}$ is the difference in work functions of the Si and the gate electrode, Q is the interface fixed charge (or trapped charge) density in the film, and K is its dielectric constant. Now, high K oxides have a large defect density, but if we assume that the density is independent of thickness, this plot will be a straight line. Extrapolating to zero t for HfO_2 gate oxide MOS capacitors gives a large $V_{\rm FB}$ value. This compares with small values for SiO_2 gate oxides. The value is of order 0.5 to 1 V for high K oxides on p-type Si and less on n-type Si. Given the rather small operating voltages now for CMOS, these values are large enough to make high K oxide devices inoperable, so the cause must be found.

A series of experiments were carried out varying the polarity of Si substrate, the polarity of poly-Si gate, the thickness of the HfO_2 gate oxide and depositing HfO_2 layers on top of SiO_2 layers, particularly by Hobbs et al. [132–135]. They indicated that the problem arises from an interaction between the HfO_2 and the poly-Si gate material. In principle, the data could be accounted for by fixed charges, dopant diffusion or interface traps [136]. However, the range of tests [132,134,137] suggests that the origin is the interaction of the gate and the HfO_2 gate oxide.

The purpose of the gate electrode in CMOS is to swing the Fermi level of the Si channel to the appropriate band edge to invert it. In the Schottky limit, a change in the gate electrode's work function of 1.1 eV would be needed to swing E_F across the 1.1 eV gap of the underlying Si channel. If we have CMOS with metal gate electrodes and in the Schottky limit, for PMOS with a n-Si channel, a metal electrode with work function 5.1 eV would invert the channel and make it strongly p-type. On the other hand, for NMOS with an initially p-type Si channel, a metal electrode with work function 4.0 eV would invert the channel and make it strongly n-type. In each case, the metal electrodes can be replaced by highly p-type and n-type poly-Si respectively. SiO_2 is a wide gap oxide, and in fact CMOS with SiO_2 or SiO_2N_x does operate close to the Schottky limit, and this is what happens.

Now consider what happens if the gate oxide is a thin layer of HfO_2. We can deposit metals of different work functions onto HfO_2 on Si. The barrier height of the metals to the HfO_2 valence band edge can be measured by photoemission, or the barrier height to the conduction band edge can be measured by tunnelling or by internal photoemission, or the band alignment can be deduced from CV measurements. These results indicate that the barrier heights change with metal by much less than the change in the work function.

As for band offsets, we can define a pinning factor as the change of VB offset divided by the change in the metal's vacuum work function,

$$S = d\phi_n/d\Phi_M. \tag{18}$$

Sayan [138] measured the VB offset by photoemission for Hf and Pt on HfO_2, as shown in Figure 37(a). Si is also included after allowing for its band gap. He found $S \sim 0.5$. Afanaseev [83] measured the Schottky barrier height of Al, Ni and Au on HfO_2 by internal photoemission, Figure 37(b), and found a similar S value. However, the actual size of the offsets are different to Sayan's, as these are also included in Figure 37(a). The barrier heights for ZrO_2 are shown in Figure 38(b), and these also give a value of $S \sim 0.5$.

Yeo [139] derived the effective work function of various metals on HfO_2 from literature data on CV measurements and tunnel barrier heights, as shown in Figure 38(a). The effective work function is defined as the barrier height to the Si CB plus the real electron affinity of Si (4.05 eV). They found an S value of about 0.5.

On the other hand, Schaeffer et al. [140] derived the flat band voltage of various metal electrodes on HfO_2/Si MOS capacitors by CV measurements. They found that V_{FB} changed by less than 0.5 of the change in metal work function. An extreme case is LaB_6 which has a very low work function of 2.6 eV. Schaeffer [140] found a pinning factor closer to 0.2 than 1. Thus their data show a much weaker dependence than that collected by Yeo et al. [139].

The experimental value of S is found to lie in the range 0.1 to 0.5, depending on experimental method used. One could argue that the photoemission measurements are direct and more reliable, while the CV measurements rely on an unproven constancy of Q in equation (17) to extract a value of Φ_{ms}. Given the disagreement between the more direct internal photoemission method and the CV method, this would argue that there is a flaw in effective work functions extracted from CV at present. On the other hand, CV does correspond to the situation in a real device.

This means that metals with a larger range of work function should be needed to drive NMOS and PMOS using HfO_2 gate dielectrics than for SiO_2. Engineers call this 'V_T shifts' when referenced to the SiO_2 case. Engineers always think in the 'Schottky limit'.

To an extent, the observed pinning behaviour is expected from the MIGS model of Schottky barriers, as the

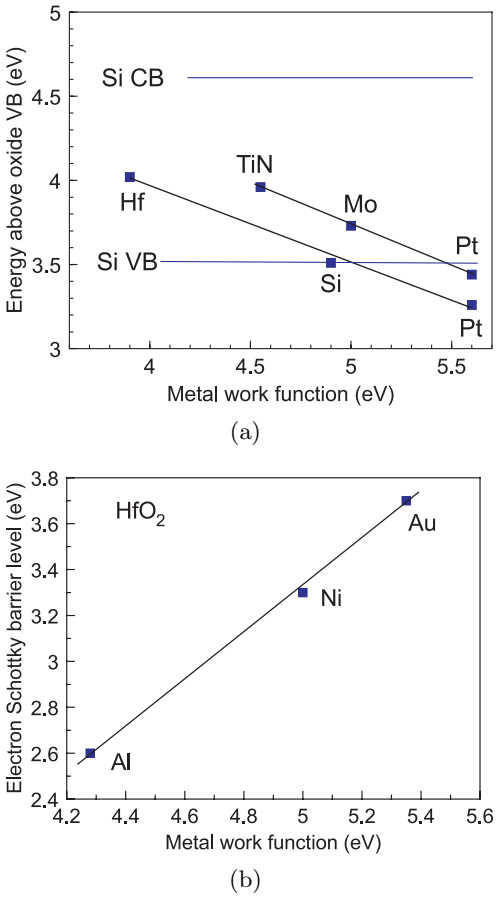

Fig. 37. (a) VB offset of Pt and Hf layers on HfO_2 films, as measured by photoemission [138]. (b) CB barrier heights for metals on HfO_2 measured by internal photoemission [83].

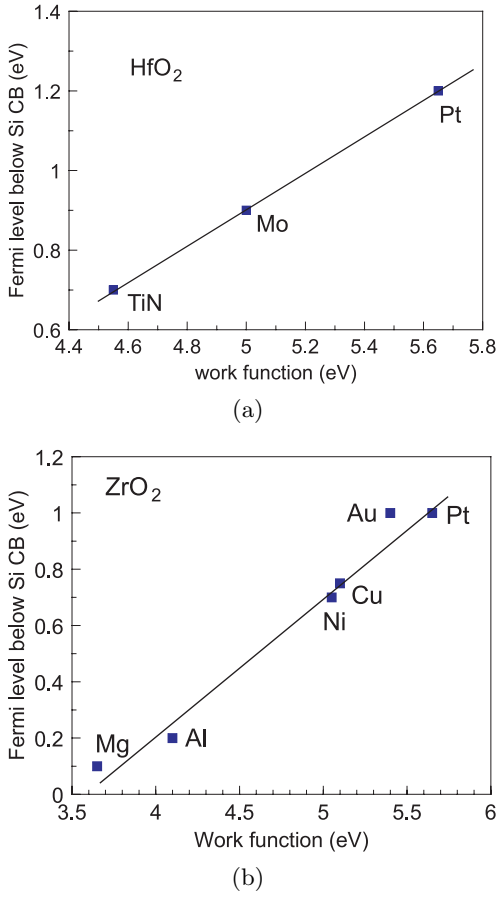

Fig. 38. Effective work function data from CV measurements of metals on HfO_2 and ZrO_2 compiled by Yeo et al. [140].

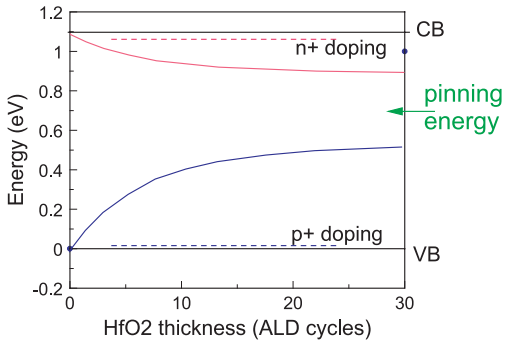

Fig. 39. Schematic of flat band voltage shifts vs HfO_2 layer thickness on SiO_2 on Si, form n-type and p-type poly-Si gate electrodes, after Hobbs [132].

pinning factor S of HfO_2 is 0.52, well below 1. Thus, the behaviour is compatible with the MIGS model. However the smaller values of S are beyond that model. Similar results are obtained for ZrO_2.

However, this is not quite what is observed in the Hobbs experiments. Figure 39 shows how the flat band shift varies for a case of 20 Å of SiO_2 layer plus a variable thickness of HfO_2 on top, for n-poly and p-poly gate electrodes [134]. The flat band shift is seen to be larger for p-poly than n-poly. It is converging towards the upper Si gap region. On the other hand, the band alignment of HfO_2 on the Si channel is such that their charge neutrality levels tend to align. The Si CNL is about 0.2 eV above its valence band edge, and thus the CNL of HfO_2 is also close to this energy, when referred to the Si gap. On the other hand, the data is being 'pinned' towards an energy in the upper gap, about 0.3 eV below the CB edge.

A possible explanation was provided by Hobbs et al. [132,135]. The SiO_2-Si interface is chemically rather simple, as it consists of only two elements. The HfO_2-Si interface is more complicated, as it contains three elements. It is assumed that an ideal, abrupt HfO_2-Si interface consists of O-terminated HfO_2 in contact with Si. It would have only Si-O bonds at the physical interface. Of course, this abrupt situation does not yet happen at the channel-oxide interface because there is usually an interlayer of SiO_2 present. In contrast, the abrupt interface is possible at the gate electrode interface, because the gate is deposited after the oxide, and there is no need for a graded layer for nucleation purposes.

If the ideal abrupt interface consists of O-terminated HfO_2 on Si, with only Si-O interface bonds, then non-ideal interfaces are those with Hf-terminated HfO_2 or with

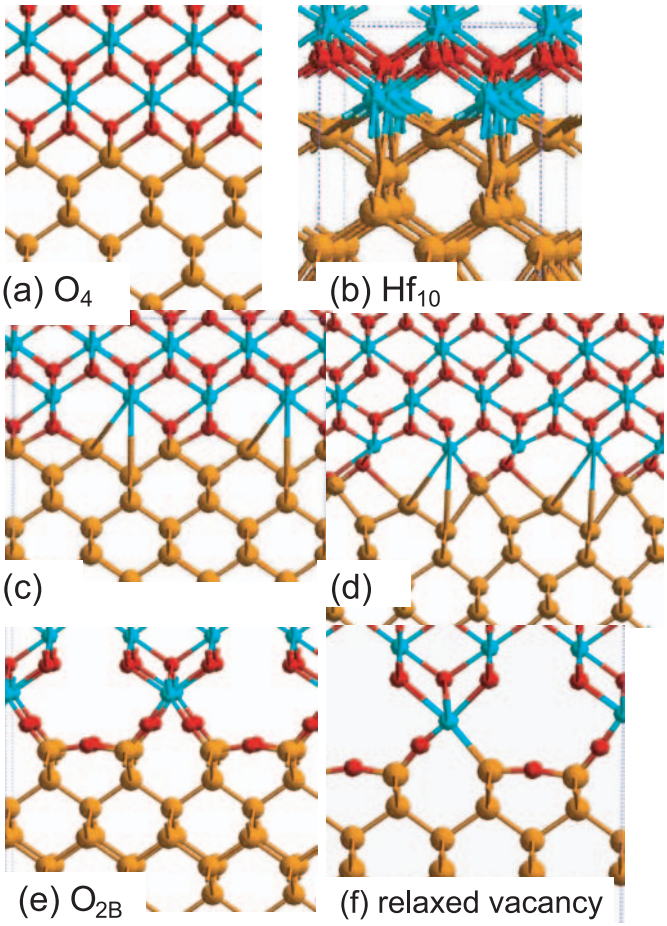

Fig. 40. (a) Ideal O_4 interface, (b) ideal Hf_{10} interface, (c) relaxed O vacancy at O_4 interface, (d) relaxed O vacancy at the O_3 interface, (e) ideal O 2B interface, (f) relaxed O vacancy at the O_{2B} interface.

mixed O and Hf termination next to Si. Both cases would place some Hf atoms next to Si and create Hf-Si bonds. Poly-Si is grown from silane, and its reducing atmosphere is likely to give an O-poor top interface and hence Hf-Si bonds. Thus, Hobbs [135] and also Chau [5] suggested that the Hf-Si bonds at the gate electrode interface lead to pinning of the Fermi level of the gate electrode.

This was supported by Fonseca's calculations reported in Hobbs et al. [135]. These calculations were extended to a much wider range of interface configurations by Xiong et al. [141]. Figure 40 compares model [100] HfO_2:Si interfaces without and with Hf-Si bonds. It was noted that the most symmetric O_4 interface could be continuously transformed into the Hf_{10} interface by removable of interface O atoms. The O_4 interface when relaxed has 2 Si-O bonds, the Hf_{10} interface has no Si-O bonds and 6 Hf-Si bonds, and is metallic. An intermediate case is shown below with 4 Hf-Si bonds and 2 Hf-O bonds. This interface structure was relaxed to minimise its total energy. The local density of states was calculated, and it was found that an interface state cause E_F to lie at about 0.3 eV below the Si CB edge. This causes a very short band bending in the poly-Si, depleting the poly-Si, so that its bulk E_F lines up with the interfacial E_F which is pinned by this interface state.

A number of other interface configurations were tried. Figure 40(e) shows the 2×1 symmetry 2-fold coordinated O-terminated interface studied by Fonseca [135], but with a better picture. An O vacancy is created, and the Hf and Si atoms are rebonded. This case also gives an interface where E_F is pinned in the upper gap. Thus, the calculations support the proposal that Fermi level pinning by Hf-Si bonds at the gate electrode-oxide interface is the cause of the large Vt shifts which appear when poly-Si gates are used with HfO_2 gate oxide. The specific interface configuration is not restrictive.

Hobbs [132] also found that poly-Si on Al_2O_3 gate oxide tended to pin E_F lower in the Si gap. This is the equivalent to the observation by Wilk et al. [1] that most high K oxides have positive fixed charge, except that Al_2O_3 has negative fixed charge. The new model attributes this effect to interaction at the gate interface, not to fixed charge. Al_2O_3 appears to behave differently because an O interface vacancy does not rebond to form Al-Si bonds but leaves a Si dangling bond. The Si DB state lies in the lower gap, about 0.2 eV above the VB.

This V_T only arises because poly-Si is not a 'real' metal. It can have dangling bond states which do not lie at its Fermi level. The problem can be removed by using real metals which can be elemental metals, or metal nitrides, silicides or metal nitride silicides. These have only a Fermi level. The metals must be chosen for their desired work function - high for PMOS and low for NMOS. On HfO_2 there still remains the problem that the work function range is reduced by intrinsic E_F pinning by S. However, this problem appears to have been circumvented by some form of interface design or 'work function engineering' by Intel, as the recent announcement shows FETs with n- and p-metal gates with low V_T offsets [5].

5.3 Charge trapping

We have already noted that high K oxides possess a larger bulk density of defects and trapped charge than SiO_2 [142]. Charge trapping leads to instability in the flat band voltage and gate threshold voltage. It is seen as hysteresis on a drive current vs. gate voltage plot. The effect can be demonstrated by charge pumping experiments. It is notable that $HfSiO_x$ gate oxides have less hysteresis than HfO_2 and also that nitrogen addition reduces it below 70 meV. The amount of trapped charge can be reduced by various annealing cycles and by design of the oxide. It would also be helped by a clearer understanding of its origin.

The origin of this trapped charge is becoming clearer. The first source is intrinsic defects in the oxide and interface traps. Zafar et al. [143] showed that trapping in HfO_2 and Al_2O_3 occurs by the filling of existing defect levels rather than the creation of new defects. This indicates that bulk defects in high K oxides are a serious problem. Kumar [144] showed that hot carriers can create additional defects, but this is an additional effect.

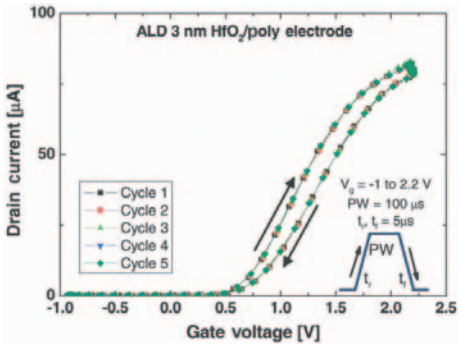

Fig. 41. Electron trapping in HfO_2 gate oxide layer. The hysteresis between the up and down ramps shows the presence of sizable trapping. The identical curves for up and down show that no new defects are created [145].

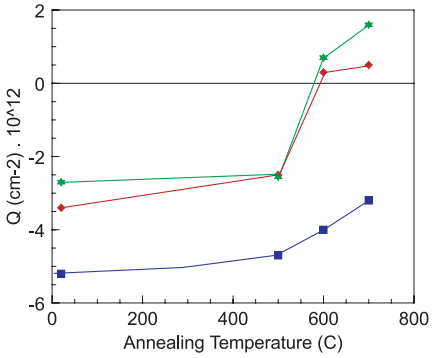

Fig. 42. Variation of trapped charge with annealing temperature, after Houssa [150].

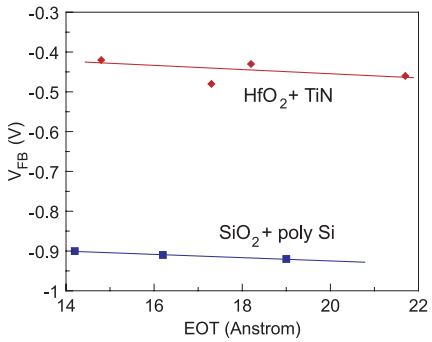

Fig. 43. Low bulk fixed charge as revealed by CV plot for HfO_2 gate oxide, after Datta [130].

Figure 41 shows the effect transient charge trapping in the gate oxide has on a device characteristics, from Bersuker et al. [145]. The gate voltage was cycled and plotted against the resulting FET drain current. The hysteresis between up and down ramps shows that the oxide traps electrons (going positively) and releases electrons (going back). The curves follow the same cycle showing that no new defect traps are formed. Kerber et al. [146] interprets this as fast trapping and detrapping in the oxide. Similar results are found by Shanware et al. [147].

The nature of intrinsic defects in ionic oxides differs from those in SiO_2. They are oxygen vacancies, oxygen interstitials, or oxygen deficiency defects. The chemical nature of the defects can be detected in their paramagnetic configuration by electron spin resonance (ESR). So far, most of the defects found by ESR have been those related to the Si dangling bond at the interface, called the P_b centre [148]. Recently, Lenahan et al. [118] identified three paramagnetic defects by ESR in bulk HfO_2 produced by ALD and subjected to corona discharging; the O vacancy, the Hf^{3+} ion (an electron trapped at Hf^{4+}) and the superoxy radical (or oxygen interstitial). These are the same centres which were previously identified in ZrO_2 powder used in catalysis [149].

The energetics and energy levels of oxygen vacancies and oxygen interstitials in ZrO_2 and HfO_2 were calculated by Forster et al. [111,112] and Xiong [113], as described in Section 4.5. Experimentally, Takeuchi et al. [117] recently used spectroscopic ellipsometry on HfO_2 films oxidised to different levels to identify an absorption band in the gap at 4.5 eV. They attribute this to transitions from the valence band to the oxygen vacancy, and so place the V_O level at 4.5 eV in the gap. Charge pumping experiments would place a defect level close to the Si conduction band, say 4.4−4.5 eV above the HfO_2 VB maximum, which is close to that found by Takeuchi. This is higher than the WDA calculation [113] and a lot higher than in Foster [111]. Kerber et al. [146] noted that the instability data were consistent with an electrical level lying just above the Si conduction band edge. Kumar's [144] data is not consistent with a level lying below the Si valence band edge.

The oxygen interstitial configuration is shown in Figure 31(b). The extra oxygen lies next to bulk oxygen, and the two form a superoxy radical, with a bond of length 1.49 Å for the neutral case. The resulting covalent O-O bond gives rise two π and π^* states at −3 eV and 0.5 eV with respect to the HfO_2 VB edge, and single σ and σ^* states at −8 eV below the main VB and at 5 eV close to the CB edge, Figure 32. The π^* states are filled and the σ^* state is empty for the neutral interstitial. The positively charge I_O^+ has a hole in one of the π^* orbitals. This orbital rises further above the VB edge. It has a unique ESR signature which has been detected in HfO_2 films by Lenahan [118].

The trapped charge can be reduced by annealing. This can be carried out in forming gas (N_2/H_2 mixture), or other nitrogen containing gases such as ammonia. The objective is to reduce the hysteresis in Figure 41 to 7 mV. This is only so far possible in the silicates. Annealing is useful for ALD films because it compacts them and removes possible impurities such as Cl, C and H. The understanding of this process is presently low. Figure 42 shows the variation of trapped charge and interface state density in ALD ZrO_2 with annealing temperature [150]. It is interesting that the trapped charge changes sign at 500 °C when annealed. Houssa [150] speculates that the positive charge can be due to protons in the oxide (that is OH^- ions). Figure 43 shows that fixed charge of 10^{11} cm^{-3} has been achieved with HfO_2 gate oxide by Datta [130].

5.4 DRAM oxides

It was noted in the introduction that replacing silicon oxynitride as the dielectric in the storage capacitor of DRAMs is an equally pressing problem. Some years ago, the roadmap was to use first Ta_2O_5 with a K of 22 and then develop (Ba,Sr)TiO_3 or BST with a very high K of order 2000 [7]. In practice, the ability to use capacitor geometries with high surface area delayed the introduction of high K oxides until recently in DRAM.

The work on gate oxides has allowed ALD as a process to mature. ALD is particularly good a coverage of complex shapes without pin-holes, a key requirement for DRAM. The ALD of Al_2O_3 is the most well developed. In addition it is realised that retaining an amorphous dielectric is very useful in DRAM, as it helps coverage and reduces possible electrical leakage paths. This also favours use of Al_2O_3. Thus the favoured dielectrics for DRAM appear to be Ta aluminate followed by Hf aluminate. The presence of more electronic defects in aluminates is less of a problem in DRAM.

It is important to realise that the requirements for a capacitor dielectric in DRAM are significantly different from those for gate dielectrics. First they are not in contact with Si. Second, the capacitor electrodes are metals, so the band offset requirement is easier. Third, the capacitor is a back end component so that it only needs to withstand lower temperature processing (600 °C). Finally, it should be resistant to hydrogen-induced degradation, and usually this requires forming a hydrogen diffusion barrier around it.

6 Achieving lower EOT

Scaling beyond 2009 requires EOT values below 0.8 nm. As well as metal gates, this will require more abrupt interfaces at the Si channel and higher K values for the bulk oxides. The group III oxides such as LaO_3 or $LaAlO_3$ have higher K than HfO_2 itself. It will also be necessary to use oxides rather than silicates. The 2003 roadmap suggests that epitaxial oxides such as $LaAlO_3$ could be used. The most well developed epitaxial oxide which is lattice matched to Si is $SrTiO_3$. The lattice matching of (001) faces involves a 45° rotation of the $SrTiO_3$ lattice on Si, so that (110)$SrTiO_3$//(100)Si. However, the problem with this interface is that $SrTiO_3$ is not thermodynamically stable next to Si, nor does it have a large enough band offset. It has been possible to create an interface of $SrTiO_3$ on Si, by passing through intermediate SrO and silicide layers [151,152]. It has even been possible to control the band offset somewhat [152]. But it unclear that this interface will pass the requirement of processing up to 1000 °C.

$LaAlO_3$ passes most of the requirements [87]; it is also closely lattice matched to Si, La and Al oxides are both stable next to Si [153], they have low oxygen diffusion coefficients, and it has a conduction band offset of about 1.8 eV. Unfortunately, it has so far not been possible to grow $LaAlO_3$ crystals directly on Si, it grows amorphous. Thus, the future is not clear.

7 Summary

This paper has reviewed the materials chemistry, bonding and electrical behaviour of oxides needed to replace SiO_2 as the gate oxide in CMOS devices. The new oxides must satisfy six conditions to be acceptable as gate dielectrics, a high enough K value, thermal stability, kinetic stability, band offsets, good interface quality with Si, and low bulk defect density. HfO_2 and Hf silicate have emerged as the preferred oxides. The necessary deposition and processing to produce working devices has been achieved. However, the oxides need to optimised substantially further, in order to achieve high performance devices. This requires improvement of flat band voltage and lower defect densities. The flat band voltage shift may be due to interface defects and interface behaviour at the gate oxide/gate electrode interface. The main defects in the oxides are oxygen vacancies and interstitials. The oxygen vacancies are most problem as they give rise to defect levels close to the Si conduction band.

The author would like to thank P.W. Peacock and K. Xiong for many calculations, P. McIntyre and S. Stemmer for illustrations, V. Afanasev for data and numerous colleagues for discussions.

References

1. G. Wilk, R.M. Wallace, J.M. Anthony, J. Appl. Phys. **89**, 5243 (2001)
2. R.M. Wallace, G.D. Wilk, Crit. Rev. Solid State **28**, 231 (2003)
3. R.M. Wallace, G. Wilk, MRS Bull. **27** (2002)
4. S.H. Lo, D.A. Buchanan, Y. Taur, W. Wang, IEEE Electr. Device L. **18**, 209 (1997)
5. R. Chau, at *International Workshop on Gate Insulator, Tokyo, 2003*, `http://www.intel.com/research/silicon/micron.htm#high`; R. Chau, IEEE ED Lett. **25**, 408 (2004)
6. R.W. Murto, M.I. Gardner, G.A. Brown, P.M. Zeitzoff, H.R. Huff, Solid State Technol. **46**, 43 (2003)
7. A.I. Kingon, A.I. Kingon, J.P. Maria, S.K. Streiffer, Nature **406**, 1032 (2000)
8. J. Robertson, J. Vac. Sci. Technol. B **18**, 1785 (2000)
9. J.D. Plummer, P.B. Griffin, Proc. IEEE **89**, 240 (2001)
10. H.J. Hubbard, D.G. Schlom, J. Mater. Res. **11**, 2757 (1996)
11. D.G. Schlom, J.H. Haeni, MRS Bull. **27**, 198 (2002)
12. M. Copel, M. Gribelyuk, E. Gusev, Appl. Phys. Lett. **76**, 436 (2000)
13. M. Gutowski, J.E. Jaffe, C.L. Liu, M. Stoker, R.I. Hegde, R.S. Rai, P.J. Tobin, Appl. Phys. Lett. **80**, 1897 (2002)
14. Y.H. Wu, M.Y. Yang, A. Chin, W.J. Chen, C.M. Kwei, IEEE Electr. Device L. **21**, 341 (2000)
15. H. Iwai et al., *Tech. Digest. Int. Electron Devices Meeting (IEEE, 2002)*
16. J. Kwo et al., Appl. Phys. Lett. **77**, 130 (2000)
17. A. Fissel, H.J. Osten, E. Bugiel, J. Vac. Sci. Technol. B **21**, 1765 (2003)

18. S. Ohmi, M. Takeda, H. Ishiwara, H. Iwai, J. Electrochem. Soc. **151**, G279 (2004)
19. G.D. Wilk, R. Wallace, J.M. Anthony, J. Appl. Phys. **87**, 484 (2000)
20. M.R. Visokay, J.J. Chambers, A.L.P. Rotondaro, A. Shanware, L. Colombo, Appl. Phys. Lett. **80**, 3183 (2002)
21. B.H. Lee, L. Kang, R. Nieh, J.C. Lee, Appl. Phys. Lett. **76**, 1926 (2000)
22. J. Robertson, C.W. Chen, Appl. Phys. Lett. **74**, 1168 (1999)
23. E.P. Gusev, D.A. Buchanan, E. Cartier, A. Kumar, S. Guha, A. Callegari, S. Zafar, P.C. Jamison, D.A. Neumayer, M. Copel, M.A. Gribelyuk, H. Okorn-Schmidt, C. D'Emic, P. Kozlowski, K. Chan, N. Bojarczuk, L.A. Ragnarsson, P. Ronsheim, K. Rim, R.J. Fleming, A. Mocuta, A. Ajmera, in *Tech. Digest. Int. Electron Devices Meeting 2001*, p. 455
24. E.P. Gusev, E. Cartier, D.A. Buchanan, M. Gribelyuk, M. Copel, Microelectron. Eng. **59**, 341 (2001)
25. S. Guha, E. Cartier, N.A. Bojarczuk, J. Bruley, L. Gignac, J. Karasinski, Appl. Phys. Lett. **90**, 512 (2001)
26. W. Tsai et al. (IMEC) *Tech. Digest. Int. Electron Devices Meeting 2003*, paper 13.2
27. Y.C. Yeo, T.J. King, C. Hu, Appl. Phys. Lett. **81**, 2091 (2002)
28. M. Ritala, in *High K gate dielectrics*, edited by M. Houssa (Inst. Physics, Bristol, UK, 2004)
29. M. Ritala, K. Kukli, A. Rahtu, P.I. Raisanen, M. Leskela, Science **288**, 319 (2000)
30. A.C. Jones, P.R. Chalker, J. Phys. D **36**, R80 (2003)
31. M.L. Green, M.Y. Ho, B. Busch, G.D. Wilk, T. Sorsch, T. Conard, B. Brijs, W. Vandervorst, P.I. Raisanen, D.A. Muller, M. Bude, J. Grazul, J. Appl. Phys. **92**, 7168 (2002)
32. L.A. Ragnarrson, L. Pantisano, V. Kaushik, S.I. Saito, Y. Shimamoto, S. DeGendt, M. Heyns, *Tech. Digest. Int. Electron Devices Meeting 2003*, paper 4.2
33. M.M. Frank, Y.J. Chabal, M.L. Green, A. Delabie, B. Brijs, G.D. Wilk, M.Y. Ho, I.J.R. Baumvol, Appl. Phys. Lett. **83**, 740 (2003)
34. M.M. Frank, Y.J. Chabal, G.D. Wilk, Appl. Phys. Lett. **82**, 4758 (2003)
35. G.B. Rayner, D. Kang, G. Lucovsky, J. Vac. Sci. Technol. B **21**, 1783 (2003)
36. J.P. Maria, D. Wicaksana, A.I. Kingon, B. Busch, H. Schulte, E. Garfunkel, T. Gustafsson, J. Appl. Phys. **90**, 3476 (2001)
37. H. Kim, P.C. McIntyre, J. Appl. Phys. **92**, 5094 (2002)
38. S. Stemmer, Z. Chen, C.G. Levi, P.S. Lysaght, B. Foran, J.A. Gisby, J.R. Taylor, Jpn J. Appl. Phys. **42**, 3593 (2003)
39. S. Stemmer, Y. Li, B. Foran, P.S. Lysaght, S.K. Streiffer, P. Fuoss, S. Seifert, Appl. Phys. Lett. **83**, 3141 (2003)
40. C. Zhao, O. Richard, E. Young, H. Bender, G. Roebben, S. Haukka, S. DeGendt, M. Houssa, R. Carter, W. Tsai, O. Van der Biest, M. Heyns, J. Non-Cryst. Solids **303**, 144 (2002)
41. M.Y. Ho, H. Gong, G.D. Wilk, B.W. Busch, M.L. Green, Appl. Phys. Lett. **81**, 4218 (2002)
42. J.C. Lee et al., *Tech. Digest. Int. Electron Devices Meeting 2003*, paper 4.4
43. J. Morais et al., Appl. Phys. Lett. **81**, 2995 (2002); **79**, 4192 (2001)
44. M.A. Quevedo-Lopez et al., Appl. Phys. Lett. **81**, 1074 (2002); **82**, 4669 (2003)
45. R.M.C. de Almeida, I.J.R. Baumvol, Surf. Sci. Rep. **49**, 1 (2003)
46. B.W. Busch, O. Pluchery, Y.J. Chabal, D.A. Muller, R. Opila, D.A. Muller, J.R. Kwo, E. Garfunkel, Mater. Res. Soc. Bull. 206 (2002)
47. S. Ferrari, G. Scarel, J. Appl. Phys. **96**, 144 (2004)
48. M.L. Green, E.P. Gusev, R. Degraeve, E.L. Garfunkel, J. Appl. Phys. **90**, 2057 (2001)
49. S. Stemmer, J. Vac. Sci. Technol. B **22**, 791 (2004)
50. J. Perriere, J. Siejka, R.P.H. Chang, J. Appl. Phys. **56**, 2716 (1984)
51. V. Naraynan, S. Guha, M. Copel, N.A. Bojarczuk, P.L. Flaitz, M. Bribelyuk, Appl. Phys. Lett. **81**, 4183 (2002)
52. M. Copel, M.C. Reuter, Appl. Phys. Lett. **83**, 3398 (2003)
53. M.A. Gribelyuk et al., J. Appl. Phys. **92**, 1232 (2002)
54. M. Copel, M.C. Reuter, P. Jamison, Appl. Phys. Lett. **85**, 458 (2004)
55. G.D. Wilk, D.A. Muller, Appl. Phys. Lett. **83**, 3984 (2003)
56. H.S. Baik, M. Kim, G.S. Park, S.A. Song, M. Varela, A. Franceschetti, S.T. Pantelides, S.J. Pennycock, Appl. Phys. Lett. **85**, 672 (2004)
57. M. Copel, Appl. Phys. Lett. **82**, 1580 (2003)
58. H. Watanabe, M. Saitoh, N. Ikarashi, T. Tatsumi, Appl. Phys. Lett. **85**, 449 (2004)
59. M. Copel, E. Cartier, V. Narayanan, M.C. Reuter, S. Guha, N. Bojarczuk, Appl. Phys. Lett. **81**, 4227 (2002)
60. R.H. French, S.J. Glass, F.S. Ohuchi, Y.N. Xu, W.Y. Ching, Phys. Rev. B **49**, 5133 (1994)
61. B. Kralik, E.K. Chang, S.G. Louie, Phys. Rev. B **57**, 7027 (1998)
62. K. Xiong, J. Robertson (unpublished), WDA bands
63. P.W. Peacock, J. Robertson, J. Appl. Phys. **92**, 4712 (2002)
64. G. Lucovsky, Y. Zhang, J.L. Whitten, D.G. Schlom, J.L. Freeouf, Microelectron. Eng. **72**, 288 (2004)
65. G.M. Rignanese, X. Gonze, A. Pasquarello, Phys. Rev. B **63**, 104305 (2001)
66. H. Kato, T. Nango, T. Miyagawa, T. Katagiri, K.S. Seol, Y. Ohki, J. Appl. Phys. **92**, 1106 (2002)
67. S.G. Lim, S. Kriventsov, T.N. Jackson, J.H. Haeni, D.G. Schlom, A.M. Balbashov, R. Uecker, P. Reiche, J.L. Freeouf, G. Lucovsky, J. Appl. Phys. **91**, 4500 (2002)
68. X. Zhao, D. Vanderbilt, Phys. Rev. B **65**, 233106 (2002)
69. G.M. Rignanese, X. Gionze, G. Jun, K.J. Cho, A. Pasquarello, Phys. Rev. B **69**, 184301 (2004)
70. A. Baldereschi, A. Baroni, R. Resta, Phys. Rev. Lett. **61**, 734 (1988)
71. C.G. van de Walle, Phys. Rev. B **39**, 1871 (1989)
72. R.T. Tung, Phys. Rev. Lett. **84**, 6078 (2000); R.T. Tung, Phys. Rev. B **64**, 205310 (2001)
73. W. Mönch, Phys. Rev. Lett. **58**, 1260 (1987)
74. W. Mönch, Surf. Sci. **300**, 928 (1994)
75. C. Tejedor, F. Flores, E. Louis, J. Phys. C **10**, 2163 (1977)
76. J. Tersoff, Phys. Rev. Lett. **52**, 465 (1984)
77. A.W. Cowley, S.M. Sze, J. Appl. Phys. **36**, 3212 (1965)
78. S.A. Chambers, Y. Liang, Z. Yu, R. Dropad, J. Ramdani, K. Eisenbeiser, Appl. Phys. Lett. **77**, 1662 (2000)
79. S. Miyazaki, J. Vac. Sci. Technol. B **19**, 2212 (2001)
80. D.J. Maria, J. Appl. Phys. **45**, 5454 (1974)

81. R. Ludeke, M.T. Cuberes, E. Cartier, Appl. Phys. Lett. **76**, 2886 (2000)
82. V.V. Afanasev, M. Houssa, A. Stesmans, M.M. Heyns, Appl. Phys. Lett. **78**, 3073 (2001)
83. V.V. Afanasev, M. Houssa, A. Stesmans, M.M. Heyns, J. Appl. Phys. **91**, 3079 (2002) (and private communication)
84. S. Sayan, E. Garfunkel, S. Suzer, Appl. Phys. Lett. **80**, 2135 (2002)
85. G.B. Rayner, D. Kang, Y. Zhang, G. Lucovsky, J. Vac. Sci. Technol. B **20**, 1748 (2002)
86. M. Oshima et al., Appl. Phys. Lett. **83**, 2172 (2003)
87. L.F. Edge, D.G. Schlom, S.A. Chambers, E. Cicerrella, J.L. Freeouf, B. Hollander, J. Schubert, Appl. Phys. Lett. **84**, 726 (2004)
88. T. Hattori, T. Yosihda, T. Shiraishi, K. Takahashi, H. Nohira, S. Joumori, K. Nakajima, M. Suzuki, K. Kimura, I. Kashiwagi, C. Ohshima, S. Ohmi, H. Iwai, Microelectron. Eng. **72**, 283 (2004)
89. A. Ohta, M. Yamaoka, S. Miyazaki, Microelectron. Eng. **72**, 154 (2004)
90. P.W. Peacock, J. Robertson, Phys. Rev. Lett. **92**, 057601 (2004)
91. J. Robertson, Microelectron. Eng. **72**, 112 (2004)
92. J. Robertson, P.W. Peacock, Phys. Stat. Sol. B **241**, 2236 (2004)
93. C.J. Forst, C. Ashman, K. Schwarz, P.E. Blochl, Nature **427**, 56 (2004)
94. X. Zhang, A.A. Demkov, H. Li, X. Hu, H. We, J. Kulik, Phys. Rev. B **68**, 125323 (2003)
95. P.W. Peacock, J. Robertson, Appl. Phys. Lett. **83**, 5497 (2003)
96. S.J. Wang, C.K. Ong, S.Y. Xu, P. Chen, W.C. Tjiu, J.W. Chai, A.C.H. Huan, W.J. Yoo, J.S. Lim, W. Feng, W.K. Choi, Appl. Phys. Lett. **78**, 1604 (2001)
97. S.J. Wang, C.K. Ong, Appl. Phys. Lett. **80**, 2541 (2002)
98. S. Guha, N.A. Bojarczuk, V. Narayanan, Appl. Phys. Lett. **80**, 766 (2002)
99. V. Narayanan, S. Guha, N.A. Bojarczuk, F.M. Ross, J. Appl. Phys. **93**, 251 (2003)
100. G. Apostolopoulos et al., Appl. Phys. Lett. **81**, 3549 (2002)
101. D. Cherns, G.R. Anstis, J.L. Hutchison, J.C.H. Spence, Philos. Mag. A **46**, 849 (1982)
102. D.R. Hamann, Phys. Rev. Lett. **60**, 313 (1988)
103. S. Satpathy, R.M. Martin, Phys. Rev. B **39**, 8494 (1989)
104. R.M. Tromp, M.C. Reuter, Phys. Rev. Lett. **61**, 1756 (1988)
105. R. Puthenkovilakam, E.A. Carter, J.P. Chang, Phys. Rev. B **69**, 155329 (2004)
106. L.R.C. Fonseca, A.A. Demkov, A. Knizhnik, Phys. Stat. Sol. B **239**, 48 (2003)
107. Y. Tu, J. Tersoff, Phys. Rev. Lett. **84**, 4393 (2000)
108. V. Fiorentini, G. Gulleri, Phys. Rev. Lett. **89**, 266101 (2002)
109. J.L. Alay, M. Hirose, J. Appl. Phys. **81**, 1606 (1997); J.W. Keister et al., J. Vac. Sci. Technol. B **17**, 1831 (1999)
110. G. Lucovsky, J. Vac. Sci. Technol. A **19**, 1553 (2001)
111. A.S. Foster, V.B. Sulimov, F.L. Gejo, A.L. Shluger, R.N. Nieminen, Phys. Rev. B **64**, 224108 (2001)
112. A.S. Foster, F.L. Gejo, A.L. Shluger, R.N. Nieminen, Phys. Rev. B **65**, 174117 (2002)
113. K. Xiong, P.W. Peacock, J. Robertson, *Mater. Res. Soc. Symp. Proc.*, Vol. 811, paper D6.4 (2004); IEEE T. Reliab. (2005)
114. B.M. Klein, W.E. Pickett, L.L. Boyer, R. Zeller, Phys. Rev. B **35**, 5802 (1987)
115. B.M. Bylander, L. Kleinman, Phys. Rev. B **41**, 7868 (1990)
116. P.P. Rushton, D.J. Tozer, S.J. Clark, Phys. Rev. B **65**, 235203 (2002)
117. H. Takeuchi, D. Ha, T.J. King, J. Vac. Sci. Technol. A **22**, 1337 (2004)
118. A.Y. Kang, P.M. Lenahan, J.F. Conley, Appl. Phys. Lett. **83**, 3407 (2003); (and unpublished)
119. S.C. Sun, J.D. Plummer, IEEE T. Electron Dev. **27**, 1497 (1980)
120. S.I. Takagi, A. Toriumi, M. Iwase, H. Tango, IEEE T. Electron Dev. **41**, 2357 (1994)
121. L.A. Ragnarsson, S. Guha, M. Copel, E. Cartier, N.A. Bojarczuk, J. Karasinski, Appl. Phys. Lett. **78**, 4169 (2001)
122. S. Guha, E.P. Gusev, H. Okorn-Schmidt, L.A. Ragnarsson, N.A. Bojarczuk, Appl. Phys. Lett. **81**, 2956 (2002)
123. M. Hiratani, S. Saito, Y. Shimamoto, K. Torii, Jpn J. Appl. Phys. **41**, 4521 (2002)
124. L.A. Ragnarsson, L. Pantisano, V. Kaushik, S.I. Saito, Y. Shimamoto, S. DeGent, M. Heyns, *Tech. Digest. Int. Electron Devices Meeting 2003*, p. 87
125. K. Onishi, S.A. Krishnan, J.C. Lee, IEEE T. Electron Dev. **50**, 384 (2003)
126. M.V. Fischetti, D.A. Neumayer, E.A. Cartier, J. Appl. Phys. **90**, 4587 (2001)
127. W. Zhu, J.P. Han, T.P. Ma, IEEE T. Electron Dev. **51**, 98 (2004)
128. L.A. Ragnarrson, N.A. Bojarczuk, J. Karaninski, S. Guha, IEEE Electr. Device L. **24**, 689 (2003)
129. Z. Ren, M.V. Fischetti, E.P. Gusev, E.A. Cartier, M. Chudzik, *Tech. Digest. Int. Electron Devices Meeting 2003*, paper 33.2
130. S. Datta et al, *Tech. Digest. Int. Electron Devices Meeting 2003*, paper 28.8
131. S. Saito, D. Hisamoto, S. Kimura, M. Hiratani, *Tech. Digest. Int. Electron Devices Meeting 2003*, paper 33.3
132. C. Hobbs et al., *VLSI Symp. (2003)*, p. 9
133. S.B. Samavedam, L.B. La, P.J. Tobin, B. White, C. Hobbs, L.C.R. Fonseca, A.A. Demkov, J. Schaeffer, E. Lukowski, A. Martinez, M. Raymond, D. Triyoso, D. Roan, V. Dhandapani, R. Garcia, S.G.H. Anderson, K. Moore, H.H. Tseng, C. Capasso, D.C. Gilnmer, W.J. Taylor, R. Hegde, J. Grant, *Tech. Digest. Int. Electron Devices Meeting 2003*, paper 13.1
134. C. Hobbs et al., IEEE T. Electron Dev. **51**, 971 (2004)
135. C Hobbs et al., IEEE T. Electron Dev. **51**, 978 (2004)
136. C.W. Yang et al., Appl. Phys. Lett. **83**, 308 (2003)
137. E. Cartier et al., *VLSI (2004)*, paper 5.4
138. S. Sayan, E. Garfunkel, J. Robertson, Proc. Electrochem. Soc. **2004-01**, 255 (2004)
139. Y.C. Yeo, T.J. King, C. Hu, J. Appl. Phys. **92**, 7266 (2002)
140. J.K. Schaeffer, S. Samavedam, L. Fonseca, C. Capasso, O. Adetutu, D. Gilmer, C. Hobbs, E. Lckowski, R. Gregory, Z.X. Jiang, Y. Liang, K. Moore, D. Roan, B.Y. Nguyen, P. Tobin, B. White, *Mater. Res. Soc. Symp. Proc. (Spring 2004)*
141. K. Xiong, P.W. Peacock, J. Robertson, Appl. Phys. Lett., to be published (2005)

142. V. Afanasev, A. Stesmans, Appl. Phys. Lett. **80**, 1261 (2002)
143. S. Zafar, A. Callegari, E. Gusev, M.V. Fischetti, J. Appl. Phys. **93**, 9298 (2003)
144. A. Kumar, M.V. Fischetti, T.H. Ning, E. Gusev, J. Appl. Phys. **94**, 1728 (2003)
145. G. Bersuker, P. Zeitzoff, G. Brown, H.R. Huff, Mats Today 7 (Jan 2004), p. 26
146. A. Kerber, E. Cartier, IEEE Lett. **24**, 87 (2003)
147. A. Shanware et al., *Tech. Digest. Int. Electron Devices Meeting 2003*, paper 38-6
148. A. Stesmans, V.V. Afanasev, Appl. Phys. Lett. **82**, 4074 (2003)
149. J. Matta et al., Phys. Chem. Chem. Phys. **1**, 4975 (1999)
150. M. Houssa, V.V. Afanasev, A. Stesmans, M.M. Heyns, Appl. Phys. Lett. **77**, 1885 (2000)
151. R.A. McKee, F.J. Walker, M.F. Chisholm, Phys. Rev. Lett. **81**, 3014 (1998)
152. R.A. McKee, F.J. Walker, M.F. Chisholm, Science **293**, 468 (2001)
153. L.F. Edge, D.G. Schlom, R.T. Brewer, Y.J. Chabal, J.R. Williams, S.A. Chambers, Y. Yang, S. Stemmer, M. Copel, B. Hollander, J. Schubert, Appl. Phys. Lett. **84**, 4629 (2004)